
Mobile Personal Healthcare Mediated by Virtual Humans

Anton Leuski

Institute for Creative Technologies, USC
12015 E Waterfront Dr.
Los Angeles, CA 90094 USA
leuski@ict.usc.edu

Rasiga Gowrisankar

College of Engg, Guindy
Sardar Patel Rd
Anna University, Chennai
Tamil Nadu, India
rasiga.g@gmail.com

Todd Richmond

Institute for Creative Technologies, USC
12015 E Waterfront Dr.
Los Angeles, CA 90094 USA
trichmond@ict.usc.edu

Ari Shapiro

Institute for Creative Technologies, USC
12015 E Waterfront Dr.
Los Angeles, CA 90094 USA
shapiro@ict.usc.edu

Yuyu Xu

Institute for Creative Technologies, USC
12015 E Waterfront Dr.
Los Angeles, CA 90094 USA
yxu@ict.usc.edu

Andrew Feng

Institute for Creative Technologies, USC
12015 E Waterfront Dr.
Los Angeles, CA 90094 USA
feng@ict.usc.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
IUI'14 Companion, February 24–27, 2014, Haifa, Israel.
ACM 978-1-4503-2729-9/14/02.
<http://dx.doi.org/10.1145/2559184.2559200>

Abstract

We demonstrate Ally – a prototype interface for a consumer-level medical diagnostic device. It is an interactive virtual character, – Virtual Human (VH), – that listens to users concern, collects and processes sensor data, offers advice, guides the user through a self-administered medical tests, and answers the users questions. The primary focus of this demo is on the VH, we describe and demonstrate the technologies for language analysis, dialogue management, response generation and presentation. The sensing and medical decision making components are simulated in the current system, but possible applications and extensions are discussed.

Author Keywords

Virtual Human; Mobile; Interface for healthcare

ACM Classification Keywords

H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces—Natural language, Voice I/O.;
H.5.1 [Information interfaces and presentation (e.g., HCI)]: Multimedia Information Systems—Artificial, augmented, and virtual realities.

Introduction

Recent advances in health sciences, technology, and manufacturing are introducing novel diagnostic devices

and self-administered test kits to the general public. These testing systems go far beyond the traditional temperature or blood pressure monitors. A consumer can now buy small bluetooth connected gadgets capable of measuring a dozen vital signs. While the makers of these devices and test kits try to simplify their use, the testing procedures still require a regular user to study and follow a list of instructions often typed in a tiny font hidden somewhere among the other printed materials supplied with the kit. Our medical subject matter experts complained that the accuracy of self-administered tests often does not exceed 10%, primarily due to confusing instructions and test guidelines. Our goal is to develop an intelligent interactive interface for diagnostic devices and test kits. Our hope is that such an interface would help people use the device or the test kit correctly, increase the accuracy and reliability of self-administered medical tests.

In this paper we introduce Ally, a Virtual Human, that listens to the user's health concerns, collects and processes data from the available medical sensors, asks the user relevant questions, guides her through self-administered tests, and answers her questions. We envision the VH as a virtual laboratory technician that gathers medical data and helps the user to find the right health practitioner to address her concerns. It is available on the user's mobile phone so it can be personal, adapt to the user's needs, monitor her health, and be readily accessible at any time.

Our demo presents a prototype Ally system and focuses on two topics: Firstly, we build the interface to both explore and demonstrate ideas of using a VH as an interface for medical diagnostic devices and test kits. Interacting with the system allows us to refine our vision of a personal health assistant and debug the technologies involved. The system is also a great tool for soliciting

feedback from medical professionals and regular folks alike. Secondly, this demo challenged us to integrate the existing VH technologies and adapt them to the limited capabilities of a smartphone. This is the first time we brought together language processing, dialogue management, and behavior generation previously used in desktop-based systems onto a mobile platform.

System Architecture

In this demo Ally addresses a user's concerns about a sore throat. She listens to the user describing the symptoms, asks questions, offers to take the user's temperature, monitors her condition, and helps her through a strep test if the conditions worsen. We focus on interaction aspects of the interface so the sensing and test result interpretation are simulated in the current system.

Figure 1 shows the overall system architecture. The system takes audio from the device microphone and sends it over the network to an automatic speech recognition (ASR) server. Our system is deployed on iOS, we use Apple's dictation services as our ASR.

The text from ASR is analyzed by the Natural Language Understanding (NLU) module Jerome. Ally has approximately three dozen predefined responses in its database that cover diagnostic steps, explanation of the strep test, and personal information. Each system response is linked with sample user utterances. Jerome is a statistical text classifier that uses relevance model approach to analyze the utterance text and retrieve a list of most appropriate responses. The unique feature of Jerome is that it evaluates both the user and the system text and constructs an implicit translation model that maps user utterances to responses. Presented with a new utterance, Jerome applies the model to compute a

statistical representation of the best possible response for the utterance. It compares the representation with available responses and selects the best match. This approach is very effective, fast, and robust to the errors in the input that can be introduced by an imperfect ASR. We used this approach in a number of both experimental and deployed systems over several years. [1, 2]

