Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Biomedical Informatics xxx (2006) xxx–xxx
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Methodological Review

Health dialog systems for patients and consumers ☆

Timothy Bickmore [a,*], Toni Giorgino [b]

[a] College of Computer and Information Science, Northeastern University, Boston, MA, USA
[b] Laboratory for Medical Informatics, Dipartimento di Informatica e Sistemistica, Università di Pavia, Pavia, Italy

Received 29 July 2005

Abstract

There is a growing need for automated systems that can interview patients and consumers about their health and provide health education and behavior change interventions using natural language dialog. A number of these health dialog systems have been developed over the last two decades, many of which have been formally evaluated in clinical trials and shown to be effective. This article provides an overview of the theories, technologies and methodologies that are used in the construction and evaluation of these systems, along with a description of many of the systems developed and tested to date. The strengths and weaknesses of these approaches are also discussed, and the needs for future work in the field are delineated.

Keywords: Dialog system; Behavioral informatics; Consumer informatics; Natural language processing

1. Introduction

One-on-one, face-to-face interaction with a health provider is widely acknowledged to be the "gold standard" for providing health education to and affecting health behavior change in patients and consumers. Automated health dialog systems—especially those which use speech and other audiovisual media—emulate this form of interaction to communicate health information to users in a format that is natural, intuitive and dynamically tailored.

A significant amount of research has been conducted over the last two decades into the automatic generation of printed materials, web pages and other static media for the purpose of providing health communication to patients and consumers. However, although these approaches have been found to be effective [1], they still fall short of the "gold standard" in several ways. For example, in static media, information cannot be rephrased if the clients do not understand it, clients cannot ask clarifying questions, and they cannot request more or less information on specific topics of interest. In addition, while many studies have demonstrated the efficacy of tailoring print or web materials based on initial characteristics of the user [2], dialog systems can allow messages to be tailored at a very fine-grained level, with each sentence of delivered information synthesized on the basis on the inferred goals and beliefs of the user at a particular moment in time, and incorporating everything that has previously been said in the conversation. When used in conjunction with speech and possibly other nonverbal conversational modalities (such as hand gesture or facial display), dialog also provides a medium through which a significant amount of information can be conveyed in addition to the linguistic content, including emphasis, affect, and attitude. For these reasons, simulated face-to-face conversation may also be an especially effective communication channel to use with individuals who have low reading or functional health literacy.

☆ Submitted to the Journal of Biomedical Informatics special issue on Dialog Systems for Health Communication.

* Corresponding author. Fax: +1 617 812 2589.
E-mail address: bickmore@ccs.neu.edu (T. Bickmore).

1532-0464/$ - see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.jbi.2005.12.004

In some ways, health dialog systems may even be better than interacting with a human provider. One problem with in-person encounters with health professionals is that all providers function in health care environments in which they can only spend a very limited amount of time with each patient [3]. Time pressures can result in patients feeling too intimidated to ask questions, or to ask that information be repeated. Another problem is that of "fidelity": providers do not always perform in perfect accordance with recommended guidelines, resulting in significant inter-provider and intra-provider variations in the delivery of health information. Finally, many people simply do not have access to the all of the health professionals they need, due to financial or scheduling constraints. Even if health dialog systems have lower efficacy than one-on-one counseling, they have the potential to reach a much greater portion of the population, resulting in greater "impact" (efficacy multiplied by reach [4]).

In addition to emulating face-to-face interaction with a health professional, dialog system technology can be used in a number of other ways to provide patients and consumers with health information. For example, real-time speech-based machine translation systems can enable a health professional to assist a patient who speaks a different language [5]. Computer games in which consumers can converse with non-player characters in natural language can be used to affect health behavior change through role playing and dialog with peer characters [6]. Thus, to be as inclusive as possible, we define health dialog systems to be those automated systems whose primary goal is to provide health communication with patients or consumers primarily using natural language dialog. While such systems can be used for a very wide range of applications—including the promotion of patient disease self-management, disease monitoring, and screening—we will focus on patient education and health behavior change applications in this paper, as these have received the most research attention to date.

The field of health dialog systems lies at the intersection of two much larger disciplines—computational linguistics (specifically work on dialog systems) and medical informatics (specifically in the area of consumer informatics). Although this intersection is still fairly small in terms of the number of active researchers and the number of systems built and deployed, it has a long history and represents a rapidly growing field. In 2004, an initial workshop was held on this topic as part of the American Association for Artificial Intelligence's Fall Symposium Series [7], and a follow-on workshop will be held in 2006, focusing specifically on automated argumentation systems for health communication [8].

This article begins with a brief review of dialog system theory followed by a discussion of what makes health dialog different from other dialog system application domains. Reviews of dialog system technologies and deployment technologies are then presented, followed by discussions of development and evaluation methodologies. Finally, a brief review is given of the efficacy of the systems fielded to date followed by a discussion of some promising areas of future research.

2. Basic concepts in dialog system theory

Linguists have traditionally decomposed the problem of understanding and generating natural language utterances into several layers of analysis (see Fig. 1) [9]. Phonetic analysis structures sequences of phonemes (the smallest units of sound) together into morphemes (roots, prefixes and suffixes). Morphology structures sequences of morphemes into words. Syntax structures sequences of words into clauses and then into sentences or utterances (when spoken). Semantics is concerned with the meaning of sentences, independent of their context of use: how words, phrases and clauses relate to the world, and how the meanings of these constituents can be combined to form the meaning of an entire utterance. Pragmatics is concerned with those elements of utterance meaning that are context-dependent, and with how language is used by people to achieve their goals.

The study of discourse and dialog falls within the realm of pragmatics. Discourse is the extended use of language to convey desires, beliefs and intentions. The pragmatics of discourse is the study of how sequences of utterances combine to form meaning, beyond that specified by the utterances in isolation. Thus, in determining the meaning of a given utterance in a conversation it is usually necessary to have some (abstracted) representation of what has been said before: the discourse context. Interlocutors are assumed to incrementally update their shared representation of this context as a conversation unfolds. Dialog is discourse between two or more parties, with the quintessential example being a conversation between two people or, in our case, between a person and a computer.

In this paper, we focus primarily on issues dealt with in the pragmatics of discourse and dialog, even though issues

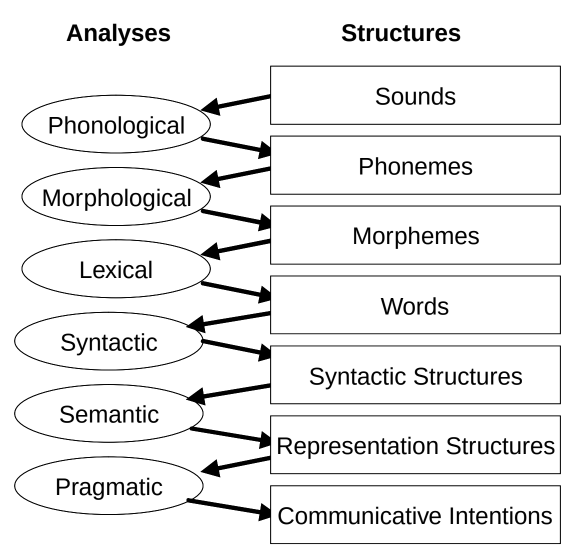

Fig. 1. Levels of linguistic analysis (adapted from [9]).

in the lower levels of analysis must also be dealt with when building dialog systems.

Discourse theory, then, is generally concerned with how multiple utterances fit together to specify meaning. Just as theories of syntax assume that sentences are composed of atomic units (words) and intermediate structures (phrases and clauses), organized according to a set of rules, theories of discourse generally assume that discourses are composed of discourse segments (consisting of one or more adjacent utterances), organized according to a set of rules. Beyond this, however, discourse theories vary widely in how they define discourse segments and the nature of the inter-segment relationships. Some define these relationships to be a function of surface structure (e.g., based on categories of utterance function, such as *request* or *inform*, called "speech acts" [10]), while others posit that these relationships must be a function of the intentions (plans and goals) of the individuals having the conversation [11,12]. In addition, researchers developing computational models of discourse have included a number of other constructs in their representation of discourse context, including: entities previously mentioned in the conversation, possibly organized into a sub-structure indicating the availability of these entities for subsequent reference; topics currently being discussed (e.g., "questions under discussion" [13]); and information structure, which indicates which parts of utterances contribute new information to the conversation as opposed to those parts which serve mainly to tie new contributions back to earlier conversation [14].

Discourse theory also seeks to provide accounts of a wide range of phenomena that occur in naturally occurring dialog including: mechanisms for conversation initiation, termination, maintenance and turn-taking; interruptions; speech intonation (used to convey a range of information about discourse context); discourse markers (words or phrases like "anyway" that signal changes in discourse context); discourse ellipsis (omission of a syntactically required phrase when the content can be inferred from discourse context); grounding (how speaker and listener negotiate and confirm the meaning of utterances through signals such as headnods and paraverbals such as "uh huh"); and indirect speech acts (e.g., when a speaker says "do you have the time?" to know the time rather than simply wanting to know whether the hearer knows the time or not).

