Delivering Automated Health Monitoring via Telephone

Rolf Schwitter[1], Robert Dale[1], Dennis Chan[2]

Affiliation of the authors:
[1] Centre for Language Technology, Macquarie University, Sydney
[2] VeCommerce Limited, Apollo Place, Lane Cove

ABSTRACT
Over the past decade, automated speech recognition technology has improved dramatically to the extent that it is now possible to successfully build and deploy telephony-based spoken language dialog systems (SLDSs) that guide users through conversations focussed on specific topics. In this feasibility study, we present the design of a SLDS that calls patients with a chronic illness at home and monitors their state of health. The structures of these dialogs can be tailored by a health care provider, using predefined dialog schemata that make it possible to take the patient's clinical conditions into account. The resulting speech application is designed to improve the patient's self-management skills and to increase their risk perception. In the normal case, no human intervention is necessary during a human–machine dialog; if, however, the application detects a problem during a conversation, then it connects the patient to a health care provider who can attend to the patient's needs.

1. INTRODUCTION

It is widely accepted that effective communication is critical in doctor–patient relationships, since it has clear benefits for the patient's health outcomes as well as for the doctor's own psychological wellbeing. Most healthcare providers identify the lack of time as the number one obstacle to effective communication, and an increasing number of patients complain about the information they receive from the health care provider and the manner of its delivery (Burley and Hampton 2003; Gorney and Bristow 2003).

In this paper, we explore the scope for using a spoken language dialog system (SLDS) in a home monitoring setting as a means of improving communication between health care providers and patients with a chronic illness. For these patients it is important to follow a prescribed therapy or rehabilitation program over a long period of time, with the aims of preventing complications and of ameliorating their overall quality of life; but it is not necessary for them to see their doctor daily. Continuous self-management of the illness and a decent risk perception are the most important skills that these patients have to learn in this context. Establishing and maintaining these skills requires a high degree of education and reinforcing support; it has been argued elsewhere (Piazza et al. 2004) that it is possible to provide this support via a SLDS.

In contrast to touchtone-based interactive voice response systems that have so far been used to monitor patients at home (see, for example, Friedman et al. 1997; Davidson 2002), our approach uses not only speech synthesis technology, but also speaker verification technology to identify the patient at the beginning of a conversation, and speech recognition technology to analyse the patient's utterances during a conversation. In contrast to more traditional SLDSs where the user has to dial a telephone number to connect to the application (see, for example, Azzini et al. 2003; Black et al. 2004, Giorgino et al. 2004), we propose a solution where the SLDS calls the patient at a given time and initiates the conversation. The patient is identified via a voiceprint that has been collected during an earlier consultation. The subsequent dialog is tailored by a health care provider before the SLDS calls the patient the first time. The health care provider selects a predefined *dialog schema* from a repository of treatment scenarios (for example, *Ischeamatic Heart Disease*) and constrains the components of this schema so that it covers the intended therapeutical goals. Prior to a session, the tailored dialog schema is instantiated with patient-specific data from the patient's electronic health record to form an executable dialog specification. In general, a dialog specification consists of a sequence of recorded prompts that require the patient to respond, and one or more specific actions that are triggered according to the information that the patient provides in their response. These actions may consist of a wide range of options, such as simple feedback, instructions, recommendations, or in the case of emergency, connecting the patient to a doctor or a nurse.

2. THE STATE OF THE ART IN SPOKEN LANGUAGE DIALOG SYSTEMS

Existing commercial deployments of speech technology range in complexity from simple call-routing 'virtual receptionists', through systems that provide information in some narrowly defined domain, to complex transactional systems that, for example, can automate the ordering of a pizza. Users of such systems are reporting high levels of satisfaction and preference in using speech over traditional touchtone interaction methods (Harris Interactive 2003).

