Griol, David; Molina, José Manuel; Callejas, Zoraida. *Modeling the user state for context-aware spoken interaction in ambient assisted living*. Applied Intelligence (2014), 40(4), 749-771.
https://doi.org/10.1007/s10489-013-0503-z

Modeling the user state for context-aware spoken interaction in ambient assisted living

David Griol · José Manuel Molina · Zoraida Callejas

Abstract Ambient Assisted Living (AAL) systems must provide adapted services easily accessible by a wide variety of users. This can only be possible if the communication between the user and the system is carried out through an interface that is simple, rapid, effective, and robust. Natural language interfaces such as dialog systems fulfill these requisites, as they are based on a spoken conversation that resembles human communication. In this paper, we enhance systems interacting in AAL domains by means of incorporating context-aware conversational agents that consider the external context of the interaction and predict the user's state. The user's state is built on the basis of their emotional state and intention, and it is recognized by means of a module conceived as an intermediate phase between natural language understanding and dialog management in the architecture of the conversational agent. This prediction, carried out for each user turn in the dialog, makes it possible to adapt the system dynamically to the user's needs. We have evaluated our proposal developing a context-aware system adapted to patients suffering from chronic pulmonary diseases, and provide a detailed discussion of the positive influence of our proposal in the success of the interaction, the information and services provided, as well as the perceived quality.

Keywords Ambient assisted living · Dialog systems · Context-awareness · Emotions · Human-computer interaction · User adaptation · Spoken interaction · Web interfaces

1 Introduction

The advances in virtual environments, monitoring technologies, intelligent decision support systems and natural interfaces constitute the basis for the new paradigm of pervasive healthcare [41, 45, 98, 106]. These new technologies create a network of communication channels and heterogeneous systems that help individuals in a variety of situations, such as detecting the needs of health professionals and patients [115, 141], healthcare interdisciplinary design [12, 93], gamification and virtual environments [13, 17, 99], physical activity and behavior evaluation [107, 126, 149], therapy and rehabilitation [143, 166], supporting the elderly [82, 91], monitoring sleep and affect [38, 147], gait and exercise evaluation [39, 126], in-home monitoring and mobile healthcare [12, 71, 76, 146, 162], supporting medical decision [101], or psych-physiological sensing [4, 92].

Conversational interfaces [86, 94, 105, 113, 118] have been proven useful for providing the general public with access to telemedicine services, promoting patients' involvement in their own care, assisting in health care delivery, and improving patient outcome [19]. Bickmore and Giorgino defined these systems as being "those automated systems whose primary goal is to provide health communication with patients or consumers primarily using natural language dialog" [19].

During the last two decades, these interfaces have been increasingly used in Ambient Assisted Living (AAL) providing services such as interviews [55, 117], counseling [57, 70], chronic symptoms monitoring [23, 56, 97, 102], medication prescription assistance and adherence [5, 22], changing dietary behavior [44], promoting physical activity [51, 120], helping cigarette smokers quit [124], speech therapy [131], and prognosis and diagnosis using different techniques [60, 88].

The proposal that we present in this paper is focused on the design of user-adapted healthcare systems in which speech is the only modality used as input and output for the system. On the one hand, speech and natural language technologies allow users to access applications in which traditional input interfaces cannot be used (e.g. in-car applications, access for disabled persons, etc.). Also, speech-based interfaces work seamlessly with small devices and allow users to easily invoke local applications or access remote information. For this reason, conversational agents are becoming a strong alternative to traditional graphical interfaces, which might not be appropriate for all users and/or applications' domains [118].

On the other hand, health dialog systems must confront social, emotional and relational issues in order to enhance patients satisfaction. For this reason, context-adaptation can play a relevant role in speech applications for AAL. As described in [75], context information can be divided into *internal* and *external*. The former describes the user state (e.g. communication context and emotional state), whereas the latter refers to the environment state (e.g. location and temporal context). Although many works emphasize the importance of considering context information to solve the tasks presented to the conversational agent by the user and to enhance the system performance in the communication task, this information is not usually considered when designing a dialog model [79, 140]. In addition, although previous works have addressed user-adapted services in AAL applications [9, 19], most of them provide some infrastructure or middleware to support the storage and management of information, but few consider how to organize it so that it can be easily adapted to new application domains.

For these reasons, our proposal is focused on the development of spoken conversational agents interacting in AAL domains providing not only a more natural and intelligent interaction, but also context-aware functionalities adapted to their location, preferences and needs. Our proposal integrates both external and internal context to provide adapted services.

With regard to external context, our proposal is based on additional agents used to capture and provide this information to the spoken conversational agent. Regarding internal context, our proposal merges the traditional view of the dialog act theory, in which communicative acts are defined as intentions or goals, with the recent trends that consider emotion as a vital part for social communication. To do so, we contribute a user state prediction module which can be easily incorporated in the typical architecture of a spoken conversational agent and that is comprised of an intention recognizer and an emotion recognizer. This way, the developed systems can anticipate the user's needs by dynamically adopting their goals and also providing them with unsolicited comments and suggestions, as well as responding immediately to interruptions and provide clarification questions.

We also describe an implementation of our proposal for the development of a context-aware dialog system designed for patients with chronic obstructive pulmonary diseases. The set of functionalities provided by the system includes assessing the patient's behavior since the last conversation, collecting data to monitor the patients' current state, providing feedback on this state, promoting medication adherence, providing personalized tips or relevant educational material, creating a self-report survey with questions assessing the patient's attitude towards the agent, and providing nearest pharmacies on duty. The emotion recognition module in the system is focused on preventing user frustration due to system malfunctions by recognizing and reacting to user negative states (concretely to *anger*, *boredom*, and *doubtfulness*). We have evaluated the developed system and assessed the influence of context information in the quality of the acquired dialogs and the information provided. The results of this detailed evaluation show that context information improves system performance as well as its perceived quality.

The rest of the paper is organized as follows. In Sect. 2 we describe the motivation of our proposal and related work. Section 3 presents in detail the design of context-aware conversational agents to provide adapted services. Section 4 shows a practical implementation of our proposal to generate a specific system. In Sect. 5 we discuss the evaluation results obtained by comparing two baseline versions of the system with a context-aware version that adapts its behavior integrating our proposal. Finally, in Sect. 6 we present the conclusions and outline guidelines for future work.

2 Related work

As described in the introduction section, there are important reasons why dialog-based technology has been applied in healthcare. Dialog systems offer an innovative mechanism for providing cost-effective healthcare services within reach of patients who live in isolated regions, have financial or scheduling constraints, or simply appreciate confidentiality and privacy. Also, as they are based on speech, they are suitable for users with a wide range of computer, reading and health literacy skills.

In general healthcare, professionals can only dedicate a very limited amount of time to each patient. Thus, patients can feel intimidated to ask questions, or to ask for information to be rephrased or simply uncomfortable to provide confidential information on face to face interviews. Many studies have shown that patients are more honest with a computer than a human clinician when disclosing potentially stigmatizing behaviors such as alcohol consumption and HIV risk behavior [2, 55]. Individuals with depression may also find a relational agent more approachable than a clinician in many situations, making it more effective at depression screening and counseling [21].

There exist different approaches to the use of natural language systems in AAL domains. For text-based systems, an interesting research line is the analysis of new valuable communication methods, such as emoticons [156]. For example, Lee et al. proposed the use of simple Smiley diagrams to help in the early detection of depression after stroke (DAS) [80]. Emoticons have a potential positive effect on various factors including enjoyment, personal interaction, perceived information richness, and perceived usefulness [69]. Ptaszynski et al. justify that although emoticons are embedded in lexical form, they convey non-linguistic information [122, 123]. However, they are promising among needy groups such as those with expressive aphasia, low-education levels or a language barrier.

Many systems have introduced relational agents and/or embodied conversational agents (ECAs) designed to establish therapeutic alliance with users over time by means of a potentially powerful and multimodal technology for delivering healthcare services [21, 34]. Most of these systems have been developed following the basis for multimodal interaction defined by important projects like Smartkom [153, 154]. Smartkom's interaction metaphor was based on the idea that the user delegates a task to the virtual communication assistant which is visualized as a life-like character. Among the input modalities considered there were spoken dialog, graphical user interfaces, gestural interaction, facial expressions, physical actions, and biometrics. In the output, it provided an anthropomorphic user interface that combined speech, gesture, and facial expressions.

More recent projects have also addressed very important aspects in multimodal human-machine interaction like the social acceptability of verbally interactive robots and agents, with a special emphasis on their applicability as assistive technologies (e.g. Sera Project [116]), or the possibility of including additional capabilities such as memory, cognition, or learning (e.g., Companions [35], Classic [164], LIREC [83], Soprano [145], Humaine [6], Callas [16], Cogniron [85], Semaine [18], or MUDIS [125]).

An important difference with respect to the type of systems in which we center our contribution, is that in our case the interaction with the user is carried out using only speech, and thus the system must rely solely in the information processed from the user speech. To this respect, the architecture designed in Smartkom centers on modality integration and synchronization and how to solve inconsistencies by using the multimodal inputs to complement one another, thus addresses dialog phenomena that arise specifically in multimodal interaction. These relevant issues are different from the challenge that we address, in which the different information sources do not correspond to different modalities, but to different agents that measure implicit information from the user's spoken input.

In spite of the described advances, a widely diffused adoption of dialog systems in the medical domain is still far from being a reality, mainly because of the inherent complexity when simulating human conversations, and also because of the complexity of this specific domain [127]. Most dialog systems in the healthcare domain maintain a continuous relation with patients through time with the main aim of either eliciting changes in patients' behaviors or habits, monitoring chronic-conditions or assisting them under a determined therapy. The interaction frequencies vary from multiple times a day, to daily, to several times per week, to once every few months, etc. In addition, the complete duration of the dialog-based treatment can extend over a long period (e.g., a month, several months, a few years or lifetime). This continuity forces dialogs to manage extensive and persistent information about the different sessions of the patient.

For these reasons, adaptation can play a relevant role in speech applications for AAL. For example, different levels of adaptation are described in [73]. The simplest one is through personal profiles in which the users make static choices to customize the interaction (e.g. whether they want a male or female system's voice), which can be further improved by classifying users into preferences' groups. Systems can also adapt to the user environment, as in the case of Ambient Intelligence applications [25, 169]. A more sophisticated approach is to adapt the system to the user specific knowledge and expertise, in which case the main research topics are the adaptation of systems to proficiency in the interaction language [112], age [161], different user expertise levels [49], and special needs [96]. Despite their complexity, these characteristics are to some extent rather static. A more complex degree of adaptation in which the system adapts to the user's intentions and state is also identified [73].

In addition, users have diverse ways of communication [59]. Novice users and experienced users may want the interface to behave completely differently, such as maintaining more guided versus more flexible dialogs. Processing context is not only useful to adapt the systems' behavior, but also to cope with the ambiguities derived from the use of natural language [140]. For instance, context information can be used to resolve anaphoric references depending on the context of the dialog or the user location. The performance of a dialog system also depends highly on the environmental conditions, such as whether there are people speaking near the system or the noise generated by other devices.

There exists a reduced number of context-aware speech interfaces in the literature and they are usually applied to very specific domains [114, 129]. In our proposal we merge context-awareness with speech interfaces in order to obtain fully accessible and personalized web services and information in hand-held devices. As stated in the previous section, one of the most important contributions of our work is to combine internal and external contextual information, given that both are essential, to provide a useful personalization and optimize the speech-based interface. Two main aspects are considered for modeling the user behavior: the user intention and the user emotional state. The next subsections discuss the state-of-the-art approaches for both types of models and the advantages of our proposal.

2.1 Modeling the user intention

Different studies have emphasized the importance of the variability in the AAL agent behavior to establish a social bond with the user and to maintain their engagement over long periods of time [19, 128]. To do this, a model of the user-agent relationship is required, methods for updating it, and its use in planning dialog and other interaction behaviors [19]. The simplest approach consists in following a fixed trajectory based on the number of interactions or total contact time with the user. However, determining automatically appropriate types and amounts of behavior variability to maintain engagement and provide user adaptation remains an important area of research [133].

Research in techniques for user modeling has a long history within the fields of language processing and speech technologies. According to Zukerman and Litman [168], very early examples of user modeling in these fields are dominated by knowledge-based formalisms and various types of logic aimed at modeling the complex beliefs and intentions of agents [27, 46, 103, 133]. Grosz and Sidner analyzed the discourse structure as composed of the linguistic structure, the intentional structure, and an attentional state [58, 65].

In more recent years, dialog systems have tended to focus on cooperative, task-oriented rather than conversational forms of dialog, so that user models are now typically less complex. It is possible to classify the different approaches with regard to the level of abstraction at which they model dialog. This can be either at the acoustic level, the word level or the intention-level. The latter is a particularly useful and compact representation of human-computer interaction. Intentions cannot be observed, but they can be described using the speech-act and dialog-act theories [139, 148].

