Butler: A Universal Speech Interface for Mobile Environments

Botond Pakucs

Centre for Speech Technology (CTT),
KTH, Royal Institute of Technology,
Lindstedtsvägen 24, 100 44 Stockholm, Sweden
botte@speech.kth.se

Abstract. Speech interfaces are about to be integrated in consumer appliances and embedded systems and are expected to be used by mobile users in ubiquitous computing environments. This paper discusses some major usability and HCI related problems that may be introduced by this development. It is argued that a human-centered approach should be employed when designing and developing speech interfaces for mobile environments. Further, the Butler, a generic spoken dialogue system developed according to the human-centered approach is described. The Butler features a dynamic multi-domain approach.

1 Introduction

Recently, the possibility to use speech interfaces in embedded products and consumer appliances in mobile and ubiquitous computing (UC) environments has begun to attract interest. The speech technology industry has already recognized the potentials of the new emerging market. If the market growth of speech interfaces is as large as expected, users will be surrounded by a multitude of speech-controlled services and appliances. However, in mobile environments the usability requirements on speech-based interfaces may increase and new, human computer interaction (HCI) and usability related problems may be introduced.

Some major usability and HCI related issues that should be considered when designing speech-based interfaces for mobile environments are discussed in Section 2 of this paper. In Section 3, it is argued for a human-centered approach and it is suggested that each user should use a single, highly individualized speech interface for accessing a multitude of appliances and services in mobile environments. In Section 4, Butler, a generic spoken dialogue system developed according to the suggested approach is described. Butler features a dynamic multi-domain approach, individualization, user modeling and context awareness.

2 Usability issues

Speech-based interaction with mobile services differs from accessing speech services through telephones or interacting with desktop computers. Users on the move, and with hands and eyes busy, have greater demands on the HCI.

Designing and building user-friendly speech-based interfaces with excellent usability properties in their own right might just not be enough. There is a need for a shift in how we think about speech-based interactions in mobile and UC environments. Usability and HCI issues should be considered for whole environments rather than for isolated services and appliances.

2.1 Diverse Speech Technology Solutions

Interface consistency is a central and well-understood concept in the HCI and usability community [1, 2]. In mobile environments however, we may expect, in the near future, a multitude of speech interfaces with various complexity, employing a range of different speech technology solutions from simple voice-triggered commands to advanced conversational dialogue systems. A lack of consistency among different speech interfaces may cause usability problems.

When encountering diverse speech interfaces, the same user may be an expert user of some speech interfaces, but still a novice user of other systems. Diverse speech technology solutions may require different interaction strategies from the user and, thus, the use of several different cognitive models. For instance, it will be hard for users to identify the currently available dialogue management strategies, voice commands, and vocabularies. It might even be hard for users to know which services and appliances can be controlled by speech.

2.2 Multiple Concurrent Speech Interfaces

As far as we know, the effects of encountering several concurrent speech interfaces at the same time have never been studied. This situation may actually occur in mobile environments, where several speech-based interfaces are listening for user commands, or even taking initiative pro-actively. Due to miss-recognitions, it is possible that several speech interfaces may be triggered by a single user utterance.

2.3 Increased Usability Requirements

In mobile and dynamically changing UC environments the user's intentions and needs may rapidly change. The user should be able to initiate a new task while waiting for some other specific service to be completed or change the parameters of some previously initiated service. Furthermore, the system itself should be able to interrupt an ongoing dialogue and direct the user's attention to some higher priority events.

For supporting a wide range of domains within one and the same dialogue and for allowing the user to transparently and seamlessly switch between several topic domains and services a *multi-domain approach* [3] is also necessary. The support for these features in current industry solutions is limited.

Consequently, to provide user-friendly speech interfaces in mobile and UC environments and to avoid the introduction of new usability related problems we need means to coordinate and control the various speech interfaces.

3 The Human-Centered Architecture Model

The currently employed speech interface architectures for desktop-based interaction all share an application-centered multi-user system design illustrated in Fig. 1A, where each service or appliance has a separate speech interface [4]. In the fast growing sector of voice portal based telephony services a centralized single entry point can be used for accessing several different services, see Fig. 1B. However, solving the usability problems discussed in Section 2 is not facilitated by these architecture models.

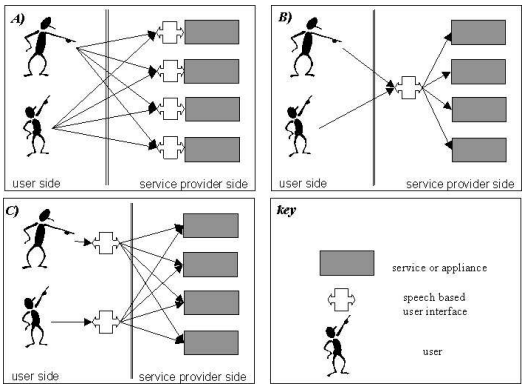

Fig. 1. Speech interface architecture models: *A)* Embedded and application-centered speech interfaces. *B)* Voice portals: - application-centered, centralized speech interfaces. *C)* human-centered and application independent speech interfaces.

The central idea proposed in this paper is the human-centered, application independent architecture for speech interfaces targeting mobile users, see Fig. 1C. Thus, every user is expected to use a *SINGLE*, highly individualized speech interface to access a multitude of services and appliances. It would be preferable if the human-centered speech interface could be integrated into some personal, wearable appliance such as a mobile phone or a PDA. In that case, the speech interface would always be accessible with all user-dependent data activated and ready to use.