While significant progress has been made in both theoretical and computational approaches to addressing most of these issues, in the most general cases these problems are far from being completely resolved, and many are known to be computationally intractable. In addition, the need for a first principles theory for these phenomena can be obviated by properly constraining a system's interaction with the user. In particular, if the range of utterances the user can make at each point in the conversation is tightly constrained, then many of the phenomena above can be designed out of the interaction (e.g., interruptions), while others can be "pre-computed" by the system designers (e.g., the meaning of indirect speech acts). Consequently, most contemporary health dialog systems—especially those which have been formally evaluated in large clinical studies—use interactions with the user that are very tightly scripted.

However, much of the ongoing research in this area is concerned with developing systems that enable user–computer conversation that more closely approximates natural and fluid human–human dialog.

3. What's unique about health dialog?

Communication between human healthcare providers and their patients is one of the most widely studied domains of communication research. Just within the field of physician–patient communication, one source lists over 3000 articles in print [15], and there are volumes written on the dialog that occurs during psychotherapy sessions. In this section, we look at a number of factors that make health communication a particularly novel and challenging application domain for dialog systems researchers. Most of these factors have yet to be definitively addressed in contemporary systems and thus represent important areas of ongoing research.

3.1. Criticality

Many health dialog systems have the potential to be used in emergency situations, for example in systems that assist patients with ambulatory care sensitive diseases or in chronic disease self-management. Several systems developed for this kind of application are designed to determine if the patient is having a life-threatening emergency as quickly as possible and either direct the patient to call 911 or immediately and automatically send a designated physician a pager message or FAX alerting them to the situation [16].

3.2. Privacy and security

Dialog content and communication media may need to be tailored based on the user's context to address privacy issues. For example, developers of applications that involve disclosure of potentially stigmatizing conditions or information should be sensitive to the user's environment and tailor content accordingly (e.g., using speech dialog systems to manage HIV medication regimen adherence).

3.3. Continuity over multiple interactions

Most health communication applications require multiple interactions with users over extended periods of time. Interaction frequencies can range from multiple times a day (e.g., in wearable monitoring applications) to daily (as in [17]) to one or more times per week (as in most TLC applications [18]), to once every few months (as in many of the health behavior change applications that use tailored documents [19]). Durations of use can span from a month (FitTrack, Section 5.3.1) to several months or a few years (most behavior change applications) to a lifetime (chronic disease monitoring and self-care). Further, these interactions are not isolated, stateless sessions (such as in a database question answering system), but require extensive information to be kept persistently between sessions for a given user, with subsequent dialog tailored on the basis of earlier conversations. This requirement for continuity over multiple interactions is found in few dialog system application domains outside of healthcare (multi-session intelligent tutoring systems being the other notable example). This requirement also drives several interesting research problems, such as determining the form and content of dialog history that is maintained between sessions, and the generation and resolution of expressions that refer to past interactions.

3.4. Language change over time

In human health provider–patient interactions language use naturally evolves over the course of time. Several studies have noted that task talk becomes more concise and takes less time as the interactants' knowledge of each other increases, while their use of social dialog generally increases as their relationship grows [20]. Some specific examples of the ways in which health behavior change dialog can evolve include: making use of information about the user's state to set behavior goals and give feedback; progressively disclosing more information about the user's condition; gradually making task language more precise; and gradually phasing out introductory how-to instructions and help messages. Maximizing conciseness in spoken output is especially important since it takes more time to communicate information in speech than in text [21]. Language change is also important just to maintain user engagement in the system. In the FitTrack study [17], several subjects mentioned that repetitiveness in the system's dialog content was responsible for their losing motivation to continue working with the system and follow its recommendations.

3.5. Managing patterns of use

One of the interesting but important ramifications of interacting with users over multiple sessions is that users' patterns of use of the system is itself is an important object of study, and may require as extensive tracking and management as the content of the intervention and the user's health behavior. Determining the optimal patterns of use for a given intervention is a difficult problem, but must be specified before a system can correctly manage interactions with its users. What is the dose–response relationship between user–system contacts and outcomes [4]? Is more frequent user–system contact always better? Is a regular contact schedule (vs. as needed by the user or as dictated by sensor data and other information) always best [22]?

3.6. Power, initiative, and negotiation

At first it may seem that conversational initiative in health communication is one feature that actually works in favor of building simpler dialog systems: as in most professional–client interactions, the professional maintains the initiative the vast majority of the time. While this is still the case in many physician–patient and therapist–patient interactions (physicians generally talk 50–100% more than patients [20]), contemporary health communication researchers have determined that the best way to motivate patients to adhere to prescribed regimens and/or change their health behavior is by moving away from this "paternalistic" style of interaction to one in which the health professional and the client work together on an equal footing to come up with a treatment plan that fits into the client's life: so-called "patient-centered" communication [23,24]. There has been a significant amount of research over the last few years on automated systems that can negotiate with users in natural language ("argumentation systems"), and this remains an active area of research.

3.7. User–computer relationship

The importance of quality relationships between health care providers and their patients is now widely recognized as a key factor in improving not only patient satisfaction, but treatment outcomes across a wide range of health care disciplines. The use of specific communication skills by physicians—including strategies for conducting patient-centered interviews and relationship development and maintenance—has been associated with improved adherence to treatment regimens improved physiological outcomes, and increased patient satisfaction, leading to recommendations for training physicians, nurses, pharmacists, and therapists in these skills [25].

Several studies have demonstrated that people respond in social ways to computers (and other media) when provided with the appropriate social cues, even though they are typically unconscious of this behavior {Reeves, 1996 #2139}. In a qualitative study of user perceptions of a telecommunications-based health behavior change intervention, Kaplan et al. found that users not only talked about the system using anthropomorphic terms (e.g., using personal pronouns), they described the system in ways indicative of having a personal relationship with it (e.g., "friend," "helper," "mentor") and seemed to be concerned about impression management (e.g., choosing to only interact with the system on days in which they met the system's health behavior goals) [26]. Milch, et al. [27] found that several subjects in their pager-based medication adherence intervention talked about their pager as a "trusted friend."

Taken together, these results indicate that an effective automated health communication system must not only be able to deploy appropriate intervention messages at the appropriate time, but must also address social, emotional, and relational issues in its communication with a user [25].

4. Dialog system technologies

A range of technologies are available for building health dialog systems. The simplest of these is a linear script that specifies the exact sequence of dialog moves the system and user will make in an interaction. State transition networks provide a more sophisticated and flexible model, allowing branches in the dialog based on what the user does in a given exchange with the computer. State transition networks can be defined hierarchically, resulting in sub-dialogs that can be factored out and re-used like subroutines: a modeling approach known as hierarchical state transition networks. Finally, plan-based dialog systems provide the potential for the greatest flexibility in dialog behavior by using action planners and plan recognition to model the underlying intentions of people in conversation. First, however, we describe pattern–response systems: a very simple, but commonly used approach for producing what appears to be flexible and coherent dialog with a computer. Table 1 presents a summary of the technologies discussed.

4.1. Pattern–response dialog systems

One of the most ubiquitous and popular methods for building systems that appear to be able to conduct coherent, intelligent dialogs with users (for primarily non-medical applications) is the use of a set of pattern–response rules. In these systems, rule patterns are matched against the sequence of words in a user utterance and, when a match is found, a corresponding system output utterance is produced. Pioneered in the ELIZA system in 1966 [28], these systems maintain little or no discourse context, but instead rely on a number of tricks to produce what is apparently coherent dialog. These tricks include: maintaining system-initiated dialog, by having most system outputs prompt the user with open-ended questions; relying on the user's sense-making ability to infer coherent explanations for the system's outputs; and reflecting the user's inputs back to them with minor wording changes in order to give the illusion of understanding what the user is saying.

An example rule in such a system is:

PATTERN: * I AM * DEPRESSED *
RESPONSE: I AM SORRY TO HEAR THAT YOU ARE DEPRESSED

where the asterisks in the pattern match zero or more words in the user's utterance. Here, the rule will match a user input of "I AM FEELING A LITTLE DEPRESSED" and produce a reasonable response. However, this same response would also be produced (not so reasonably) for user inputs of "I AM NOT REALLY DEPRESSED" and "MY BROTHER THINKS I AM DEPRESSED".

Unfortunately, since the user's inputs are unconstrained and there is no linguistic analysis or discourse model that could enable the system to truly understand what the user is talking about in all situations, these systems cannot be relied upon for critical applications in health communication in which errors in understanding user input can have dire consequences. However, this type of interaction has proven effective for emulating the behavior of a Rogerian psychotherapist (the purpose for which this type of dialog system was originally developed), and has been proven effective for therapy in which the system is essentially prompting a patient to think aloud and work through his or her own problems [29]. In these applications, significant errors in understanding user input or in producing incoherent system output can often be tolerated, as the primary function of the system is just to keep the user engaged in the interaction.

4.2. State-based dialog systems

The most common technology used for health dialog systems is a state machine in which each dialog move the system can make (utterance or discourse segment) is represented by a state, and arcs between states represent possible state transitions, with all of the arcs leading out of a given state (typically) representing alternative user inputs that are allowed in that state. In a state machine in which each state has only either zero or one next state, this represents an inflexible linear script such as the one shown in Fig. 2, for a simplified physical activity promotion system.