The notion of a dialog act plays a key role in studies of dialog, in particular in the interpretation of communicative behavior of dialog participants, in building annotated dialog corpora, and in the design of dialog management systems for spoken human-computer dialog. A dialog act has two main components: a communicative function and a semantic content. A standard representation for dialog act annotation is proposed in [29], which makes uniform the semantic annotation of dialog corpora. Thus, it provides a standard representation for the output provided by the SLU module in dialog systems and its communication with the dialog manager.

In recent years, simulation on the intention-level has been most popular [133]. This approach was first used by [84] and has been adopted in later work on user simulation by most research groups [42, 54, 119, 136]. Modeling interaction on the intention-level avoids the need to reproduce the enormous variety of human language on the level of speech signals [8, 158] or word sequences [53, 87].

The main purpose of a user intention model in this field is to improve the usability of a conversational agent through the generation of corpora with interactions between the system and the user model [108], reducing time and effort required for collecting large samples of interactions with real users. The user model can be employed to evaluate different aspects of a conversational agent, particularly at the earlier stages of development, or to determine the effects of changes to the system's functionalities (e.g., evaluate confirmation strategies or introduce errors or unpredicted answers in order to evaluate the capacity of the dialog manager to react to unexpected situations).

Two main approaches can be distinguished to the creation of user intention models: rule-based and data or corpus-based. In a rule-based user model, different rules determine the behavior of the system [40, 87]. In this approach the researcher has complete control over the design of the evaluation study. However, these proposals are usually designed ad-hoc for their specific domain using models and standards in which developers must specify each step to be followed by the user model. This way, the adaptation of the hand-crafted designed models to new tasks is a time-consuming process that implies a considerable effort.

Corpus-based approaches use probabilistic methods to generate the user input, with the advantage that this uncertainty can better reflect the unexpected behaviors of users interacting with the system. Statistical models of user intention have been suggested as the solution to the lack of the data that is required for training and evaluating dialog strategies. Using this approach, the conversational agent can explore the space of possible dialog situations and learn enhanced strategies [133].

In [48], Eckert, Levin and Pieraccini introduced the use of statistical models to predict the next user action by means of a n-gram model. The proposed model has the advantage of being both statistical and task-independent. Its weak point is that it approximates the complete history of the dialog by a bigram model. In [84], the bigram model is modified by considering only a set of possible user answers following a given system action (the Levin model). Both models have the drawback of assuming that every user response depends only on the previous system turn. Therefore, they allow that the user model changes objectives continuously or repeats information previously provided.

Georgila, Henderson and Lemon propose the use of HMMs, defining a more detailed description of the user states and considering an extended representation of the history of the dialog [54]. A dialog is represented as a sequence of *Information States* [26, 105]. Two different methodologies are described to select the next user action given a history of information states. The first method uses n-grams [48], whereas the second is based on the use of a linear combination of 290 characteristics to calculate the probability of every action for a specific state.

In [134], a new technique is presented for user modeling based on explicit representations of the user goal and the user agenda. The user agenda is a structure that contains the pending user dialog acts that are needed to elicit the information specified in the goal. In [135], the agenda-based simulator is used to train a statistical Partially Observable MDP (POMDP)-based dialog manager [151]. The main drawback of this approach is that the large state space of practical conversational agents makes its direct representation intractable [165]. Another disadvantage of the POMDP methodology is that the optimization process is free to choose any action at any time.

A data-driven user intention simulation method is presented in [74] that integrates diverse user discourse knowledge (cooperative, corrective, and self-directing). User intention is modeled based on logistic regression and Markov logic framework. Human dialog knowledge is designed into two layers: domain and discourse knowledge, and it is integrated with the data-driven model in generation time. Recent studies have also addressed important points related to the use of semantic agents, multilevel concepts and behavioral model [47], automatic user-profile generation [50], or using physiological signals to detect natural interactive behaviors [89, 100]. A statistical user model supported by a R-Tree structure and several search spaces is presented in [31]. This work also describes a framework for the comparative evaluation of statistical user models. User models have been recently applied to the web in a variety of applications, such as discovering user behavior patterns and interests [77, 144], building semantic social network-based expert systems [43], designing recommendation systems [52], or developing personalized e-news systems [36].

As will be described in Sect. 3.1, our proposed user intention simulation technique is based on a classification process that considers the complete dialog history by incorporating several knowledge sources, combining statistical and heuristic information to enhance the dialog model. The proposed technique presents important differences with respect to user intention modeling in multimodal systems (e.g. Smartkom), in which the main objective is to rank the remaining interpretation hypotheses for the different input modalities (confidence in the speech recognition result, confidence in the gesture recognition result, confidence in the speech understanding result, planning act, and object reference) and obtain a completely instantiated domain object.

2.2 Modeling the user emotional state

Although emotion is receiving increasing attention from the dialog systems community, most research described in the literature is devoted exclusively to emotion recognition. For example, a comprehensive and updated review can be found in [10, 138].

Emotions affect the explicit message conveyed during the interaction and is frequently mentioned in the literature as the most important factor in establishing a working alliance in AAL applications [20, 22]. They change people's voices, facial expressions, gestures, and speech speed; this is a phenomenon addressed as emotional coloring [1]. This effect can be of great importance for the interpretation of the user input.

Emotions can also affect the actions that the user chooses to communicate with the system. According to [159], emotion can be understood more widely as a manipulation of the range of interaction affordances available to each counterpart in a conversation. They have also been recently considered as a very important factor of influence in the decision making processes. For instance, a context-aware model of emotions that can be used to design intelligent agents endowed with emotional capabilities is described in [90]. The study is complemented by also modeling personalities and mood [130].

Despite its benefits, the recognition of emotions in dialog systems presents important challenges which are still unresolved. The first challenging issue is that the way a certain emotion is expressed generally depends on the speakers, their culture, and their environment [24]. Most work has focused on monolingual emotion classification, making an assumption there is no cultural difference among speakers. However, the task of multi-lingual classification has also been investigated [68].

Another problem is that some emotional states are long-term (e.g. sadness), while others are transient and do not last for more than a few minutes. As a consequence, it is not clear which emotion the automatic emotion recognizer will detect: the long-term emotion or the transient one. Thus, it is not trivial to select the categories being analyzed and classified by an automatic emotion recognizer. Linguists have defined extensive inventories of daily emotional states. A typical set is given by Schubiger [137] and O'Connor and Arnold [111], which contains 300 emotional states. However, how to classify such a large number of emotions, or even if it is tractable or practical, remains an open research question.

Also there is not a clear agreement about which speech features are most powerful in distinguishing between emotions. The acoustic variability introduced by the existence of different sentences, speakers, speaking styles, and speaking rates adds another obstacle because these properties directly affect most of the common extracted speech features such as pitch, and energy contours [11].

Related to these problems, some corpus developers prefer the number of utterances for each emotion to be almost the same in order to properly evaluate the classification accuracy. While balanced utterances are useful for controlled scientific analysis and experiments, they may reduce the validity of the data. For this reason, many other researchers prefer that the distribution of the emotions in the database reflect their real-world frequency [104, 163]. In this case, the number of neutral utterances should be the largest in the emotional speech corpus. In addition, the recorded utterances in most emotional speech databases are not produced in the conversational domain of the system [81]. Therefore, utterances may lack some naturalness since it is believed that most emotions are outcomes of our response to different situations.

Very recently, other authors have developed effective dialog models which take into account both emotions and dialog acts. The dialog model proposed by [121] combined three different submodels: an emotional model describing the transitions between user emotional states during the interaction regardless of the data content, a plain dialog model describing the transitions between existing dialog states regardless of the emotions, and a combined model including the dependencies between combined dialog and emotional states. Then, the next dialog state was derived from a combination of the plain dialog model and the combined model. In our proposal, we employ statistical techniques for inferring user acts, which makes it easier to port it to different application domains. Also the proposed architecture is modular and thus makes it possible to employ different emotion and intention recognizers, as the intention recognizer is not linked to the dialog manager as in [121].

Van de Wal and Kowalczyk have recently presented a system that automatically measures changes in the emotional state of the speaker by analyzing their voice [155]. The system was evaluated using natural non-acted human speech of 77 speakers. Chen et al. have also recently introduced an approach that combines acoustic information and emotional point information by means of SVMs, HMMs, and a soft decision strategy [37].

Bui et al. [28] based their model on POMDPs that adapt the dialog strategy to the user actions and emotional states, which are the output of an emotion recognition module. Their model was tested in the development of a route navigation system for rescues in an unsafe tunnel in which users could experience five levels of stress. In order to reduce the computational cost required for solving the POMDP problem for dialog systems in which many emotions and dialog acts might be considered, the authors employ decision networks to complement POMDPs. As will be described in Sect. 3.2, we propose an alternative to this statistical modeling which can also be used in realistic conversational agents and evaluate it in a less emotional application domain in which emotions are produced more subtly.

Different works on audiovisual emotion recognition [66, 95, 157, 167] have shown that facial expression is a better indicator than voice for most emotions. Thus, being able to disambiguate one with the other in a multimodal system produces better results. For example, in SmartKom the results of a recognizer of emotional prosody [14] are merged with the results of a recognizer for affective facial expression [152].

In our case, we count only with the acoustic channel; we carry out a prosody processing procedure like in SmartKom, but additionally, we consider other sources in order to obtain better recognition rates (as we cannot rely on other modalities). This is particularly interesting in systems in which the dialog is less flexible, where the length of the user utterances may be insufficient to enable other knowledge sources like linguistic information to be employed. That is why we propose to take into account information from the user model as well as information related to the context of the dialog that may influence the user's emotional state. This way, restricting a multimodal approach to a single modality (only voice) is not equivalent to our proposal, as we include additional sources of information that deal with the specific challenges of unimodal emotional processing.

3 Context-aware conversational agents to provide adapted services

Our proposal is built on top of the multiagent architecture proposed in [63], adding the capabilities of context management and user adaptation. As it can be observed in Fig. 1, our architecture consists of different types of agents that cooperate to provide an adapted service. *User agents* access the system by means of mobile devices or PDAs. *Provider Agents* supply the different services in the system and are bound to *Conversational Agents* that provide the specific services. A *Facilitator Agent* links the different positions to the providers and services defined in the system. A *Positioning Agent* communicates with the ARUBA positioning system to extract and transmit positioning information to other

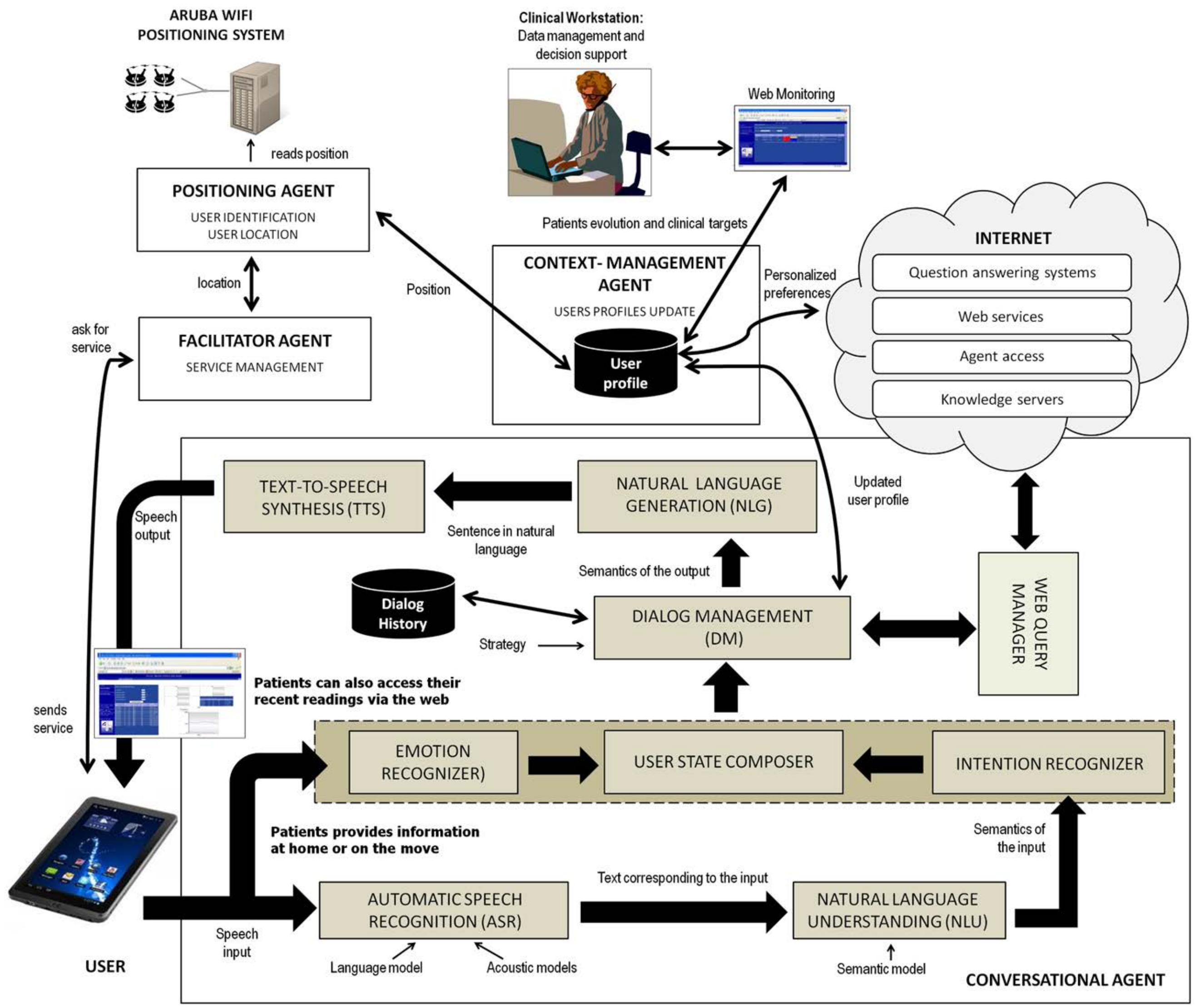

Fig. 1 Architecture proposed to provide context-aware information and services by means of conversational agents

agents in the system. Finally, a *Context-Management Agent* generates and updates user profiles that are used by the Conversational Agents to dynamically adapt their behavior, taking into account the preferences detected in the users' previous dialogs and the information related to the environment for the current interaction.

