JASPIS – AN ADAPTIVE SPEECH APPLICATION ARCHITECTURE

Markku Turunen

Speech based and Pervasive Interaction Group

Tampere Unit for Computer Human Interaction

Department of Computer Sciences

University of Tampere, Finland

mturunen@cs.uta.fi

Abstract

Speech can be an efficient and natural way for communication between humans and computers. In speech applications various user groups, such as users with special needs, can have very different needs. Therefore, the development of speech applications needs suitable techniques, methodology and development tools for flexible and adaptive interaction methods which take into account the needs of different users and different environments. This paper gives an overview of adaptive speech applications and presents the Jaspis speech application architecture that has been designed to support adaptive and flexible human-computer interaction techniques. We also present several applications constructed with the Jaspis architecture.

Keywords: Spoken Dialogue Systems, Speech Interfaces, Human-Computer Interaction, System Architectures

INTRODUCTION

The development of speech applications has been an active research area for long. Despite the numerous research prototypes and commercial services, the development of speech applications is still not a mature field, and the full potential of speech-based interaction has not been utilized. The key issues in contemporary speech applications are adaptive and flexible interaction methods. These are needed in many application areas, such as in multilingual and multimodal applications. In the emerging applications areas, such as pervasive computing applications, these needs are urgent. In order to facilitate the development of advanced speech applications, we need advanced techniques, models, methodology and tools. In particularly, we need system frameworks which support the issues mentioned, since with all technical variation, the key issue is the integration of components into a working system (McTear, 2002).

In this paper the focus is on advanced speech applications and their system architectures. We start by presenting various types of speech applications. Next, we will present the Jaspis speech application architecture that has been designed to support adaptive and flexible human-computer interaction techniques. Finally, we present several Jaspis-based applications for various domains.

SPEECH APPLICATIONS

Successful speech applications have been constructed in many research projects and several commercial applications are also available. Next we provide a summary of various speech application areas. Here we divide speech applications into conventional, multilingual, multimodal and pervasive applications. In addition, we present how adaptivity may be applied in these different application areas.

Conventional Speech Applications

The first speech applications were telephone-based interactive voice response (IVR) systems, which used speech outputs and telephone keys for interaction. These applications can be very useful for users if they introduce new services which are not possible or affordable with human operators (or in any other way). In addition to IVR applications, many other forms of telephony applications have dominated the field. These include information services, such as timetable, weather forecasting and banking services, e-mail applications and voice portals.

Desktop applications form another popular area of speech applications. Most desktop applications are either targeted at dictation or the control of existing graphical applications. Dictation applications are relatively popular within special groups. Command and control applications make it possible to accomplish tasks which are usually done by using the mouse and the keyboard. Such applications may become more popular when computers become small and embedded in the environment (personal digital assistants, ubiquitous computing).

Multilingual Speech Applications

Translation systems are one example of multilingual speech applications. These systems translate spoken utterances between users. Typical applications include booking systems and other similar applications. Another form of multilingual applications is systems allowing users to access information services, such as forecasting information, using multiple languages.

A third type of multilingual applications uses multiple languages with the same user. The need for multiple languages arises from the application domain. Typical examples of this approach are speech-based e-mail systems, such as the Mailman system (Turunen & Hakulinen, 2000b). The use of multiple languages is essential in these applications. For example, it would be hard to imagine a monolingual e-mail system for Finnish users, since a typical mailbox contains messages written in multiple languages.

Multimodal Speech Applications

From the historical perspective, multimodality offers promising opportunities, as presented by the famous Put-That-There system (Bolt, 1980). Among the most studied multimodal speech applications are so-called "talking heads" or "speaking agents". In these systems the main interface element is audio-visual speech synthesis, which uses anthropomorphic figures to convey facial expressions and head-movements. In addition to talking heads, systems with animated interactive characters have been constructed. Another active area of research in multimodal applications is information kiosks (intelligent kiosks). In these systems modalities such as speech and haptics (e.g., touch-screens) are used to provide interface for users in public places.

