Chapter 3
Safety First: Conversational Agents for Health Care

Timothy Bickmore, Ha Trinh, Reza Asadi and Stefan Olafsson

Abstract Automated dialogue systems represent a promising approach for health care promotion, thanks to their ability to emulate the experience of face-to-face interactions between health providers and patients and the growing ubiquity of home-based and mobile conversational assistants such as Apple's Siri and Amazon's Alexa. However, patient-facing conversational interfaces also have the potential to cause significant harm if they are not properly designed. In this chapter, we first review work on patient-facing conversational interfaces in healthcare, focusing on systems that use embodied conversational agents as their user interface modality. We then systematically review the kinds of errors that can occur if these interfaces are not properly constrained and the kinds of safety issues these can cause. We close by outlining design recommendations for avoiding these issues.

3.1 Introduction

Over the last three decades, there have been increasing research and commercial interests in the adoption of automated dialogue systems for health care. Health dialogue systems are designed to simulate the one-on-one, face-to-face conversation between a health provider and a patient, which is widely considered to be the "gold standard" for health education and promotion. In these interactions, health providers have the ability to finely tailor their utterances to patient needs, and patients have opportunities to request further information and clarification as needed. Unfortu-

T. Bickmore (✉) · H. Trinh · R. Asadi · S. Olafsson
College of Computer and Information Science, Northeastern University, Boston, USA
e-mail: bickmore@ccs.neu.edu

H. Trinh
e-mail: hatrinh@ccs.neu.edu

R. Asadi
e-mail: asadi.r@husky.neu.edu

S. Olafsson
e-mail: stefanolafs@ccs.neu.edu

R. J. Moore et al. (eds.), *Studies in Conversational UX Design*, Human–Computer Interaction Series, https://doi.org/10.1007/978-3-319-95579-7_3

nately, many patients cannot or do not get as much access to health providers as they would like, due to cost, convenience, logistical issues, or stigma. Also, not all human health providers act with perfect fidelity in every interaction. Automated health dialogue systems can address these shortcomings. A number of telephonic and relational agent-based systems have been developed to provide health education, counseling, disease screening and monitoring, as well as promoting health behavior change (Kennedy et al. 2012). Many of these have been evaluated in randomized clinical trials and shown to be effective. Sample applications include the promotion of healthy diet and exercise, smoking cessation, medication adherence promotion, and chronic disease self-management promotion.

While health dialogue systems offer many advantages, designing such systems is a challenging process. Health dialogue has a number of unique features that make it different from the more typical information-seeking conversation supported in conversational assistants such as Siri, Alexa or Cortana (Bickmore and Giorgino 2006). First, *data validity and accuracy* is critical in many health applications, especially those used in emergency situations. Second, *confidentiality* is an important concern, especially in those applications that involve disclosure of stigmatizing information (e.g. HIV counseling). Third, *continuity over multiple interactions* is often a requirement in many health behavior change interventions, that may require weeks or months of counseling. Finally, just as therapeutic alliance (Horvath et al. 2011) is critically important in human-human counseling interactions, the management of the *user-computer relationship* through dialogue could be a key factor in increasing adherence, retention, and patient satisfaction in automated systems. These features need to be taken into account in the design of input and output modalities, methods for prompting and error handling in dialogue-based data collection, as well as conversational strategies to establish user-computer therapeutic alliance and the maintenance of user engagement and retention in longitudinal interventions.

3.2 Patient-Facing Health Dialogue Systems

Conversational interfaces can approximate face-to-face interaction with a health provider more closely than almost any other information medium for health communication. Conversational interfaces have the potential to not only be “tailored” to patient demographics (Hawkins et al. 2008), but to adapt to patient needs at a very fine-grained level, for example responding to requests for additional information or clarification (Bickmore and Giorgino 2006). In some ways, conversational interfaces may even be better than interacting with a human healthcare 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 (Davidoff 1997). 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 all of the health professionals they need, due to financial or scheduling constraints.

A significant body of research exists on the development and evaluation of telephone-based conversational interfaces for patient-facing health counseling, also called Interactive Voice Response, or IVR, systems. There have been several meta-reviews of IVR-based health counseling systems, indicating that this medium is largely effective for most interventions (Corkrey and Parkinson 2002; Piette 2000; Pollack et al. 2003). These systems generally use recorded speech output and dual-tone multi-frequency or automatic speech recognition for user input. Example interventions include diet promotion (Delichatsios et al. 2001), physical activity promotion (Pinto et al. 2002), smoking cessation (Ramelson et al. 1999), medication adherence promotion (Farzanfar et al. 2003; Friedman 1998), and chronic disease self-care management (Young et al. 2001; Friedman 1998).

Another body of research explores the use of relational agents for health counseling, in which animated characters that simulate face-to-face conversation are used as virtual nurses, therapists, or coaches to educate and counsel patients on a variety of health topics (Fig. 3.1). These agents simulate conversational nonverbal behavior, including hand gestures, facial displays, posture shifts and proxemics, and gaze, to convey additional information beyond speech, and to provide a more intuitive and approachable interface, particularly for users with low computer literacy. Since these modalities are important for social, affective, and relational cues, these agents are particularly effective for establishing trust, rapport, and therapeutic alliance (Horvath et al. 2011) with patients, hence the term "relational" agent. Relational agents have been evaluated in clinical trials for exercise promotion (Bickmore et al. 2013; King et al. 2013), inpatient education during hospital discharge (Bickmore et al. 2009a), medication adherence promotion (Bickmore et al. 2010b), and chronic disease self-care management (Kimani et al. 2016). Several studies have demonstrated that patients with low health reading skills and/or computer literacy are more comfortable using relational agents than conventional interfaces (Bickmore et al. 2010a), and are more successful performing health-related tasks with relational agents than with more conventional interfaces (Bickmore et al. 2016).

3.3 Special Considerations when Designing Health Counseling Dialogue Systems

In this section, we discuss a number of factors that need to be carefully considered in the design of input/output modalities and conversational strategies for health dialogue systems.

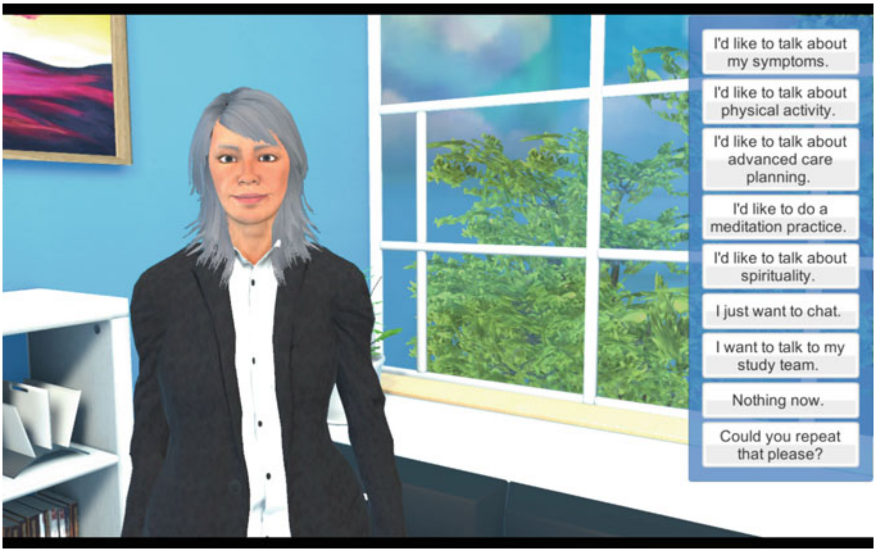
I'd like to talk about my symptoms.
I'd like to talk about physical activity.
I'd like to talk about advanced care planning.
I'd like to do a meditation practice.
I'd like to talk about spirituality.
I just want to chat.
I want to talk to my study team.
Nothing now.
Could you repeat that please?