These systems operate by using both speech recognition and speech synthesis (or recorded speech) to enable an interactive dialog with a user, generally enforced to follow a particular structure by means of a predefined call flow which determines what is said by the system at each point, and the action that is taken dependent upon the user's response. In addition to the recognition and synthesis technologies, an additional technology, speaker verification, is becoming increasingly popular in SLDS deployments (in Australia there are only two publicised deployments, at DoFA and Austar; around the globe, speaker verification is employed at Intrust, Charles Schwab, SwissCom, Bradesco and The Hartford to name a few). This is a process whereby a voiceprint, analogously to that of a fingerprint, is used to verify whether a speaker is in fact who they claim to be. In our proposed solution, speaker verification is used to ensure that the system is conversing with the appropriate person, so as to avoid violating the patient's privacy.

Complex SLDSs still predominantly require manual design and configuration of every dialog turn, since every application has differing underlying business needs. In our solution, the healthcare provider can semi-automate this process for a specific patient by selecting appropriate predefined medical dialog schemata from a repository of reusable components. The result is a predefined dialog that is tailored to the patient's needs, with minimal manual configuration.

In the last five or so years, emerging standards for the development and specification of SLDSs (specifically VoiceXML and SALT) have become widespread; these standards make it possible to construct reusable dialog components that are independent of particular vendors' platforms. Reusable dialog components offer pre-packed functionality 'out-of-the-box', provide user interface consistency, and improve ease of application development. It is much easier for a developer to specify a set of parameters for a reusable dialog component than to develop a piece of code from scratch. Reusable dialog components are also interesting from an economical viewpoint, since they reduce the overall costs of creating, developing, and managing speech solutions. There are various commercial speech toolkit providers (such as IBM and Nuance) that are developing entire libraries of such reusable dialog components.

3. HOW THE SPOKEN LANGUAGE DIALOG SYSTEM WORKS

Figure 1 shows the data flow between the health care provider, who tailors the speech application according to the patient's clinical conditions and the therapeutic goals, and the patient, who communicates with the SLDS via a standard telephone that does not require any technical skills.

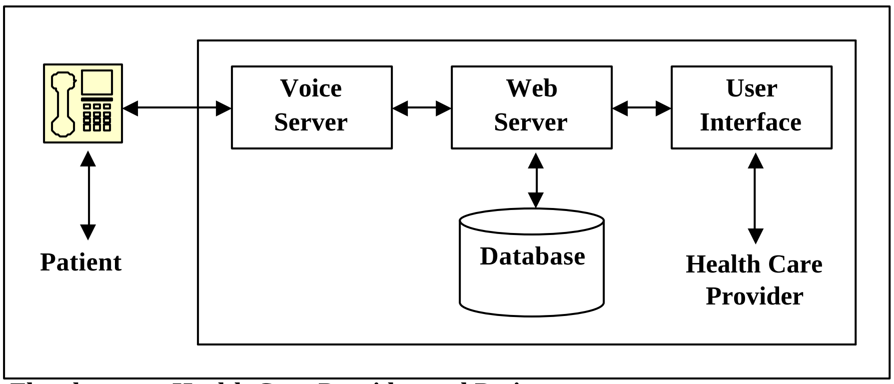

Figure 1: Data Flow between Health Care Provider and Patient

The *User Interface* is a web-based graphical user interface that allows the doctor to register and upload the voiceprint for identifying the patient, and to compose the call flow of the speech application using predefined VoiceXML dialog schemata from a repository. Once the structure of the dialog is complete, the dialog schemata can be further instantiated with information that is available from the patient's health care record stored in the database. The instantiated dialog specification is then stored on the Web Server and is ready for deployment.

The *Web Server* stores an instantiated dialog specification and a voiceprint for each user, along with information that causes the Web Server to connect to the *Voice Server* at specified times to initiate a call. The Web Server sends the dialog specification and associated grammar files used for speech recognition to the *Voice Server*, and deals with database updates that are requested by the Voice Server on the basis of the user's input.

The *Database* contains the health records of the patients and the configured dialog schema. Information that is provided by a patient during a face-to-face doctor–patient session is stored in the health record and can be consulted by the doctor via the *User Interface*.

The *Voice Server* consists of a voice browser that interprets the dialog specification, a speech recogniser that transforms a patient's utterances into text, a speech synthesiser that transforms text into audio prompts and feedback, a telephony interface that allows for making outbound calls to patients, and alternatively a touchtone recogniser that enables patients to answer audio prompts by pressing keys on their telephones.