We have implemented the Facilitator System using the Appear IQ Platform (AIQ).[1] The platform features a distributed modular architecture that supports multiple network configurations and can be deployed as a distributed system. It consists of two main modules: the Appear Context Engine (ACE) and the Appear Client (AC).

The ACE implements a rules engine, where the domain-specific rules that are defined determine what should be available to whom, and where and when it should be available. These rules are fired by a context-awareness runtime environment, which gathers all known context information about a device and produces a context profile for that device. In our system, the defined context parameters include physical location, date/time, device type, network IP address, and user language.

The ACE is installed in a server, while the ACs are included in the users' devices. The network management is carried out by the Appear Context Proxy (ACP), which eliminates unnecessary traffic thus ensuring bandwidth for new user requests, and keeps a cache of active user sessions and most accessed information and services. When a wireless device enters the network, it immediately establishes the connection with a local proxy that evaluates the position of the client device and initiates a remote connection with the server. Once the client is in contact with the server, it pro-

[1] http://www.appearnetworks.com.

vides the set of applications the user can access depending on his physical position.

Given the number of operations that must be carried out by a conversational agent, the scheme used for the development of these systems usually includes several generic modules that deal with multiple knowledge sources and that must cooperate to satisfy the user's requirements. With this premise, a dialog system can be described in terms of the following modules. The *Automatic Speech Recognition module* (ASR) transforms the user utterance into the most probable sequence of words. The *Natural Language Understanding module* (NLU) provides a semantic representation of the meaning of the sequence of words generated by the ASR module. The *Dialog Manager* determines the next action to be taken by the system following a dialog strategy. The *Web Query Manager* receives requests for web services, processes the information, and returns the result to the dialog manager. The *Natural Language Generator module* (NLG) receives a formal representation of the system action and generates a user response that can include multimodal information (video, data tables, images, gestures, etc.). Finally, a *Text to Speech Synthesizer* (TTS) generates the audio signal transmitted to the user.

We propose a framework for predicting the user state that can be integrated in the architecture of a conversational agent as shown in Fig. 1. As can be observed, the framework is placed between the natural language understanding and the dialog management phases. The framework is comprised of an emotion recognizer, an intention recognizer and a user state composer. The emotion recognizer detects the user emotional state by extracting an emotion category from the voice signal and the dialog history. The intention recognizer takes the semantic representation of the user input and predicts the next user action. Then, in the user state composition phase, a data structure is built from the emotion and intention recognized and passed on to the dialog manager.

Our proposal for developing the intention and emotion recognizers allows designing transmutable systems that can engage in many different types of tasks in various usage scenarios. This feature also facilitates the portability and scalability of the developed systems to a wide range of hardware platforms. As in the case of the important projects and architectures previously described in Sect. 2, our proposal combines valuable aspects that model human-human interaction, which is focused on spoken interaction in our proposal.

The complete interaction with the proposed architecture is as follows. Once the user is detected in the network by the ACP, it evaluates the position of the client device and initiates a remote connection with the ACE. The ACE gathers all known information about the device, the user and his context including physical location, date/time, device type, user roles, network IP address range, user locale and other customized context providers. The Context Profile is generated by the Context Engine and then transmitted to the client. This transmission is shown as icons on the hand-held device of the user interface. Then, the client decides which service to pull by clicking on the corresponding icon. Once the resource is available on the device, the installation proceeds.

The user selects the spoken communication interface to receive the information. Immediately, the ACE informs the conversational agent about his identification and current location. Using such information, the conversational agent selects the profile of the recognized user and communicates this information to the different modules of the dialog manager. Each module uses this knowledge to load its specific information and models.

The user starts the interaction with the conversational agent. Throughout the interaction, each module can update the active user profile Depending on the information that is modified the Context-Management Agent sends the value of the new features only to the modules in the conversational agent that require such information. At the end of the interaction, the user profile is updated using the information acquired during the last dialog session. The service is discarded when a user leaves the network or if a context condition has changed for a service.

3.1 The user intention recognizer

The methodology that we have developed for modeling the user intention extends our previous work in statistical models for dialog management [62]. We define user intention as the predicted next user action to fulfill their objective in the dialog. It is computed taking into account the information provided by the user throughout the dialog history, and the last system turn. The formal description of the proposed model is as follows. Let A_i be the output of the dialog system (the system response) at time i, expressed in terms of dialog acts. Let U_i be the semantic representation of the user intention. We represent a dialog as a sequence of pairs (*system-turn*, *user-turn*)

$$(A_1, U_1), \ \ldots, \ (A_i, U_i), \ \ldots, \ (A_n, U_n)$$

where A_1 is the greeting turn of the system, and U_n is the last user turn.

We refer to a pair (A_i, U_i) as S_i, the state of the dialog sequence at time i. Given the representation of a dialog as this sequence of pairs, the objective of the user intention recognizer at time i is to select an appropriate user response U_i. This selection is a local process for each time i, which takes into account the sequence of dialog states that precede time i and the system answer at time i. If the most likely user intention level U_i is selected at each time i, the selection is made using the following maximization rule:

$$\hat{U}_i = \operatorname*{argmax}_{U_i \in \mathcal{U}} P(U_i | S_1, \ldots, S_{i-1}, A_i)$$

where set $\mathcal{U}$ contains all the possible user answers.

As the number of possible sequences of states is very large, we establish a partition in this space (i.e., in the history of the dialog up to time i). Let UR_i be what we call user register at time i. The user register can be defined as a data structure that contains information about concepts and attributes values provided by the user throughout the previous dialog history. The information contained in UR_i is a summary of the information provided by the user up to time i. That is, the semantic interpretation of the user utterances during the dialog and the information that is contained in the user profile

The user profile is comprised of the user's:

- Id and user's name, which he can use to log in to the system.
- Gender.
- Experience, which can be either 0 for novel users (first time the user calls the system) or the number of times the user has interacted with the system.
- Skill level, estimated taking into account the level of expertise, the duration of their previous dialogs, the time that was necessary to access a specific content, and the date of the last interaction with the system. A low, medium, high, or expert level is assigned using these measures.
- Most frequent objective of the user.
- Medical profile configured by the specialist, including primary medications and prescribed doses, changes in the level of physiological parameters (e.g. blood pressure), time of day the user is likely to use the system, whether patient is able to engage in brisk walking as a form of physical activity, name and social relation of a friend or family member who can provide support to the user if needed.
- Reference to the location of all the information regarding the previous interactions and the corresponding objective and subjective parameters for the user.
- Parameters of the user neutral voice as will be explained in Sect. 3.2.

The partition that we establish in this space is based on the assumption that two different sequences of states are equivalent if they lead to the same UR. After applying the above considerations and establishing the equivalence relations in the histories of dialogs, the selection of the best U_i is given by:

$$\hat{U}_i = \operatorname*{argmax}_{U_i \in \mathcal{U}} P(U_i | UR_{i-1}, A_i)$$

To recognize the user intention, we assume that the exact values for the attributes provided by the user are not significant. They are important for accessing the databases and constructing the system prompts. However, the only information necessary to determine the user intention and their objective in the dialog is the presence or absence of concepts and attributes. Therefore, the values of the attributes in the UR are coded in terms of three values 0, 1, 2, where each value has the following meaning:

- 0: The concept is not activated, or the value of the attribute has not yet been provided by the user.
- 1: The concept or attribute is activated with a confidence score that is higher than a given threshold (between 0 and 1). The confidence score is provided during the recognition and understanding processes and can be increased by means of confirmation turns.
- 2: The concept or attribute is activated with a confidence score that is lower than the given threshold.

Based on our previous work on dialog management, we propose the use of a classification process to predict the user intention following the previous equation. Specifically, we use a multilayer perceptron (MLP) for the classification where the input layer received the current situation of the dialog, which is represented by the term (UR_{i-1}, A_i). The values of the output layer can be viewed as the a posteriori probability of selecting the different user intention given the current situation of the dialog.

3.2 The emotion recognizer

As our architecture has been designed to be highly modular, different emotion recognizers could be employed within it. We propose to use an emotion recognizer based solely in acoustic and dialog information because in most application domains the user utterances are not long enough for the linguistic parameters to be significant for the detection of emotions.

Our recognition method, based on the previous work described in [32, 33], firstly takes acoustic information into account to distinguish between the emotions which are acoustically more different, and secondly, it uses dialog information to disambiguate between those that are more similar. We are interested in recognizing negative emotions that might discourage users from employing the system again or even lead them to abort an ongoing dialog. Concretely, we have considered three negative emotions: anger, boredom, and doubtfulness, where the latter refers to a situation in which the user is uncertain about what to do next.

Following the proposed approach, our emotion recognizer employs acoustic information to distinguish anger from doubtfulness or boredom and dialog information to discriminate between doubtfulness and boredom, which are more difficult to discriminate only by using phonetic cues.

This process is shown in Fig. 2. As can be observed, the emotion recognizer always chooses one of the three negative emotions under study, not taking neutral into account. This is due to the difficulty of distinguishing neutral from emotional speech in spontaneous utterances when the application domain is not highly affective. This is the case of most spoken dialog systems, in which a baseline algorithm which always chooses "neutral" would have a very high accuracy, which is difficult to improve by classifying the rest of emotions that are very subtly produced. Instead of considering neutral as another emotional class, we calculate the most likely non-neutral category and then the dialog manager employs the intention information together with this category to decide whether to take the user input as emotional or neutral.

The first step for emotion recognition is feature extraction. The aim is to compute features from the speech input which can be relevant for the detection of emotion in the users' voice. We extracted the most representative selection from the list of 60 features shown in Table 1. The feature selection process is carried out from a corpus of dialogs on demand; this way, when new dialogs are available, the selection algorithms can be executed again, and the list of representative features can be updated. The features are selected by majority voting of a forward selection algorithm, a genetic search, and a ranking filter using the default values of their respective parameters provided by Weka [160].

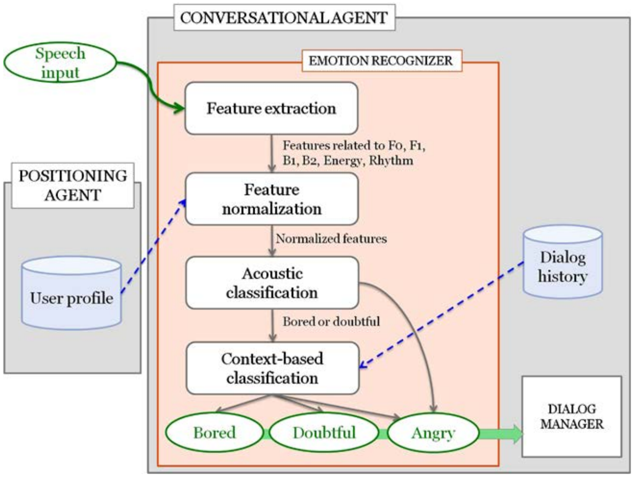

Fig. 2 Schema of the emotion recognizer

The second step of the emotion recognition process is feature normalization, with which the features extracted in the previous phase are normalized around the user's neutral speaking style. This enables us to make more representative classifications, as it might happen that a user 'A' always speaks very fast and loudly, while a user 'B' always speaks in a very relaxed way. Then, some acoustic features may be the same for 'A' neutral as for 'B' angry, which would make the automatic classification fail for one of the users if the features are not normalized.