Pervasive Speech Applications

Pervasive applications, i.e. ubiquitous and mobile computing applications, introduce both new opportunities and challenges for speech as an interface element. Speech is a suitable element for pervasive applications because of the size of the devices, and the lack of large visual displays and familiar interaction devices, such as the mouse and the keyboard. By embedding systems into everyday environments ("ambient intelligence") the focus of speech applications shifts from strict turn-based, user-initiative conversational systems in a more proactive and distributed direction.

In the area of mobile applications some work has been done in assistive technologies, which enable disabled users to access mobile devices by using speech (Manaris et al., 1999). Speech-based telephone interface can be also used as an alternative for mobile devices (Tang et al., 2001).

Adaptive applications

As presented, speech applications form a very diverse field and although the

main focus is still on traditional applications, it is reasonable to expect that this situation will change. Multilingual, multimodal and pervasive applications will change the focus from the development of basic technologies to interface issues and new interaction methods and techniques are needed. The key issues are adaptiveness and flexible interaction.

One type of adaptive applications is systems which are targeted at special groups, such as disabled users. In this context adaptivity refers to ways which help different users to use applications by using customized interaction methods and techniques. Special groups constitute strong motivation for adaptive applications.

From the system viewpoint, the need for flexibility and adaptivity means that there should be either numerous very specific architectures or speech architectures must be suitable for a wide range of application areas and support adaptivity and distributed and concurrent dialogues. These are strong motivations for the Jaspis architecture.

JASPIS ARCHITECTURE

The Jaspis architecture (Turunen & Hakulinen, 2000a; Turunen, 2004) is both a conceptual model and a concrete speech application development framework. The architecture provides in particular a flexible interaction coordination model, an explicit system level adaptation mechanism, a shared system knowledge and an efficient distribution model. The architecture introduces the agents – managers – evaluators paradigm. Compact *agents* support highly modular systems and reusability, while *managers* coordinate agents in a flexible way, allowing highly distributed systems to be constructed.

The Jaspis architecture contains a general adaptation mechanism that can be used across system modules and applications. In a nutshell, each manager uses a set of *evaluators* to select the most suitable agent for each situation. Each evaluator gives a floating point score between zero and one for each agent. The agent with the highest overall score is selected to handle the current task.

The general system level adaptation mechanism can be applied to various tasks. It is noteworthy that there is no single evaluator, nor any single component in general, which selects agents for each situation, but instead the selection is always both dynamic and distributed. This makes it possible to keep the program control dynamic and adaptive at the architectural level. Furthermore, new features can be added to applications without modifications to existing components. For example, adaptive features, such as guidance prompts and interactive tutoring, can be included to the system in this way (Hakulinen et al., 2003).

The usefulness of the Jaspis architecture has been demonstrated in many practical applications, as presented in the next section. In addition, a Jaspis-based bus timetable service is in daily use. In order to provide better support for pervasive computing applications, and more natural and flexible applications in general, a new version of the architecture was designed. It introduces triggers, transactions and various other improvements in the architecture. The resulting

Jaspis[2] architecture allows concurrency, but preserves the coordinated architecture level adaptation of the original architecture (Turunen & Hakulinen, 2003; Turunen, 2004).

In order to facilitate development of advanced speech applications for both research and commercial purposes we offer the open source distribution of the Jaspis architecture. The architecture is released under GNU Lesser General Public License. In practice this means that it is an open source software that can be used for the development of commercial applications as well. The Jaspis distribution consists of the core infrastructure and numerous extensions for various purposes.

JASPIS APPLICATIONS

Next we present several applications implemented using the Jaspis architecture. These applications cover many ares, including information services, multilingual systems and pervasive computing applications.

E-mail systems

Our research group has been involved in the development of several speech-based e-mail applications. Mailman a bilingual (Finnish and and English) speech-based e-mail application. It allows the user to access his/her mailbox using a standard mobile or fixed-line telephone and provides the most common e-mail client functions. The system can be used in multimodal fashion (speech and telephone keys). AthosMail is based on the Mailman application and in the EU-funded DUMAS project new and elaborated components have been produced for various purposes. During the design, development and evaluation of these systems we have been working in collaboration with visually impaired users.