The patient communicates with the SLDS via a standard telephone or mobile phone. The *Voice Server* calls the patient and if the patient does not answer a call after a predefined number of attempts, a notification message is sent to the *User Interface* of the health care provider.

4. TAILORING OF DIALOGS

A health care provider can tailor the dialog structure to be used with a specific patient via a graphical user interface. The dialog schemata are predefined in a repository for the most common cases such as *Ischaemic Heart Disease*, *Chronic Liver Disease*, and *Bronchiectasis*. Each dialog schema subsumes a number of components (= subdialog schemata) that can be configured by the health care provider. Some of these components (for example, *Measuring Blood Pressure*) belong to more than one dialog schema. Such reusable components promote consistency across applications.

4.1 Selecting a Case

In a first step, the health care provider selects a case (for example, *Ischaemic Heart Disease*) from the repository, and the system then displays the labels of the subdialog schemata that are defined for this case, such as:

- *Measuring Blood Pressure*
- *Measuring Heart Rate*
- *Measuring Blood Sugar*
- *Checking Weight*
- *Checking Dietary*
- *Checking Medication*
- *Changing Smoking Habits*
- …

The interface allows the health care provider to inspect the contents of these subdialogs and to configure them. For example, clicking on *Measuring Blood Pressure* displays the actual speech prompt that is defined for this task (for example, *What is your most recent blood pressure reading?*) and shows the different possible actions that may be triggered depending upon the patient's response in combination with the information available in the patient's health record.

4.2 Tailoring the Dialog

Once the health care provider has selected a case, he or she can add the name and the unique identifier of the patient to the interface, configure the subdialog schemata, and specify their sequential order to be used in the required dialog schema. The subdialog schema for *Measuring Blood Pressure* contains predefined feedback (best practice) for various possible situations and can be overwritten by the health care provider, for example:

Diagnosis:	**Those values are better than last week.**
Education:	**You must have done your walking exercise as recommended.**
Prevention:	**You should now gradually increase your pace and distance; aim for at least thirty minutes per session, three days a week.**
Diagnosis:	**Those values are the same as last week.**
Education:	**Remember you should do your walking exercise as recommended by Dr. Gibson.**
Prevention:	**You should go for a twenty-minute walk twice a week.**

The actual feedback that the system provides is depending upon the patient's response in combination with the information available in the patient's health record and needs to be configured via a set of parameters.

Once the call flow is complete, the health care provider can test the application. Before the application is executed, the underlying dialog schemata are instantiated with information from the patient's health record.

4.3 A Dialog Excerpt

The following is an excerpt of a conversation between the computer and a patient that shows how the patient is first identified by the SLDS, and then illustrates how the dialog might progress. In this example, there is a requirement that the patients should measure their blood pressure daily, and they should do this prior to the SLDS calling in. This obviously requires a prior education stage; written guidelines to aid in system usage, such as *It will help to write your blood pressure values down before the system calls you*, are also very important.

Computer:	**Good morning, this is Doctor Norman Gibson's automatic assistant calling Peter Hartmeier for a follow-up check. Please say your name and date of birth so that I can identify you.**
Patient:	**Peter Hartmeier, 19 February, 1953.**
Computer:	**Hi, Peter, how are you today?**
Patient:	**Hmm, not too bad.**
Computer:	**That doesn't sound too convincing. Would you like to speak with Doctor Gibson?**
Patient:	**No, no that's not necessary.**
Computer:	**Okay, I'll now ask you a number of routine questions to check your state of health.**
Computer:	**First, what's your most recent blood pressure reading?**
Patient:	**162 over 105.**
Computer:	**Those values are better than last week. You must have done your walking exercise as recommended. You should now gradually increase your pace and distance; aim for at least thirty minutes per session, three days a week.**

The system identifies the patient by asking for his or her name and date of birth, and then matches the subsequent input against the expected voice profiles. In addition to this identification process, speaker verification can be carried out periodically throughout the conversation to verify that the identified patient has not been replaced by another speaker. This process consists in comparing voice features with the patient's voice profile, and can be done during the conversation in a way that is not noticeable for the patient. The subsequent dialog shows that the system offers the patient the opportunity to speak with a doctor whenever the input is vague or does not follow certain expected rules. After greeting and identifying the patient, the system briefly explains what the purpose of the call is and then starts with the first routine question, in this case asking the patient for their blood pressure. The values provided by the patient are compared with the values available from the patient's electronic health record; in the example, because the values have improved since the last conversation and because the patient now only has a moderate hypertension, the system assumes that the patient has followed a previous recommendation to do walking exercises, and encourages the patient to continue with the rehabilitation program by giving concrete advice.