To provide variations in system behavior based on user input (and other factors such as physiological measurements, user characteristics or information gleaned from a

Table 1
Summary of health dialog system technologies

Dialog system technology	Discourse context representation	Use for
Pattern–response	None	Entertainment, engagement of user
State-based linear	Current state	Very short series of questions (e.g., screening)
State transition network	Current state	Brief dialog with some branching
Hierarchical state transition network	Stack of states	Partitioning extended dialog, or dialog with reusable sub-dialogs
Augmented transition network	Stack of states, database	Multiple extended dialogs, or dialogs in which branching is based on several earlier responses
Plan-based	Many possible representations encompassing beliefs and intentions of system and user	Generating dialog from deep knowledge of domain and natural language

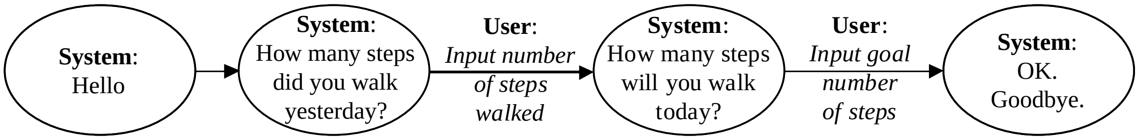

Fig. 2. Example linear dialog script.

user in previous dialogs), the linear script can be generalized to a State Transition Network, in which dialog states can have more than one next state, as shown in Fig. 3.

Often, dialog state machines need to be created for a variety of situations in which fragments of the state machine are repeated. For example, a different top-level dialog network may be developed for every contact with a user, but every contact includes a sub-dialog for assessing the user's health behavior in the same way. For this reason, and also to reduce the complexity of very large dialog networks, it becomes desirable to factor out commonly used dialog fragments and arrange for them to be invoked in a hierarchical manner, like subroutines in a software program. This model—as depicted in Fig. 4—is referred to as a hierarchical state transition network, in which the boxes represent invocation of sub-networks which are run to completion before the parent network is resumed. Execution of these networks thus requires a run-time stack to keep track of the suspended (invoking) networks and return states.

Linguists have previously proposed using grammars to represent general dialog structure, based on the observation that there are many sequencing regularities among utterances in human conversation, for example "adjacency pairs" such as a question typically being followed by an answer [30]. However, there have also been many arguments against the use of dialog grammars for representing natural human conversation. For example, the fact that a given utterance can perform multiple conversational functions makes a single next state impossible to specify [31].

The expressive power of hierarchical state transition networks can further be extended by allowing the actions taken upon user input recognition to include storing and retrieving information from a persistent database, and allowing network branches to be (partially) conditioned on this stored information. For example, in a physical activity promotion system, information about whether a user likes to exercise alone or with others can be obtained early in a conversation with a user and later used to determine whether to invoke a social support sub-dialog or not. Hierarchical state transition networks augmented in this manner are called "Augmented Transition Networks," and were originally developed for sentence parsing [32]. Augmented transition networks remain the most commonly used technology for implementing health dialog systems, and is the model underlying the VoiceXML dialog system standard [33].

4.3. Plan-based systems

The ultimate goal for many applications in dialog systems research is the development of systems that allow users to have as much freedom as possible to conduct an

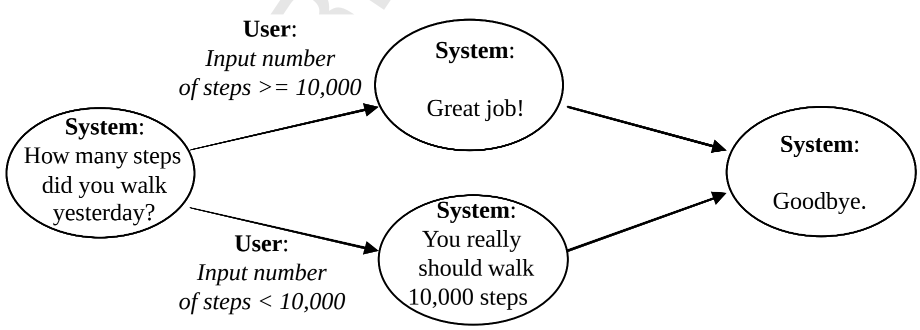

Fig. 3. Example state transition network dialog model.

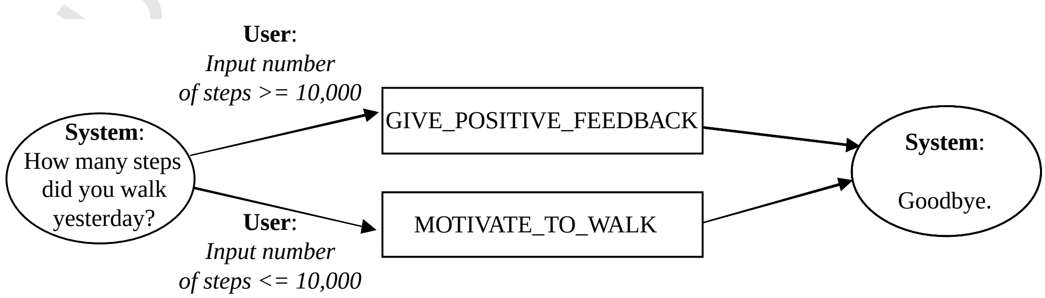

Fig. 4. Example hierarchical state transition network dialog model.

unconstrained conversation with a system, including all of the behavior observed in natural human–human conversations. This behavior includes: unconstrained user input; mixed-initiative dialog, in which either the user or the system can take control of the conversation at any time; proper handling of interruptions and requests for clarifications; indirect speech acts; and, ultimately, the proper recognition, display and use of nonverbal conversational behavior such as hand gesture.

The predominant approach taken to building these sophisticated dialog systems involves representing and reasoning about the intentions that underlie system and user utterances, inferring the user's goals and task plan, and dynamically synthesizing the system's task plan. Inferring a user's goals and task plan is necessary because, as exemplified by indirect speech acts, people's utterances do not always correspond directly to their communicative intent (e.g., as in "Do you have the time?"). Thus, plan-based theories of communicative action and dialog assume that the speaker's speech acts are part of a plan, and the listener's task is to infer it and respond appropriately to the underlying plan, rather than just to the utterance [34]. Synthesizing system task plans, including communicative and other actions, is necessary in complex applications in which all possible conversational contingencies (and their possible orderings) cannot be anticipated and scripted, but must be addressed in an incremental, reactive manner.

Dynamic planning and plan inference can be computationally very complex, and thus have not been used much to date in fielded health dialog systems. However, they remain active areas of research in Artificial Intelligence, and a handful of health dialog systems that use these techniques have been developed for the application of clinical guidelines [35], for the automatic generation of reminders for older adults with cognitive impairment [36], for medication advice [37], and for diet promotion [38]. Plan recognition, and especially dialog planning systems have been developed to consider several types of information in sequencing dialog segments including task dependencies, rhetorical strategies, and conversational conventions. Some research has also been conducted into machine learning of dialog plans [39], but these approaches require large samples of sample dialogs and have only been used for relatively simple planning problems to date.

4.3.1. Example: COLLAGEN

As an example of a plan-based computational model of discourse, we briefly review the theory developed by Grosz and Sidner [11], later elaborated by Grosz and Kraus and Lochbaum [40,41], and implemented in the COLLAGEN dialog engine [42]. In this theory, discourse context is represented by three elements:

- Linguistic structure—the structure of the utterances that comprise a discourse, partitioned into discourse segments, where the utterances in each segment are grouped according to intention (the Discourse Segment Purpose or DSP, representing the goal that the utterances relate to).
- Intentional structure—represents relationships among the DSPs and the overall goal of the discourse (the Discourse Purpose, DP). These relationships can be either sub-goal relationships (e.g., to conduct a conversation you need a greeting, a body and a farewell) or precedence relationships (e.g., the greeting precedes the body which precedes the farewell).
- Attentional state—is an abstraction of the participants' focus of attention as their discourse unfolds. It is dynamic, recording the entities (typically objects referred to in noun phrases) that are salient at each point in the discourse. It is represented as a stack of ⟨DSP, focus space⟩ pairs, where the focus space represents the entities under discussion ("in focus") during pursuit of the DSP. With each new discourse segment, a new pair is pushed onto the stack (possibly after other focus spaces are first popped off). One of the primary roles of the focus space is to constrain the range of DSPs to which a new DSP can be related, thus greatly simplifying the problem of plan recognition [43].