As described in Sect. 3.1, the values for all features in the neutral style are stored in a user profile. They are calculated as the most frequent values of the user's previous utterances which have been annotated as neutral. This can be done when the user logs in to the system before starting the dialog. If the system does not have information about the identity of the user, we take the first user utterance as neutral assuming that he is not placing the telephone call already in a negative emotional state.

Once we have obtained the normalized features, we classify the corresponding utterance with a multilayer perceptron (MLP) into two categories: *angry* and *doubtful_or_bored*. The precision values obtained with the MLP are discussed in detail in [33], where we evaluated the accuracy of the initial version of this emotion recognizer. If an utterance is classified as angry, the emotional category is passed to the user state composer, which merges it with the intention information to represent the current mental state of the user. If the utterance is classified as *doubtful_or_bored*, it is passed through an additional step in which it is classified according to two dialog parameters: depth and width. Dialog context is considered for emotion recognition by calculating these parameters.

Table 1 Features employed for emotion detection from the acoustic signal [15, 67, 104, 150]

Groups	Features	Physiological changes related to emotion
Pitch	Minimum value, maximum value, mean, median, standard deviation, value in the first voiced segment, value in the last voiced segment, correlation coefficient, slope, and error of the linear regression	Tension of the vocal folds and the sub glottal air pressure
First two formant frequencies and their bandwidths	Minimum value, maximum value, range, mean, median, standard deviation and value in the first and last voiced segments	Vocal tract resonances
Energy	Minimum value, maximum value, mean, median, standard deviation, value in the first voiced segment, value in the last voiced segment, correlation, slope, and error of the energy linear regression	Vocal effort, arousal of emotions
Rhythm	Speech rate, duration of voiced segments, duration of unvoiced segments, duration of longest voiced segment and number of unvoiced segments	Duration and stress conditions

Depth represents the total number of dialog turns up to a particular point of the dialog, whereas width represents the total number of extra turns needed throughout a subdialog to confirm or repeat information. This way, the emotion recognizer has information about the situations in the dialog that may lead to certain negative emotions, e.g. a very long dialog might increase the probability of boredom, whereas a dialog in which most turns were employed to confirm data can make the user angry.

The computation of depth and width is carried out according to the dialog history, which is stored in log files Depth is initialized to 1 and incremented with each new user turn, as well as each time the interaction goes backwards (e.g. to the main menu). Width is initialized to 0 and is increased by 1 for each user turn generated to confirm repeat data or ask the system for help.

Then, the dialog manager tailors the next system answer to the user state by changing the help providing mechanisms, the confirmation strategy, and the interaction flexibility. The conciliation strategies adopted are, following the constraints defined in [30], straightforward and well delimited in order not to make the user lose focus on the task. They are as follows:

– If the recognized emotion is doubtful and the user has changed his behavior several times during the dialog, the dialog manager changes to a system-directed initiative and adds at the end of each prompt a help message describing the available options. This approach is also selected when the user profile indicates that the user is non-expert (or if there is no profile for the current user), and when his first utterances are classified as doubtful.
– In the case of anger, if the dialog history shows that there have been many errors during the interaction, the system apologizes and switches to DTMF (Dual-Tone Multi-Frequency) mode. If the user is assumed to be angry but the system is not aware of any error, the system's prompt is rephrased with more agreeable phrases and the user is advised that they can ask for help at any time.
– In the case of boredom, if there is information available from other interactions of the same user, the system tries to infer from those dialogs what the most likely objective of the user might be. If the detected objective matches the predicted intention, the system takes the information for granted and uses implicit confirmations For example, if a student always asks for subjects of a certain degree, the system can directly disambiguate a subject if it is in several degrees.
– In any other case, the emotion is assumed to be neutral, and the next system prompt is decided only on the basis of the user intention and the user profile (i.e., considering his preferences, previous interactions, and expertise level).

4 Evaluation scenario: patients with domiciliary oxygen therapy

Domiciliary oxygen therapy has been used during the last five decades to alleviate reduced arterial oxygenation (hypoxemia) and its consequences [7, 72]. It is considered to be the only therapeutic approach that can prolong survival in patients with chronic pulmonary diseases. This therapy is also aimed at relieving dyspnea and improving exercise capacity and sleep quality. Patients have portable cylinders, concentrators and portable liquid systems as well a pulse oximeter that monitors the oxygen saturation of a patient's blood and changes in blood volume in the skin. The pulse oximeter is usually incorporated into a multiparameter patient monitor, which also monitors and displays the pulse rate and blood pressure [142].

We have applied our proposed architecture to develop and evaluate an adaptive system that provides context-aware functionalities oriented to these patients. The system is capable of greeting the patient, conducting a chat, assessing the patient's behavior since the last conversation, collecting data to monitor the patients' current state, providing feedback on this behavior, setting new behavioral goals for the patient to work towards before the next conversation, promoting medication adherence, providing personalized tips or relevant educational material, creating a self-report survey with questions assessing the patient's attitude towards the agent, providing nearest pharmacies on duty, and personalized farewell exchanges. The information offered to the patient is extracted from different web pages. Several databases are also used to store this information and automatically update the data that is included in the application.

The greeting and farewell functionalities have been designed to achieve the personalization of the system right from the beginning of the interaction, modifying the structure of the initial and ending prompts to incorporate not only the name of the patient, but also additional functionalities like encouraging them to follow personalized advices.

Given that continuous control and monitoring is a key factor for these diseases, this is one of the main functionalities of the system. The data collected by the system using VoiceXML files [2] are the patient's oxygen saturation level, heart rate, and blood pressure (systolic and diastolic values). The system validates and analyzes the data, providing some immediate feedback to the patients regarding their current progress as well as communicating the results to doctors at the hospital who are able to review the patient's progress graphically and deal with alerts generated by the system concerning abnormal developments.

The evolution of the patient is also taken into account in the personalized tips functionality (e.g., "drink often to

[2] http://www.w3.org/TR/voicexml20/.

Input sentence:
I would like to know my last oxygen saturation level and pharmacies on duty today.

Semantic interpretation:
(*Survey*)
Query_type: OxygenSaturation_Level
(*Pharmacies_Duty*)
Date: Today

Fig. 3 An example of the labeling of a user turn for the described system

avoid dehydration", "keep a varied and balanced diet", "try to keep in the same weight", "avoid caffeine and salty food", "eat in a relaxed environment without hurry", "visit the doctor at the first evidence of cold or influenza" etc.). Also for these patients it is important to receive support, as they sometimes suffer from anxiety and diminished self-esteem because the illness deeply affects their social life.

The chat functionality extends this goal by means of personalized forms related to educational hints explaining details of their illness (e.g., how the respiratory system works and what are the consequences of their treatment so that they can better face them). The medication adherence functionality emulates previous works [5, 22] to remind patients to take all their medications as prescribed in the medical profile Finally, the pharmacies functionality is based on dynamic information automatically provided by the system and related to the current location of the terminal and the daily updated list of pharmacies on duty.

We have defined ten concepts to represent the different queries that the user can perform (*Greeting*, *Chat*, *Logbook*, *Feedback*, *Tips*, *Goals*, *Medication*, *Pharmacies*, *Survey*, and *Farewell*). Three task-independent concepts have also been defined for the task (*Affirmation*, *Negation*, and *Not-Understood*). A total of 109 system actions (dialog acts) were defined taking into account the information that is required by the system to provide the requested information for each one of the queries. For instance, for the logbook functionality, users must provide four attributes (*Oxygen-Saturation*, *Heart-Rate*, *Systolic-Pressure*, and *Diastolic-Pressure*).

An example of the semantic interpretation of a user utterance is shown in Fig. 3.

The *UR* defined for the task is a sequence of 123 fields corresponding to:

- The 10 concepts defined for the dialog act representation.
- The total of 109 possible attributes for the concepts.
- The 3 task-independent concepts that users can provide (*Acceptance*, *Rejection*, and *Not-Understood*).
- A reference to the user profile Configuration parameters in the user profile are entered via a form, and consist of the different items described in Sect. 3.1.

As the intention recognizer and the emotion recognizer modules in the system iteratively improve and adapt their operation as the number of user interactions with the system increases, users are required to complete the information corresponding to their user profile before the first interaction with the system. As it has been described in Sect. 3.1, this profile includes generic features (as the user's name and gender) and task-dependent features (in our case, the patient's medical profile)

Regarding the emotion recognizer, we have carried out a study in which three human annotators have tagged the emotions they detected in the interactions between the users and the system from a list of the 12 emotions considered in the paper. As a result, they did not detect a significant number of positive emotions or negative emotions different from the ones used in this study. The block of parameters related to the user's neutral voice is especially important for this module. For this reason, users have also been required to train the system to their specific characteristics of their voice before interacting with the system for the first time. As it has been explained in Sect. 3.2, these parameters are updated in the following dialogs with the system. To avoid over-specialization, the application adapts to each user, storing their specific models in each terminal.

The architecture of the system used for the experimental setup includes a VoiceXML-compliant IVR platform, which also provides the ASR and TTS modules, and telephone access. The system prompts and the grammars for automatic speech recognition are also implemented in VoiceXML-compliant formats (e.g., Java Speech Grammar Format or JSGF, and Speech Recognition Grammar Specification or SRGS); and the VoiceXML files include each specific system prompt defined for the system and a reference to a grammar that defines the valid user inputs for the corresponding prompt. The emotion recognizer, the intention recognizer and the user state composer for the context-aware system are stored in an external web server and includes the data structure corresponding to the User Register and the trained neural networks. The result generated by the dialog manager informs the IVR platform about the most probable system prompt to be selected for the current dialog state. Then, the platform selects the corresponding VoiceXML file and reproduces it to the user.

A previously developed automatic user simulation technique [61] has been employed to generate the dialog corpus required for learning the neural networks for the emotion recognizer and the intention recognizer in the context-aware system. This simulator carries out the functions of the ASR and NLU modules. An additional error simulator module is used to perform error generation and the addition of ASR confidence scores. The number of errors that are introduced in the recognized sentence can be modified to adapt the error simulator module to the operation of any ASR and NLU modules.

For these experiments, we have adapted this simulator to generate simulated user intentions following the semantics defined for the system. As in the intention recognizer, the user simulation generates the user intention level; that is, the user simulator provides concepts and attributes that represent the intention of the user utterance. Additionally, we have added as a novel function the simulation of the output of the emotion recognizer. In order to do so, the selection of the possible users' emotions coincides with the set described for the development of our emotion recognizer for the system (boredom, anger, doubtfulness, and neutral).

To generate the emotion label for each turn of the simulated user, we employ a rule-based approach based on dialog information similar to the threshold method employed as a second step in the emotion recognizer described in Sect. 3.2. In each case, the method chooses randomly (0.5 probabilities) between an emotion (doubtful, bored, or angry) and neutral. The probability of choosing the emotion rises to 0.7 when the same emotion was chosen in the previous turn, which allows simulating moderate changes of the emotional state. Although the simulated users resemble the behavior of the real users in the initial corpus acquired for the task (the changes in the emotional state correspond to the same transitions observed in the dialog states), they are more emotional, as the probability of neutral in this corpus was 0.85. This way, it is possible to obtain different degrees of emotional behavior with which to evaluate the benefit of our proposal.

A user request for closing the dialog is selected once the system has provided the information defined in the objective(s) of the dialog. The simulated dialogs that fulfill this condition before a maximum number of turns are considered successful. The dialog manager considers that the dialog is unsuccessful and decides to abort it when the following conditions hold: (i) the dialog exceeds the maximum number of user turns, specified taking into account real dialogs for the task; (ii) the answer selected by the dialog manager corresponds with a query not required by the user simulator; (iii) the web query manager generates an error warning because the user simulator has not provided the mandatory information needed to carry out the query; (iv) the natural language generation module generates an error when the selected answer involves the use of a data not provided by the user simulator. The user simulation technique was used to acquire a total of 2,000 successful dialogs for the task.

5 Evaluation process and discussion of results

For comparison purposes, we have developed two additional systems that do not include the modules and agents proposed: the baseline and simulated context-aware systems. The baseline system does not carry out any adaptation to the user, while the simulated context-aware system adapts the dialog considering some of the strategies described in Sect. 3.2, without being context-aware. In order to do so, it decides the dialog strategy for the different emotional states by taking into account the dialog state in terms of dialog length and number of confirmations and repetitions, but it does not incorporate the phonetic emotion recognizer or the external context recognizer.

To assess the benefit of our proposal, we have evaluated the context-aware system and compared it to the baseline and simulated context-aware systems. In order to do so, 90 recruited users participated in the evaluation, aged 51 to 69 (mean 57.2), 67 % male, five with chronic pulmonary diseases. Although not all users suffered from them, they were recruited taking into account the age range which is more affected by these disorders. Additionally, the design of the application and its functionalities was carried out with the continuous feedback of several patients and the medical personal that treats them.

