Tampere
University
Computer
Human
Interaction
Group

MAILMAN - a Multilingual Speech-only E-mail Client Based on an Adaptive Speech Application Framework

Markku Turunen, Jaakko Hakulinen

University of Tampere

Department of Computer and Information Sciences

Workshop on Multi-Lingual Speech Communication
Kyoto, Japan, 11.10.2000.

Contents

Introduction 1

Multilingual speech application development suffer from a lack of proper tools

We need to support multilingual:

- speech outputs (not only synthesizers)
- speech inputs (not only recognizers)
- dialogue management strategies

We must support:

- multiple components for different languages
- components which can be shared between languages

Introduction 2

Interesting and problematic questions:

- when to use what language in different situations?
- how to model and present multilingual information?
- how to handle prosody in different languages?
- how to adapt dialogue handling to different languages?
- how to interpret multilingual speech inputs?

We need *interaction models*, *techniques* and *guidelines*

Introduction 3

Electronic mail as a multilingual domain

- any mixture of languages in unpredictable combinations
- structural elements, such as e-mail addresses, lists and tables
- multilingual inputs even within a single user
- different dialogue management strategies for different languages and cultures

Introduction 4

Common multilingual issues

- proper use of multiple synthesized voices
- detection accuracy
- frequent changes in voice can lead to problems - can also be used as an advantage
- structured elements: how to read common elements?

All of these problems are *very common* and if not properly handled, can destroy the usefulness of an application

Mailman 1

Our approach

We have studied how problematic questions in multilingual speech application can be solved

Mailman (Postimies), a speech-only e-mail client

To support multilingual issues we developed *Jaspis*, an open architecture for adaptive speech application development