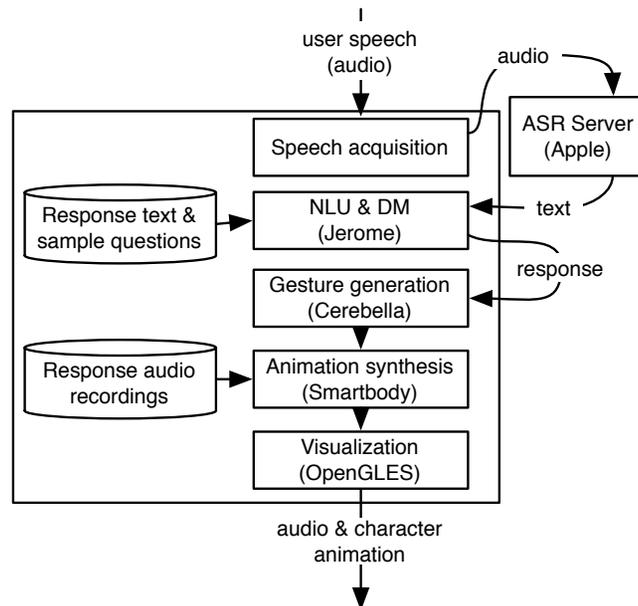


Figure 1: The Ally system architecture.

Ally divides the interaction into several states such as greeting, initial assessment and analysis, indirect monitoring, followback, and test-taking. The character designer defines the states and assigns responses to them. Some responses can be assigned to multiple states.

Transitions between states are triggered by user utterances or system responses. The dialogue manager (DM), a part of Jerome, maintains the current dialogue state, observes the user's utterances, processes the responses selected by the NLU, and picks the final response based on the current state of the conversation and the information already delivered to the user. If the NLU module finds no matching response, the DM attempts to clarify the user's input by prompting the user to repeat or rephrase her question. Ally's domain knowledge is limited: she can deal with a sore throat and she can answer a few personal questions. If a conversation steps over this boundary, Ally attempts to bring it back to the domain.

While, other approaches for NLU and DM exist that allow for deeper language understanding and more flexible dialogue strategies [6], our design is compact, robust, easy to integrate, and requires neither significant computational resources nor sophisticated linguistic expertise to author the system knowledge. In fact, the whole system was put together by a summer student in the course of six weeks. Such a design also gives us precise control over the system responses and allows us to sidestep the lack of expressiveness in automatically generated speech – we pre-record all of the system responses using a voice actor. That way Ally can sound more involved and concerned about the user's condition.

Once the response text is chosen, a set of verbal and nonverbal behaviors are automatically generated by inferring the mental state and emotional content from the utterance (Cerebella) [3]. Cerebella produces a set of high level behaviors including head movements (nodding and shaking), gesturing (metaphorical, beat), gaze shifts and facial expressions synchronized with the words in the utterance. The behaviors are then interpreted through an

animation system (SmartBody) [5] that converts them into animation curves and resolves conflicts, such as coarticulating gestures that are close in time to each other. SmartBody also synchronizes the lip animation with the audio of the utterance [7]. The rendering engine (OpenGL ES) uses the animation curves to control the 3D character on the mobile device display.

Discussion and Future Work

A VH offers numerous advantages over voice-only interfaces. Studies have shown that visible speech increases the comprehension of speech signals [4]. Watching a VH say and behave in an appropriate manner would likely increase speech comprehension. Also, the VH's appearance can be inexpensively customized for a particular user, changing gender, ethnicity or spoken language. In addition, the use of a VH, as compared against displaying a video of a live human actor, allows the ability to change the content of the interaction without having to recreate the entire character performance. It also easier to create interactions where the user can interrupt, correct, or query the VH, without requiring an explicit response to every anticipated user interaction.

While we have been developing interactive VHs for more than over a decade, due to prior technical constraints this is our first attempt at bringing a VH onto a mobile device. The current and future generation mobile devices (phones and tablets) provide a viable platform for VH application development. The medical domain has widely adopted these devices, though largely in the practitioner sphere. Most consumer applications are in the fitness domain, and often suffer from usability, especially with non-technophile users. One advantage to VHs is the ability to put a “friendly face” on a device/application, hopefully encourage use, and lead to better outcomes with patients.

We plan to continue and extend the current work both by exploring this specific medical application domain and advancing the underlining technology.

References

- [1] Leuski, A., and Traum, D. Practical language processing for virtual humans. In *Proceedings of the Twenty-Second Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-10)* (July 2010), 1740–1747. Winner of the 2010 IAAI Deployed Application Award.
- [2] Leuski, A., and Traum, D. NPCEditor: Creating virtual human dialogue using information retrieval techniques. *AI Magazine* 32, 2 (2011), 42–56.
- [3] Marsella, S., Xu, Y., Lhommet, M., Feng, A., Scherer, S., and Shapiro, A. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2013), 25–35.
- [4] McGurk, H., and MacDonald, J. Hearing lips and seeing voices. *Nature* 264 (Dec 1976), 746–748.
- [5] Shapiro, A. Building a character animation system. In *Motion in Games*, J. Allbeck and P. Faloutsos, Eds., vol. 7060 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2011, 98–109.
- [6] Traum, D., Swartout, W., Gratch, J., Marsella, S., Kenney, P., Hovy, E., Narayanan, S., Fast, E., Martinovski, B., Bhagat, R., Robinson, S., Marshall, A., Wang, D., Gandhe, S., and Leuski, A. Dealing with doctors: Virtual humans for non-team interaction training. In *Proceedings of the 6th annual SIGDIAL Conference* (2005).
- [7] Xu, Y., Feng, A. W., Marsella, S. C., and Shapiro, A. A practical and configurable lip sync method for games. In *ACM SIGGRAPH Motion in Games* (Dublin, Ireland, Nov. 2013).