A total of 900 dialogs were recorded from the interactions of the recruited users. 30 users employed the real context-aware system, 30 users employed the simulated context-aware system, and 30 users employed the baseline version of the system. The users were provided with a brochure describing the scenarios that they were asked to complete and main functionalities of the system. A total of 40 scenarios were defined to consider the different queries that may be performed by users. Each scenario specified a set of objectives that had to be fulfilled by the user at the end of the dialog, and they were designed to include and combine the complete set of functionalities previously described for the system. An example of the defined scenarios is as follows:

```
User name: Luis Martín
Location: Reina Street
Date and Time: 2012-11-18, 15:35pm
Device: SmartPhone 00-00-45-5A-02-D9
Objective: Introduce new data in the logbook
and ask for the nearest pharmacy on duty.
```

For illustrative purposes, Fig. 4 shows a dialog corresponding to the previous scenario acquired with the baseline and real context-aware systems. Turns with S refer to system turns, and turns with *U* refer to user turns. As it can be observed, the real context-aware system shows a tendency of providing the required services with higher agility and using more natural answers than the baseline system.

To compare the different versions of the system, we computed the mean value for the evaluation measures shown in Table 2, which we extracted from different studies [3, 64, 132]. We then used two-tailed *t*-tests to compare the means

S1: Welcome to the system. Please say or enter your four digit pin.	*No prompts personalization and user's adaptation.*
U1: 1234.	
S2: How can I help you?	*No information about preferred interests.*
U2: I want to access the logbook functionality.	
S3: Let's begin with your oxygen saturation level. What is the percentage?	
U3: 93.	
S4: What is your heart rate?	
U4: 76.	
S5: And finally, for your blood pressure you will be asked for the top number and then the lower number. What is the top number of your blood pressure?	
U5: 160.	
S6: And what is the lower number?	
U6: 86.	
S7: Your readings have been successfully updated. According to them, your oxygen saturation level is lower than yesterday. It can affect your physical ability. Anything else?	*No tips functionality.*
U7: Which is the nearest pharmacy on duty?	
S8: Tell me the name of the street.	*No information about current location.*
U8: Reina Street	
S9: Do you know the number?	
U9: 86.	
S10: The nearest pharmacy on duty is at Harbor Avenue. Anything else?	
U10: No, thank you.	
S11: Thanks for contacting the system. Have a nice day!	

S1: Good afternoon Luis. Do you want to access the logbook functionality?	*Personalized prompt and the categories preferred by the user are highlighted.*
U1: Yes.	
S2: Let's begin with your oxygen saturation level. What is the percentage?	
U2: 93.	
S3: What is your heart rate?	
U3: 76.	
S4: And finally, for your blood pressure you will be asked for the top number and then the lower number. What is the top number of your blood pressure?	
U4: 160.	
S5: And what is the lower number?	
U5: 86.	
S6: Your readings have been successfully updated. According to them, your oxygen saturation level is lower than yesterday. Do not forget to carry out your daily exercises, they will help you to augment your pulmonary capacity. Anything else?	*Personalized tips functionality activation.*
U6: Which is the nearest pharmacy on duty?	
S7: Given your current position, the nearest pharmacy on duty is at Harbor Avenue. Anything else?	*Current location automatically provided.*
U7: No, thank you.	
S8: Have a nice day! Do not forget to take your medication at the end of the day.	*Personalized prompt taking into account previous interactions.*

Fig. 4 An example of a dialog for the practical domain using the baseline system (*above*) or the real context-aware system (*below*)

across the different types of scenarios and users as described in [3]. The significance of the results was computed using the SPSS software [78] with a significance level of 95 %.[3]

In addition, we asked the recruited users to complete a questionnaire to assess their subjective opinion about system performance. The questionnaire had eight questions: (i) Q1: How well did the system understand you?; (ii) Q2: How well did you understand the system messages?; (iii) Q3: Was it easy for you to get the requested information?; (iv) Q4: Was the interaction rate adequate?; (v) Q5: If the system made errors, was it easy for you to correct them?; (vi) Q6: How much did you feel that the system cares about you?; (vii) Q7: How much did you trust the system?; (viii) Q8: With which frequency would you continue working with the system? The possible answers for each one of the questions were the same: Never/Not at all, Seldom/In some measure, Sometimes/Acceptably, Usually/Well, and Always/Very Well. All the answers were assigned a numeric value between one and five (in the same order as they appear in the questionnaire). The following subsections present the results obtained for the four types of evaluation metrics previously described.

[3] The degrees of freedom that SPSS employs for t-tests are $N - 1$ in case the compared groups have the same number of samples (N), and $N1 + N2 - 1$ when they differ in the number of samples ($N1$ and $N2$).

Table 2 Evaluation measures based on the interaction parameters

Dialog success
Dialog success rate (*%success*). The percentage of successfully completed tasks. In each scenario, the user has to obtain one or several pieces of information, and the dialog success depends on whether the system provides the correct data (according to the aims of the scenario) or incorrect data to the user
Average number of corrected errors per dialog (*nCE*)
The average of errors detected and corrected by the dialog manager. We have considered only the errors that modify the values of the attributes and that could cause dialog failure
Average number of uncorrected errors per dialog (*nNCE*). The average of errors not corrected by the dialog manager. Again, only errors that modify the values of the attributes are considered
Error correction rate (*%ECR*) is the percentage of corrected errors, computed as $nCE/(nCE + nNCE)$
High-level dialog features
Average number of turns per dialog (*avgturns/dial*)
Percentage of different dialogs (*%diff*)
Number of repetitions of the most seen dialog (*#repMS*)
Number of turns of the most seen dialog (*#turnsMS*)
Number of turns of the shortest dialog (*#turnsSh*)
Number of turns of the longest dialog (*#turnsLo*)
Ratio users vs. system actions (*usAct/sysAct*)
Dialog style/cooperativeness measures
System dialog acts: Confirmation of concepts and attributes, Questions to require information, and Answers generated after a database query
Confirmation rate (*%confirm*) was computed as the ratio between the number of explicit confirmation turns (*nCT*) and the number of turns in the dialog (*nCT/nT*)
User dialog acts: Request to the system, Provide information, Confirmation Yes/No answers, and Other answers
Goal directed actions vs. grounding actions: Goal directed actions are requesting and providing information, while grounding actions are explicit and implicit confirmations dialog formalities (greetings, instructions, etc.) and unrecognized actions

5.1 Dialog success and high-level dialog features

By means of high-level dialog features, we evaluated the duration of the dialogs, how much information was transmitted in individual turns, and how active the dialog participants were. These dialog features cover the following statistical properties: (i) Dialog length, measured as the mean and shape of the distribution of the number of turns per task, the number of turns of the shortest dialog, the number of turns of the longest dialog, and the number of turns of the most frequent dialog; (ii) Percentage of different dialogs in each corpus and the number of repetitions of the most frequent dialog; (iii) Turn length, measured by the number of actions per turn; and (iv) Participant activity as a ratio of system and user actions per dialog. Table 3 shows the comparison of the different high-level measures for the baseline and context-aware versions of the system.

Table 3 Results for the high-level dialog features

	Baseline	Simulated context-aware system	Real context-aware system
%success	85 %	89 %	94 %
nCE	0.83	0.87	0.90
nNCE	0.18	0.12	0.09
%ECR	81 %	88 %	92 %
avgturn/dial	13.4	12.6	8.6
%diff	76 %	80 %	89 %
#repMS	5	10	8
#turnsMS	9	9	7
#turnsSh	7	7	7
#turnsLo	17	15	15

As can be observed, the different systems could interact correctly with the users in most cases, achieving success rates higher than 85 % in a difficult domain in which only spoken interaction is provided. However, the real context-aware system obtained a higher success rate, improving the baseline system results by 9 % absolute. The simulated context-aware system improved the success rate of the baseline system by 4 %. In our opinion, this can be explained by the use of the dialog context to modify the dialog initiative (from mixed-initiative to system-initiative) and alternate spoken and DTMF interaction modes. Combining this data with that generated by the recognizer is very useful, given that the real context-aware system improves the simulated context-aware system by 5 %. This difference showed a significance value of 0.025 in the two-tailed *t*-test.

On the other hand, although the error correction rates were also improved in absolute values by using the real context-aware system, this relationship was not significant in the *t*-test. Both results are explained by the fact that we have not designed a specific strategy to improve the recognition or understanding processes and decrease the error rate, but rather our proposal for adaptation to the user state overcomes these problems during the dialog once they are produced. From the results obtained for these measures, it can also be observed the importance of considering the dialog context for error detection and correction. This way, the simulated context-aware system achieves values for this set of measures that are closer to the corresponding results obtained by the real context-aware system.

Regarding the number of dialog turns, the real context-aware system produced shorter dialogs (8.6 turns in average) compared to the number of turns of the baseline system (13.4). This result was obtained with a 0.000 significance value in the *t*-test when compared to the number of turns of the baseline system. Nevertheless, the comparison of the results obtained for this measure in the real context-aware system and the simulated context-aware system was not significant in the *t*-test. Besides the low difference obtained in the average number of turns for both systems (13.4 and 12.6 respectively), an in-depth study showed that the simulated context-aware system generated longer dialogs than the baseline system in several scenarios. This could be due to the simulation of the user's emotional state in the simulated context-aware system, which caused false positives in the detection of the angry emotional state and the consequent activation of the system-initiative or DTMF interaction mode when they were not really required.

The dialogs acquired with the baseline system have a higher standard deviation (4.57) given that the proportion of number of turns per dialog is more diverse. The dialogs gathered with the context-aware system have a smaller deviation (3.84) since the successful dialogs are usually those which require the minimum number of turns to achieve the objective(s) predefined in the scenarios. The simulated context-aware system obtained an intermediate value for the standard deviation (4.06).

As shown in Table 3, this general reduction in the number of turns is generalized also to the case of the longest, shortest and most seen dialogs for the real context-aware system. This might be because users have to explicitly provide and confirm more information using the baseline system, whereas the real context-aware system automatically adapted the dialog to the user and the dialog history. This way, users have more variability in order to provide the information that is needed to access the different services in the baseline system. Although the reduction in the average number of dialog turns between the real context-aware system and the baseline system was significant (significance value of 0.018), the comparative between the results of the longest, shortest and most seen dialogs for the three systems provided a non-significant value in the *t*-test.

Table 4 Percentage of different dialogs obtained

Percentage of different dialogs	Baseline	Simulated context-aware system	Real context-aware system
Difference at intention level only	76 %	80 %	85 %
Difference at user-state level (intention + emotion)	76 %	82 %	89 %

Table 4 sets out the results regarding the percentage of different dialogs obtained. When we considered the dialogs to be different only when a different sequence of user intentions was observed, the percentage was lower using the context-aware system, due an increment in the variability of ways in which the users can provide the data required using the context-aware system. This is consistent with the fact that the number of repetitions of the most observed dialogs is higher for the baseline system. When emotions were also taken into account, i.e. when even with the same sequence of intentions two dialogs were considered different if the emotions observed were different, we obtained a higher percentage of different dialogs. However, this difference was low because our user state recognizer tends to classify utterances as emotional rather than neutral, as described in Sect. 3.2. The difference between the values obtained for the simulated context-aware system and the real context-aware system can be explained because of the personalization of the dialogs that is achieved with the introduction of the intention recognizer and the user profiles

Regarding the dialog participant activity, the dialogs acquired with the real context-aware version of the system have a higher proportion of system actions, as less confirmations are required using this system. There is also a slight reduction in the mean values of the turn length; these dialogs are statistically shorter, as they provide 1.26 actions per user turn instead of the 1.49 actions provided by the baseline dialogs and the 1.44 actions provided by the simulated context-aware system. This is again because the users

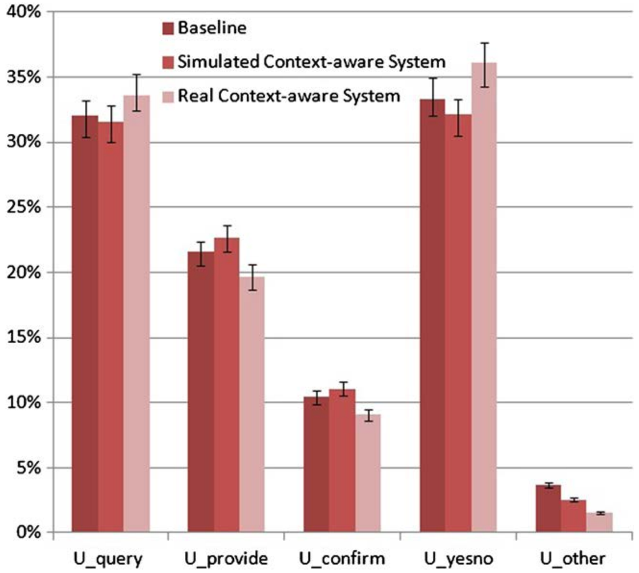

Fig. 5 Percentages of the different types of user dialog acts

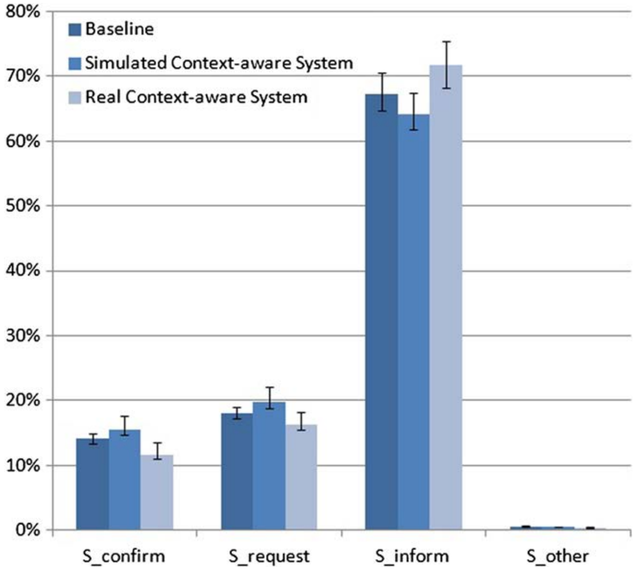

Fig. 6 Percentages of the different types of system dialog acts in the three systems

have to explicitly provide and confirm more information in the baseline system. The results of the *t*-test in a comparative analysis of this measure showed a significant difference just in the comparison of the real context-aware system and the baseline system (significance value of 0.029).