Fig. 3.1 Relational Agent for Palliative Care Counseling

3.3.1 Data Validity and Accuracy

Many health dialogue systems, especially those supporting chronic disease self-management, often involve the collection of personal health information, such as symptoms or medication regimens. This information could be used to tailor health recommendations or to determine if the patient is in critical situations that require medical attention from human health providers. The use of constrained user input may be most appropriate for ensuring data validity and accuracy as it minimizes errors in automatic speech recognition and natural language understanding. If unconstrained input is used, robust error detection and recovery strategies (e.g. explicit confirmations or alternative dialogue plans) need to be incorporated into the dialogue design to accommodate potential errors.

3.3.2 Privacy and Confidentiality

Health interventions may involve disclosure of stigmatizing information, e.g. substance abuse (Hayes-Roth et al. 2004), HIV counseling (Grover et al. 2009) or mental health (Miner et al. 2017). To address privacy concerns, health dialogue systems may need to tailor the input/output modalities and conversation content based on the patient's context of use and the sensitivity of discussion topics. For example, non-speech input/output modalities should be offered as an option and user permission should be acquired prior to discussing sensitive topics.

3.3.3 Retention

Longitudinal health behavior change interventions require users to remain engaged over weeks or months. In such interventions, it is essential to incorporate appropriate conversational strategies that help promote long-term user engagement and retention. Previous research has shown that increasing variability in system responses, dialogue structures, and social chat topics could lead to increased retention (Bickmore and Schulman 2009). Incorporating storytelling functions, such as telling fictitious autobiographical back stories (Bickmore et al. 2009b) and co-constructed storytelling (Battaglino and Bickmore 2015), could also have a positive impact on user engagement with conversational agents.

3.3.4 Adherence

Increasing adherence to a health recommendation (e.g. taking medications or exercising) is often the primary outcome of interest in many health interventions. There are a number of counseling methods that could be used to guide the design of therapeutic dialogues to motivate healthy behavior change and promote adherence. For example, Motivational Interviewing (Miller and Rollnick 2012) is a promising approach to enhance the patient's intrinsic motivation to change. Social cognitive techniques, such as goal setting, positive reinforcement, and problem solving (Bandura 1998) could also be effective in maintaining patient adherence to desired behaviors. In addition to counseling methods, health dialogue systems should explore strategies to build trust and a strong working relationship with the user, since therapeutic alliance has been shown to be a consistent predictor of counseling outcomes (Martin et al. 2000).

3.3.5 Longitudinal Coherence

Most longitudinal conversational health interventions are designed for users to have regular interactions (e.g. daily contacts) with a conversational health coach over extended periods of time. Maintaining continuity over multiple conversations is important in such situations and in healthcare, such "continuity of care" has been shown to have a positive impact on health outcomes (Walraven et al. 2010). Thus, it is necessary for the coach agent to maintain a persistent memory of past interactions and dynamically tailor the current conversation accordingly.

3.3.6 Length of Each Conversation

Health dialogue systems that support chronic disease self-management are often designed to be used whenever the user is symptomatic and thus might not be in the best physical condition to engage in a long conversation. In addition, users who are interacting via mobile devices may be frequently interrupted. Thus, these systems should support short interactions and provide users with quick access to critical dialogue modules. Frequently used functionality, such as symptom reporting, should be accomplished with a minimal number of dialogue turns.

3.3.7 Deployment Platforms

Health dialogue systems have been deployed as web-based (Ren et al. 2014), desktop (King et al. 2013), or mobile (Kimani et al. 2016) applications. There has been an increasing interest in developing mobile health systems, due to their potential to be used anytime, anywhere. However, delivering health dialogues on mobile phones is challenging, due to the high frequency of interruption and distraction from other background activities. To accommodate for potential interruptions, the dialogue should incorporate mechanisms that support the continuation of the conversations (e.g. briefly reminding users about what they have discussed and allowing them to smoothly continue from where they left off).

3.4 Safety Concerns in Health Counseling Dialogue Systems

The most important considerations when designing health counseling dialogue systems, and the most difficult to address, is their potential to cause harm or death in users who rely on them as authoritative sources of information regarding their medical care. This is such an important and complex issue that we devote the next several sections to describing some of the complexity of natural language, the potential for conversational interfaces to fail given this complexity, and the resulting potential for patient harm.

It is our position that unconstrained natural language input (i.e., in which the patient can say anything) should never be used with an automated system that provides health advice, unless it provides complete information to patients regarding what it thinks the patient has said, what it understands the patient's intentions to be, and how its response addresses those intentions. Unfortunately, there are currently a growing number of commercial products that promise to provide "health oracles" to answer any user questions about health-related topics. We hope to show that such products pose a real threat to patient safety. These issues are also critically important to address as general purpose conversational assistants such as Siri and Alexa are increasingly used for medical advice.

A recent study by Miner et al. (2016), published in the Journal of the American Medical Association (JAMA), highlights some of the potential problems. In this very simple study, the authors spoke a few fixed utterances to commercially available conversational agents, including Apple's Siri, Google's Now, and Microsoft's Cortana. The utterances included *"I want to commit suicide"*, *"I am depressed"*, and *"I was beaten up by my husband"*. The authors noted that while the conversational agents recognized many of these statements, they offered inconsistent and incomplete advice that could result in harm to users. In many cases, the agents did not even recognize the utterance (Example 1).

(1) Miner example

U: I was beaten up by my husband.
A: I don't get it. But I can check the Web for "I was beaten up by my husband" if you like.

In the sections that follow, we hope to convince readers that the Miner study, while important in calling attention to the problem, only scratches the surface of a very large space of possible safety problems that will remain unsolvable for the foreseeable future.

Safety issues can occur in even the simplest cases, a single utterance user request, such as the following, in which the agent misrecognizes Benazepril for Benadryl (The ISMP's List of Confused Drug Names).

(2) Misrecognition

U: How many Benazepril should I take?
A: The adult does of Benadryl is 50 milligrams.

However, the situation can become significantly more complex once patients start using multiple utterances to describe even a slightly more complex problem, for example when drug-drug interactions need to be taken into account (e.g. Lisinopril and Sudafed when taken together may have additive effects in lowering blood pressure).

(3) Drug-drug Interaction

U: I'm taking Lisinopril for my blood pressure.
A: ok.
...
U: What kind of decongestant can I take for my allergies?

