The New Navigators: from Professionals to Patients
R. Baud et al. (Eds.)
IOS Press, 2003

Automated Spoken Dialog System for Home Care and Data Acquisition from Chronic Patients

Ivano Azzini[a], Daniele Falavigna[b], Toni Giorgino[a], Roberto Gretter[b], Silvana Quaglini[a], Carla Rognoni[a], Mario Stefanelli[a]

[a] *Laboratory of Medical Informatics, Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy*
[b] *Istituto per la Ricerca Scientifica e Tecnologica ITC – irst, Panté di Povo, Trento, Italy*

Abstract

Recent advances in automatic speech recognition and related technologies allow computers to carry on conversations by telephone. We developed an intelligent dialog system that interacts with hypertensive patients to collect data about their health status. Patients thus avoid the inconvenience of going for frequent visits that monitor some clinical variables they can easily measure at home; the physician is facilitated in acquiring and reviewing patient information and related risk indicators, which are evaluated from the data according to noted guidelines. The system described here is a prototype of future configurable and component-based dialog systems, which may allow a new modality for users and physicians to access electronic health records.

Keywords:
Telemedicine; Data Acquisition; Electronic Patient Record; Databases; Human Factors

1. Background and setting

Hypertension is a chronic disease whose management requires carefulness in monitoring the blood pressure values; it is also important that the patient follows the prescribed therapy, unless she is affected by side effects, which need to be reported as soon as possible [1]. Thus, new technology making data collection easier and faster may increase the quality of the care of patients with hypertension [2].

Recently, Dual Tone Multi Frequency (DTMF) based systems have been used as a supplement to a visit performed by a physician. When a user dials a DTMF-based system, she listens to a series of pre-recorded audio messages, which, at each step, prompt her to type some data on the telephone's keypad. 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 showed such DTMF systems to be successful in home monitoring patients with chronic diseases, like hypertension [3;4].

Following the motivations above, in the E.U. research project "Homey" we built an intelligent dialogue system to monitor and manage patients with essential hypertension. We tried to keep the dialogue close to the usual interaction between a physician and a patient; it also gives advice, by issuing alerts and prompts. The domain knowledge (e.g., the evaluation of risk factors) was derived from a set of world-widely accepted guidelines for the hypertension and dyslipidemia [5]. The data acquisition process may happen both during the traditional encounter with the physician, and via an automatic dialog engine that talks to the patient on the phone. The information acquired is stored in a purposely-designed Electronic Health Records (EHR), whence it can be queried via a web interface.

We have involved into the project several Italian hospitals willing to cooperate in the design of the system and to adopt it. The monitoring service they provide normally requires each patient to meet her physician approximately every 2–4 weeks to monitor blood pressure values, habits, and other variables, and estimate some risk indicators. The physician takes account of these results in order to prescribe or modify the pharmacological therapy and to possibly prompt the patient to make changes to her life style.

2. Architecture

Figure 1 depicts a simplification of the data flow inside our dialog system application. Two actors are allowed to enter data into the EHR. In the first place, the physician uses a conventional (graphics, keyboard and mouse) interface to store and update patient information. On the other hand, the patient is also allowed to enter the data that she can acquire at home by herself and transfer them to the care-providing centre, by the means of a telephone.

To accomplish this task, the patient is instructed to dial a toll free number once every a chosen amount of time. When this happens, she is connected to a call center; the dialog is then directed by a *dialog manager*. The system interprets a *dialog description,* written in a high level language. The dialog description specifies the spoken interaction: which steps should be taken, what questions are asked to the user, what possible utterances are allowed. According to this program, the user is first authenticated with a numeric password. The dialog engine then reads a *state vector,* associated to the patient, which represents her health state and other associated information.