Dialogues in the e-mail domain are rather open, and especially there are no clearly defined tasks like in many other application areas (e.g. in timetable systems). Therefore, a collaborative user initiated dialogue strategy is applied. The e-mail systems support the user by providing context-sensitive help and guidance. For example, AthosMail uses adaptive prompts, integrated tutoring and universal commands to help the user to know what to do. The system monitors the user and adjusts the amount of guidance. Still, the dialogue strategy is based on user-initiative approach, and experienced users can perform operations efficiently.

Bus timetable systems

In addition to e-mail systems, our research group has been working in various projects with several bus timetable systems. The Busman system provides Tampere area bus route and timetable information. The functionality of the system is similar to other timetable services, such as MALIN (Dahlbäck & Jönsson, 1999). The user may request information such as bus routes (e.g., “which bus goes to the university hospital?”) and timetables (e.g., “when does the next one leave?”). Like the Mailman and AthosMail applications, the Busman system has a speech interface and can be used with mobile and desktop telephones, but unlike

the e-mail systems, the Busman system is unimodal, i.e. it does not use DTMF keys.

The Interact system developed in the national Finnish research project is similar system providing Helsinki area bus timetable information. The system also has multimodal extensions, which include a graphical touchscreen interface for information kiosk style usage, and a short-message interface for mobile non-speech usage.

The third timetable system ("*Pysäkkimies*" in Finnish) is targeted for users who want to know timetables for specific bus stops. This service is in daily use in Tampere. In an ongoing project it will be expanded with the guidance information. This is targeted especially for visually impaired users and to the mobile use.

Other applications

The Jaspis architecture has been used for development of several pervasive computing applications. One example is Doorman. The Doorman system serves staff members and visitors in an office environment. The system controls access to our premises by identifying staff members and helping visitors to find the place or person that they are looking for. The system gives speech and audio-based guidance to visitors about how to reach their destination in our premises.

CONCLUSIONS

In this paper we have presented the Jaspis architecture to address the challenges of flexible and adaptive speech applications. We presented the main principles of the Jaspis architecture, and explained how it addresses the challenges of both traditional and conventional speech applications. Finally, we presented examples of several Jaspis-based applications.

During the development of the Jaspis architecture and its various applications we have worked in collaboration with users groups with special needs. We have worked mostly with visually impaired users. In the future we like to extend our work to cover more user groups with special needs. The adaptive and flexible nature of the Jaspis architecture provides strong support to this.

REFERENCES

Bolt, R. "put-that-there": Voice and gesture at the graphics interface. Computer Graphics, 14(3): 262-270, 1980.

Dahlbäck, N., Jönsson, A. Knowledge Sources In Spoken Dialogue Systems. In Proceedings of Eurospeech'99, Budapest, Hungary: 1523-1526, 1999.

Hakulinen, J., Turunen, M., Salonen, E.-P. Agents for Integrated Tutoring in Spoken Dialogue Systems. In Proceedings of Eurospeech 2003: 757-760, 2003.

Manaris, B., MacGyvers, V., Lagoudakis, M. Universal Access to Mobile Computing Devices through Speech Input. In Proceedings of 12th International

Florida AI Research Symposium (FLAIRS-99), Orlando, FL: 286-292, 1999.

McTear, M. Spoken Dialogue Technology: Enabling the Conversational Interface. ACM Computing Surveys, 34, 1, March 2002: 90-169, 2002.

Tang, J., Yankelovich, N., Begole, J., Van Kleek, M., Li, F., Bhalodia, J. ConNexus to Awarenex: Extending awareness to mobile users. In Proceedings of CHI 2001, 2001.

Turunen, M., Jaspis - A Spoken Dialogue Architecture and its Applications, PhD dissertation, University of Tampere, Department of Computer Sciences A-2004-2, 2004.

Turunen, M., Hakulinen, J. Jaspis - A Framework for Multilingual Adaptive Speech Applications. In Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000), 2000.

Turunen, M., Hakulinen, J. Mailman - a Multilingual Speech-only E-mail Client based on an Adaptive Speech Application Framework. In Proceedings of Workshop on Multi-Lingual Speech Communication (MSC 2000), 2000: 7-12.

Turunen, M., Hakulinen, J. Jaspis^2 - An Architecture For Supporting Distributed Spoken Dialogues. In Proceedings of the Eurospeech 2003: 1913-1916.