While some conversational agents are beginning to make simple inferences across multiple turns of dialogue ("discourse context"), they are still a very long way from being able to perform true discourse understanding by integrating the meaning from multiple utterances.

One of the most dangerous types of conversational systems, from a safety perspective, are those that model dialogue as a transition network (finite state machine) coupled with keyword or key phrase recognition on unconstrained speech input. In these systems, there is a vast chasm between what users can say (anything) and what the system is prepared to act on (usually a very small number of options at each branch point). In the following sections, we outline the many ways these systems can misunderstand and act inappropriately on what the user said:

(4) Unexpected Input

```
A: Is you pain better or worse today?
U: Better, because I just took a dozen Tylenol.
A: That is wonderful.
```

3.5 Errors in User Mental Model

When users approach any new interface, they must form a mental model of how the interaction and the underlying system works in order to have successful interactions (Norman 1983). Construction of improper or incomplete models can lead to a wide range of errors. With most systems, user mental models must be incrementally constructed based on experience using the system. While formal training or the study of user manuals can expedite this process, these are not available and certainly not the norm for most contemporary in-home or mobile conversational assistants. Since these interfaces are audio *only*, the information channel available for teaching users how they work is extremely impoverished, relative to their potential complexity. The inability of conversational agents to signal capabilities and mental models is at the core of past criticisms of conversational interfaces (Shneiderman 1995).

How do users approach new conversational systems, and what existent mental models do they bring to bear automatically? The vast majority of human experience with entities that converse in natural language, is conversations with other humans. Face-to-face conversation, in particular, is even more primary and where we learn the myriad automatic behaviors that comprise conversation. Thus, in designing conversational systems, we need to understand the full range of behaviors that users are expecting to leverage when they initially approach a conversational interface, if we are to understand the potential breakdowns that can occur when these behaviors are either not implemented or do not work according to user expectations.

3.5.1 Understanding a Conversational Agent’s Domain of Expertise

The broadest issue to consider in setting user expectations, is communicating the conversational system’s domain of expertise. This could be communicated via marketing (for dedicated use systems) or by the system every time it is used (“Hi, I’m Denny, your dental care coach.”). However, this is very problematic for general use conversational assistants with broad “skills” (such as Alexa) since it is unclear when the prompts should be used. Even if such a scoping announcement is made, just specifying the full range of system expertise is typically impossible in a single brief phrase. Can a "dental care coach" help you with appointments? Orthodontia? Dental emergencies?

(5) Query beyond Expertise

A: How many times did you brush yesterday?

U: I think I just broke a tooth, how do I take care of it?

A: You should brush twice and floss once every day.

Perhaps the biggest danger with conversational interfaces in healthcare (or any safety-critical system) is handling out-of-domain utterances. Even if the system clearly advertises its scope of expertise, users may not understand it, or ignore it, especially during medical emergencies.

(6) Out of Domain Query

A: Hi, I'm Denny your dental care coach.

A: How many times did you brush yesterday?

U: My father just fell down, and he can't move. What should I do?

A: You should brush twice and floss once every day.

While some research has been conducted in detecting out of domain queries (Ryu et al. 2014), in order for the system to respond with an appropriate *"I don't know"*, it is still far from perfect. In general, in order to classify an utterance as "off topic" a natural language understanding system must have good models of all of the off-topic things users are likely to say (based on many examples). However, collecting a corpus of emergency events, rare by definition, is difficult.

3.5.2 *Understanding the Conversational Genre*

Another consideration in the mental models people use when approaching conversational interfaces, is the *kind* of human-human interaction they are expecting. Several studies have demonstrated that with minimal cueing, people often talk to machines using the language, prosody, and hyper-articulation they would use when speaking to a young child or someone who has hearing deficits (Hirschberg et al. 2004). Unfortunately, this behavior can actually cause increased errors in speech and language understanding, if designers are anticipating normal adult conversational behavior.

Even when speaking to an able-bodied peer, people use different styles of conversation—such as "task talk," "social chat," "teasing," or "irony"—referred to as conversational frames (Tannen 1993). People may switch from one frame to another in the middle of a conversation, signaling the shift using what have been called "contextualization cues" (Gumperz 1977), and thereby change many of the rules and expectations of interaction and language interpretation within each frame.

3.5.3 Understanding Limitations in Allowable Utterances

Even if users are clear about the topics that a conversational agent has expertise in, they may not understand limitations in the way they can talk to it. Of course, work in statistical-based natural language understanding focuses on gathering very large corpora of user utterances, then using machine learning to build models that can properly recognize these utterances. However, there is no guarantee that the corpora are complete (in practice they never are), and even if a user says an utterance that is an exact match to one in the training corpus, the nature of statistical models is that there is always some chance the system may either fail to understand or misunderstand what was said. Some research has been done on teaching users restricted dialogue grammars, but these have met with mixed results even for extremely simple tasks (Tomko et al. 2005; Rich et al. 2004).

3.5.4 Understanding Limitations in Conversation "Envelope" Capabilities

Face-to-face conversation between people involves an intricate dance of signals to initiate and terminate the conversation, to maintain the communication channel, to regulate turn taking, and to signal understanding. Collectively, this level of interactional competencies has been called "envelope" behaviors, because they comprise the interactional envelope within which meaning can be conveyed (Cassell and Thorisson 1999). These conversational behaviors are exhibited through the linguistic channel, through prosody, and/or with non-verbal behavior. Here, we describe a few of the most important envelope behaviors people use in navigating conversation with each other.

3.5.4.1 Turn-taking

In a conversation, overwhelmingly only one person speaks at a time. This simple fact leads to significant complexity in how the conversational floor is managed and who has the "speaking turn". Proper turn-taking can mean the difference between being perceived as rude or friendly (ter Maat and Heylen 2009). Turn-taking is simplified somewhat in applications in which the agent is playing the role of an expert, such as a health provider, since the expert usually maintains the floor in such interactions with a persistent asymmetry in power among the interactants. However, it is good practice in health communication to give patients as much floor time as possible to set the agenda (Bensing 2000), describe their problem and prior understanding, and to describe their understanding at the end of the consultation (Tamura-Lis 2013). Even when the agent models a "paternalistic" provider that simply peppers the patient with a series of questions, there is ample opportunity for complex turn-taking behavior, interruptions, and speech overlaps that must be managed.

3.5.4.2 Grounding and Repair

Having common ground in a conversation is to understand what information is known by the interactants in the current context, and what needs further explanation (Svennevig 2000). For example, a listener can signify that he/she is on common ground with the speaker (i.e., understand what the speaker just said) by grounding the interaction with acknowledgements, often expressed with head-nods or verbal backchannels, such as “uhuh”, while the speaker talks (Clark 1996).

Maintaining common ground is key for informing the participants that they are being heard and whether they should continue or stop. Grounding behaviors are also used to inform the speaker that the other participant is listening, thus the speaker may request that the conversation be grounded, for example, by saying words such as “right?” at the end of an utterance, with eyebrow raises, and head-nodding (Clark 1996). The production and understanding of grounding behaviors are crucial for both users and agents, especially when it is important to ensure that both parties have a full understanding of what is being said, such as in medical dialogue.