5. CONCLUSIONS

In this paper, we have presented a novel approach to tailoring speech dialogs and delivering automated health monitoring via telephone. The key features here are as follows:

- The use of automatic speech recognition and speech synthesis technology offers great potential for automating routine check-up messages, while at the same time providing the patient with a more natural conversational interface.
- The use of speaker verification provides an essential degree of privacy protection.
- The use of dialog schemata and their constituents built in VoiceXML enables easy tailoring of customised dialog structures in accordance with the doctor's goals, while at the same time freeing the doctor from the need to have a deep technical understanding of how the technology works.
- The automatic population of patient-specific data allows for dialogs to take account of and be tailored to individual patient circumstances.

The resulting system allows a health care provider to select a predefined dialog schema from a repository of treatment scenarios and to constrain the components of this schema to enforce the intended therapeutical goals. Dialog specifications for individual patients can be automatically executed at a specified time, with the voice server contacting the patient and monitoring the patient's state of health in an unobtrusive way. The voice dialogs are designed to improve the patients' self-management skills and to enforce their risk perception; this approach allows the provision of tailored information with clarity.

REFERENCES

AZZINI, I., FALAVIGNA, D., GIORGINO, T., GRETTER, R., QUAGLINI, S., ROGNONI, C., STEFANELLI, M., (2003). Automated Spoken Dialog System for Home Care and Data Acquisition from Chronic Patients. In: *Studies in Health Technology and Informatics*, 95:146-151.

BLACK, L., MCTEAR, M., BLACK, N., HARPER, R., LEMON, M. (2004). Talking Telemedicine: Is the Interactive Voice-Logbook Evolving into the Cornerstone of Diabetes Healthcare. In: T. Bickmore (ed.), Dialog Systems for Health Communication, AAAI Fall Symposium, Technical Report FS-04-04, October 21-24, Arlington, Virginia, 13-17.

BURLEY, L., HAMPTON, A. (2003). *Talking Matters* - developing the communication skills of doctors, available at <http://www.scotland.gov.uk/library5/health/tmcsd-00.asp>.

DAVIDSON, R. A. (2002) A home monitoring service was more effective than usual care in patients with essential hypertension. In: *ACP Journal Club*, 136(1):13.

FRIEDMAN, R. H., STOLLERMAN, J. E., MAHONEY, D. M., ROZENBLYUM, L. (1997). The Virtual Visit: Using Telecommunications Technology to take Care of Patients. In: *Journal of the American Medical Informatics Association*, 4:413-425.

GIORGINO, T., QUAGLINI, S., STEFANELLI, M. (2004). Evaluation and Usage Patterns in the Homey Hypertension Management Dialog System. In: T. Bickmore (ed.), Dialog Systems for Health Communication, AAAI Fall Symposium, Technical Report FS-04-04, October 21-24, Arlington, Virginia, 32-35.

GORNEY, M., BRISTOW, J. (2003). Effective Physician Communication Skills. *The Doctors Company*, available at <http://www.thedoctors.com/risk/general/communication/ J4238.asp> or <https://secure.thedoctors.com/sourcebook/J4238.pdf>.

HARRIS INTERACTIVE (2003). Harris Interactive Speech Satisfaction Study, available at <http://www.nuance.com/forms/speechstudy.html> or <http://www.omilia.com/pdf/ speech_study.pdf>

Piazza, M., Giorgino, T., Azzini, I., Stefanelli, M., Luo, R. (2004). Cognitive Human Factors for Telemedicine Systems. In: *Studies in Health Technology and Informatics*, 107:974-978.