The Voice-Logbook: Introducing Human Factors for a Chronic Care System

Black, L.¹, McTear, M.², Black N.¹, Harper, R.², and Lemon, M.²

¹Faculty of Engineering, University of Ulster, Jordanstown, Belfast, Northern Ireland. ²Diabetes Centre, Ulster Hospital, Dundonald, Northern Ireland
la.black@ulster.ac.uk

Abstract

We are developing a tele-care system for the management of hypertensive type 2 diabetes mellitus (T2DM) patients using spoken dialogue technology. This paper discusses the human factors which have been considered in the development and design of ‘DI@L-log’, with regard to the potential complexities posed by speech systems in the medical domain. To accomplish this goal, fine tuning of event handlers, error mitigation and the appropriateness of speech as an interface to chronic care are examined.

1. Introduction

Spoken Dialogue Technology enables humans to communicate with machines via speech interaction [1]. Commercial speech systems proliferate the global marketplace; however the application of spoken dialogue systems (SDS) to the domain of medicine is relatively less mature, multifaceted, and contentious. A SDS for patient information acquisition and retrieval by health professionals is riddled with ethical and privacy issues. Moreover, to acquire accurate home monitored patient data, a consistent and simple application is required if it is to extend to multiple, diverse and individual end users. The foundations of our system are based on some recent research on health related dialogue systems, such as the HOMEY intelligent dialogue project [2].

2. System Overview and Motivation

The size and scale of managing T2DM is a global concern. Recent trials advocate measuring blood sugar and blood glucose in order to help patients become actively involved and responsible in their care. In between clinic visits, communication between doctors and patients is scarce. Therefore a communication platform is necessary to autonomously facilitate patient monitoring in their home environment. The motivation of the proposed system, as described in [3], enables potentially infirm, disabled and elderly hypertensive T2DM patients to have more frequent interaction with the hospital via telecommunications technology. Recorded data using SDS will provide medical professionals with up-to-date patient profiles, affording more timely intervention and prevention of the onset of serious micro and macrovascular complications associated with the disease. Our research aims to determine whether the intended system promotes a better care delivery service by replacing the static paper logbook traditionally used by patients to document home monitored data.

2.1. Study Design

The telephone is deemed a cost effective, user-friendly, practical, and ubiquitous solution to enable patients to regularly communicate context-sensitive data because users do not require special training or need to be computer literate. Furthermore, as argued by [4], speech is described as the most inherently and natural interactive modality for humans. The use of artificial rather than trained ‘live’ call centre agents, such as those used in NHS Direct to regularly consult with patients, shifts the conventional diabetes care process by overcoming constraints of space, time and financial resources to facilitate a more active communication strategy for patients and providers of healthcare.

DI@L-log is envisaged as an intelligent voice or ‘V’-Logbook, which captures ambulatory patient data weekly, provides support and feedback to the patient, and in turn provides updated information and decision support directly to the point of care. The study is being designed in conjunction with a group of patients and health professionals at the Ulster Community Hospitals Trust (UCHT) in Northern Ireland. Ethical approval has been granted from the relevant governing bodies. Consent forms have been signed by selected patients who have agreed to be involved in the study. Three medically determined patient data (weight, blood sugar and blood pressure) have been requested as the parameters that UCHT health professionals would like to obtain from patients in order to keep track of their health status. These parameters are modelled on clinical practice guidelines and target protocols for weight, blood glucose and hypertension respectively.

2.2. Psychosocial factors

In the design of DI@L-log it is imperative to understand the psychosocial factors that affect diabetes patients in order to create a system that is sensitive towards patients needs as well as the underlying clinical contexts and goals. Structured questionnaires and focus group sessions regarding the management and control of the diabetes have been conducted in order to gain a better understanding of how our system can best meet user needs. An initial ethnographic study of the user population established that 2061 T2DM patients are served by the UCHT, of which at least 1092 have co-morbid hypertension and could potentially benefit from our proposed system. These patients are characteristically from lower income backgrounds, unemployed/retired and over 50 years of age. There are 1145 males and 916 females, and the average age group for these patients is 66 years.
Initially, 32 T2 patients were randomly assigned a questionnaire to assess their attitudes towards diabetes, medication/lifestyle compliance, communication with the point of care, and the potential feasibility of a telephone-based system to help augment care from home. Results illustrated that 67% of patients felt they would like more regular support or interaction from the hospital. All patients questioned had a fixed telephone line, and 31% had both a landline and a mobile telephone. Of the 97% of patients who had previous experience of an IVR (interactive voice response) system, 41% said they neither liked nor disliked such systems. Furthermore, 83% of patients thought it would be beneficial to extend home care using a telephone-based contact system for their home monitored data, and to receive monthly reports on their progress. Of these, 79% would like such a system to feature an alert function to notify the doctor of abnormal results. This survey shows how patients as health consumers now demand more of a voice in the control of their condition. From this data a generic user profile for the system was established and WOZ experiments were designed.