5.2 Dialog style and cooperativeness measures

Dialog style and cooperativeness measures show the frequency of different speech acts and reflect the proportion of actions that are goal-directed (i.e. not indexed in dialog formalities). For dialog style features, we defined and counted a set of system/user dialog acts. On the system side, we measured the confirmation of concepts and attributes, questions to require information, and system answers generated after a database query. On the user side, we measured the percentage of turns in which the user carries out a request to the system, provides information, confirms a concept or attribute, provides Yes/No answers, and gives other responses not included in the previous categories. Finally, we have measured the proportion of goal-directed actions (request and provide information) versus the grounding actions (confirmations and rest of actions.

Figures 5 and 6 respectively show the frequency of the most predominant user and system dialog acts in the dialogs acquired with the three systems. On the system side, *S_request*, *S_confirm*, and *S_inform* indicate actions through which the system respectively requests, confirms or provides information. *S_other* stands for other types of system prompts (e.g, Waiting and Not-Understood dialog acts). On the user side, *U_provide*, *U_query*, *U_confirm*, and *U_yesno* respectively identify actions by which the user provides, requests, or confirms information or gives a yes/no answer, while *U_other* represents all other user actions (e.g, dialog formalities or out of task information).

In both cases, it can be observed that there are significant differences in the distribution of dialog acts. On the one hand, Fig. 5 shows that users need to provide less information using the real context-aware system. This explains the higher proportion for the rest of user actions with regard to the baseline and the simulated context-aware systems (both differences significant over 98 %). There is also a higher proportion of yes/no actions for the context-aware dialogs, which was not significant in the *t*-test. These actions are mainly used to confirm that the specific queries have been correctly provided using context information. The comparison between the simulated context-aware and the baseline systems showed that there is a slight increment of the user's turns providing and confirming information. This might be again due to the activation of the system-directed initiative and DTMF modes in the simulated context-aware system mainly to provide or confirm item by item.

On the other hand, Fig. 6 shows that there is a reduction in the system requests when the real context-aware system is used. This explains a higher proportion of the inform and confirmation system actions when this system is used in comparison with the simulated context-aware and baseline systems (both differences significant over 98 %). The increment of the query and confirmation system actions in the simulated context-aware system has been considered as an explanation of the increment of user actions providing information and user's confirmations using this system. This difference was non-significant in the two tailed *t*-test.

Additionally, we grouped all user and system actions into three categories: “goal directed” (actions to provide or request information), "grounding" (confirmation and negations), and "other". Table 5 shows a comparison between these categories. As can be observed, the dialogs provided by the real context-aware system have a better quality, as the proportion of goal-directed actions is higher than the values obtained for the simulated context-aware and baseline systems. This difference showed a significance value of 0.029 in the two-tailed *t*-test.

5.3 Users satisfaction and emotional behavior

Finally, Table 6 shows the average results obtained with respect to the subjective evaluation carried out by the recruited users. As can be observed, the three systems correctly understand the different user queries and obtain a similar evaluation regarding the user observed easiness in correcting errors made by the ASR module. However, the real context-aware system has a higher evaluation rate regarding the user observed easiness in obtaining the data required to fulfill the complete set of objectives defined in the scenarios, as well as the suitability of the interaction rate during the dialog. Ratings of satisfaction, ease of use, trust, and desire to continue using the system were also improved by the real context-aware system. Together, these results indicate that the conversational agent represents a viable and promising medium for helping patients with the described diseases.

The following main conclusions can also be extracted from the analysis of the results obtained for the different questions and systems. With regard to questions Q1 and Q2 (users understanding system responses and system understanding users responses), the analysis of the results showed that there were not significant differences between the three systems. This might be because the three systems integrated the same ASR, NLU and TTS modules. A similar conclusion can be extracted from the analysis of the facility of correcting errors (question Q5), for which there is an agreement between the results obtained for this subjective assessment and the statistical analysis obtained for the error correction rates (Sect. 5.1). In both cases, this relationship was not significant in the *t*-test.

Regarding the ease of obtaining information (question Q3) and the adequacy of the interaction rate (question Q4), the real context-aware system improves the results obtained with the simulated context-aware and the baseline systems. These results agree with the reduction in the average number of dialog turns and highest percentage of goal-directed actions achieved by the real context-aware system.

The same conclusion can be extracted from the analysis of the users' perception about the credibility and concern transmitted by the system (questions Q6 and Q7). Users also significantly prefer to continue working with the real context-aware, which obtained the highest mean and lowest standard deviation for question 8. In our opinion, this can be explained by the user's adaptation achieved by the introduction of our proposal in the real context-aware system. With regard the simulated context-aware system, a study of the relationship between the subjective and objective evaluations showed that there was a set of dialogs in which users provided a negative opinion about the system due to the activation of the system-directed initiative and the use of system apologies when they were not required. For this reason, the simulated context-aware system achieves the highest standard deviation for these questions in the subjective assessment.

Table 5 Proportions of dialog spent on-goal directed actions, ground actions and other possible actions

	Baseline	Simulated context-aware system	Real context-aware system
Goal directed actions	66.85 %	68.27 %	74.56 %
Grounding actions	31.97 %	30.42 %	24.38 %
Rest of actions	1.18 %	1.31 %	1.06 %

6 Conclusions

The use of conversational agents for Ambient Assisted Living in hand-held mobile devices implies very important challenges to handle information about the user's activity and dynamically control and adapt the system behavior accordingly. In this paper, we contribute a framework which

Table 6 Results of the subjective evaluation with real users (for the mean value M: 1 = worst, 5 = best evaluation)

	Baseline	Simulated context-aware system	Real context-aware system
Q1	M = 4.62, SD = 0.37	M = 4.54, SD = 0.42	M = 4.82, SD = 0.34
Q2	M = 3.65, SD = 0.24	M = 3.82, SD = 0.36	M = 3.93, SD = 0.27
Q3	M = 3.84, SD = 0.56	M = 3.91, SD = 0.77	M = 4.36, SD = 0.34
Q4	M = 3.43, SD = 0.28	M = 3.54, SD = 0.39	M = 4.24, SD = 0.29
Q5	M = 3.27, SD = 0.59	M = 3.43, SD = 0.36	M = 3.34, SD = 0.57
Q6	M = 3.71, SD = 0.41	M = 3.64, SD = 0.91	M = 4.30, SD = 0.35
Q7	M = 4.22, SD = 0.46	M = 4.05, SD = 1.03	M = 4.51, SD = 0.26
Q8	M = 3.80, SD = 0.42	M = 3.93, SD = 0.65	M = 4.47, SD = 0.36

can be used to develop context-aware conversational agents that can be easily integrated in hand-held mobile devices, facilitating permanent and personalized access to healthcare services. The framework proposed is comprised of an architecture in which different systems and modules cooperate to provide adapted responses.

A method for predicting user states in conversational agents has been incorporated to the architecture of the system. These states are defined as the combination of the user emotional state and the predicted intention according to their objective in the dialog.

The proposed method is implemented as a module comprised of an emotion recognizer and an intention recognizer. The emotion recognizer obtains the user emotional state from the acoustics of his utterance as well as the dialog history. The intention recognizer decides the next user action and their dialog goal using a statistical approach that relies on the previous user inputs and system prompts.

To store and share context information we have defined a data structure that manages user profiles. These profiles include not only information external to the users, such as their location, but also specific information about their pathologies (e.g., primary medications and prescribed doses, changes in the level of physiological parameters) and their needs and preferences using the system. Some of these parameters are automatically extracted from previous interactions.

We have provided a complete implementation of our architecture in a system that provides personalized context-aware services for patients suffering from chronic pulmonary diseases. To develop this system we have defined the complete requirements for the task and developed the different modules and the necessary information to be incorporated in the user profile. From a set of dialogs acquired with recruited users, we have studied the influence of context information on the quality of the services that are provided by the system.

The results show that context information not only allows a higher success rate in the provision of the adapted services, but using context information, the time required to provide the information can be reduced. In addition, the quality of the interaction between the user and the system is increased, as context-aware dialogs present a better ratio of goal-directed actions selected by the system to successfully provide the different services. This way, actions that might discourage users (e.g., confirmations or re-request of information) are reduced.

As future work, we want to carry out a detailed study with a large number of patients in a continuous use of the system during several months. We are also interested in extending and evaluating our proposal considering additional features for the intention and emotion recognizers. With regard to emotion recognition, an interesting extension of the described proposal is based on its combination with sentiment analysis approaches analyzing the text transcription hypothesis provided by the ASR module [109, 110]. Regarding the intention recognizer, we want to also consider the communication and misunderstanding errors previously detected as an additional feature for each user to avoid them in future interactions.

Acknowledgements This work was supported in part by Projects MINECO TEC2012-37832-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485).

References

1. Acosta J, Ward N (2009) Responding to user emotional state by adding emotional coloring to utterances. In: Proc Interspeech’09, pp 1587–1590
2. Ahmad F, Hogg-Johnson S, Stewart D, Skinner H, Glazier R, Levinson W (2009) Computer-assisted screening for intimate partner violence and control: a randomized trial. Ann Intern Med 151(2):93–102
3. Ai H, Raux A, Bohus D, Eskenazi M, Litman D (2007) Comparing spoken dialog corpora collected with recruited subjects versus real users. In: Proc SIGdial’07, pp 124–131
4. Alamudun F, Choi J, Khan H, Ahmed B, Gutierrez-Osuna R (2012) Removal of subject-dependent and activity-dependent variation in physiological measures of stress. In: Proc Pervasive-Health’12
5. Allen J, Ferguson G, Blaylock N, Byron D, Chambers N, Dzikovska M, Galescu L, Swift M (2006) Chester: towards a personal medication advisor. J Biomed Inform 39(5):500–513
6. Andre E, Bevacqua E, Heylen D, Niewiadomski R, Pelachaud C, Peters C, Poggi I, Rehm M (2011) Non-verbal persuasion and communication in an affective agent. In: Emotion oriented systems. The humaine handbook. Cognitive technologies. Springer, Berlin, pp 585–608
7. Antoniu S (2006) Outcomes of adult domiciliary oxygen therapy in pulmonary diseases. Expert Rev Pharmacoecon Outcomes Res 6(1):9–66
8. Araki M, Watanabe T, Doshita S (1997) Evaluating dialogue strategies for recovering from misunderstandings. In: Proc IJCAI workshop on collaboration cooperation and conflict in dialogue systems, pp 13–18
9. Augusto J, Huch M, Kameas A, Maitland J, McCullagh P, Roberts J, Sixsmith A, Wichert R (2012) Handbook of ambient assisted living. IOS Press, Amsterdam
10. Ayadi ME, Kamel M, Karray F (2011) Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recognit 44:572–587
11. Banse R, Scherer K (1996) Acoustic profiles in vocal emotion expression. J Pers Soc Psychol 70(3):614–636
12. Bardhan I, Thouin M (2013) Health information technology and its impact on the quality and cost of healthcare delivery. Decis Support Syst 55(2):438–449
13. Basole R, Bodner D, Rouse W (2013) Healthcare management through organizational simulation. Decis Support Syst 55(2):552–563
14. Batliner A, Huber R, Niemann H, Noth E, Spilker J, Fischer K (2000) The recognition of emotion. In: Verbmobil: foundations of speech-to-speech translation. Springer, Berlin, pp 122–130
15. Batliner A, Steidl S, Schuller B, Seppi D, Vogt T, Wagner J, Devillers L, Vidrascu L, Aharonson V, Kessous L, Amir N (2011) Whodunnit: searching for the most important feature types signalling emotion-related user states in speech. Comput Speech Lang 25(1):4–28