The speech recognition and dialog technologies we used in our application are provided by the SPINET system (Speech Into Enriched Text). The system was developed by ITC-irst; it includes both the engine which performs the recognition, and the dialog manager [6]. The dialog system we have developed leverages two advances in dialog technology, as it is *mixed initiative* and *adaptive.*

Dialogue initiative

Mixed-initiative dialog systems try to mimic the behavior of a telephone caller, whose counterpart anticipates answers to questions that have not been formulated yet. Mixed initiative dialogues are opposed to simpler *form filling* interactions, in which the

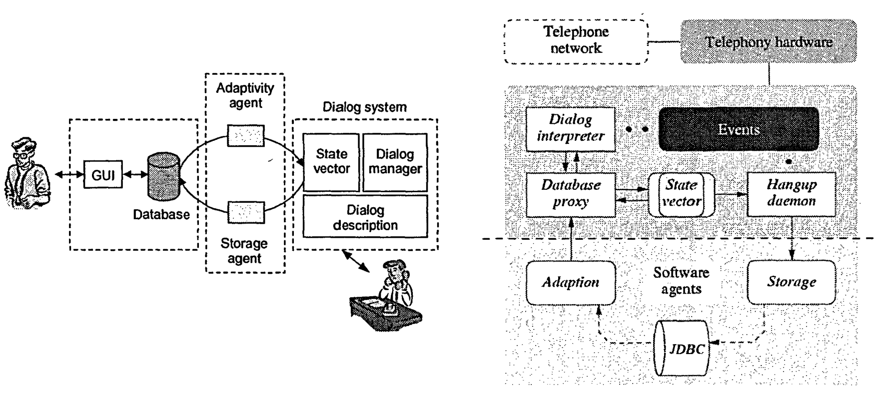

Figure 1: Architecture overview: data flow (left) and block diagram (right).

speaker is only allowed to answer questions posed by the system, one at a time, as if the user was filling a form by dictating fields' content on the telephone:

System:	Please tell me the heart rate you have measured today.
Patient:	90.
System:	Now, please, say your maximum arterial pressure.
Patient:	My pressure is 120.

When talking to a mixed initiative system [7] the user has a greater degree of control in the progression of questions. At certain points, she may answer to a query only partially: e.g., she is requested two values, but chooses to give only one, or she may want to anticipate some answer, providing data that was not yet asked like in the following example:

System:	Please tell me your heart rate.
Patient:	I have not measured it, but today my weight is almost 85 kilograms.
System:	So your heart rate is missing, and your weight is 85 kilograms?
Patient:	Yes.

The mixed initiative approach may be pushed to an extreme to leave the speaker free, in principle, to give the information she desires, in any order. This approach, although exciting, has drawbacks. For instance, such an open prompt may give to a naive caller the false impression that the system will understand anything she would say, which is not true.

The syntax of sentences that the user is allowed to speak, and the places whence to extract the relevant data, are expressed by a *regular grammar* (Figure 2). Grammars are similar to regular expression patterns – flexible specifications which the computer tries to match to what the user inputs. They are represented by means of graphs whose transitions correspond either to terminal symbols (i.e. words) or to non-terminal symbols (i.e. grammars themselves); the graphs related to a given dialogue application represent the Language Models for the Automatic Speech Recognition System adopted. Grammars should match exactly the user's utterances; a less restrictive approach also employed, especially useful for complex sentences, foresees the use of *bigram models,* which provide a statistical description of the allowed sentences.

Adaptability

Choosing the initiative strategy and designing appropriate grammars is an important part of the effort to build an effective dialog system. Such choices are usually made on a trial-and-error basis, resorting to field tests. We shall now discuss how the problem was tackled in this project, by taking into account a possible increase of a user's performances between calls.

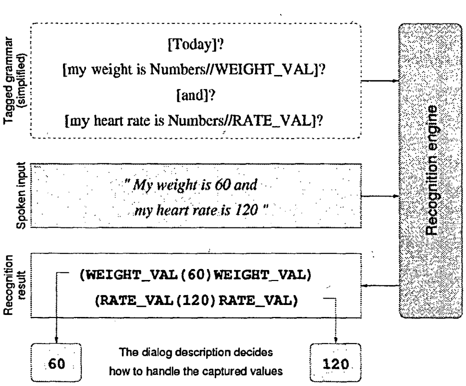

Figure 2: A sample mixed initiative grammar. Question marks denote optional segments and parenthesized expressions are semantic tags.