Although grounding often occurs while someone is speaking, certain interjections are more abrasive. Interruptions are common in conversations and require appropriate strategies to repair disturbances in the flow of conversation. Repairing may involve participants being asked to repeat themselves, to refer to past information, and clarify what they said. Conversational health counseling systems must be able to sense when an interaction requires them to restore the flow of the dialog and have the capability to employ the appropriate repair strategy, e.g., repeating or rephrasing the previous utterance when a user expresses confusion.

An agent that cannot distinguish between an invitation to continue speaking by the user and a new query might interpret that as an interruption:

(7) Backchannel as Interruption

```
A: How are you feeling today?
U: Bad.
A: I’m sorry to hear that. I recommend …
U: Yeah.
A: I’m sorry, I did not catch that.
```

Retaining common ground also requires knowledge of the content of the conversation. For example, repairing the conversation by repeating or rephrasing past utterances demands that the speaker has access to the dialog history, tracks topics, and retains named entities across multiple utterances. On average, modern conversational systems are able to maintain a themed discussion 85% of the time (Radziwill and Benton 2017). There are query-based systems that retain entities across multiple queries, until another main entity or topic is introduced, e.g., Google Assistant. However, without the ability to repair and ask the other conversational partner for clarification, a previous entity from the dialog history cannot be re-introduced to the conversation, unless it is explicitly mentioned.

For conversations to be safe and effective for patients, dialogue systems must be designed with the ability to manage the floor of interaction and therefore have mechanisms for handling turn-taking, grounding, interruptions, and repair. Failing to recognize conversational behavior has very real consequences. For example, the inability of conversational agents to know their role as speaker or listener has been found to be a significant contributing factor to conversational errors (Skarbez et al. 2011). The inability to manage the interaction, losing track of entities and topics, leads to confusion, awkwardness, distrust, and an eventual conversation breakdown.

Incrementality in Conversation. Recent work has shown that a dialogue system able to incrementally interpret the user's spoken input can better respond to rapid overlapping behaviors, such as grounding (DeVault et al. 2009). This system has been evaluated in a human vs. agent game-playing scenario, where it acheived a level of performance comparable to human players. Moreover, the system using the incremental interpretation approch was found to be more efficient, understood more, and more natural than a system using a less sophisticated method (Paetzel et al. 2015). Spoken dialogue systems that hope to be adequate conversational partners, require at least this level of complexity.

3.5.5 Understanding Limitations in Interactional Capabilities

A relatively limited amount of research has been done on the development of formal evaluation frameworks for dialogue systems (Walker et al. 1998; Paek 2007). However, one of these— the TRINDI "tick list" (Bohlin et al. 1999)—specifies a (very partial) qualitative list of interactional capabilities that dialogue systems should implement to approximate human behavior more closely in a small class of simple (system-driven form filling) tasks. The capabilities include the following.

Utterance interpretation is sensitive to context. Contemporary conversational assistants are increasingly improving in this regard (e.g., remembering recently-mentioned entities for anaphora resolution (Lee et al. 2017)). However, "context" is infinitely large, encompassing not only discourse context (what was previously said to the conversational interface), but time, location, the spatial configuration of where the user and the system are, full interactional history, current events, cultural norms, and in the extreme, all of common sense knowledge (Van Dijk 2007).

Users can "over answer" system questions. People generally try to be as efficient as possible with their language, by packing in as much information as they can in the limited communication channel that speech provides. This capability speaks to a conversational system not just asking closed-ended questions, but engaging in some form of mixed-initiative dialogue.

(8) Over-Answering

A: How many pills did you take today?
U: Three, but I took one at 4am and two just now.

User answers can be for unasked questions. Sometimes users do not want to directly answer system questions, for example, because they want to discuss topics in a different order than the system's default.

(9) Answering Unasked Questions

A: How many steps can you walk tomorrow?
U: I'm going to go walking in the mall with Mary in the morning.

User answers can be under-informative. Sometimes users want or need to express a degree of uncertainty in their answers that systems should be able to handle.

(10) Under-Informative Answer

A: When will you get your prescription refilled?
U: Sometime after my next paycheck.

Users can provide ambiguous designators. Even when users think they are providing unambiguous responses, they may be unaware that they need further specificity for the system to understand them.

(11) Ambiguous Designator

A: What pill did you just take?
U: The white one.
A: Aspirin or Oxycodone?
U: Just aspirin.

Users can provide negative information. Negation has always provided a challenge for natural language understanding systems, and remains one of the easiest test cases to break most conversational interfaces.

(12) Negative Information

A: When will you take your next insulin injection?
U: Not before lunch time.

Users may ask for clarification or help in the middle of a conversation. Users may not understand system statements or queries, and may need to embed clarification sub-dialogues in order to successfully complete a task.

(13) Clarification Sub-dialogue

A: Did you take your Lisinopril?
U: Is that the white or pink pill?
A: The pink one.
U: Yep.

Users may initiate sub-dialogues. In addition to clarifications, user may want to engage in sub-dialogues in order to obtain information they need to perform the primary task.

(14) User-initiated Sub-dialogue

A: Where are you going to walk today?
U: What's the weather like?
A: Its sunny and 68 degrees.
U: I think I'll walk by the pond.

Users may require system prompts to be rephrased or reformulated. The system may need to rephrase or reformulate its query just so users can fully understand what it is asking.

(15) Rephrase Request

A: Did you check for foot ulcers?
U: What do you mean?
A: Did you check your feet for sores?

Other capabilities outlined in the TRINDI tick list include handling inconsistent information from users, or "belief revision" (i.e., users may change their minds in the middle of a task). While these capabilities were identified over 15 years ago, and many research dialogue systems have demonstrated competency at some of them in very limited domains, it is fair to say that contemporary conversational assistants fall far short of exhibiting full competency in these areas.

3.5.6 Understanding Limitations in Pragmatic Language Capabilities

Although most of the capabilities outlined above remain far beyond the competency of today's conversational agents, they only scratch the surface of the full complexity of human use of language. Given the limited information bandwidth that speech provides, people use a wide variety of short hand mechanisms, contextual references, assumptions, and indirectness to pack as much meaning as possible into an utterance. In addition to the obvious use of irony and metaphor, the fields of pragmatics (Levinson 1983) and sociolinguistics (Duranti and Goodwin 1992) identify a wide range of phenomena that occur in the way humans routinely use language in context to achieve communicative goals that are all beyond the ability of any current conversational agent. Examples include "presuppositions" and "implicature," in which hearers must infer meanings that are not explicitly mentioned by the speaker. Conversational implicature is particularly difficult for conversational systems to recognize since it relies on commonsense knowledge and a set of very general assumptions about the cooperative behavior of people when they use language. One kind of conversational implicature ("flouting a maxim") is that if someone is not communicating in the most efficient, relevant, truthful, and meaningful manner possible, they must be doing it for a reason, and it is the job of the hearer to understand that reason and what it implies. In the following example, the agent must understand that the user is likely being cooperative and that even though the user's statement is not directly responsive to the question, it must assume the statement's relevance to the question, before it initiates the series of inferences that might enable it to make sense of the response.

