

DUMAS – Adaptation and Robust Information Processing for Mobile Speech Interfaces

Kristiina Jokinen

University of Art and Design Helsinki
Hämeentie 135 C
FIN-00560 Helsinki Finland
kjokinen@uiah.fi

Björn Gambäck

SICS, Swedish Institute of Computer Science AB
Box 1263
SE – 164 29 Kista, Sweden
gamback@sics.se

Abstract

In this paper we present the EU-IST project DUMAS (Dynamic Universal Mobility for Adaptive Speech Interfaces), and discuss adaptation and robust information processing as realized in AthosMail, a speech-based multilingual email application developed within the project. AthosMail allows users to read and manipulate their mailbox via a mobile phone. One of the goals of the research conducted in the project has been to develop a spoken interactive mail system with components that would make the user's interaction with the system more flexible and natural. This paper gives an overview of the project as well as the AthosMail components that support adaptation both on the system level and functionality towards the user.

1. Introduction

Recent advances in human language technology have made spoken-dialogue systems a commercial possibility which can be used in several interactive applications. The state of the art speech technology is already on such a high level of accuracy and precision that the users can dictate text or execute simple control commands to direct system operations, and also have short conversations with the system to search for information or to complete well-defined tasks like hotel room booking. However, current speech-based applications, interaction techniques and application development architectures still lack many features necessary for natural multilingual interaction. The three main areas where the current technology falls short and needs improvement are:

- 1) Incapability to process structured text with different presentation formats and with different languages: e.g. electronic text processing requires new facilities that can cope with various different formats like tables, lists, URLs and email addresses, possibly in different languages while speech interface requires intelligent error handling and compensation of the possible misunderstandings.
- 2) Limited conversational abilities: present-day dialogue systems designed for various information services (train schedule, hotel

information) do not satisfactorily cope with requests that are syntactically incomplete and/or semantically creative, or with requests that refer to information that results from previous requests within the same interaction situation (“go to the previous paragraph”, “try Brussels instead”).

- 3) Limited user models: incapability to adapt to the user's personalised needs and to take into consideration previous interactions with the user, or to learn the user's profile through repeated interactions with him/her.

The main ambitious goal in the DUMAS project is to furnish electronic systems with intelligent spoken interaction capabilities. In particular, we have investigated adaptive multilingual interaction techniques to handle both spoken and text input and to provide coordinated linguistic responses to the user. Moreover, future communication with electronic systems requires dynamic and adaptive capabilities, and the project has thus also explored possibilities for building systems that can learn through interaction and adapt their behaviour to different users and different situations.

The project has built AthosMail, an e-mail application that allows users to read and check their emails via a mobile phone. AthosMail is multilingual, and can understand Finnish,

English and Swedish, and its functionality can be adapted to different users.

The applicability of the framework is also planned to be investigated on other applications, such as speech-based document retrieval, speech user interfaces for SMS messages, text television and radio stations. Applications for disabled people, such as services for newspaper reading, are already on the way.

2. The DUMAS Consortium

The DUMAS consortium consists of eight partners from four different countries. The project is coordinated by the Swedish Institute of Computer Science (SICS), one of Europe's leading NLP sites. SICS offers a wide range of resources within language technology, and is mainly responsible for the linguistic and speech processing parts of the system's input analysis, as well machine learning techniques. The scientific coordinator of the project is the University of Art and Design Helsinki (UIAH) and the Intelligent Dialogue Interaction Systems research group at Medialab. UIAH focuses on the project's user modelling and adaptation parts, and collaborates especially with SICS on the machine learning techniques (neural networks) in order to model the complexity of user-specific features such as the user's skill level and preferences for message prioritizing.

The three other research partners in the project are the University of Tampere (UTA), Finland, the University of Manchester Institute of Science and Technology (UMIST), England, and the Centre for Speech Technology at the Royal Institute of Technology (KTH), Sweden. UTA specializes in innovative user interfaces, and they bring knowledge about the design, implementation and evaluation of speech-based applications to the consortium. UTA also provides the basis for the system architecture. UMIST is the foremost centre in the UK working on multilingual aspects of language engineering, and covers pure and applied research in various areas of language studies, linguistics and computational linguistics. They specialise in dialogue management, utterance planning and tactical generation, as well as usability studies and the needs of blind and sight-impaired users. KTH works on the user modelling and adaptation aspects of spoken interaction.