A requirement for our dialog application was to allow for *adaptive* dialogues. To be perceived as intelligent, in fact, a dialogue system should be able to modify its behavior to match the context of the call and the ability of the human party. The system we have designed changes its behavior depending on both the progress of the current call, and the clinical history of the caller. This is achieved by preserving a representation of call data, called a state vector, which is made up of a set of variables; some of them affect the dialog flow (e.g. if the user is not a smoker, she will not be further asked about the number of cigarettes she smokes). Others are output variables, which may be strings, integers, or Boolean values, which are filled with the recognition results during the progress of the call. When the telephone is finally hung up, the values of the output variables reflect what the user has uttered during the call. Conversely, the database holds other values that should affect the dialog (e.g. whether the patient has been prescribed to follow a diet, or the date of the next visit). When a new call is set up for a certain patient, these values are extracted from the database and the corresponding state vector is prepared by an "adaptivity agent". The name stems from the fact that this agent will also perform some of the adaptations to make the dialog more effective and friendly.

Table 1 - EHR sections

Section	Fields
Anamnesis	50
Life style and risk factors	30
Drugs and tests prescribed	130
Physical examination	85
Report	10

The dilemma of open (mixed initiative) versus closed (system-directed) questions is then solved if one lets beginner users talk their way through a pedantic but error-proof sequence of questions. Once they have successfully completed a chosen amount of calls, the prompts may become more concise, possibly giving hints on how to further shorten the dialogue providing multiple data per utterance. Once a user becomes "expert" according to a suitable criterion, she would plausibly know in advance what data the system is expecting her to provide, and an open question like "Please say the values you have measured today" would not be too confusing.

The electronic health record

The EHR has the function of storing both the data entered by physicians during visits and those told on the phone by patients themselves; it will then generate reports that can be reviewed by health care personnel. Reports include risk indicators, evaluated according to standard clinical guidelines [5], and alerts, which could be issued graphically or via other means such as SMS.

The user interface to the database was designed with the collaboration of physicians (Figure 3). It is web-based, so that GP's and patients can access it with a standard web browser. The interface is made up by forms, grouped in sections as shown in Table 1.

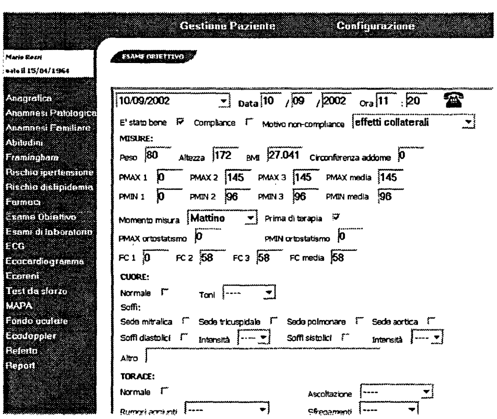

Figure 3: The graphical interface as shown to the GP. The telephone icon, which appears in the form, reminds the reviewer that the data displayed were entered by the patient via the spoken dialog system.

3. Evaluation

The dialog application should be, at the same time, easy to use and understand, and robust. A number of issues in the design of our application were therefore analyzed by the means of on-field trials.

The objectives of the evaluation phases were:

- Testing the reliability of the system

- Extension of grammars and lexicon
- Reformulation of questions' wording
- Extraction of patient's learning curve
- Assessing the clinical effectiveness

Table 2 - Some evaluation results

Transcibed dialogues	91
Total dialogue time	519 min.
Achieved goal	96 %
Questions per dialogue	4–39
Avg. time per call	5.6 min
Avg. time per question	15 s

Grammar	OK	FAIL	OOV
Cough	81%	10%	9%
Compliance	41%	48%	11%

To debug our dialog application according to the goals listed above, we designed a thorough evaluation plan to happen in two phases. The first phase (*internal trial*) involved a group of volunteers, which were assigned a realistic disease profile. The assigned profile included therapy, average blood pressure values and so on; the profile, in turn, affected the inquiry of side effects during the dialogue. All the phone calls were recorded and, later, they were transcribed and checked. The first phase, therefore, mainly addressed technical and usability issues. It involved about 15 people and collected 150 dialogues, amounting to approximately 500 minutes of conversation, of which 150 were human speech.