16. Bee N, Wagner J, André E, Charles F, Pizzi D, Cavazza M (2010) Multimodal interaction with a virtual character in interactive storytelling. In: Proc AAMAS'10, pp 1535–1536
17. Berkovsky S, Coombe M, Freyne J, Bhandari D, Baghaei N (2010) Physical activity motivating games: virtual rewards for real activity. In: Proc CHI'10, pp 243–252
18. Bevacqua E, Mancini M, Pelachaud C (2008) A listening agent exhibiting variable behaviour. Lect Notes Comput Sci 5208:262–269
19. Bickmore T, Giorgino T (2006) Health dialog systems for patients and consumers. J Biomed Inform 39(5):556–571
20. Bickmore T, Caruso L, Clough-Gorrb K, Heeren T (2005) It's just like you talk to a friend' relational agents for older adults. Interact Comput 17:711–735
21. Bickmore T, Mitchell S, Jack B, Paasche-Orlow M, Pfeifer L, O'Donnell J (2010) Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 22:289–298
22. Bickmore T, Puskar K, Schlenk E, Pfeifer L, Sereika S (2010) Maintaining reality: relational agents for antipsychotic medication adherence. Interact Comput 22:276–288
23. Black LA, McTear MF, Black ND, Harper R, Lemon M (2005) Appraisal of a conversational artefact and its utility in remote patient monitoring. In: Proc CBMS'05, pp 506–508
24. Boehner K, DePaula R, Dourish P, Sengers P (2007) How emotion is made and measured. Int J Hum-Comput Stud 65:275–291
25. Bonino D, Corno F (2011) What would you ask to your home if it were intelligent? Exploring user expectations about next-generation homes. J Ambient Intell Smart Environ 3(2):111–116
26. Bos J, Klein E, Lemon O, Oka T (2003) DIPPER: description and formalisation of an information-state update dialogue system architecture. In: Proc SIGdial'03, pp 115–124
27. Bretier P, Sadek MD (1996) A rational agent as the kernel of a cooperative spoken dialogue system: implementing a logical theory of interaction. In: Proc ATAL'96, pp 189–203
28. Bui T, Poel M, Nijholt A, Zwiers J (2009) A tractable hybrid DDN-POMDP approach to affective dialogue modeling for probabilistic frame-based dialogue systems. Nat Lang Eng 15(2):273–307
29. Bunt H, Alexandersson J, Carletta J, Choe J, Fang A, Hasida K, Lee K, Petukhova V, Popescu-Belis A, Romary L, Soria C, Traum D (2010) Towards an ISO standard for dialogue act annotation. In: Proc LREC'10, pp 2548–2555
30. Burkhardt F, van Ballegooy M, Engelbrecht K, Polzehl T, Stegmann J (2009) Emotion detection in dialog systems—usecases, strategies and challenges. In: Proc ACII'09, pp 1–6
31. Calle J, Castano L, Castro E, Cuadra D (2013) Statistical user model supported by R-Tree structure. J Appl Intell 39(3):545–563
32. Callejas Z (2008) On the development of adaptive and portable spoken dialogue systems: emotion recognition, language adaptation and field evaluation. PhD thesis, University of Granada, Spain
33. Callejas Z, López-Cózar R (2008) Influenc of contextual information in emotion annotation for spoken dialogue systems. Speech Commun 50(5):416–433
34. Cassell J (2000) More than just another pretty face: embodied conversational interface agents. Commun ACM 43(4):70–78
35. Cavazza M, de la Cámara RS, Turunen M (2010) How was your day? A companion ECA. In: Proc AAMAS'10, Toronto, Canada, pp 1629–1630
36. Chen CM, Liu CY (2009) Personalized e-news monitoring agent system for tracking user-interested Chinese new events. J Appl Intell 30(2):121–141
37. Chen L, Mao X, Wei P, Xue Y, Ishizuka M (2012) Mandarin emotion recognition combining acoustic and emotional point information. J Appl Intell 37(4):602–612
38. Chen Z, Lin M, Chen F, Wang R, Li T, Campbell A (2013) Unobtrusive sleep monitoring using smartphones. In: Proc PervasiveHealth'13
39. Chittaro L, Zuliani F (2013) Exploring audio storytelling in mobile exergames to change the perception of physical exercise. In: Proc PervasiveHealth'12
40. Chung G (2004) Developing a flexible spoken dialog system using simulation. In: Proc ACL'04, pp 63–70
41. Coronato A, Pietro GD (2010) Pervasive and smart technologies for healthcare: ubiquitous methodologies and tools. Medical Information Science Reference
42. Cuayahuitl H, Renals S, Lemon O, Shimodaira H (2005) Human-computer dialogue simulation using hidden Markov models. In: Proc ASRU'05, pp 290–295
43. Davoodi E, Kianmehr K, Afsharchi M (2013) A semantic social network-based expert recommender system. J Appl Intell 39(1):1–13
44. Delichatsios H, Friedman R, Glanz K, Tennstedt S, Smigelski C, Pinto B (2000) Randomized trial of a talking computer to improve adults eating habits. Am J Heal Promot 15:215–224
45. Delmastro F (2012) Pervasive communications in healthcare. Comput Commun 35:1284–1295
46. Dols F, van der Sloot K (1992) Modelling mutual effects in belief-based interactive systems. In: Proc 3rd int workshop on user modeling, pp 3–19
47. Dourlens S, Ramdane-Cherif A, Monacelli E (2013) Multi levels semantic architecture for multimodal interaction. J Appl Intell 38(4):586–599
48. Eckert W, Levin E, Pieraccini R (1997) User modeling for spoken dialogue system evaluation. In: Proc ASRU'97, pp 80–87
49. Evanini K, Hunter P, Liscombe J, Suendermann D, Dayanidhi K, Pieraccini R (2008) Caller experience: a method for evaluating dialog systems and its automatic prediction. In: Proc SLT'08, pp 129–132
50. Eyrharabide V, Amandi A (2012) Ontology-based user profil learning. J Appl Intell 36(4):857–869
51. Farzanfar R, Frishkopf S, Migneault J, Friedman R (2005) Telephone-linked care for physical activity: a qualitative evaluation of the use patterns of an information technology program for patients. J Biomed Inform 38:220–228
52. Felfernig A, Friedrich G, Isak K, Shchekotykhin K, Teppan E, Jannach D (2009) Automated debugging of recommender user interface descriptions. J Appl Intell 31:1–14
53. Filisko E, Seneff S (2005) Developing city name acquisition strategies in spoken dialogue systems via user simulation. In: Proc SIGdial'05, pp 144–155
54. Georgila K, Henderson J, Lemon O (2005) Learning user simulations for information state update dialogue systems. In: Proc Eurospeech'05, pp 893–896
55. Ghanem K, Hutton H, Zenilman J, Zimba R, Erbelding E (2005) Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. Sex Transm Infect 81(5):421–425
56. Giorgino T, Azzini I, Rognoni C, Quaglini S, Stefanelli M, Gretter R, Falavigna D (2004) Automated spoken dialogue system for hypertensive patient home management. Int J Med Inform 74:159–167
57. Glanz K, Shigaki D, Farzanfar R, Pinto B, Kaplan B, Friedman R (2003) Participant reactions to a computerized telephone system for nutrition and exercise counseling. Patient Educ Couns 49:157–163
58. Gnjatovic M, Janev M, Delic V (2012) Modeling attentional information in task-oriented human-machine. J Appl Intell 37(3):305–320
59. González-Rodríguez M, Manrubia J, Vidau A, González-Gallego M (2009) Improving accessibility with user-tailored interfaces. J Appl Intell 30(1):65–71

60. González-Vélez H, Mier M, Juliá-Sapé M, Arvanitis T, García-Gómez J, Robles M, Lewis P, Dasmahapatra S, Dupplaw D, Peet A, Arús C, Celda B, Van-Huffel S, Lluch-Ariet M (2009) HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. J Appl Intell 30(3):191–202
61. Griol D, Hurtado L, Sanchis E, Segarra E (2007) Acquiring and evaluating a dialog corpus through a dialog simulation technique. In: Proc SIGdial'07, pp 29–42
62. Griol D, Hurtado L, Segarra E, Sanchis E (2008) A statistical approach to spoken dialog systems design and evaluation. Speech Commun 50(8–9):666–682
63. Griol D, Sánchez-Pi N, Carbó J, Molina J (2010) An architecture to provide context-aware services by means of conversational agents. Adv Intell Soft Comput 79:275–282
64. Griol D, Molina J, Callejas Z (2012) Bringing together commercial and academic perspectives for the development of intelligent AmI interfaces. J Ambient Intell Smart Environ 4(3):183–207
65. Grosz B, Sidner C (1986) Attention, intentions and the structure of discourse. Comput Linguist 12(3):175–204
66. Guan L, Xie Z (2013) Multimodal information fusion of audio emotion recognition based on kernel entropy component analysis. Int J Semant Comput 7(1):25–42
67. Hansen J (1996) Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. Speech Commun 20(2):151–170
68. Hozjan V, Kacic Z (2003) Context-independent multilingual emotion recognition from speech signal. Int J Speech Technol 6:311–320
69. Huang A, Yen D, Zhang X (2008) Exploring the potential effects of emoticons. Inf Manag 45(7):466–473
70. Hubal R, Day R (2006) Informed consent procedures: an experimental test using a virtual character in a dialog systems training application. J Biomed Inform 39:532–540
71. Isern D, Moreno A, Sánchez D, Hajnal A, Pedone G, Varga L (2011) Agent-based execution of personalised home care treatments. J Appl Intell 34(2):155–180
72. Jindal S (2008) Oxygen therapy: important considerations. Indian J Chest Dis Allied Sci 50(1):97–107
73. Jokinen K (2003) Natural interaction in spoken dialogue systems. In: Proc workshop ontologies and multilinguality in user interfaces, pp 730–734
74. Jung S, Lee C, Kim K, Lee D, Lee G (2011) Hybrid user intention modeling to diversify dialog simulations. Comput Speech Lang 25(2):307–326
75. Kang H, Suh E, Yoo K (2008) Packet-based context aware system to determine information system user's context. Expert Syst Appl 35:286–300
76. Karan O, Bayraktar C, Gümüskayab H, Karlik B (2012) Diagnosing diabetes using neural networks on small mobile devices. Expert Syst Appl 39(1):54–60
77. Kim HR, Chan P (2008) Learning implicit user interest hierarchy for context in personalization. J Appl Intell 28(2):153–166
78. Kirkpatrick L (2012) Creating a dynamic speech dialogue: how to implement dialogue initiatives and question selection strategies with VoiceXML agents. Wadsworth, Belmont
79. Ko J, Murase F, Mitamura T, Nyberg E, Tateishi M, Akahori I (2006) Context-aware dialog strategies for multimodal mobile dialog systems. In: Proc of AAAI int workshop on modeling and retrieval of context, pp 7–12
80. Lee A, Tang S, Yu G, Cheung R (2008) The smiley as a simple screening tool for depression after stroke: a preliminary study. Int J Nurs Stud 45(7):1081–1089
81. Lee C, Narayanan S (2005) Toward detecting emotions in spoken dialogs. IEEE Trans Speech Audio Process 13(2):293–303
82. Lee H, Lee S, Ha K, Jang H, Chung W, Kim J, Chang Y, Yoo D (2009) Ubiquitous healthcare service using Zigbee and mobile phone for elderly patients. Int J Med Inform 78(3):193–198
83. Leite I, Pereira A, Castellano G, Mascarenhas S, Martinho C, Paiva A (2012) Modelling empathy in social robotic companions. Adv User Model 7138:135–147
84. Levin E, Pieraccini R, Eckert W (2000) A stochastic model of human-machine interaction for learning dialog strategies. IEEE Trans Speech Audio Process 8(1):11–23
85. Li S, Wrede B (2007) Why and how to model multi-modal interaction for a mobile robot companion. In: Proc AAAI spring symposium 2007 on interaction challenges for intelligent assistants, pp 72–79
86. López-Cózar R, Araki M (2005) Spoken, multilingual and multimodal dialogue systems: development and assessment. Wiley, New York
87. López-Cózar R, de la Torre A, Segura J, Rubio A (2003) Assessment of dialogue systems by means of a new simulation technique. Speech Commun 40:387–407
88. Maglogiannis I, Zafiropoulo E, Anagnostopoulos I (2009) An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers J Appl Intell 30(1):24–36
89. Malatesta L, Raouzaiou A, Karpouzis K, Kollias S (2009) Towards modeling embodied conversational agent character profile using appraisal theory predictions in expression synthesis. J Appl Intell 30(1):58–64
90. Marreiros G, Santos R, Ramos C, Neves J (2010) Context-aware emotion-based model for group decision making. IEEE Intell Syst 25(2):31–39
91. Martin A, Jones J, Gilbert J (2013) A spoonful of sugar: understanding the over-the-counter medication needs and practices of older adults. In: Proc PervasiveHealth'13
92. Matic A, Osmani V, Maxhuni A, Mayora O (2012) Multimodal mobile sensing of social interactions. In: Proc PervasiveHealth'12
93. McGee-Lennon M, Smeaton A, Brewste S (2012) Designing home care reminder systems personalisable: Lessons learned through co-design with older users. In: Proc PervasiveHealth'12
94. McTear MF (2004) Spoken dialogue technology: towards the conversational user interface. Springer, Berlin
95. Metallinou A, Lee S, Narayanan S (2008) Audio-visual emotion recognition using gaussian mixture models for face and voice. In: Proc 10th IEEE int symposium on multimedia, pp 250–257
96. Miesenberger K, Klaus J, Zagler W, Karshmer A (2010) Computers helping people with special needs. In: Proc ICCHP 2010. Lecture Notes on Computer Science, vol 4061. Springer, Berlin
97. Migneault JP, Farzanfar R, Wright J, Friedman R (2006) How to write health dialog for a talking computer. J Biomed Inform 39(5):276–288
98. Mihailidis A, Bardram J (2007) Pervasive computing in healthcare CRC Press, Boca Raton
99. Miller A, Pater J, Mynatt E (2013) Design strategies for youth-focused pervasive social health games. In: Proc PervasiveHealth'13
100. Mohammad Y, Nishida T (2010) Using physiological signals to detect natural interactive behavior. J Appl Intell 33(1):79–92
101. Montani S (2008) Exploring new roles for case-based reasoning in heterogeneous AI systems for medical decision support. J Appl Intell 28(3):275–285
102. Mooney K, Beck S, Dudley W, Farzanfar R, Friedman R (2004) A computer-based telecommunication system to improve symptom care for women with breast cancer. Ann Behav Med Annu Meet Supplement(27):152–161
103. Moore R (1977) Reasoning about knowledge and action. In: Proc IJCAI'77, pp 223–227
104. Morrison D, Wang R, DeSilva L (2007) Ensemble methods for spoken emotion recognition in call-centres. Speech Commun 49(2):98–112