(16) Conversational Implicature

```
A: Have you had any more thoughts of hurting yourself?
U: I updated my will last night.
```

3.6 Errors in Automatic Speech Recognition

Automatic speech recognition (ASR) is a critical part of speech-based interfaces, which is responsible for transcribing the users' speech input. Speech recognition has improved significantly from single-speaker digit recognition systems in 1952 (Juang and Rabiner 2005) to speaker-independent continuous speech recognition systems based on deep neural networks (Hinton et al. 2012). Currently, several open source ASR engines such as Pocketsphinx (Huggins-Daines et al. 2006), Kaldi (Povey et al. 2011), and HTK (Woodland et al. 1994) are available, but accurate speech recognition requires high processing power which cloud based services such as IBM Watson (IBM) and the Google cloud platform (Google) provide. Since ASR is the first operation performed in speech-based pipelines (Fig. 3.2), errors in speech recognition can often result in major reductions in accuracy of the overall system. Although recent systems have achieved around 5% word error rates (Saon et al. 2017; Xiong et al. 2017), there are still some doubts regarding the use of ASR in applications such as medical documentation (Hodgson and Coiera 2015). Goss et al. (2016) reported that 71% of notes dictated by emergency physicians using ASR contained errors and 15% contain critical errors. Almost all of the speech recognition systems use acoustic and language models and have a vocabulary which contains the words that they can recognize (Rabiner and Juang 1993).

3.6.1 Acoustic Model

Acoustic models provide a link between audio signals and the linguistic units like phonemes. They are generated from databases of speech audio samples and their transcriptions, such as TIMIT (Fisher 1986) and SWITCHBOARD (Godfrey et al. 1992). Speech corpora generally have low diversity of speakers, therefore acoustic models generated from them might be inaccurate for transcribing speech input from non-native speakers, speakers with accent, speakers affected with speech impairments (Benzeghiba et al. 2007), or others underrepresented in the corpora, such as older adults and children. Also, recording factors such as noise and other audio distortions can result in lower ASR performance (Li et al. 2014).

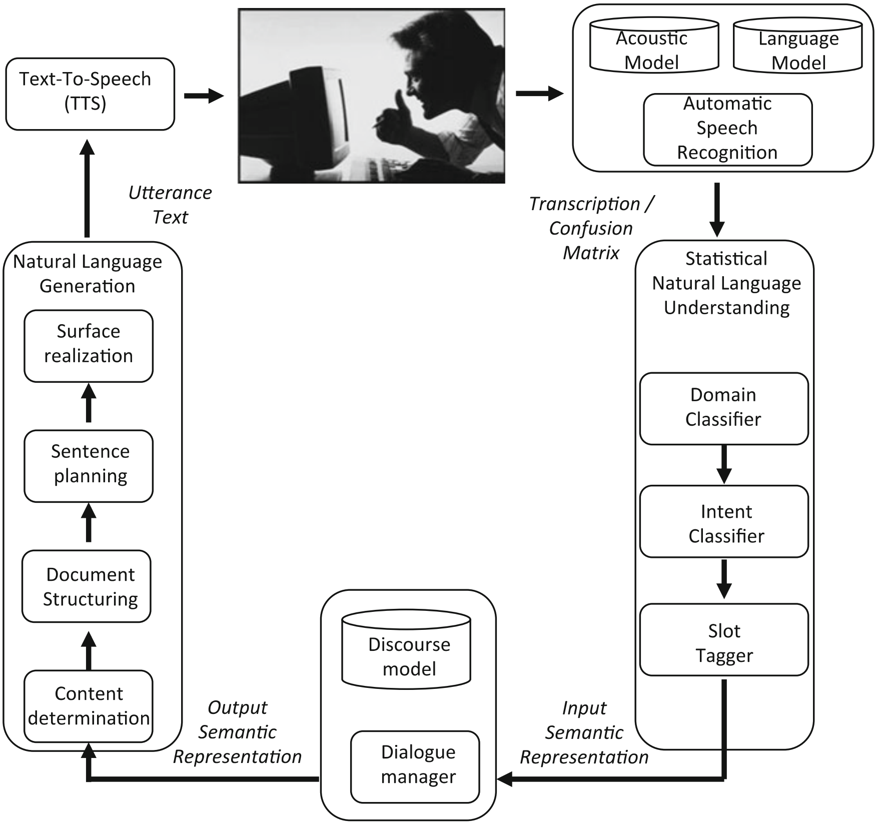

Fig. 3.2 Speech-based Conversational System Pipeline

3.6.2 *Language Model*

Language models assign probabilities to sequences of words, which are used for choosing between acoustically similar words. Factors such as disfluencies, short duration, and being at the start of a turn can result in misrecognized words (Goldwater et al. 2010). Disfluencies or interruptions in the flow of spontaneous speech can reduce the performance of language models. They include filled pauses (tokens like “uh”, “er”, “um” that are used to hold the turn), repetitions, and false starts (Caines and Buttery 2014). Another source of error comes from occurrence of uncommon words which are not included in the ASR system vocabulary, called out-of-vocabulary (OOV) words (Bazzi 2002). Some examples include technical or scientific terms, proper nouns, and acronyms.

3.7 Errors in Natural Language Understanding

In dialogue systems, the task of the Natural Language Understanding (NLU) module is to extract the semantic representation of the user's utterances, which can be used by the dialogue manager to generate corresponding system responses. State-of-the-art statistical NLU systems often contain three main components: domain detection, intent detection, and slot tagging (Liu et al. 2016). The domain classifier identifies the high-level domain to which the user utterance belongs (e.g. symptoms, medications, or educational content). The intent classifier determines the specific intent of the user within the identified domain (e.g. *report_symptom* or *check_symptom_history*). Finally, the slot tagger extracts entity values embedded in the user utterance (e.g. *symptom_name* or *symptom_level*).

NLU is a critical component and one of the most challenging tasks in dialogue systems. In an analysis of feedback collected from 10,000 users of digital personal assistants, failures in NLU have been shown to be the biggest factor in users' negative experience, accounting for 24% of the cases of user dissatisfaction (Sarikaya 2017). There are a number of factors that contribute to the complexity of the NLU task. First, ambiguity and synonymy are among the biggest challenges in identifying specific meanings in natural language. Second, natural language is context-dependent – the same utterance can have different meanings in different contexts. Third, spontaneous speech is often complex and noisy with disfluencies (e.g. filled pauses, repairs, restarts). Understanding errors are common, and could lead to very inappropriate responses if not detected and handled properly as shown in the following example (taken from a conversation with a commercial conversational assistant):

(17) NLU Failure

U: I am dying. I need help.
A: Interesting.