3. DI@L-log Engineering

Human factors engineering strives to ensure technologies are used to enhance human productivity and task performance safely and with minimum error. DI@L-log is a mixed mode prototype tool that allows patients to explicitly enter their medical data (namely weight, blood sugar and blood pressure) using spoken language or DTMF (Dual Tone Multi-Frequency). A top-level view of the system architecture is shown in Figure 1.

Real subjects (n=20 T2DM patients) have agreed to participate and assist developers in the iterative DI@L-log engineering process. From the first ideas in the prototype through WOZ testing and evaluation, we will continue to collaborate with both the patients and consultants at the UCHT. This will authenticate and clinically validate its feasibility by reflecting real life users and their needs and preferences as much as possible. Patients selected are already familiar with weight and regular blood glucose monitoring, however training was provided for the simple use of the Omron R5-1 automatic blood pressure cuff. High-level design of system prompts and their order was in accordance with a special paper data sheet given to the patients for a one month period prior to the trial to familiarise documenting data on a weekly basis.

3.1. Task Hierarchy

Due to the nature of DI@L-log as a medical SDS application, a system directed approach rather than a mixed initiative approach was adopted. The majority of data input by patients will consist of numerical data input and confirmation utterances in a predetermined sequence. The temporal, transient nature of speech means that the user can create a mental model of the application and the tasks to be completed. In a sense, this will impose a greater cognitive load on the patient compared with the traditional paper logbook because the system will necessitate extra data and on a more regular basis. However to accommodate this, this paper sheet is consistent with the order presented in the system prompts, where weight is requested first, then blood sugar, followed by blood pressure. From several focus group sessions with patients, a small set of grammars from a sample corpus were written to encompass possible user utterances and encoded with the Nuance SGF grammar format used by the voice service provider (VSP). The user will be asked to telephone the system each week, and will be required to enter and confirm a unique PIN number before information is sequentially elicited as indicated in Figure 2.
3.2. Technology Employed

The dialogue flow for the system was initially designed using the CSLU Toolkit, a graphical authoring environment for creating and testing spoken dialogue systems [5]. An initial version of the application was coded in VoiceXML using the IBM WebSphere Voice Server Software Developers Kit Release 3.1.1 [6]. Apache Tomcat and JSP (Java Server Pages) were used to provide dynamic content and a connection to the MS Access database. The next version of the system will be developed in VoxBuilder, an online site provided by Vopilot for developing and hosting VoiceXML applications that can be tested in a more realistic environment using the telephone [7].

![High Level Task Hierarchy](image)

Figure 2: High Level Task Hierarchy.

3.3. System Characteristics and Analysis

The design reflects the background status of the user population; many patients are less educated and from lower social classes, therefore simple lay-man vocabulary has been employed. Property confidence levels, the minimal classification needed to execute an action, are set to a 0.7 threshold, consequently leading to the prompt ‘Sorry I didn’t understand, please repeat that’ if the input does not match the system’s grammars. DI@L-log automatically defaults to DTMF mode when the system recognition confidence is low, or after several unrecognised inputs. Default nomatch/noinput takes effect when the system either does not recognize the input, or no input was given. Global help and exit commands are always available and confirmation is given after each input prompt. The help menus give explicit examples of what can be spoken by the patient, for example:

**System:** OK, now can you give your blood sugar?
**User:** Err, help please.

**System:** To enter a blood sugar reading - for example, seven point two, simply say the value, or otherwise use the telephone keypad to enter 7 then the star key for the decimal point, followed by the 2.

**User:** 10.1.
**System:** You said 10.1. Is that correct?
**User:** No I didn’t.
**System:** Please re-enter you blood sugar.
**User:** 11.1.
**System:** You said 11.1. Is that right?
**User:** Yes, that’s right.

Errors are less likely to be propagated in this way, therefore the user can become quickly familiarised with the consistent nature of the system over time. Longer prompt timeouts have been used to give the patient (especially older patients) enough time to enter data. This reduces the pressure and intimidation that may occur if the system appears to be rushing the user and seemingly becoming inpatient. Instead of an undo function, a confirmation protocol is necessary after each piece of inputted data to verify in the user’s mind that the system correctly understood their utterance. Auditory feedback after each input will repeat and confirm their measurements and give the patient the opportunity to change the values, if required. This is also extremely important in crafting the patient’s trust in the application. Timeouts become slightly shorter with each no-input/no-match event so that the user is not left to feel unsure whether the system is still ‘listening’. Full duplex functionality exists throughout the application for the more experienced user to enter data before prompts have been played in full. The caller environment is assumed to be predominantly from the home and rarely in an office as most patients are retired, unemployed or unable to work due to illness. They will also require a quiet atmosphere and the security issue of entering private data means that most patients will want to do so in the comfort of their own home.