105. Moubaiddin A, Obeid N (2009) Partial information basis for agent-based collaborative dialogue. J Appl Intell 30(2):142–167
106. Mukhopadhyay S, Postolache O (2013) Pervasive and mobile sensing and computing for healthcare: technological and social issues. Springer, Berlin
107. Munson S, Consolvo S (2012) Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In: Proc PervasiveHealth'12
108. Möller S, Englert R, Engelbrecht K, Hafner V, Jameson A, Oulasvirta A, Raake A, Reithinger N (2006) MeMo: towards automatic usability evaluation of spoken dialogue services by user error simulations. In: Proc Interspeech'06, pp 1786–1789
109. Nasukawa T, Yi J (2003) Sentiment analysis: capturing favorability using natural language processing. In: Proc 2nd int conference on knowledge capture
110. Neviarouskaya A, Prendinger H, Ishizuka M (2010) EmoHeart: conveying emotions in second life based on affect sensing from text. Adv Hum-Comput Interact 1(1):1–13
111. O'Connor G, Arnold J (1973) Intonation in colloquial English. Longman, Harlow
112. Ohkawa Y, Suzuki M, Ogasawara H, Ito A, Makino S (2009) A speaker adaptation method for non-native speech using learners' native utterances for computer-assisted language learning systems. Speech Commun 51(10):875–882
113. O'Shea K (2012) An approach to conversational agent design using semantic sentence similarity. J Appl Intell 37(4):558–568
114. Paek T, Pieraccini R (2008) Automating spoken dialogue management design using machine learning: an industry perspective. Speech Commun 50:716–729
115. Patel R, Hartzler A, Pratt W, Back A (2013) Visual feedback on nonverbal communication: a design exploration with healthcare professionals. In: Proc PervasiveHealth'13
116. Payr S (2010) Closing and closure in human-companion interactions: analyzing video data from a field study. In: Proc IEEE RO-MAN'10, pp 476–481
117. Pfeifer L, Bickmore T (2010) Designing embodied conversational agents to conduct longitudinal health interviews. In: Proc intelligent virtual agents'10, pp 4698–4703
118. Pieraccini R (2012) The voice in the machine: building computers that understand speech. MIT Press, Cambridge
119. Pietquin O (2004) A framework for unsupervised learning of dialogue strategies. PhD thesis, Faculte Polytechnique de Mons
120. Pinto B, Friedman R, Marcus B, Kelley H, Tennstedt S, Gillman M (2002) Effects of a computer-based, telephone-counseling system on physical activity. Am J Prev Med 23:113–120
121. Pittermann J, Pittermann A, Minker W (2010) Emotion recognition and adaptation in spoken dialogue systems. Int J Speech Technol 13:49–60
122. Ptaszynski M, Dybala P, Shi W, Rzepka R, Araki K (2009) Towards context aware emotional intelligence in machines: computing contextual appropriateness of affective states. In: Proc IJCAI'09
123. Ptaszynski M, Maciejewski J, Dybala P, Rzepka R, Araki K (2010) CAO: a fully automatic emoticon analysis system based on theory of kinesics. IEEE Trans Affect Comput 1(1):46–59
124. Ramelson H, Friedman R, Ockene J (1999) An automated telephone-based smoking cessation education and counseling system. Patient Educ Couns 36:131–143
125. Rehrl T, Geiger J, Golcar M, Gentsch S, Knobloch J, Rigoll G, Scheibl K, Schneider W, Ihsen S, Wallhoff F (2013) The robot ALIAS as a database for health monitoring for elderly people. In: Proc AAL'13, pp 414–423
126. Reiss A, Stricker D (2013) Towards robust activity recognition for everyday life: Methods and evaluation. In: Proc PervasiveHealth'12
127. Rojas-Barahona L (2009) Health care dialogue systems: practical and theoretical approaches to dialogue management. PhD thesis, Universita degli Studi di Pavia
128. Rojas-Barahona L, Giorgino T (2009) Adaptable dialog architecture and runtime engine (AdaRTE): a framework for rapid prototyping of health dialog systems. Int J Med Inform 785:56–68
129. Rouillard J (2007) Web services and speech-based applications around VoiceXML. J Netw 2(1):27–35
130. Santos R, Marreiros G, Ramos C, Neves J, Bulas-Cruz J (2011) Personality, emotion, and mood in agent-based group decision making. IEEE Intell Syst 26(6):58–66
131. Saz O, Yin SC, Lleida E, Rose R, Vaquero C, Rodríguez WR (2009) Tools and technologies for computer-aided speech and language therapy. Speech Commun 51(10):948–967
132. Schatzmann J, Georgila K, Young S (2005) Quantitative evaluation of user simulation techniques for spoken dialogue systems. In: Proc SIGdial'05, pp 45–54
133. Schatzmann J, Weilhammer K, Stuttle M, Young S (2006) A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. Knowl Eng Rev 21(2):97–126
134. Schatzmann J, Thomson B, Weilhammer K, Ye H, Young S (2007) Agenda-based user simulation for bootstrapping a POMDP dialogue system. In: Proc HLT/NAACL'07, pp 149–152
135. Schatzmann J, Thomson B, Young S (2007) Statistical user simulation with a hidden agenda. In: Proc SIGdial'07, pp 273–282
136. Scheffer K, Young S (2001) Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning. In: Proc HLT'02, pp 12–18
137. Schubiger M (1958) English intonation: its form and function. Niemeyer Verlag, Tübingen
138. Schuller B, Batliner A, Steidl S, Seppi D (2011) Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge. Speech Commun 53(9–10):1062–1087
139. Searle J (1969) Speech acts. An essay on the philosophy of language. Cambridge University Press, Cambridge
140. Seneff S, Adler M, Glass J, Sherry B, Hazen T, Wang C, Wu T (2007) Exploiting context information in spoken dialogue interaction with mobile devices. In: Proc IMUx'07, pp 1–11
141. Shaban-Nejad A, Riazanov A, Charland K, Rose G, Baker C, Tamblyn R, Forster A, Buckeridge D (2012) HAIKU: a semantic framework for surveillance of healthcare-associated infections. Proc Comput Sci 10:1073–1079
142. Shah N, Ragaswamy H, Govindugari K, Estanol L (2012) Performance of three new-generation pulse oximeters during motion and low perfusion in volunteers. J Clin Anesth 24(5):385–391
143. Shi W, Wang X, Zhao X, Prakash V, Gnawali O (2013) Computerized-eyewear based face recognition system for improving social lives of prosopagnostics. In: Proc PervasiveHealth'13
144. Shie BE, Yu P, Tseng V (2013) Mining interesting user behavior patterns in mobile commerce environments. J Appl Intell 38(3):418–435
145. Sixsmith A, Meuller S, Lull F, Klein M, Bierhoff I, Delaney S, Savage R (2009) SOPRANO—an ambient assisted living system for supporting older people at home. In: Proc ICOST'09, pp 233–236
146. Tartarisco G, Baldus G, Corda D, Raso R, Arnao A, Ferro M, Gaggioli A, Pioggia G (2012) Personal health system architecture for stress monitoring and support to clinical decisions. Comput Commun 35(11):1296–1305
147. Toscos T, Conelly K, Rogers Y (2013) Designing for positive health affect: Decoupling negative emotion and health monitoring technologies. In: Proc PervasiveHealth'13

148. Traum D (1999) Speech acts for dialogue agents. In: Foundations of rational agency. Kluwer Academic, Norwell, pp 169–201
149. Treur J (2011) A virtual human agent model with behaviour based on feeling exhaustion. J Appl Intell 35(3):469–482
150. Ververidis D, Kotropoulos C (2006) Emotional speech recognition: resources, features and methods. Speech Commun 48:1162–1181
151. Vien N, Ertel W, Dang VH, Chung T (2013) Monte-Carlo tree search for Bayesian reinforcement learning. J Appl Intell 39(2):345–353
152. Wahlster W (2006) Dialogue systems go multimodal: the SmartKom experience. In: SmartKom: foundations of multimodal dialogue systems cognitive technologies. Springer, Berlin, pp 3–27
153. Wahlster W, Reithinger N, Blocher A (2001) Smartkom: towards multimodal dialogues with anthropomorphic interface agents. In: Proc status conference: lead projects human-computer interaction, pp 22–34
154. Wahlster W (2006) SmartKom: foundations of multimodal dialogue systems. Springer, Berlin
155. van der Wal C, Kowalczyk W (2013) Detecting changing emotions in human speech by machine and humans. J Appl Intell 39(4):675–691
156. Walther J, D'Addario K (2001) The impacts of emoticons on message interpretation in computer-mediated communication. Soc Sci Comput Rev 19(3):324–347
157. Wang Y, Guan L, Venetsanopoulos AN (2011) Audiovisual emotion recognition via cross-modal association in kernel space. In: Proc ICME'11, pp 1–6
158. Watanabe T, Araki M, Doshita S (1998) Evaluating dialogue strategies under communication errors using computer-to-computer simulation. IEICE Trans Inf Syst E81-D(9):1025–1033
159. Wilks Y, Catizone R, Worgan S, Turunen M (2011) Some background on dialogue management and conversational speech for dialogue systems. Comput Speech Lang 25(2):128–139
160. Witten I, Frank E (2005) Data mining: practical machine learning tools and techniques. Morgan Kaufmann, San Mateo
161. Wolters M, Georgila K, Moore J, Logie R, MacPherson S (2009) Reducing working memory load in spoken dialogue systems. Interact Comput 21(4):276–287
162. Wu I, Li J, Fu C (2011) The adoption of mobile healthcare by hospital's professionals: an integrative perspective. Decis Support Syst 51:587–596
163. You M, Chen C, Bu J, Liu J, Tao J (1997) Getting started with susas: a speech under simulated and actual stress database. In: Proc Eurospeech'97, vol 4, pp 1743–1746
164. Young S (2011) Cognitive user interfaces. IEEE Signal Process Mag 27(3):128–140
165. Young S, Schatzmann J, Weilhammer K, Ye H (2007) The hidden information state approach to dialogue management. In: Proc ICASSP'07, pp 149–152
166. Yuan B, Herbert J (2012) Fuzzy CARA—a fuzzy-based context reasoning system for pervasive healthcare. Proc Comput Sci 10:357–365
167. Zeng Z, Hu Y, Roisman G, Wen Z, Fu Y, Huang T (2007) Audio-visual spontaneous emotion recognition. Lect Notes Comput Sci 4451:72–90
168. Zukerman I, Litman D (2001) Natural language processing and user modeling: synergies and limitations. User Model User-Adapt Interact 11:129–158
169. Ábalos N, Espejo G, López-Cózar R, Callejas Z, Griol D (2010) A multimodal dialogue system for an ambient intelligent application in home environments. Lect Notes Artif Intell 6231:484–491