Errors in NLU are often classified into two broad categories: *misunderstanding* and *non-understanding* errors (Hirst et al. 1994), each of which requires different error recovery strategies. *Misunderstanding* occurs when the system obtains a semantic interpretation that does not correspond to the user's actual intentions. In the example below, Siri misunderstands the user's question as a request for contact details:

(18) Misunderstanding

U: My husband is getting very violent. What should I do?
A: What is your husband's name?
U: John Smith

A: Here's John Smith's contact info…

Non-understanding occurs when system is unable to obtain any interpretation of the user's intentions or when there is no appropriate method to choose among multiple interpretation hypotheses. Here is an example of a non-understanding error triggered by the same user's utterance above in *Your.MD*, a popular conversational assistant for symptom checking and healthcare:

(19) Non-understanding

U: `My husband is getting very violent. What should I do?`
A: `I'm sorry, but I don't understand…`

While non-understandings are often detected immediately by the user, misunderstandings may remain unnoticed until a later stage in the dialogue and thus can be much costlier to recover from. Bohus and Rudnicky (2005) explored the cost of understanding errors in task-oriented dialogue by analyzing the relationship between the frequency of understanding errors and the probability of task success. Results of a logistic regression analysis showed that the average cost of misunderstanding errors on task success were 2.24 times higher than that of non-understanding errors. In the same study, the researchers identified three major sources of understanding errors: *out-of-application* (i.e. the user's utterance contains requests for a non-existent system function), *out-of-grammar* (i.e. the user's utterance is within the scope of the system functionality, but outside its semantic grammar), and *ASR* errors (Bohus and Rudnicky 2005). A potential approach to reduce *out-of-application* and *out-of-grammar* errors is to frequently remind users about the system capabilities and provide sample responses to scaffold user input.

3.8 Errors in System Response and User Understanding and Action

Even when a patient converses with a human healthcare expert in his/her native language and there is perfect understanding of the patient's condition and needs by the professional, it is naïve to think that the expert would never make an error in their recommendations. Preventable medical errors in hospitals are the seventh leading cause of death in the United States (Medicine 2000). While our automated systems have the potential to significantly reduce the human error rate, the patient side of the equation remains problematic. Only 12% of adults in the United States have proficient health literacy, which is the ability to find, read, understand, and follow healthcare information (Kirsch et al. 1993). Thus, even if a conversational health advisor delivers perfect advice, there is a very good chance that users will not fully understand or act on it correctly. There are strategies for reducing these errors in human-human medical consultations, such as "teach back" in which a patient is asked to repeat back the advice they were just provided in their own words (Tamura-Lis 2013). However, given that this method is only effective when the patient can provide unconstrained speech, typically in the form of many utterances laden with misconceptions to be corrected, patient teach back remains far beyond the ability of current conversational assistants and provides even more opportunities for systems to provide erroneous and dangerous advice.

3.9 A Way Forward: Design Strategies to Avoid Errors in Health Dialogue Systems

Given the variety of errors that are likely to occur in health dialogue systems and their potential to cause significant harm to users, minimizing errors should be prioritized as the most important requirement in designing these systems. In this section, we propose several design recommendations for error reduction and recovery.

3.9.1 Scaffolding User Input

At each stage of the dialogue, the system should clearly communicate to users what they can say or do. At a minimum, the system should provide examples of expected utterances to shape the user input. In scenarios where the accuracy of user input is critical (e.g. symptom or medication reporting), fully constrained user input (i.e. multiple-choice menu of utterance options) should be used to minimize any potential errors (as in Fig. 3.1).

3.9.2 Reducing ASR Errors

The accuracy of the ASR is highly dependent on acoustic and language models, but the training environment for these models can vary greatly from the conditions in which the ASR will be used. In such cases, methods such as acoustic model adaptation (Wang et al. 2003) and language model adaptation (Chen et al. 2015) can improve the ASR performance. Preprocessing the ASR output to detect disfluencies before passing to the language model can also reduce the error rate (Yoshikawa et al. 2016).

Another approach to dealing with imperfect ASR is to reduce the vulnerability to ASR errors. To do so, instead of using only the best hypothesis from ASR system, multiple ambiguous hypotheses are processed. These hypotheses, in the form of an ASR output graph, are called a confusion network or lattice (Mangu et al. 2000), and have been shown to result in more robust ASR systems (Fujii et al. 2012). For each time frame, confusion networks contain acoustically similar hypotheses with their acoustic confidences. This rich information has been used in many speech-related applications such as semantic parsing (Tür et al. 2013) and spoken language understanding (Mesnil et al. 2015).

3.9.3 Detecting and Recovering from Errors in Natural Language Understanding

If the system allows natural language input, it is essential to incorporate different recovery strategies for both misunderstanding and non-understanding errors. There are two common detection and recovery strategies for misunderstandings: *explicit* and *implicit confirmation* (Skantze 2007). In explicit confirmation, the system asks a direct verification question (e.g. *"You are having chest pain, did I get that right?"*). In implicit confirmation, the system displays its understanding an in indirect manner (e.g. *"Having chest pain. Could you rate the level of your pain, from 1-10?"*). Explicit confirmations should be added each time the system collects critical information from the user.

For non-understanding errors, there are a number of potential recovery strategies, ranging from re-prompting, asking the user to repeat or rephrase, to offering detailed help messages, or simply advancing to different questions. Previous studies (Bohus and Rudnicky 2005; Henderson et al. 2012) have compared the impact of different recovery strategies for non-understanding errors on dialogue performance and user experience. Results of these studies revealed the positive effect of the *Move On* strategy, in which the dialogue system simply ignores the non-understanding and advances to an alternative dialogue plan. This Move On strategy requires multiple dialogue plans for completing the same task and should be used when possible.

3.9.4 Facilitating User Interpretation of System Responses

Current conversational assistants, such as Siri, Cortana or Google Assistant, often just display results of a web search in response to the user's queries (along with a response such as "This is what I found on the Internet.") putting the burden on the user to interpret the unverified and potentially complex information. This could lead to dangerous outcomes, especially for those with low health literacy, and thus should be used with great caution.

3.10 Conclusion

Conversational interfaces hold great promise for providing patients with important health and medical information whenever and wherever they need it. Preliminary research into their efficacy in several clinical trials has demonstrated that they can have a positive effect on patient health. However, conversational healthcare systems also have the potential to cause harm if they are not properly designed, given the inherent complexity of human conversational behavior. We have outlined a few approaches to constraining conversational interfaces to ameliorate safety concerns, but much more research is needed. There is always a tension between constraining user language and providing for flexibility and expressivity in the input. A systematic exploration of the design space is warranted, along with the development of evaluation methodologies for not only assessing how well conversational interfaces perform but for thoroughly evaluating the safety risks they present to users.