An example showing the state of a discourse in progress is given in Fig. 5. The discourse involves a physical activity promotion system, involving: a greeting (Opening); review of a client's previous day's exercise (DiscussPreviousDay); setting goals for the next day (DiscussNextDay); and presenting and discussing a self-monitoring graph depicting exercise progress over time (ShowGraph, DiscussGraph). The linguistic structure on the right shows (an excerpt) of the dialog, its partition into discourse segments, and the

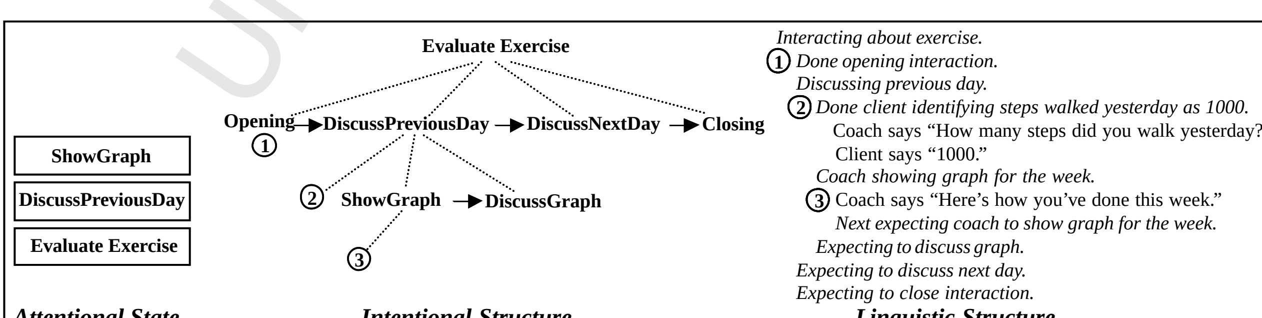

Fig. 5. Example discourse context in Grosz & Sidner's model.

embedding relationships among them. The intentional structure in the middle shows the relationship among the DSPs corresponding to the discourse segments (with arrows representing the sequencing relationships among the DSPs and dashed lines representing decomposition relationships). The attentional state on the left shows the stack of DSP/focus space pairs at position (3) in the dialog.

The theory (and the COLLAGEN implementation) also includes algorithms for determining the user's task goals on the basis of their utterances and other actions (plan recognition) and the planning of system actions (including utterances) required to collaborate with the user on the task being performed.

4.4. Utterance understanding and generation

Although the focus of this paper is on the discourse level of analysis in dialog systems, the issues of how individual user and system utterances will be recognized and produced must be addressed in the course of their development. In this section, we provide a brief overview of the approaches to these functions most commonly used in fielded systems.

4.4.1. Utterance understanding

Understanding user communicative intent on the basis of speech, text, and other input modalities, taking into account discourse context and world knowledge, is the single most difficult problem in developing dialog systems, and is thus the aspect that is typically the most tightly constrained. One of the ways this is usually accomplished is by providing users a discourse context in each dialog state in which their choices of possible responses are obvious and small in number, such as when a system asks closed-ended (e.g., yes/no) questions. Given this, however, there are still a range of approaches to mapping user inputs onto the range of input options the system is able to handle.

The simplest way to constrain user responses to system prompts is to provide users with an exhaustive multiple choice list of input options. An input context-free grammar, usually specified for each dialog state, allows significantly more flexibility in specifying allowed user inputs. This format is typically used for recognizing everything from individual numbers and dates up to phrases and sentences, and is commonly used in Automatic Speech Recognition (ASR) systems. More sophisticated parsing techniques using more powerful grammars and probabilistic/empirical techniques are available, but tend to not be used in dialog systems in which the focus is on discourse issues and high accuracy in understanding user intent. Multi-modal input understanding—in which either nonverbal conversational behavior, such as hand gesture or alternative input modalities, such as stylus gesture [44] are used—represents another active area of dialog system research, although little work has been done in the medical domain.

4.4.2. Utterance generation

Text generation is the problem of transforming a logical representation into a natural language utterance [45]. The simplest form of utterance generation involves simply indexing a fixed string or pre-recorded speech utterance and producing this for the user. A slightly more sophisticated technique—and the one most often used in fielded systems—is template-based generation, in which a string is annotated with variables whose values are determined at run-time (e.g., "YOU WALKED ⟨NumSteps⟩ STEPS TODAY."). In the most general case, text generation can involve word-by-word synthesis of utterances based on a grammar and dictionary, discourse context and world knowledge, although this level of sophistication is typically not required for most dialog system applications. Research has also been conducted into generation of multi-modal system outputs (speech or text plus accompanying nonverbal behavior or graphics) although, as with multi-modal input understanding, this has not been used widely in health dialog systems to date.

5. Deployment technologies

Health dialog systems may be deployed using a range of communication media. In this section, we provide an overview of the technologies that have been used.

5.1. World wide web

Among the deployment media for automated dialog systems, the Internet offers a number of attractive features. The main issue of deployment of automated dialog systems is what technology to use at the user's endpoint. The more advanced communication medium one chooses, the more complex (and costly) is the deployment process and its maintenance if it requires any special "receiver" technology. This applies both to hardware (whatever device patients are required to physically interact with), as well as to any user-visible software possibly involved.

Technologies that make use of client–server architectures are therefore preferable in situations in which ease of deployment is the most important factor. Among Internet-based technologies, web pages allow for very straightforward implementation of *questionnaires* and written turn-based dialogs. Deployment is straightforward because web pages only require a web browser to be displayed at the client site, and this software is available more or less universally. The limiting factor may still be availability of Internet connection and computers themselves, especially for certain user groups (e.g., low income, older adult, etc.).

While the most natural deployment medium for speech-only dialog systems is via telephony, Internet technologies support multimodal interfaces featuring speech with simultaneous graphical output, enabling the use of pictures, diagrams, and animations. Proposed solutions for multimodal browsing can be divided into server- and client-side speech recognition. In the former, the bulk of the speech recognition process happens at the remote server site, by transmitting the voice signal over the internet [46]. Client side recognition, instead, performs speech recognition on the user device; it therefore requires less bandwidth for the transmission of voice, but higher processing power. Client side recognition is endorsed by the W3C via the XHTML + Voice profile, related to VoiceXML [47]. Multimodal browsing is especially attractive for mobile devices, although still in its infancy.

5.2. Speech and telephony

A natural, technologically mature way to provide direct access to health communication interventions to patients from home is via their telephone, dialed into a specially equipped server computer. These systems are known as Interactive Voice Response (IVR). While it is possible to set up an inexpensive IVR system for relatively simple, low call volume applications, complex dialogue systems targeted at high volume applications can be very expensive to develop and deploy. Systems are typically built to deal with incoming calls (dial in)—but in some cases they can be deployed to automatically dial out connections and process them (once callee's privacy issues are addressed, of course).

IVR systems can communicate with users by playing messages over the telephone line. Such messages, or prompts, must be either pre-recorded by voice actors and stored inside the computer system or dynamically synthesized. Recorded prompts are usually natural and intelligible; however, the messages cannot be altered after being recorded, but only combined sequentially. This is a major drawback if one needs to convey to the user information that is evaluated at runtime: for example, large numbers, or even names that were not foreseen at the time when the system was built.

Text to speech (TTS) systems are a viable alternative for prerecorded voice prompts. TTS systems are able to transform an arbitrary text string into a sound signal, which can be played over the telephone line [48]. Since the synthesis process starts from the string, any utterance can be generated, and TTS is required when system utterances are dynamically generated.

Users can communicate with IVR systems by pressing keys on touch tone phones. The vast majority of current telephones, including cellular phones, produce a known frequency combination when each key is depressed. The frequencies, commonly known as Dual Tone Multi Frequency (DTMF) or touch-tones, can be transmitted over channels made for carrying voice, and reliably detected by algorithms built into telephony hardware or software. For these reasons, DTMF signaling became a sensible means to acquire user input in IVR, allowing users to provide feedback, for example, selecting items in a menu structure presented during the progress of an automated call. The data that can be entered are necessarily limited to numeric quantities or codes and navigation is usually restricted to a tree-like structure. Despite this somewhat cumbersome usage, controlled studies have shown such DTMF systems to be successful for in-home monitoring of patients with chronic diseases such as hypertension [49–51] and diabetes [52].

Automatic speech recognition (ASR) technology is now widely available and has been integrated into many IVR systems as an alternative to DTMF. The accuracy of ASR is still far from perfect, especially for certain types of users (e.g., for those with non-standard accents, older adults, or children) or dialog. Thus, speech input grammars—specifying what users can say at each dialog state—must be carefully designed, often using DTMF as a fallback. Unconstrained spoken input is possible, in principle, in dictation systems—but in practice it is not usable for IVR, since dictation systems need a lengthy training on the specific speaker (speaker-dependent recognition) to achieve satisfactory performance, and even with this, accuracy is usually too low to be useful for health communication. Grammars, instead, restrict the input space of utterances and make speaker-independent recognition of sentences over the telephone reliable enough for practical use.

A significant advance in the deployment of IVR systems, both keypad- and voice-based, has been the standard endorsed by the W3 Consortium (W3C). The standardization activity has yielded a dialog planning language, VoiceXML, and also standardized *grammar definition* languages, such as the Speech Recognition Grammar Format (SRGF). The W3C Voice Interaction group proposed an architecture for IVR systems which closely resembles that for standard web-based applications, the main difference being that the visual web browser (client), is replaced by a *voice browser*, which interprets a dialog description written in VoiceXML and conducts the interaction [33]. Dialog description and its linked grammars are served over the internet or intranet in a manner analogous to HTML pages and linked images. Detailed discussion of the languages and standards is outside of the scope of this paper; further details can be found, e.g., in [53]. Programming VoiceXML can be cumbersome, resulting in a growing number of commercial tools for authoring VoiceXML documents and approaches to dynamically generating these documents [54].

5.2.1. Example: HOMEY

The HOMEY project was funded in 2001 by the European Union with the aim to advance research in spoken dialog systems applied to enhance communication between specialist health centers and patients with chronic diseases [55]. The project resulted in three demonstrators: (1) one for monitoring patients affected by hypertension [55], (2) a second for studying automated dialog planning from ontologies and computerized guidelines [35], and (3) a PDA-based multimodal electronic patient record interface [46]. This section gives a short account of the first system; the second is addressed by Beveridge and Fox in a separate paper in this issue.