Mailman 2

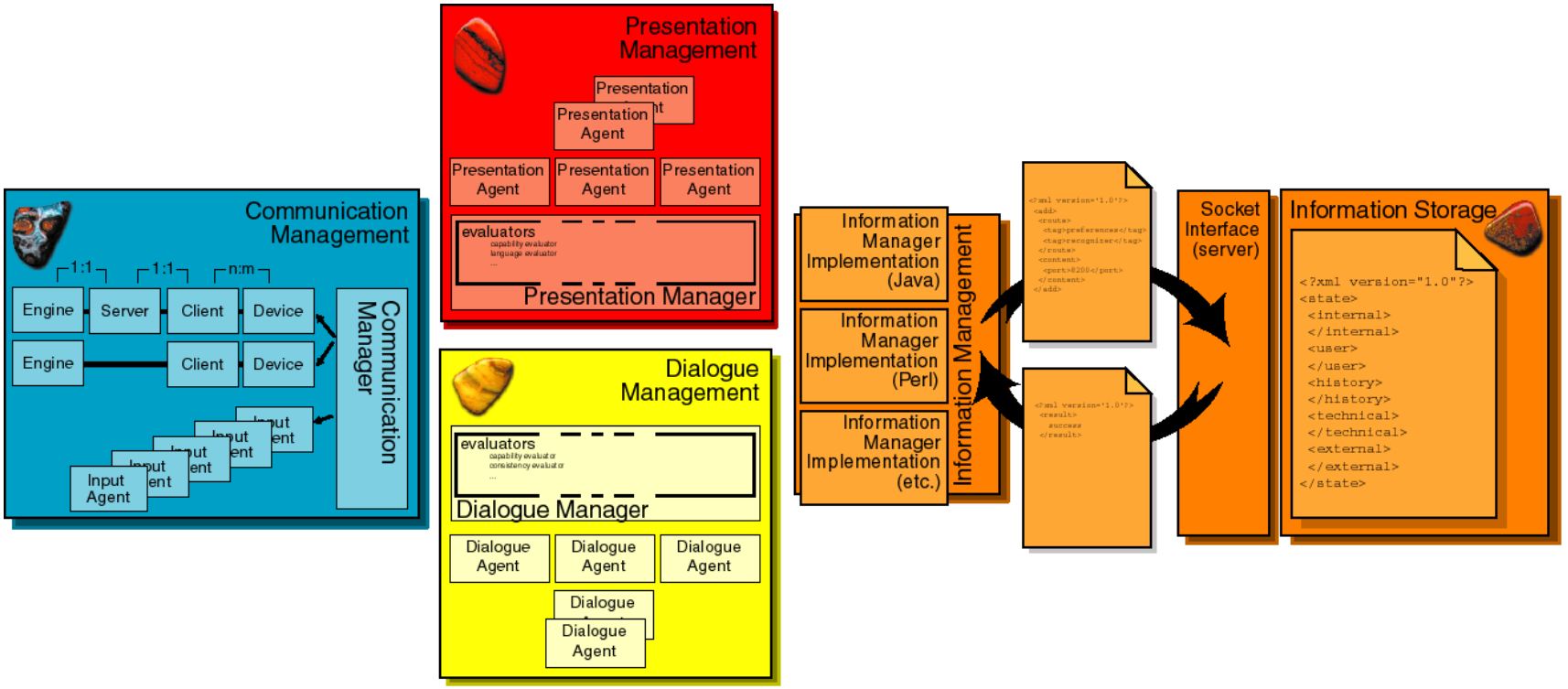
Features

- basic e-mail reading capabilities for telephone users
- currently bilingual (Finnish and English)
- rather minimalist and extendable than hard-tuned and specialized: basic support for all languages
- extensible for multiple languages and interaction strategies
- this far we have studied mainly speech outputs
- in daily use

Mailman 3

Mailman uses Jaspis architecture

- shared information storage
- *agents* - *managers* - *evaluators* -architecture
- *presentation*, *dialogue* and *communication* modules



Mailman 4

Information presentation

- a conceptual, language independent information presentation
- crucial especially for multilingual speech outputs
- every component can utilize all the information

Agents, Managers and Evaluators

- agents handle various interaction situations
- managers choose the most suitable agent to handle each situation
- evaluators evaluate available agents

Multilingual Speech Outputs 1

Speech outputs

- three kinds of speech outputs: *system utterances*, *views of e-mails* (i.e. different structures) and the e-mail messages

Presentation framework

- multiple agents are used to convert (conceptual) messages to speech outputs

- for every output the most suitable agent is selected based on the message, context, user model etc.

Multilingual Speech Outputs 2

Presentation agents

- ”natural language generation” (for different languages)
- speech synthesis control, including the selection of synthesizers etc.
- prosodic information (according to language)
- spealized agents for certain situations (reusability)
- agents for same situations (adaptivity)

Multilingual Speech Outputs 3

Presentation evaluators

- each agent has a profile that tells about its features – evaluators use these to check in what the agent is capable of
- other evaluation sources include user profile, current dialogue history, output capabilities (such as synthesizers etc.)
- evaluators are specialized for certain evaluations – in this way reusable evaluators can be built

Multilingual Dialogue Handling 1

- multiple dialogue agents specialized for certain kind of dialogue turns
- support for different communication strategies needed by different languages
- we can also support different dialogue strategies of different users (system-initiative, mixed-initiative)

Multilingual Dialogue Handling 2

Dialogue agents

- language or culture specific agents
- language independent dialogue components can be shared
- toolbox of common dialogue situations (error correction etc.)

Dialogue evaluators

- the style of the communication is explicitly managed

Multilingual Input Handling

Input agents

- handle, combine and coordinate different input streams
- language-aware input agents
- language independent components for common interaction techniques
- can use multiple recognizers, grammars and error correction strategies for different languages

Example 1

Simple example situation:

1. user*:* "*read the third message*"
   "*lue kolmas posti*" (in Finnish)

2. conceptualized to form "READ [3]" by an input agent which understand either English or Finnish

3. a dialogue agent, which handles the e-mail reading is selected. It checks if there is at least three messages in the current folder - if message does not exist, an error message is produced (NO MESSAGE [3]).

Example 2

example continues…

4. an output agent specialized for error-messages is selected according to the language of the user

5. system: "*There is no <SLOW>third</SLOW> message at the <EMP>first</EMP> folder*.

"*<EMP>Ensimmäisessä</EMP> kansiossa ei ole <SLOW>kolmatta</SLOW> viestiä.*"

Conclusions and Future Work 1

Multilingual speech applications contain many problematic interaction situations

- if not properly handled, destroy the usefulness of the system
- technological approaches, such as machine translation, multilingual synthesizers and recognizers are not enough
- we need adaptive interaction methods which support multilingual inputs, dialogue strategies and outputs

Conclusions and Future Work 2

Our solution:

- Mailman uses presentation, dialogue and input agents to handle multilingual issues in e-mail domain
- in future we focus especially on multilingual inputs
- Jaspis is presented in ICSLP 2000
  - ”Jaspis - A Framework for Multilingual Adaptive Speech Applications”
  - Wednesday 18th
  - Hall E
  - 13:30-15:15 (17:00)

http://www.cs.uta.fi/research/hci/SUI/

speech@cs.uta.fi