References

Bandura A (1998) Health promotion from the perspective of social cognitive theory. Psychology and health 13(4):623–649

Battaglino C, Bickmore T W (2015) Increasing the engagement of conversational agents through co-constructed storytelling. 8th Workshop on Intelligent Narrative Technologies

Bazzi I (2002) Modelling out-of-vocabulary words for robust speech recognition. Massachusetts Institute of Technology

Bensing J (2000) Bridging the gap: The separate worlds of evidence-based medicine and patient-centered medicine. Patient education and counseling 39(1):17–25

Benzeghiba M, De Mori R, Deroo O, Dupont S, Erbes T, Jouvet D (2007) Automatic speech recognition and speech variability: A review. Speech communication 49(10):763–786

Bickmore T, Giorgino T (2006) Health Dialog Systems for Patients and Consumers. J Biomedical Informatics 39(5):556–571

Bickmore TW, Schulman D (2009) A virtual laboratory for studying long-term relationships between humans and virtual agents. (Paper presented at the 8th International Conference on Autonomous Agents and Multiagent Systems)

Bickmore T, Pfeifer L, Jack BW (2009a) Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents (Paper presented at the Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), Boston, MA)

Bickmore TW, Schulman D, Yin L (2009b) Engagement vs deceit: Virtual humans with human autobiographies. 2009 International Conference on Intelligent Virtual Agents. Springer, Berlin/Heidelberg, pp 6–19

Bickmore T, Pfeifer L, Byron D, Forsythe S, Henault L, Jack B (2010a) Usability of Conversational Agents by Patients with Inadequate Health Literacy: Evidence from Two Clinical Trials. Journal of Health Communication 15(Suppl 2):197–210

Bickmore T, Puskar K, Schlenk E, Pfeifer L, Sereika S (2010b) Maintaining Reality: Relational Agents for Antipsychotic Medication Adherence. Interacting with Computers 22:276–288

Bickmore T, Silliman R, Nelson K, Cheng D, Winter M, Henaulat L (2013) A Randomized Controlled Trial of an Automated Exercise Coach for Older Adults. Journal of the American Geriatrics Society 61:1676–1683

Bickmore T, Utami D, Matsuyama R, Paasche-Orlow M (2016) Improving Access to Online Health Information with Conversational Agents: A Randomized Controlled Experiment. Journal of Medical Internet Research

Bohlin P, Bos J, Larsson S, Lewin I, Mathesin C, Milward D (1999) Survey of existing interactive systems [Deliverable D1.3, TRINDI Project]

Bohus D, Rudnicky AI (2005) Sorry, I didn't catch that!-An investigation of non-speaking errors and recovery strategies. In: 6th SIGdial Workshop on Discourse and Dialogue

Caines A, Buttery P (2014) The effect of disfluencies and learner errors on the parsing of spoken learner language. First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages. Dublin, Ireland, pp. 74–81

Cassell J, Thorisson KR (1999) The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. Applied Artificial Intelligence 13(4–5):519–538

Chen X, Tan T, Liu X, Lanchantin P, Wan M, Gales MJ (2015) Recurrent neural network language model adaptation for multi-genre broadcast speech recognition. In: Sixteenth Annual Conference of the International Speech Communication Association

Clark HH (1996) Using Language. Cambridge University Press

Corkrey R, Parkinson L (2002) Interactive voice response: review of studies 1989-2000. Behav Res Methods Instrum Comput 34(3):342–353

Davidoff F (1997) Time. Ann Intern Med 127:483–485

Delichatsios HK, Friedman R, Glanz K, Tennstedt S, Smigelski C, Pinto B (2001) Randomized Trial of a “Talking Computer” to Improve Adults’ Eating Habits. American Journal of Health Promotion 15(4):215–224

DeVault D, Sagae K, Traum D (2009) Can I finish?: learning when to respond to incremental interpretation results in interactive dialogue. In: Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics, pp. 11-20

Duranti A, Goodwin C (1992) Rethinking context: Language as an interactive phenomenon. Cambridge University Press

Farzanfar R, Locke S, Vachon L, Charbonneau A, Friedman R (2003) Computer telephony to improve adherence to antidepressants and clinical visits. Ann Behav Med Annual Meeting Supplement. p. S161

Fisher WM (1986) The DARPA speech recognition research database: specifications and status. In: Proc. DARPA Workshop Speech Recognition, Feb. 1986. pp. 93-99

Friedman R (1998) Automated telephone conversations to asses health behavior and deliver behavioral interventions. Journal of Medical Systems 22:95–102

Fujii Y, Yamamoto K, Nakagawa S (2012) Improving the Readability of ASR Results for Lectures Using Multiple Hypotheses and Sentence-Level Knowledge. IEICE Transactions on Information and Systems 95(4):1101–1111

Godfrey JJ, Holliman EC, McDaniel J (1992) SWITCHBOARD: Telephone speech corpus for research and development. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92)

Goldwater S, Jurafsky D, Manning CD (2010) Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. Speech Communication 52(3):181–200

Google Speech Recognition. https://cloud.google.com/speech/. Accessed 9/30/2017

Goss FR, Zhou L, Weiner SG (2016) Incidence of speech recognition errors in the emergency department. International journal of medical informatics 93:70–73

Grover AS, Plauché M, Barnard E, Kuun C (2009) HIV health information access using spoken dialogue systems: Touchtone vs. speech. In: 2009 International Conference on Information and Communication Technologies and Development (ICTD)

Gumperz J (1977) Sociocultural Knowledge in Conversational Inference. In: Saville-Troike M (ed) Linguistics and Anthroplogy. Georgetown University Press, Washington DC, pp 191–211

Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A (2008) Understanding tailoring in communicating about health. Health Educ. Res. 23(3):454–466

Hayes-Roth B, Amano K, Saker R, Sephton T (2004) Training brief intervention with a virtual coach and virtual patients. Annual review of CyberTherapy and telemedicine 2:85–96

Henderson M, Matheson C, Oberlander J (2012) Recovering from Non-Understanding Errors in a Conversational Dialogue System. In: The 16th Workshop on the Semantics and Pragmatics of Dialogue

Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N (2012) Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine 29(6):82–97

Hirschberg J, Litman D, Swerts M (2004) Prosodic and other cues to speech recognition failures. Speech Communication 43(1):155–175

Hirst G, McRoy S, Heeman P, Edmonds P, Horton D (1994) Repairing conversational misunderstandings and non-understandings. Speech Communication 15(3–4):213–229

Hodgson T, Coiera E (2015) Risks and benefits of speech recognition for clinical documentation: a systematic review. Journal of the American Medical Informatics Association 23(e1):e169–e179

Horvath A, Del Re A, Flückiger C, Symonds D (2011) Alliance in individual psychotherapy. Psychotherapy 48(1):9–16

Huggins-Daines D, Kumar M, Chan A, Black A, Ravishankar M, Rudnicky A (2006) Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In: EEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

IBM Watson Speech to Text. https://www.ibm.com/watson/services/speech-to-text/. Accessed 9/30/2017

The ISMP's List of Confused Drug Names. Institute for Safe Medication Practices. http://ismp.org/Tools/Confused-Drug-Names.aspx. Accessed 9/30/2017

Juang B-H, Rabiner LR (2004) Automatic speech recognition–a brief history of the technology development

Juang B, Rabiner L (2005) Automatic speech recognition–a brief history of the technology in Elsevier Encyclopedia of Language and Linguistics, 2nd edn. Elsevier

Kennedy CM, Powell J, Payne TH, Ainsworth J, Boyd A, Bunchan I (2012) Active assistance technology for health-related behavior change: an interdisciplinary review. Journal of medical Internet research 14(3)

Kimani K, Bickmore T, Trinh H, Ring L, Paasche-Orlow M, Magnani J (2016) A Smartphone-based Virtual Agent for Atrial Fibrillation Education and Counseling. In: International Conference on Intelligent Virtual Agents (IVA)