The HOMEY hypertension system enables patients to self-report clinical values and possible medication side effects via a telephone-based, mixed initiative spoken dialog system. It also provides simple educational messages and serves as a reminder for clinical tests and scheduled appointments. Data entered by patients is reported to physicians through a web-based electronic medical record, which is integrated with the system. This self-reported data is stored and displayed along with data entered by physicians from face-to-face encounters.

Hardware and speech recognition software, and the proprietary dialog scripting language, were provided by project partners, while the development of the application itself (the dialog scripts) and the web-based patient record has been co-designed together with knowledge engineers and medical specialists.

The hypertension prototype was subject to two pre-deployment tests with volunteers, which were used to assess ergonomic aspects, including dialog adaptation and refinements of language models. The system was finally used by two hospitals in a controlled clinical trial that lasted approximately one year (6 months between enrollment and follow-up for each patient). Results indicated that 24-h averaged blood pressure values decreased more in the dialog-system treatment group compared to a control group ($p < 0.1$).

5.3. Embodied conversational agents

Embodied Conversational Agents (ECAs) are animated humanoid computer-based characters that use speech, eye gaze, hand gesture, facial expression, and other nonverbal modalities to emulate the experience of human face-to-face conversation with their users [56]. Such agents can provide a "virtual consultation" with a simulated health provider, offering a natural and accessible source of information for patients. These agents represent one form of multimodal dialog system, in which the nonverbal modalities are recognized and produced in addition to accompanying text or speech, to more fully understand the user's communicative intent. In addition to carrying additional factual information, nonverbal behavior is also used in face-to-face conversation to regulate the interaction structure itself, for example, gaze and intonation to regulate turn-taking behavior, body position and orientation to regulate conversation initiation and termination.

In addition to the FitTrack system described below, several ECAs have been developed for use in health dialog systems, for applications spanning training in human subjects consenting procedures [57], training in coping skills for caregivers of children with cancer (deployed on both desktops and PDAs [58]), and diet behavior change. These systems vary greatly in their linguistic capabilities, input modalities (most are mouse/text/speech input only), and task domains, but all share the common feature that they attempt to engage the user in natural, full-bodied (in some sense) conversation.

5.3.1. Example: FitTrack

The FitTrack system was developed to investigate the ability of an ECA to establish and maintain a long-term therapeutic alliance with users, and to determine if these relationships could be used to increase the efficacy of health communication and health behavior change programs delivered by the agent [59,60]. An ECA was expected to be particularly effective at relational communication, given that most human relationships are formed and maintained in face-to-face conversation where nonverbal behavior can be used to communicate and assess the social aspects of the interaction. In the FitTrack system, the ECA uses nonverbal behavior to convey propositional, interactional, affective and attitudinal information in addition to the speech channel.

The ECA, named "Laura," played the role of an exercise advisor who motivated sedentary adults to obtain the minimum level of physical activity recommended by current public health guidelines [61] over a two-month period of time. The dialog was modeled using augmented transition networks, with dynamic multiple choice inputs by users and embodied conversational agent output (synthesized speech and synchronized nonverbal conversational behavior displayed by an animated agent). The system was designed to run on standard home desktop computers so that participants could interact with the system on a daily basis.

The appearance and nonverbal behavior of the exercise advisor was based on a review of relevant literature and a series of pre-test surveys. Fig. 6 shows the character and user interface. The system used the BEAT text-to-embodied-speech translator [62] to generate nonverbal behavior for the agent, including hand gestures, posture shifts, head nods, gaze and eyebrow behavior, immediacy behavior (liking or disliking of one's conversational participant demonstrated through nonverbal behaviors such as proximity and gaze [63,64]) and nonverbal signaling of different conversational frames [65] (health dialog, social dialog, empathetic dialog, and motivational dialog).

FitTrack was successfully used in two randomized clinical trials, one involving MIT students and the second an urban, older adult population.

5.4. Robots

There is an emerging interest in developing autonomous, mobile robotic systems that can interact with users to perform various health-related tasks. Many of these robots include some speech-based natural language dialog capability, although they appear to be mostly very simplistic from a dialog systems perspective. Example applications include robotic nurse spirometry assistants for post-cardiac surgery patients [66], arm motion rehabilitation for stroke patients [67], and eldercare [68].

6. Development methodologies

The development methodologies used in dialog systems research depends very heavily upon the type of technology

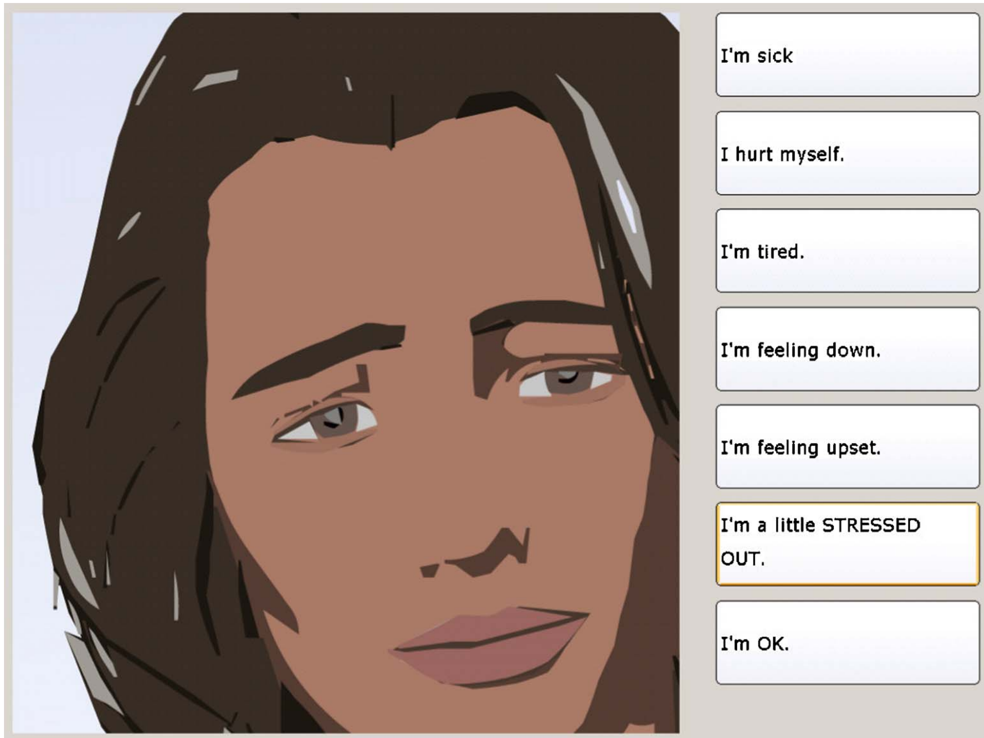

Fig. 6. FitTrack embodied conversational agent.

and underlying models employed. Development of all kinds of dialog systems often begins with the collection and analysis of sample dialogs between real people (e.g., between health providers and patients). The resulting recordings (audio or video) are transcribed and subjected to discourse analysis [69]. This analysis results in a characterization of the range of concepts, terms, and syntax typically used in patient–provider communication, in addition to the range of topics discussed, the types of questions asked, and the overall conversation structure and sub-dialog structure used. Much of this process is analogous to the knowledge engineering methodology followed in the development of expert systems. Typically, full characterization of dialogs is achieved through a combination of literature review, discourse analysis, and direct authoring of scripts by expert providers.

Another method that is widely used in dialog system development is the "Wizard-Of-Oz" technique, in which (unbeknownst to test subjects) a human confederate replaces some or all of a dialog system's functionality during live interactions between subjects and the system [70]. Dialog from these sessions is recorded and analyzed for several purposes, including: early characterization of domain dialogs; characterization of user responses in particular contexts of interest; assessment of user acceptance of and attitude towards a planned system; and assessment of utility and efficacy of a planned system. Although ideally, user–system interaction will closely follow provider–patient interaction, it has been observed that in many situations users speak and otherwise behave differently when interacting with a computerized system than with another human (e.g., they simplify their speech patterns) [71]. In these situations, Wizard-of-Oz testing is particularly important, since the study of provider–patient interaction will not correctly characterize these dialogs.

The underlying model to be built into the dialog system also influences development. State-based and grammar-based dialog systems are designed with a focus on characterizing the surface level of the dialog and a small number of relatively large-grained variations in dialog structure. This effort can proceed from the collected corpora, from one or more providers who author the grammars or networks directly, or by a linguist/knowledge engineer who interviews one or more providers and develops the grammar. Development of plan-based dialog systems is much more involved, and requires deeper modeling of relevant ontologies and knowledge structures in the domain, as well as the development of dialog plan fragments.

Finally, development of dialog systems that are going to be fielded, for example for use in a clinical trial, requires extensive pre-testing and iterative refinement to ensure that the resulting system is both functional and natural.