The industrial partners of the project include ETeX Sprachsynthese AG, Germany, and Connexor Oy and Timehouse Oy, Finland. Connexor brings in expertise in natural language text analysis and has established technology for part-of-speech tagging and syntactic analysis for several languages. ETeX and Timehouse produce and sell technology based on the use of a speech synthesis of highest quality (Text-to-Speech, TTS). Timehouse develops its own speech synthesis applications, while ETeX uses products developed by Elan text-to-speech in Toulouse. ETeX and Timehouse provide the consortium with the speech expertise and the user-groups (such as the Finnish Blind Association) among their customers, and are interested in the exploitation of the results.

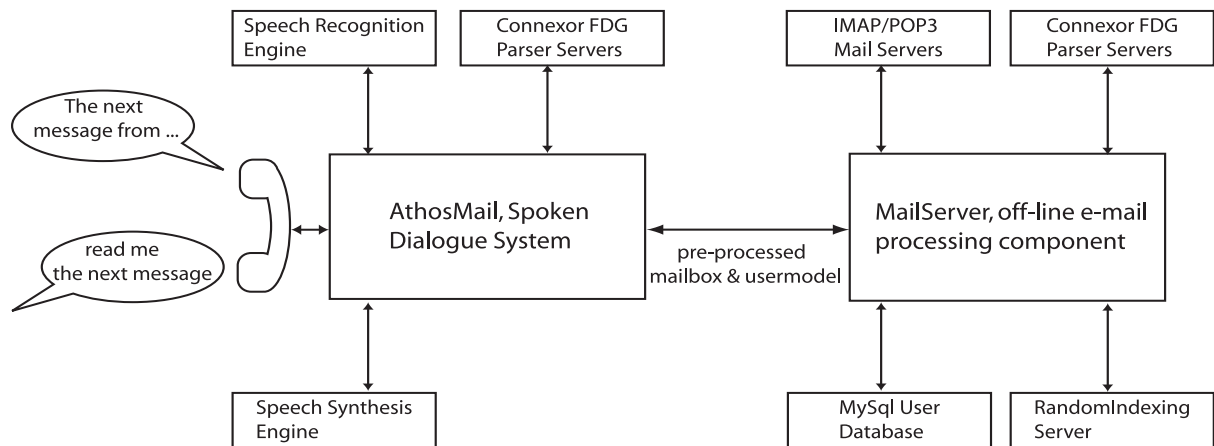
The consortium is in an excellent position to produce truly multi-lingual applications, since it can experiment with four different languages (English, Finnish, Swedish and German, with the first three being specifically addressed in the project).

3. System Architecture

The AthosMail application is a speech-based multilingual e-mail application. It is based on the existing Mailman application (Turunen and Hakulinen, 2000a), and provides new components for text-processing, dialogue management (Black et al., 2003), semantic template construction (Cheadle and Gambäck, 2003), user modelling (Jokinen et al., 2002), and random indexing (Sahlgren, 2003). The new components deal with functionality such as information retrieval, information extraction, user preferences, message prioritizing, logical form building, response planning, and response generation.

The AthosMail functionality allows the users to check their e-mail by a mobile phone. Besides the read command, the users can also manipulate the content of their mail-box and use meta-commands that deal with requests for help and repetition, with marking of a message important (for some future use) and with cancelling of the previous command.

In order to provide an efficient interface for telephone use, messages are automatically organized into manageable groups, and divided into sections in which the user can navigate when a message is being read. The system allows for both spoken input and touch-tone (DTMF) keys over the phone, as



well as for typed keyboard input.

AthosMail is built as an instance of a more general, flexible framework for spoken dialogue applications. This framework (“Athos”) allows for having different agents working in parallel on both the same and different tasks (e.g., multiple speech recognizers for one language and for different languages).

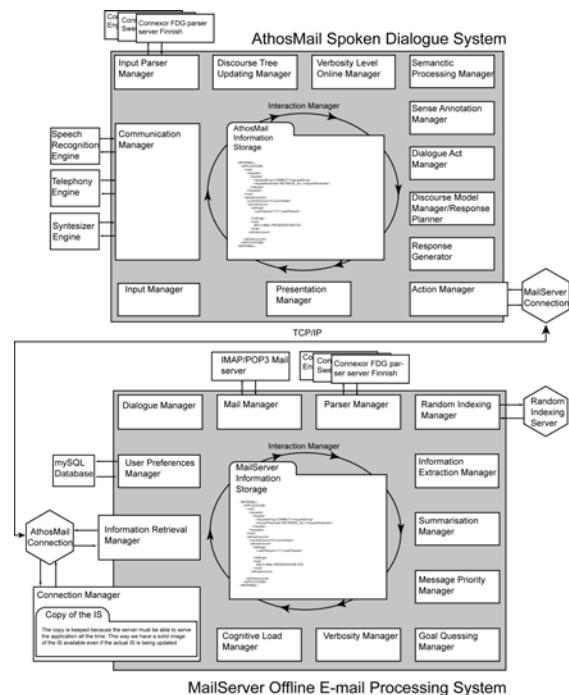
Athos is in turn an extension of Jaspis (Turunen and Hakulinen, 2000b; Turunen and Hakulinen, 2003), an architecture supporting highly distributed – but coordinated – components, shared system knowledge, and system-level adaptation. The system consists of several managers that are under central coordination. Evaluators are used to choose between different agents. This feature plays a major role in the system adaptation. The system architecture is distributed so that different managers and even agents can run on different computers and platforms.

