Chapter X
Natural Human–System Interaction Using Intelligent Conversational Agents

Yacine Benahmed
Université de Moncton, Canada

Sid-Ahmed Selouani
Université de Moncton, Canada

Habib Hamam
Université de Moncton, Canada

ABSTRACT

In the context of the prodigious growth of network-based information services, messaging and edutainment, we introduce new tools that enable information management through the use of efficient multimodal interaction using natural language and speech processing. These tools allow the system to respond to close-to natural language queries by means of pattern matching. A new approach which gives the system the ability to learn new utterances of natural language queries from the user is presented. This automatic learning process is initiated when the system encounters an unknown command. This alleviates the burden of users learning a fixed grammar. Furthermore, this enables the system to better respond to spontaneous queries. This work investigates how an information system can benefit from the use of conversational agents to drastically decrease the cognition load of the user. For this purpose, Automated Service Agents and Artificial Intelligence Markup Language (AIML) are used to provide naturalness to the dialogs between users and machines.

INTRODUCTION

Nowadays, most interfaces incorporating speech interaction fall into three broad categories. The first category includes Command and Control (C&C) interfaces that rely on a fixed task dependant grammar

to provide user interaction (Paek, 2007). Their main advantage is their ease of implementation and high command recognition rate. However, their downside is the high cognitive load required to learn and use the system because of its lack of flexibility and lack of uniform command sets. The Universal Speech Interface project tries to fix some of the problems tied to C&C and natural language processing. This is done by providing a general task independent vocabulary for interaction (Nichols, 2007). However, the user is still limited in his choice of utterance since the system is strict on form.

The second category is based on interactive voice response (IVR) that guides users by the means of prompts in order to validate the utterance at every step (Shah, 2007). This style of interaction is mostly used in menu navigation such as that found with phone and cable companies. Its relative lack of efficiency for fast interaction makes it a poor choice for every day use.

Finally, the third category uses natural language processing (NLP) to parse the user's utterance and to determine the goal of the request. This can be done through multiple ways such as semantic and language processing and filtering (Sing, 2006). Hence, to be effective, due to "limitless" vocabulary, this type of interface needs an accurate Automatic Speech Recognition (ASR) system. Another disadvantage of this system is the relatively steep development cost. This is mainly due to the complexity of parsing spontaneous utterances that might not follow conventional grammar.

In this chapter we demonstrate that an effective speech interfaces can be created by following a simple architectural plan, and combining Artificial Intelligence Markup Language (AIML) (ALICE, 2005) for language processing as well as a novel automatic learning framework capable of learning new utterance patterns tied to the current context and to the user profile and preferences.

BACKGROUND

Spoken dialogue constitutes the most natural and powerful means to interact with computers. Systems based on natural spoken dialogue start to appear feasible with the recent improvements in computer engineering and in speech and language processing. Speech-based interfaces increasingly penetrate into environments that can benefit from hands-free and/or eyes-free operation. In the context of the growth of network-based information services, messaging and edutainment, or the demand for personalized real-time services, automatic speech recognition and speech synthesis are highly promising. They are considered as sufficiently mature technologies to allow their inclusion as effective modalities in both telephony and multimodal Human-Computer Interaction (HCI) (Deng & Huang, 2004). Some applications of Internet searching and navigating are currently known; Opera version 9 has a basic voice interface (Opera, 2007). However, its recognition engine is pretty basic and cannot be trained; also, its synthesized voice is robotized. Google provides this kind of functionality through Google Voice Search which is still in a demo state. To use it, a user needs to make a phone call. This is not very convenient for users (Google, 2006). Another application is the one developed by Dr. Meirav Taieb-Maimon and colleagues from Ben-Gurion University of the Negev where car drivers can consult the Internet through voice commands (Sommer, 2005). Lyons et al. (2004) introduced a concept of a dual-purpose speech interaction that provides meaningful input to the computer in order to manage calendar and other communication tools for users. In these applications, the limitation of automatic speech recognition engines is outlined as the main obstacle to the efficiency of the interaction.

One of the main concerns for both companies and research teams is the lack of efficient and affordable natural language based Human-Computer Interfaces. While companies such as Nuance (http://www.