7. Evaluation methodologies

There are three broad approaches to the evaluation of health dialog systems (as compared with other kinds of systems in medical informatics [72]). First, qualitative and quantitative evaluation of a single user–system conversation—focusing on issues such as accuracy, efficiency, and subjective user evaluation—can be performed using a variety of methods and instruments. Second, and perhaps unique to health dialog systems, is the analysis of usage patterns over time—how often users choose to conduct interactions, whether these taper off over time, etc.—and how these patterns are affected by features of the dialog system and how, in turn, they affect health outcomes. Finally, evaluation of the efficacy of health dialog systems can be established through standard randomized clinical trial methodologies. In practice, researchers whose backgrounds are in the medical professions tend to focus primarily on the last type of evaluation, while those in computational linguistics tend to focus primarily on the first. Ideally, multiple forms of evaluation should be used throughout the development lifecycle to ensure the most efficacious system.

In addition to these task- and outcome-oriented assessments, it may also be important to evaluate the psychological aspects of interactions between users and a health dialog system. Very little work has been to date in this area. It may be important to assess user attitudes towards a system after some period of use: qualitative methods (as in [26]) and standardized measures of patient–provider relationship (as used in [73]) may be used for this purpose. We know of no cognitive evaluations of conversations between users and health dialog systems (e.g., of the form done in [74]). However, as these systems move away from scripting technologies and incorporate dialog planners that synthesize language from explicit knowledge representations (as discussed in Section 4.3), "cognitive analysis" of the machine's knowledge should become a simple matter of inspection.

7.1. Dialog performance evaluation

One of the most mature methods for evaluating dialog system performance is provided by the PARADISE framework [75]. PARADISE uses a decision-theoretic framework to combine evaluations of system accuracy (success rate at achieving desired conversational outcomes) with the "costs" of using a system—comprised of quantitative efficiency measures (number of dialog turns, conversation time, etc.) and qualitative measures (e.g., number of repair utterances)—to yield a single quality measure for a given interaction. Weights for the various elements of the evaluation are determined empirically from overall assessments of user satisfaction for a sample set of conversations, and the evaluation formula can be applied to sub-dialogs as well as to entire conversations to enable identification of problematic dialog fragments.

Two other qualitative evaluation methods were developed on the TRINDI and DISC projects. They provide criteria for evaluating a dialog system's competence in handling certain dialog phenomena. The TRINDI Tick-List consists of three sets of questions that are intended to elicit explanations describing the extent of a system's competence [76]. The first set consists of eight questions relating to the flexibility of dialog that a system can handle. For example, the question "Can the system deal with answers to questions that give more information than was requested?" assesses whether the system has any ability to handle mixed-initiative dialog. The DISC Dialog Management grids [77] include a set of nine questions, similar to the Trindi Tick-List, that are intended to elicit some factual information regarding the potential of a dialog system.

Since it is desirable to perform extensive evaluation of health dialog systems prior to using them in expensive clinical trials, they are often evaluated by volunteers who are given scripts and asked to interact with a system to perform a series of "real life tasks." These users have to find their way through the system interaction in order to accomplish the task.

Evaluation may also be conducted on the basis of call logs in telephony systems that record conversations between users and the system. These recordings can be listened to and annotated by human expert evaluators, but at the expense of effort and time. Woodbridge [78] describes how telemedicine interactions can be scored via a handcrafted algorithm, while Giorgino [55] proposes to apply supervised machine learning algorithms to reproduce human-provided numeric annotations, based on attributes that can be gathered automatically.

7.2. Evaluating patterns of use

Health communication applications in general, and health behavior change applications in particular, require multiple contacts with a user over extended periods of time. In these systems, it is the user's decision whether to conduct a given conversation with the system or not, even if the conversations are system initiated. Acquisition of such usage data for many users over extended periods of time results in datasets that can be analyzed to determine: typical usage patterns; correlations between system or user characteristics and usage; and correlations between system usage and outcomes (dose–response relationships). These are important objects of study, because they can inform the design of future systems that users like interacting with (maximizing usage) or which are most efficacious (maximizing outcomes) or, ideally, both.

This is a nascent area of research, but there have already been a few published studies. Farzanfar partitioned users of a telephone-based physical activity promotion system into five usage groups: (1) those who adhered to the recommended call schedule (twice weekly for 3 months) at least 80% of the time; (2) those who used the system throughout the three months but intermittently; (3) those who used the system consistently for a while but then discontinued use; (4) those who only used the system zero or one time; and (5) those who had one or more incomplete calls [22]. Differences between these groups were found in both outcomes and self-reported system evaluations. For example individuals in the intermittent group (2) had the highest ratio of satisfied users and better reported outcomes both in terms of physical activity levels and perceived benefits, compared to the other groups. Giorgino made similar observations in analysis of the call data from the HOMEY system [55].

7.3. Randomized clinical trials

As the ultimate objective of the majority of health dialog systems is to affect the health of its users, the evaluation of these systems involves randomized clinical trials in which they are compared (typically) to standard-of-care conditions and evaluated using the same outcome measures that would be used in a trial involving any other health intervention technology or method. The vast majority of NIH-funded health dialog systems have been evaluated in this manner. The only differences between a study involving an automated dialog system and one involving human health providers are: study eligibility criteria usually specify that subjects must speak a particular language (since most projects do not have the resources to produce multi-lingual systems); subjects have access to the terminal device required (phone, home computer, etc.) or are provided one for the study; and they have the cognitive and physical ability to use the system. Subjects in dialog systems studies are also either provided an initial training session and/or printed materials describing how to access and use the system initially. Given the amount of longitudinal data typically collected in these studies, longitudinal data analysis methodologies are normally employed in addition to standard before-and-after (or baseline/end-of-intervention/follow-up) comparisons [79].

8. Efficacy of formally evaluated systems

A number of health dialog systems to deliver health education or effect health behavior change have been developed and successfully evaluated in randomized clinical trials, with the results generally demonstrating significant improvements in health outcomes over standard-of-care or no-intervention control conditions, and in many cases demonstrating outcomes equivalent to similar interventions by human health providers.

Revere and Dunbar conducted a meta-review of 37 evaluation studies involving generation of print-based health educational materials, and telephone-based and computer-based health dialog system interventions [80]. These systems provide health behavior change information to users based on a wide variety of health behavior theories (e.g., the stages of change model [81], the health belief model [82], and social cognitive theory [83]), and were applied to a number of health behaviors (physical activity promotion, diet adherence, medication regimen adherence, smoking cessation, chronic disease self-management, and others). The authors found that 33 of the 37 studies reported improved outcomes and 20 of these (60.6%) were statistically significant. The authors also concluded that tailored interventions—those whose messages are based on a specific individual's characteristics—generally outperformed interventions that were generic, targeted (developed for a specific subgroup of the population), or just personalized (included the user's name in the messages). Of the studies reviewed, only 13 could be considered true dialog systems (i.e., communicated using interactive utterance exchanges with a user), but of these 11 (85%) reported statistically significant improvements in health outcomes.

8.1. Evaluation of IVR systems

One meta-review, specifically focused on outcome studies of IVR-based systems published during years 1989–2000 is provided by [84].The reviewers exhaustively took into consideration 54 studies concerning health-related DTMF systems, published in peer-reviewed journals. (It is however not clear how many distinct systems they are related to.) The first interesting point of the review is that the papers were grouped by intervention area, thus providing a useful synopsis of the intervention types to which these systems have been applied. Authors also identified common features which make IVR systems applicable for healthcare interventions, including: absence of interviewer bias, low cost per interview, automatic and continuous operation, and greater confidentiality. Positive outcomes were reported according to different intervention areas: change in screening habits and self-reported satisfaction with the system for telephone-based information services; increased treatment compliance and child immunization rates for reminder calls about children immunization and other appointments; reduced hemoglobin readings for diabetic patients in chronic disease monitoring; and more faithful reporting of misbehaviors in behavior assessment. Not all of the studies examined were controlled, and some interventions which were did not show statistically significant improvements. Insufficient IVR compliance was noted in several studies.

Another review [85] explicitly focused on IVR interventions for management of chronic disease conditions. This review concludes that, while there are still few peer-reviewed evaluations of the impact of IVR-supported disease management systems, "those that have been conducted indicate that some outcomes can be moderately improved."

Finally, the clinical effectiveness of educational voice messages has been assessed by another recent meta-review [36], which concludes that among 19 studies considered (of which 16 were controlled), "more than 80% of studies showed significant impact upon measurable health outcomes."

One series of IVR systems and studies deserve special mention: the Telephone-Linked Care (TLC) systems developed by Friedman and colleagues at Boston University over the last 20 years. These systems are developed primarily using two-level augmented transition networks, recorded speech output, and either DTMF or ASR for user input. TLC behavior change applications have been applied to changing dietary behavior [86], promoting physical activity [87], smoking cessation [88], and promoting medication adherence in patients with depression [89] and hypertension [18]. TLC chronic disease applications have been developed for chronic obstructive pulmonary disease (COPD) [90], and coronary heart disease, hypercholesternemia, and diabetes mellitus [18]. All of these systems have been evaluated in randomized clinical trials and most were shown to be effective on at least one outcome measure, compared to standard-of-care or non-intervention control conditions.