The data collected about grammars resemble that shown in the bottom of Table 2, which displays statistics for grammars associated to two questions. Grammar *Cough* asked what kind of cough one experienced as a side effect (when, and how strong), while *Compliance* was more open-ended and inquired the reason why a prescription was not taken (presence of side effects, missing drugs, etc. were among the allowed answers). Both grammars capture information that is directly useful for the clinician. The percentages given in the table show the fraction of correct recognitions (OK), incorrect (FAIL) and out-of-vocabulary (OOV), the latter meaning the user's utterance was not foreseen in the grammar. Not surprisingly, OOV incidence is higher for the open question; this is an indication that the lexicon for that grammar needs extending with more words. Similar data was collected for all of the 127 grammars used by our application.

The second test phase will take place after the changes suggested by the first evaluation are performed, and will involve real patients (approximately 200) and their clinicians.

4. Future developments

The prototype discussed stimulated ideas for further work in this field. Three directions to extend the system are discussed below.

1. The callers' ease of use could be increased by putting more "intelligence" into the adaptive agents, for example to shorten the dialogue duration. One way could be to generate shorter prompts spoken to users considered "expert". It would also be possible to extract statistics on the confirmation prompts' outcomes: users that are often understood by the system could have such confirmation questions reduced to a more general summary confirmation, like "...is all of the above information correct?" In case of a negative answer, the dialogue description could fall back to question-by-question confirmations. A further source of adaptation could provide some learning of user's answers, to avoid repeatedly asking selected questions whose answers have been constant in some time span (e.g. "do you still do swimming?").
2. The separation between database and dialog engine may allow the adaptation programs to develop from simple procedures into agencies capable of coping with the users' preferences via machine learning techniques.

3. A more ambitious enhancement would be to try to map interesting concepts (values to be acquired, prompts, and so on) into an appropriate ontology. This would allow more elaborate reasoning and potentially interfacing with clinical tools like guideline servers.

The goal of the project was to explore the capabilities of spoken dialog systems in the medical domain. While the resulting prototype shows satisfying performance, its development was not as straightforward as one would expect for such techniques to be adopted in mainstream practice. We are currently studying a framework which will leverage reuse of pre-built components for recurrent tasks (like patient authentication, database access) and common questions (like weight, compliance to the therapy, etc.). Such a framework would enable a rapid prototyping and deployment of dialog applications in this domain.

5. Conclusions

We have developed and described a prototype home monitoring system for hypertensive patients. The system keeps an EHR of the managed subjects; the data may be updated either directly by the physician, or by the patients themselves, via a dialogue on the telephone with an intelligent system. This system allows them to avoid the inconvenience of going for the visit to record those values that can easily measured at home. The physician is able to review all the information entered, along with the derived risk indicators, to make informed decisions.

6. Acknowledgments

This research has been partially funded by EU under the Fifth Framework IST project HOMEY, number IST-2001-32434.

7. References

[1] Mulrow CD. Evidence-based Hypertension. BMJ Books, 2001.

[2] Singh M, Singh MP, Barnett J. Spoken dialogue interfaces for patient care. 2002.

[3] Friedman RH, Stollerman JE, Mahoney DM, Rozenblyum L. The virtual visit: using telecommunications technology to take care of patients. J Am Med Inform Assoc 1997; 4(6):413-425.

[4] Davidson RA. A home monitoring service was more effective than usual care in patients with essential hypertension. ACP J Club 2002; 136(1):13.

[5] 1999 World Health Organization-International Society of Hypertension Guidelines for the Management of Hypertension. Guidelines Subcommittee. J Hypertens 1999; 17(2):151-183.

[6] Falavigna D, Gretter R, Orlandi M. A mixed language model for a dialogue system over the telephone. Proceedings of ICSLP 2000; Beijing, China: 16 October 2000.

[7] Azzini I, Orlandi M, Lanzola G, Gretter R, Falavigna D. First steps toward an adaptive spoken dialogue system in medical domain. EUROSPEECH 2001; Aalborg, Denmark: September 2001.

Electronic correspondence for the authors may be addressed to toni.giorgino@unipv.it.