The high-level system architecture is depicted in Figure 1. The e-mail interfacing system consists of an on-line and an off-line part, that communicate over an XML-RPC interface, as shown in more detail Figure 2.

The on-line AthosMail application is the actual dialogue system the user is interacting with. The off-line MailServer processes e-mail messages continuously and interacts with AthosMail. This is necessary because some of the techniques used for message processing are resource intensive, and it is not possible to perform them in real-time.

The off-line system coordinates components related to the handling of e-mail messages. It receives requests from the on-line system and performs operations on the user’s mailbox. For

this purpose it utilizes components for communicating with IMAP and POP3 servers, transforming text messages into XML documents, filtering out unwanted messages, and modifying the contents of messages. Messages are prioritised and categorized into meaningful groups. The MailServer also contains components for user modelling, adaptation, and information refinement.



4. User Modelling in AthosMail

In recent years, the notion of adaptivity has become more important when building user interfaces that take various users into account. Adaptivity is often realised in a static and mechanical way as personalised interfaces where user preferences take the form of colour or sound choices, and characteristics are listed in personal profiles.

In DUMAS, the departure point for adaptation and user modelling is in flexible human-computer interaction, and the research has focussed on online adaptation, as opposed to the common state of affairs where the systems require the user's adaptation to the system. The project has especially explored system capabilities to adapt and adjust its functionality dynamically according to various types of users and user actions. Furthermore, for the mobile users of AthosMail, it is important that the content of their mail-box is presented in such a way that it best serves the user's current interests and preferences. Message prioritizing is thus one of the important aspects of user modelling, and for this we have especially worked on robust information processing and machine-learning techniques.

The AthosMail User Model consists of three components: Message Priority, Goal Guessing, and Cooperativity Component. Together, these components take care of the system's adaptation to the users: they record the user characteristics and actions, and give recommendations to the system to tailor its responses so that the responses follow the assumed skill levels of the user and her expectations about natural and enjoyable interaction. The purpose of the user modelling is three-fold:

1. to provide flexibility and variation in the system utterances,
2. to allow the users to interact with the system in a more natural way, and
3. to allow developers to implement and test machine learning techniques.

The system monitors the user's actions in general, but also specifically on each possible system act. Thus the system can provide help tailored with respect to the user's familiarity with individual acts. For instance, the user may need more help with commands that she does not use so often.

The UM components produce recommendations for the Dialogue Manager, and the text planning and generator components use these recommendations when producing system responses. For instance, if the user is a novice, the system can provide longer and more explicit utterances than when the user is familiar with the system and its functionality. The UM can also be used in the interpretation of the user utterances to give expectations of the user's vocabulary and likely next actions. In the beginning of the interaction the default user preferences are loaded into the system from the UM.

5. Robust Information Analysis

The project has developed methods for the analysis and interpretation of spoken language contributions, focusing especially on structured text and the use of conversational context. Methods for intelligent text processing have also been investigated.

AthosMail has two types of semantic analysis tasks to address: the interpretation of the (spoken) user commands to the system, and the interpretation of the documents in the application, the (written) e-mails. The language of e-mails in many ways resembles that of speech, with shorter and grammatically incomplete utterances. To handle this type of input we need a very robust interpretation strategy; thus we aim to complement classical deep-level, logical-form-based language processing with a more template-based, domain-dependent semantic interpretation.

We have thus investigated methods for the three main information refinement tasks: Information Retrieval, Information Extraction, and Text Summarisation. These methods are necessary the AthosMail application in order to be able to retrieve some relevant e-mails for a user query, to extract a particular piece of knowledge from a set of e-mails, or for summarising some e-mails selected by the user. Here, the e-mails are the documents, and the user's query is expanded into a format that can be matched with the representation of the document set. The documents that best match the query are retrieved, ranked and presented to the user, usually as a list of possibly relevant documents. The Information Retrieval engine is then mainly used as a supporting agent in the information extraction and text summarisation tasks.