8.2. Evaluation of ECA systems

An evaluation of the FitTrack physical activity advisor agent was conducted in a randomized study comparing college-aged subjects who conducted daily dialogs with the agent with subjects who simply kept track of their physical activity (time estimates and pedometer steps) [25]. Subjects who interacted with the agent increased their number of days per week during which they had 30 min or more of moderate-or-greater intensity physical activity, compared to subjects in the CONTROL condition, $t(86) = 1.98$ $p < .05$. A second study evaluated FitTrack for an urban, older adult population, in which subjects who interacted with the agent were compared to a standard of care (print materials and pedometer only) control group [91]. The estimated slope of pedometer steps over the two-month study duration (increase per week in mean weekly steps walked) was significantly greater for the intervention group than the control group ($p = 0.004$).

9. Conclusion and future directions

There is a growing body of research on the development and evaluation of systems which can interview patients and consumers about their health and provide health information and counseling using natural language dialog. The formal evaluation of many of these systems has demonstrated that they are effective compared to standard-of-care controls and, in some cases, are as effective as human health providers (e.g., [92]). These systems have the potential to reach large numbers of users at relatively low cost, resulting in the potential for high impact on population health. At the same time, health dialog represents a challenging and important application domain for dialog system researchers, with many features—such as repeated contacts over extended periods of time—relatively unique to the domain.

There are many future directions for research in health dialog systems that are currently being pursued. One of the most important is the further development of plan-based dialog systems that incorporate medical and behavioral ontologies and deep knowledge of health communication strategies. The use of standard, underlying ontologies will allow the theory-level knowledge in these systems to be shared and validated, and to be directly compared in a meaningful manner. On a more practical level, the lack of model-based representations in these systems limits their scalability, tailorability, and adaptability, and requires that every new intervention be developed from scratch, requiring months of duplicated effort when teams of behavioral scientists write dialog scripts for a new application, even if it is only a slight variant of a previously developed system.

Other promising directions of research include the increasing use of multi-modal dialog, including both embodied conversational agents and other systems that support elements of natural face-to-face conversation, as well as systems that use other modalities such as speech and pen-based input. Properly conducting the affective and empathic dimensions of provider–patient communication represents a significant challenge, as is the maintenance of engagement over many interactions.

Multi-party dialog is understudied in both linguistics and computational linguistics, but represents a potentially important area of future research for health dialog systems. Some health behavior change systems have already been developed that interact with multiple members of a household (e.g., to increase medication adherence in childhood asthma [93]), and this type of intervention represents a promising avenue for effecting change through social support. Systems to support case management nurses in their telephone consultations with patients have also been developed, and the development of systems that can support 3-way, real-time conversations between nurses, patients, and a dialog system that can offload routine parts of these interactions also represents an interesting area of inquiry. However, much more work remains to be done in this area.

Finally, the use of mobile devices (e.g., cellular phones) provides the opportunity for automated systems to dialog with patients "anywhere, anytime." When coupled with real-time sensors, these systems can provide pro-active health messaging at the time of need (e.g., when a user is starting a bout of exercise or lighting up a cigarette). Developing health behavior change systems that can maintain a persistent and continuous dialog with patients about their health behavior, incorporating awareness of the user and their environment, providing comfort and empathy in addition to tailored and theory-driven pragmatic advice, and tying in human health providers when needed may still be science fiction, but it represents a grand goal to work towards.

Acknowledgments

Thanks to Candy Sidner for providing the COLLAGEN example, and to Jennifer Smith, Candy Sidner, Rob Friedman, Daniel Mauer, and Daniel Schulman for their many helpful comments on this paper.

References

[1] de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: Introduction to a new approach. Patient Educ Couns 1999;36:99–105.
[2] Counsel PE. Special Issue on computer-generated tailored health behavior change interventions. Patient Educ Couns 1999;36(2).
[3] Davidoff F. Time. Ann Intern Med 1997;127:483–5.
[4] Velicer W, Prochaska J, Fava J, Laforge R, Rossi J. Interactive versus noninteractive interventions and dose–response relationships

for stage-matched smoking cessation programs in a managed care setting. Health Psychol 1999;18(1):21–8.

[5] Narayanan S, Ananthakrishnan SR, Belvin EE, Gandhe S, Ganjavi S, Georgiou PG, et al. The Transonics Spoken Dialogue Translator: An aid for English-Persian Doctor-Patient Interviews. In: AAAI Fall Symposium on Dialog Systems for Health Communication. Washington, DC; 2004.

[6] Luperfoy S. Retrofitting Synthetic Dialog Agents to Game Characters for Lifestyle Risk Training. In: AAAI Fall Symposium on Dialogue Systems for Health Communication. Washington, DC; 2004.

[7] AAAI. Fall Symposium on Dialog Systems for Health Communication; 2004.

[8] AAAI. Spring Symposium on Argumentation for Consumers of Healthcare; 2006.

[9] Winograd T. Language as a Cognitive Process: Volume 1—Syntax. Reading, MA: Addison-Wesley; 1983.

[10] Searle J. Speech Acts: An essay in the philosophy of language. Cambridge University Press; 1969.

[11] Grosz B, Sidner C. Attention, intentions, and the structure of discourse. Comput Linguist 1986;12(3):175–204.

[12] Allen J, Perrault CR. Analyzing intention in utterances. In: Grosz BJ, Jones KS, Webber BL, editors. Readings in natural language processing. Los Altos, CA: Morgan Kaufmann Publishers, Inc.; 1986. p. 441–58.

[13] Larsson S, Ljunglof P, Cooper R, Engdahl E, Ericsson S. GoDIS—An Accommodating Dialogue System. In: ANLP/NAACL-2000 Workshop on Conversational Systems; 2000. p. 7–10.

[14] Prince EP. Toward a taxonomy of given-new information. In: Cole, editor. Radical Pragmatics. New York: Academic Press; 1981. p. 223–55.

[15] Parks M, Floyd K. Meanings for closeness and intimacy in friendship. J Social Personal Relationships 1996;13(1):85–107.

[16] Finkelstein J, Friedman R. Potential role of telecommunication technologies in the management of chronic health conditions. Dis Manage Health Outcomes 2000;8(2):57–63.

[17] Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns (to appear).

[18] Friedman R. Automated telephone conversations to asses health behavior and deliver behavioral interventions. J Med Syst 1998;22:95–102.

[19] Velicer W, Prochaska J. An expert system intervention for smoking cessation. Patient Educ Couns 1999;36:119–29.

[20] Graugaard P, Holgersen K, Eide H, Finset A. Changes in physician–patient communication from initial to return visits: a prospective study in a haemotology outpatient clinic. Patient Educ Couns 2004.

[21] Kuppevelt JV, Heid U. Best practice in spoken language dialogue systems engineering. Nat Lang Eng 2000;6(3&4).

[22] Farzanfar R, Frishkopf S, Migneault J, Friedman R. Telephone-linked care for physical activity: a qualitative evaluation of the use patterns of an information technology program for patients. J Biomed Inform 2005;38(3):220–8.

[23] Bensing J. +Bridging the gap. The separate worlds of evidence-based medicine and patient-centered medicine. Patient Educ Couns 2000;39:17–25.

[24] Stewart M. Effective physician–patient communication and health outcomes: a review. Can Med Assoc J 1995;152:1423–33.

[25] Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns (to appear).

[26] Kaplan B, Farzanfar R, Friedman R. Personal relationships with an intelligent interactive telephone health behavior advisor system: a multimethod study using surveys and ethnographic interviews. Int J Med Inform 2003;71(1):33–41.

[27] Milch R, Ziv L, Evans V, Hillebrand M. The effect of an alphanumeric paging system on patient compliance with medicinal regimens. Am J Hosp Palliat Care 1996;13:46–8.

[28] Weizenbaum J. Eliza—a computer program for the study of natural language communication between man and machine. Commun ACM 1966;9(1):36–45.

[29] Slack W. Patient-computer dialogue: a review. Yearbook Med Inform 2000:71–8.

[30] Sacks H, Schegloff EA, Jefferson G. A simplest systematics for the organization of turn-taking for conversation. Language 1974;50:696–735.

[31] Levinson S. Some pre-observations on the modelling of dialogue. Discourse Process 1981;4(1):93–116.

[32] Woods WA. Transition network grammars for natural language analysis. In: Grosz BJ, Jones KS, Webber BL, editors. Readings in natural language processing. Los Altos, CA: Morgan Kaufmann Publishers, Inc.; 1986. p. 71–88.

[33] Beskow J, McGlashan S. Olga: a conventional agent with gestures. In: IJCAI 97; 1997.

[34] Cohen P. Dialogue modeling. In: Cole R, editor. Survey of the state of the art in human language technology. National Science Foundation; 1996.

[35] Beveridge M, Millward D. Combing Task Descriptions and Ontological Knowledge for Adaptive Dialogue. In: Proceedings of the 6th international conference on text, speech and dialogue (TSD-03); 2003.

[36] Pollack ME, Brown L, Colbry D, McCarthy CE, Orosz C, Peintner B, et al. Autominder: an intelligent cognitive orthotic system for people with memory impairment. Robotics Autonomous Syst 2003;44:273–82.

[37] Ferguson G, Allen J, Blaylock N, Byron D, Chambers N, Dzikovska M, et al. The Medicaion Advisor Project Preliminary Report. Technical Report. Rochester, NY: Dept of Computer Science, University of Rochester; 2002. Report No.: 776.

[38] Grasso F, Cawsey A, Jones R. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. Int J Hum-Comput Stud 2000;53(6):1077–115.