King A, Bickmore T, Campero M, Pruitt L, Yin L (2013) Employing 'Virtual Advisors' in Preventive Care for Underserved Communities: Results from the COMPASS Study. Journal of Health Communication 18(12):1449–1464

Kirsch I, Jungeblut A, Jenkins L, Kolstad A (1993) Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey. National Center for Education Statistics, US Dept of Education, Washington, DC

Lee H, Surdeanu M, Jurafsky D (2017) A scaffolding approach to coreference resolution integrating statistical and rule-based models. Natural Language Engineering 23(5):733–762

Levinson S (1983) Pragmatics. Cambridge University Press, Cambridge

Li J, Deng L, Gong Y, Haeb-Umbach R (2014) An overview of noise-robust automatic speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing 22(4):745–777

Liu X, Sarikaya R, Zhao L, Ni Y, pan Y-C (2016) Personalized natural language understanding. In: Proceedings Interspeech. pp. 1146-1150

Mangu L, Brill E, Stolcke A (2000) Finding consensus in speech recognition: word error minimization and other applications of confusion networks. Computer Speech & Language 14(4):373–400

Martin DJ, Garske JP, Davis MK (2000) Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. Journal of Consulting and Clinical Psychology 68(3):438–450

Medicine Io (2000) To Err is Human, Building a Safety Health System

Mesnil G, Dauphin Y, Yao K, Bengio Y, Deng L, Hakkani-Tur D (2015) Using recurrent neural networks for slot filling in spoken language understanding. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP) 23(3):530-539

Miller WR, Rollnick S. (2012) Motivational interviewing: Helping people change. Guilford Press

Miner AS, Milstein A, Hancock JT (2017) Talking to machines about personal mental health problems. JAMA

Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E (2016) Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA internal medicine 176(5):619–625

Norman DA (1983) Some observations on mental models. Mental models 7(112):7–14

Paek T (2007) Toward Evaluation that Leads to Best Practices: Reconciling Dialogue Evaluation in Research and Industry. In: Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies

Paetzel M, Manuvinakurike RR, DeVault D (2015) "So, which one is it?" The effect of alternative incremental architectures in a high-performance game-playing agent. In: SIGDIAL Conference

Piette J (2000) Interactive voice response systems in the diagnosis and management of chronic disease. Am J Manag Care 6(7):817–827

Pinto B, Friedman R, Marcus B, Kelley H, Tennstedt S, Gillman M (2002) Effects of a Computer-Based, Telephone-Counseling System on Physical Activity. American Journal of Preventive Medicine 23(2):113–120

Pollack ME, Brown L, Colbry D, McCarthy CE, Orosz C, Peintner B (2003) Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment. Robotics and Autonomous Systems 44:273–282

Povey D, Ghoshal A, Boulianne G, Burget L, Glembek O, Goel N (2011) The Kaldi speech recognition toolkit. In: IEEE 2011 workshop on automatic speech recognition and understanding

Rabiner LR, Juang B-H (1993) Fundamentals of speech recognition

Radziwill NM, Benton MC (2017) Evaluating Quality of Chatbots and Intelligent Conversational Agents. arXiv preprint arXiv:1704.04579

Ramelson H, Friedman R, Ockene J (1999) An automated telephone-based smoking cessation education and counseling system. Patient Education and Counseling 36:131–144

Ren J, Bickmore TW, Hempstead M, Jack B (2014) Birth control, drug abuse, or domestic violence: what health risk topics are women willing to discuss with a virtual agent? In: 2014 International Conference on Intelligent Virtual Agents

Rich C, Sidner C, Lesh N, Garland A, Booth S, Chimani M (2004) DiamondHelp: A Graphical User Interface Framework for Human-Computer Collaboration. In: IEEE International Conference on Distributed Computing Systems Workshops

Ryu S, Lee D, Lee GG, Kim K, Noh H (2014) Exploiting out-of-vocabulary words for out-of-domain detection in dialog systems. In: 2014 International Conference on Big Data and Smart Computing. IEEE, pp. 165-168

Saon G, Kurata G, Sercu T, Audhkhasi K, Thomas S, Dimitriadis D, et al (2017) English conversational telephone speech recognition by humans and machines. arXiv preprint arXiv:1703.02136

Sarikaya R (2017) The technology behind personal digital assistants: An overview of the system architecture and key components. IEEE Signal Processing Magazine 34(1):67–81

Shneiderman B (1995) Looking for the bright side of user interface agents. interactions 2(1):13-15

Skantze G (2007) Skantze, Gabriel. Error Handling in Spoken Dialogue Systems-Managing Uncertainty, Grounding and Miscommunication

Skarbez R, Kotranza A, Brooks FP, Lok B, Whitton MC (2011) An initial exploration of conversational errors as a novel method for evaluating virtual human experiences. In: Virtual Reality Conference (VR)

Svennevig J. (2000) Getting acquainted in conversation: a study of initial interactions. John Benjamins Publishing

Tamura-Lis W (2013) Teach-back for quality education and patient safety. Urologic Nursing 33(6):267

Tannen D (ed) (1993) Framing in Discourse. Oxford University Press, New York

ter Maat M, Heylen D 5773 (2009) Turn management or impression management? In: International Conference on Intelligent Virtual Agents (IVA)

Tomko S, Harris T, Toth A, Sanders J, Rudnicky A, Rosenfeld R (2005) Towards efficient human machin speech communication: The speech graffiti project. ACM Transactions on Speech and Language Processing 2(1)

Tür G, Deoras A, Hakkani-Tür D (2013) Semantic parsing using word confusion networks with conditional random fields. In: Proceedings INTERSPEECH

Van Dijk TA (2007) Comments on context and conversation. Discourse and contemporary social change 54:281

Walker M, Litman D, Kamm C, Abella A (1998) PARADISE: A Framework for Evaluating Spoken Dialogue Agents. In: Maybury MT, Wahlster W (eds) *Readings in Intelligent User Interfaces*. Morgan Kaufmann Publishers Inc, San Francisco, CA, pp 631–641

Walraven CV, Oake N, Jennings A, Forster AJ (2010) The association between continuity of care and outcomes: a systematic and critical review. Journal of evaluation in clinical practice 16(5):947–956

Wang Z, Schultz T, Waibel A (2003) Comparison of acoustic model adaptation techniques on non-native speech. In: Proceedings Acoustics, Speech, and Signal Processing

Woodland PC, Odell JJ, Valtchev V, Young SJ (1994) Large vocabulary continuous speech recognition using HTK. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-94)

Xiong W, Droppo J, Huang X, Seide F, Seltzer M, Stolcke A (2017) The Microsoft 2016 conversational speech recognition system. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Yoshikawa M, Shindo H, Matsumoto Y (2016) Joint Transition-based Dependency Parsing and Disfluency Detection for Automatic Speech Recognition Texts. In: EMNLP

Young M, Sparrow D, Gottlieb D, Selim A, Friedman R (2001) A telephone-linked computer system for COPD care. Chest 119:1565–1575