6. Dialogue Management

The interaction model of the AthosMail application is based on a user-initiative dialogue strategy with some mixed-initiative features. In such open-ended dialogues tasks are not well structured, the user may interact with the system fairly freely, and his/her goals may differ between sessions according the mailbox content. However, system takes the initiative in special situations, such as in the login procedure and in error situations. Depending on the user input, the dialogue is handled by a suitable set of agents. Information is provided to the user according his/her expertise level.

7. Data Collection

In the initial design phase, we needed information for the functional specification on how the users would interact with the envisaged AthosMail system, and data on vocabulary and language use of the future users, for the dialogue and language model specification. The actual data collection was performed in a role-playing, scenario-based Wizard of Oz setting (Kanto et al., 2003).

The WOz set-up entailed two novel features: firstly, instead of doing solo sessions with a static mailbox, our test users communicated with each other in groups of six using a simulated speech-based e-mail system via telephone. The e-mail system was controlled by a Wizard, allowing the subjects to dictate and receive messages, arrange them in folders, etc. Secondly, the communication took place over several sessions in a period of several days during which the subjects played a role defined for them in the scenario. This was to gain information on the effects of user accommodation and system adaptivity, focusing on the development of user expertise, user strategies, and linguistic accommodation.

We chose to let several persons interact with each other, in a group scenario, in order to create a more authentic experimental setting, since we wanted the participants to be motivated by the interaction within the group rather than by the interaction with the e-mail system. The scenario generated active e-mail traffic, even heated discussions, and the subjects seemed committed to playing their characters. The messages were generally short compared to written e-mails, as had been anticipated due to the modality of the

interaction. The participants had been advised to call at least twice a day, resulting in a total of around 60 dialogues (about 6 hours of speech) per language.

8. Conclusions and Future Work

In this paper we described AthosMail, an interactive speech-based email application developed in the DUMAS project. By providing a generic application development environment for multilingual speech application, the project has opened up new substantial commercial prospects for natural speech dialogue applications. Furthermore, by constructing AthosMail we have developed not only a product, but also a set of methods than can be used in other domains as well. By introducing prototypes of new innovative applications we have made groundwork for the concrete adaptive multilingual speech applications in many areas.

9. Acknowledgements

This research was carried out within the European Union's Information Society Technologies project DUMAS (Dynamic Universal Mobility for Adaptive Speech Interfaces), IST-2000-29452. We thank all the project participants for cooperation in the DUMAS project.

10. References

- William Black, Paul Thompson, Adam Funk, and Andrew Conroy. 2003. Learning to classify utterances in a task-oriented dialogue. In K. Jokinen, Y. Wilks, B. Gambäck, W. Black, and R. Catizone, editors, *Proceedings of the EACL Workshop on Dialogue Systems: Interaction, Adaptation and Styles of Management*, Budapest, Hungary, April. ACL.
- Maria Cheadle and Björn Gambäck. 2003. Robust semantic analysis for adaptive speech interfaces. In C. Stephanidis, editor, *Universal Access in HCI: Inclusive Design in the Information Society*, volume 4, pages 685-689, Mahwah, New Jersey, June. Lawrence Erlbaum Associates.
- Kristiina Jokinen, Jyrkki Rissanen, Heikki Keränen, and Kari Kanto. 2002. Learning interaction patterns for adaptive user interfaces. In N. Carbonell and C. Stephanidis, editors, *7th Workshop on User*

- Interfaces for All*, Paris, France, October. ERCIM.
- Kari Kanto, Maria Cheadle, Björn Gambäck, Preben Hansen, Kristiina Jokinen, Heikki Keränen, and Jyrki Rissanen. 2003. Multi-session group scenarios for speech interface design. In C. Stephanidis and J. Jacko, editors, *Human-Computer Interaction: Theory and Practice (Part II)*, volume 2, pages 676-680, Mahwah, New Jersey, June. Lawrence Erlbaum Associates.
- Magnus Sahlgren. 2003. Content-based adaptivity in multilingual dialogue systems. In *Proceedings of the 14th Nordic Conference of Computational Linguistics*, University of Iceland, Reykjavík, Iceland, May.
- Markku Turunen and Jaakko Hakulinen. 2000a. Mailman – a multilingual speech-only e-mail client based on an adaptive speech application framework. In *Proceedings of the Workshop on Multilingual Speech Communication*, pages 7-12, Kyoto, Japan, October.
- Markku Turunen and Jaakko Hakulinen. 2000b. Jaspis – a framework for multilingual adaptive speech applications. In *Proceedings of the 6th International Conference on Spoken Language Processing*, Beijing, China, October.
- Markku Turunen and Jaakko Hakulinen. 2003. Jaspis2 – an architecture for supporting distributed spoken dialogues. In *Proceedings of the 8th European Conference on Speech Communication and Technology*, pages 1913-1916, Geneva, Switzerland, September. ISCA.