[39] Walker M, Fromer J, Narayanan S. Learning optimal dialogue strategies: a case study of a spoken dialogue agent for email. In: ACL/COLING 98; 1998.

[40] Lochbaum K. A collaborative planning model of intentional structure. Comput Linguist 1998;24(4):525–72.

[41] Grosz B, Kraus S. The Evolution of SharedPlans. In: Rao A, Wooldridge M, editors. Foundations and Theories of Rational Agency; 1998.

[42] Rich C, Sidner CL. COLLAGEN: A collaboration manager for software interface agents. Int J User Model User-Adapted Interact 1998;8(3–4):315–50.

[43] Lesh N, Rich C, Sidner C. Using plan recognition in human-computer collaboration. In: Proceedings of the Conference on User Modelling. Banff, Canada: Springer Wien New York; 1999. p. 23–32.

[44] Oviatt S, DeAngeli A, Kuhn K. Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction. In: CHI 97; Atlanta, GA; March 22–27, 1997. p. 415–22.

[45] Reiter E, Dale R. Building natural language generation systems. Cambridge: Cambridge University Press; 2000.

[46] Armaroli C, Azzini I, Ferrario L, Giorgino T, Nardelli L, Orlandi M, et al. An architecture for a multi-modal web browser. In: ICSLP; Denver, CO; 2002. p. 2553–6.

[47] Axelsson J, Cross C, Lie HW, McCobb G, Raman TV, Wilson L. XHTML+Voice Profile. In: W3C; 2001.

[48] Dutoit T. An Introduction to Text-to-Speech Synthesis. Kluwer Academic; 1997.

[49] Rogers MA, Small D, Buchan DA, Butch CA, Stewart CM, Krenzer BE, et al. Home monitoring service improves mean arterial pressure in patients with essential hypertension. A randomized, controlled trial. Ann Intern Med 2001;134(11):1024–32.

[50] Young M, Sparrow D, Gottlieb D, Selim A, Friedman R. A telphone-linked computer system for COPD care. Chest 2001;119(5):1565–75.

[51] Friedman R, Stollerman J, Mahoney D, Rozenblum L. The virtual visit: using telecommunications technology to take care of patients. JAMA 1997;4:413–25.

[52] Bellazzi R, Arcelloni M, Bensa G. Design, methods, and evaluation directions of a multi-access service for the management of diabetes mellitus patients. Diabetes Technol Ther 2003;5(4):621–9.

[53] Mittendorfer M, Niklfeld G, Winiwarter W. Making the VoiceWeb Smarter—integrating intelligent component technologies and VoiceXML. In: Second international conference on web information systems engineering (WISE'01); 2001. p. 126–31.

[54] Mittendorfer M, Niklfeld G, Winiwarter W. Evaluation of intelligent component technologies for VoiceXML applications: SCCH; 2001.

[55] Piazza M, Giorgino T, Azzini I, Stefanelli M, Luo R. Cognitive human factors for telemedicine systems. In: Medinfo, San Francisco; 2004, to appear.

[56] Cassell J, Sullivan J, Prevost S, Churchill E. Embodied conversational agents. Cambridge: MIT Press; 2000.

[57] Visscher WA, Hubal RC, Guinn CI, Studer EJ, Sparrow DC. A Synthetic Character Application for Informed Consent. In: AAAI fall symposium on dialog systems for health communication. Washington, DC; 2004.

[58] Johnson W, LaBore C, Chiu Y. A pedagogical agent for psychosocial intervention on a handheld computer. In: AAAI fall symposium on dialog systems for health communication. Washington, DC; 2004.

[59] Bickmore T. Relational agents: effecting change through human-computer relationships. Cambridge, MA: MIT; 2003.

[60] Bickmore T, Picard R. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact (to appear).

[61] Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health: a recommendation from the centers for disease control and prevention and the American College of Sports Medicine. JAMA 1995;273(5):402–7.

[62] Cassell J, Vilhjálmsson H, Bickmore T. BEAT: The behavior expression animation toolkit. In: SIGGRAPH '01. Los Angeles, CA; 2001. p. 477–86.

[63] Argyle M. Bodily communication. New York: Methuen & Co. Ltd; 1988.

[64] Richmond V, McCroskey J. Immediately. In: Nonverbal behavior in interpersonal relations. Boston: Allyn & Bacon; 1995. p. 195–217.

[65] Tannen D. Introduction (framing in discourse). In: Tannen D, editor. Framing in discourse. New York: Oxford University Press; 1993. p. 3–13.

[66] Kang K, Freedman S, Mataric M, Cunningham M, Lopez B. A hands-off physical therapy assistance robot for caridac patients. In: IEEE international converence on rehabilitation robotics (ICORR-05). Chicago, IL; 2005.

[67] Eriksson J, Mataric M, Winstein C. Hands-off assistive robotics for post-stroke arm rehabilitation. In: IEEE international conference on rehabilitation robotics (ICORR-05). Chicago, IL; 2005.

[68] Pollack M, Engberg S, Matthews J, Thrun S, Brown L, Colbry D, et al. Pearl: a mobile robotic assistant for the elderly. In: AAAI workshop on automation as Eldercare; 2002.

[69] Brown G, Yule G. Discourse analysis. Cambridge: Cambridge University Press; 1983.

[70] Dahlback N, Jonsson A, Ahrenberg L. Wizard of Oz studies: why and how. In: IUI 93; 1993. p. 193–9.

[71] Oviatt S. Predicting spoken disfluencies during human-computer interaction. Comput Speech Lang 1995;9:19–35.

[72] Shortliffe E, Friedman CP, Wyatt JC, Smith AC, Kaplan B. Evaluation methods in medical informatics. Berlin: Springer; 1997.

[73] Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005;59(1):21–30.

[74] Patel V, Arocha J, Kushniruk A. Patients' and physicians' understanding of health and biomedical concepts: relationship to the design of EMR systems. J Biomed Inform 2002;35(1):8–16.

[75] Walker M, Litman D, Kamm C, Abella A. PARADISE: a framework for evaluating spoken dialogue agents. In: Maybury MT, Wahlster W, editors. Readings in intelligent user interfaces. San Francisco, CA: Morgan Kaufmann Publishers, Inc.; 1998. p. 631–41.

[76] Bohlin P, Bos J, Larsson S, Lewin I, Mathesin C, Milward D. Survey of existing interactive systems. 1999.

[77] Bernsen N, Dybkjaer L. A methodology for evaluating spoken language dialogue systems and their components. In: Second international conference on language resources and evaluation; 2000. p. 183–8.

[78] Woodbridge P, Lowery K, Gates M, Woodbridge N, Moore R. Telehealth case management scoring and display algorithms. In: Medinfo; 2004. p. 1909.

[79] Fitzmaurice G, Laird N, Ware J. Applied longitudinal analysis. Hoboken, NJ: Wiley; 2004.

[80] Revere D, Dunbar P. Review of computer-generated outpatient health behavior interventions: clinical encounters "in Absentia. JAMA 2001;8:62–79.

[81] Prochaska J, Marcus B. The transtheoretical model: applications to exercise. In: Dishman R, editor. Advances in exercise adherence. Champaign, IL: Human Kinetics; 1994. p. 161–80.

[82] Glanz K, Lewis F, Rimer B. Health behavior and health education: theory, research, and practice. San Francisco, CA: Jossey-Bass; 1997.

[83] Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1997;84:191–215.

[84] Corkrey R, Parkinson L. Interactive voice response: review of studies 1989–2000. Behav Res Methods Instrum Comput 2002;34(3):342–53.

[85] Piette J. Interactive voice response systems in the diagnosis and management of chronic disease. Am J Manag Care 2000;6(7):817–27.

[86] Delichatsios HK, Friedman R, Glanz K, Tennstedt S, Smigelski C, Pinto B, et al. Randomized trial of a "Talking Computer" to improve adults' eating habits. Am J Health Promot 2001;15(4):215–24.

[87] Pinto B, Friedman R, Marcus B, Kelley H, Tennstedt S, Gillman M. Effects of a computer-based, telephone-counseling system on physical activity. Am J Prevent Med 2002;23(2):113–20.

[88] Ramelson H, Friedman R, Ockene J. An automated telephone-based smoking cessation education and counseling system. Patient Educ Couns 1999;36:131–44.

[89] Farzanfar R, Locke S, Vachon L, Charbonneau A, Friedman R. Computer telephony to improve adherence to antidepressants and clinical visits. In: Ann Behav Med Annual meeting supplement; 2003. p. S161.

[90] Young M, Sparrow D, Gottlieb D, Selim A, Friedman R. A telephone-linked computer system for COPD care. Chest 2001;119:1565–75.

[91] Bickmore T, Caruso L, Clough-Gorr K, Heeren T. "It's just like you talk to a friend" relational agents for older adults. Interacting With Computers (to appear).

[92] King A, Friedman R, Marcus B, Napolitano M, Castro C, Forsyth L. Increasing regular physical activity via humans or automated technology: 12-month results of the CHAT trial. In: 25th Annual meeting of the soceity of behavioral medicine; 2004; Baltimore, MD.

[93] Adams W, Fuhlbrigge A, Miller C, Panek C, Gi Y, Loane K, et al. TLC-Asthma: an integrated information system for patient-centered monitoring, case management, and point-of-care decision support. In: AMIA; 2003. p. 1–5.