Service and application-specific data, including dialogue management capabilities, domain knowledge etc., has to be encoded in *service descriptions* and stored locally at the service provider side. Whenever the user enters a new environment, the available, distributed service descriptions have to be dynamically loaded into the personalized speech interface through some *ad-hoc* and wireless communication solution.

The human-centered, *single user and multiple application* approach to speech interfaces would be an appropriate solution for coordinating and controlling various speech based interfaces. This approach would facilitate the handling of the usability problems discussed in the previous section.

A human-centered approach would facilitate the building of advanced user and domain knowledge models which could provide support for *context awareness* [5]. However, collecting data on the user's behavior, speech patterns etc. is a delicate issue. We believe that a single human-centered interface, because it is controlled by the user, provides better *security and integrity* properties than a multitude of different embedded and distributed systems, which are outside the user's control.

By employing a human-centered solution, it would also be unnecessary for the users to learn and adapt to several different interfaces. The impact of some major challenges for spoken dialogue systems [6] can also be reduced. The speaker variation can be reduced significantly through the possibility to use speaker dependent and speaker adaptive speech recognition. This way the amount of speech recognition errors could be decreased substantially. Addressing challenges such as the variability in channel conditions or background noise could also be facilitated by consistent use of a personalized microphone solutions.

3.1 The SesaME Dialogue Manager

One of the major challenges for the human-centered approach is the dialogue management. SesaME [3] is a generic, task-oriented dialogue manager specially designed for the human-centered approach as well as for mobile environments. Special attention has been given to support adaptive interaction methods and context awareness. In SesaME, a content-based solution [7] is employed for performing the user modeling. This way a simultaneous adaptation to an individual user and to the user's current situation is supported [8].

One of the key-issues in the SesaME architecture is to support a *dynamic multi-domain approach*. The locally available service descriptions, including dialogue descriptions and grammars has to be dynamically loaded and activated on the fly. For handling these requirements, a dynamic plug-and-play functionality of the dialogue management capabilities has been developed [9]. The XML-based service descriptions are distributed through the HTTP protocol however, generic service discovery is also supported.

4 The Butler

Currently, the evaluation of the Butler, a new multi-domain application based on SesaME, is being conducted. The main goal is to evaluate the support for individualization and context awareness, however, speech user interface related problems, such as protecting privacy of the user, disturbance to other people will be also studied. The Butler provides speech-based multi-domain information services through telephones or PDAs. The services provided by Butler can be categorized in three main categories, *public services* such as accessing commuter and subway train timetables, menu information for the nearby restaurants, *accessing personal information* from calendars and *accessing workplace related information*, such as time and location of meetings and seminars.

For identifying the users, telephone number-based or speaker verification is used. The back-end information for all of these services is based on publicly available web-based services. The domain descriptions necessary for the Butler and SesaME are dynamically generated and processed at runtime.

5 Summary and Future Work

Some usability and HCI related problems, which may arise when speech interfaces are integrated in mobile and UC environments have been discussed in this paper. Based on these issues, a novel human-centered approach is proposed for speech interfaces in mobile environments. Further, SesaME, a generic multi-domain dialogue manager, built according to the human-centered approach, has been described. The SesaME dialogue manager is employed in the framework of the Butler demonstrator. By employing a dynamic multi-domain approach, the Butler acts as an individualized universal speech interface.

The suggested approach creates novel possibilities for supporting personalization, context awareness and user modeling in dialogue management. These features will be studied in an upcoming long-term user-study.

Acknowledgments This research was carried out at the CTT, a competence center at KTH, supported by VINNOVA (The Swedish Agency for Innovation Systems), KTH and participating Swedish companies and organizations. This work was also sponsored by the European Union's IST Programme under contract IST-2000-29452, DUMAS (Dynamic Universal Mobility for Adaptive Speech Interfaces).

References

1. Grudin, J.: The case against user interface consistency. Communications of the ACM **32** (1989)
2. Nielsen, J.: 5.4 Consistency. In: Usability Engineering. Morgan Kaufmann, Inc., San Francisco, CA, USA (1994)
3. Pakucs, B.: Towards dynamic multi-domain dialogue processing. In: Proceedings of the Eurospeech'03. Volume 1., Geneva, Switzerland (2003) 741–744
4. Larson, J.A.: Speech-enabled appliances. Speech Technology Magazine (2000) http://www.speechtek.com/.
5. Dey, A.K.: Understanding and using context. Personal and Ubiquituos Computing **5** (2001) Special issue on Situated Interaction and Ubiquitous Computing.
6. Glass, J.R.: Challenges for spoken dialogue systems. In: Proceedings of 1999 IEEE ASRU Workshop, Keystone, CO, USA (1999)
7. Zukerman, I., Albrecht, D.W.: Predictive statistical models for user modeling. User Modeling and User-Adapted Interaction **11** (2001) 5–18
8. Pakucs, B.: SesaME: A Framework for Personalised and Adaptive Speech Interfaces. In: Proceedings of EACL-03 Workshop on Dialogue Systems: Interaction, Adaptation and Styles of Management, Budapest, Hungary (2003) 95–102
9. Pakucs, B.: VoiceXML-based dynamic plug and play dialogue management for mobile environments. In: Proceedings of ISCA T&R Workshop on Multi-Modal Dialogue in Mobile Environments, Kloster Irsee, Germany (2002)