3.4. WOZ Testing

Prototyping and usability testing are essential in the early stages of the speech interface design. This is simulated through Wizard of Oz (WOZ) techniques [8], whereby a human (the wizard) simulates the role of the computer by playing certain prompts and eliciting responses from the user. Alternative design decisions and details can be readily implemented, from the system persona to timing and prosodic elements. Such tests are valuable in identifying user preferences and the utterances most likely to be spoken during the interaction, to represent how users typically respond to prompts, and to detect whether users are easily confused or frustrated by certain requests or events. Identifying these traits can reduce recognition error - a crucial factor in the usability evaluation metrics for efficiency of use, task completion and user satisfaction.

We recorded WOZ tests carried out at the Ulster Hospital with 5 patients using a Dictaphone for later analysis. Our WOZ experiments involved telephoning the patient in the consultant’s room in the hospital. The wizard was based in the receptionist’s room. An additional diabetes nurse observed and documented the patient interaction. User expectations of the system were slightly biased towards previous encounters with telephone-based applications, and possibly attributed to the highly sophisticated portrayal speech systems have in the entertainment industry. Furthermore, patients had a lack of knowledge about computers and had assumed that the system could understand more than what they had expected. However, after several empirical tests the users quickly became aware of the limitations of the system.
Each dialogue took around 2-3 minutes to complete. We tested a variety of scripts to gauge a preference for an anthropomorphic rather than a formal system. A sample dialogue from the WOZ experiments is shown below:

**System:** Welcome to the Dialogue Interactive Service. You can say exit at anytime to quit the system, or help if you need assistance. Please enter or say your 4 digit pin.
**User:** 1 9 7 4.
**System:** You said 1 9 7 4 is that right? Say yes or no.
**User:** Yes.
**System:** This is the patient record for John Mann. Are you ready to give your measurements?
**User:** Yes.
**System:** Thank you. What is your weight today?
**User:** I’m 12.2 stone today.
**System:** Sorry, I didn’t get that, what is your weight today?
**User:** I said 12.2.
**System:** I got 12 stone 2 pounds, OK?
**User:** Yes.
**System:** And your blood sugar?
**User:** Blood sugar of 9.
**System:** You said nine, right?
**User:** That’s right.
**System:** Finally what is your blood pressure?
**User:** It’s 145 over 90.
**System:** Your pressure is 145 over 90. Is that correct?
**User:** Yes.
**System:** Please wait one moment John, while the system updates your records.

The prompts may seem tedious, but given that this is a medical application, acquisition of accurate data is essential. We found that lengthy prompts led patients to a high level of user frustration. We also found that the user tended to mimic the system behaviour by using the same vocabulary that was used in the prompt and preferred to hear a female voice.

Appropriate feedback will be given at the end of the call to briefly summarise the health status of the patient based on conditional logic rules in the knowledge base. Feedback is intended to be imparted in a simple, positive manner, encouraging the user, for example:

*Your blood sugar has increased slightly since your last call. Please try to reduce sugar intake and have 5 portions of fruit and vegetables daily. Keep working hard and contact the system again next week.*

The end of the session tells the user that their information is being processed and how much difference, if any, the user’s results were from the preceding week. As a rule, the system will not record the values unless all three data are given. This gives the patient an incentive to work hard at control, and will empower them, a feature that is not possible with the static paper logbook used by patients. For ethical reasons, DI@L-log is not designed to give specific advice on medication regimens.

### 4. Discussion

One of our main challenges is to make the system confidently recognise when the user has input a wrong piece of data, for example, if they input a number as their weight data instead of their blood sugar reading. The system will use grounding to ask the user to confirm the information, by repeating what the user said. This gives the patient the opportunity to correct the system. If the user confirms the wrong data and is unaware of the mistake, the system function is designed to consult a database containing previous clinic and DI@L-log readings to check their previous dialogue readings in order to confidently estimate if the input was ambiguous. If one piece of data is incorrectly recognised this will affect the subsequent error recovery of the system, which is inevitably more difficult to recover from.

Substantial work is further needed to assess our generic user models within the context of system efficiency and robustness. More mixed initiative approaches will also be iteratively developed for the more experienced user. In addition we hope to create system events for handling missing data, and data visualization at the point of care. Workflow practice integration for the health professional need to be further explored, and these individuals will need to be prepared to embrace such new technologies accordingly.

### 5. Conclusion

Our experiences in creating the DI@L-log system illustrate the limiting complexities associated with the design of a context sensitive application in the medical domain. We have deliberately kept the dialogue system-directed to help guide the user population who are primarily older and not familiar with speech applications. To mitigate the likelihood of system errors we have employed event handlers when the system detects a misunderstood or no input event from the user. Errors are inevitable until both users and designers have a clear understanding of the nature of the system goals and appreciate the limitations of the technology, which can potentially extend successfully to better care solutions for the chronic consumers of healthcare.

### 6. References

7. [http://www.voxpilot.com](http://www.voxpilot.com)