Multimodal Stack Widget – a Proactive User Interface

ETSI Workshop on Multimodal Interaction on Mobile Devices
18 -19 November 2008 - ETSI, Sophia Antipolis, France

Dr. Robert van Kommer, Swisscom AG
SCS-SIS-ICC-MME
Robert.vankommer@swisscom.com
Outline

• Introduction
  – Road map of telephony user interfaces
  – The interaction pattern of interest
  – Multimodal stack widget – An overview
• Stack widget’s technologies with implementation examples
  I. Multimodal user interface \(\rightarrow\) Freedom of modality
  II. Meta-user model \(\rightarrow\) A user-centered data modeling
  III. Intelligent user interface \(\rightarrow\) Subjective filtering
• Conclusion

Annexes

Abstract
Learning performance and validation
Road map of telephony user interfaces

It informs personalizes, recommends. It is proactive and pervasive.

Emotional intelligent UI
Affective computing
Mood detection and its usage
Towards a more human-like UI

Today’s possible usage mix on an iPhone: TUI 35% - VUI 3% - GUI 50% - MUI 10% - IUI 2% - EUI 0%
The interaction pattern of interest

- Interaction pattern:
  - **Input:** A user request (e.g. a search string)
  - **Output:** A ranked list of items

- Pattern examples:
  - Search engines
  - Directory assistance
  - Timetable information
  - Media retrieval systems
  - Hit-parades
  - News
  - List of emails
  - Recommendation lists
  - Consuming RSS feeds
Multimodal stack widget – An overview

- Multimodal User Interface (MUI)
- Meta-User Model (MUM)
- Intelligent User Interface (IUI)

Web portal

Users’ Meta Data

User: Alice
Model storage
Service 1
Service 2

Info service 1
Service 2

This content is relevant for Alice!
MUI – A modality choice at each interaction node of applications

Communication services/ Content access
MUI – Multimodal User Interface concepts

- **Freedom**: Users are now free to choose their preferred modality
  - Depending on users’ situation contexts (e.g. driving a car, being in a conference room, running to catch a train)
  - Depending on a disability (e.g. visual impairment)

- Improved **robustness** is achieved through redundancy in available modalities as compared to e.g. “voice-only” interfaces

- Given the chosen modality, **efficiency** is increased in terms of completion rate (e.g. speech input, display output, as opposed to, speech input and speech output)

- **Further potential improvements** could be achieved by combining/synchronizing several modalities
  - E.g. autocompletion upon combined speech and text inputs
MUI – Implementation example on the Nokia N95

Multimodal stack widget

Text input modality

Voice input modality

Speech recognition N-best output

Stack output: At each user interaction, the best matching and ranked results are presented on top of the stack.
MUI – Voice user interface example

- State-of-the-art speech recognition engine (Sphinx 4, an open source Java software running on the web portal)

- Speech recognition models are available in French, German, Italian, English.

- J2ME Java at the client side and Tomcat & Apache as web server
Why → From a service provider’s perspective, better knowing the user enables to improve quality of service

How → To create one meta-user model that supersedes or eventually complements several existing data access models

The bottom line:

*How many user’s interactions are needed by the system to learn a good meta-user model?*
MUM – How to model users?

• Making the process as transparent as possible for users

• Boosting the learning efficiency and therefore reducing the needed quantity of user’s interactions

• Further efficiency is achieved by grouping personalization and modeling information into one hierarchical structure called ePersona

• User’s preferences and the meta-user model are therefore easily propagated to new services with similar content (solving the “cold-start” issue of standard recommendation engines)

• Finally, a user-centered model empowers users to control their personal data and the related access rights of third parties. Security and privacy is now under user’s control.
Besides the current modeling scheme (Support Vector Machine), many potential improvements exist:

- At the start, when almost no user’s interactions exist, the very first **content selections, keywords or associated tags** are to be used to predict user’s interests
- **Post-rating/ranking** of the above-mentioned approach
- **Active learning** proactively asks users to explicitly rate items that are hard to classify automatically
- **Online learning**: Learned models are updated at each user interactions
- **Hidden Markov Model**: Enables to learn/predict users’ state sequences
- **Support Vector Machine** for learning a better ranking of stack items
- **Reinforcement learning**: Automatic learning and discovering of better dialog structures and strategies
MUM – Implementation example on a mobile phone

- Implicit versus explicit users’ data collection
MUM – Implementation example at the portal side

Getting

Enables the user to register to personalized content’s sources

Sharing

Users can join the community and share content as well as create new content events

Managing

A full range of tools offer to the user an in-depth control over his/her personalization and data

One-stop-shop for getting, sharing and managing my meta data
IUI – Intelligent User Interface - “to be informed”

- “To be informed” about my truly relevant content
- Economy of user’s attention through intelligent man-machine interactions
IUI – Intelligent User Interface

• By using the previous described meta-user model, IUI enables **subjective filtering** of new or of existing content (content discovery)
• **Proactively** informs by pushing filtered content on my mobile phone
• **Warning!** Any user-meta modeling performance is dependent on data quality!  **Garbage in => garbage out!**

The learning scheme (Support Vector Machine) has now been validated on several data sets:
– Reuters articles (English language, see annexes)
– Bilanz articles (A Swiss-German financial magazine)
– Flickr personalization (Using both images’ content and tags)
– Movie Multilens data set (Using users’ rating and movie data)
– Picture profiling for advertisements (Currently under test)
IUI – Implementation example

IUI offers new perspectives
• No need to type a text
• No need to speak
• No haptic input needed
Content of personal interest is pushed proactively on my phone

A relevant item is pushed right on top of the stack
Conclusion: Multimodal stack widget contributions

- **MUI - A robust user interface available at each interaction node**
- **MUM - Learned personalization**: Each single user gets his/her own dedicated meta model and therefore a more personalized service
- **IUI - Proactive user interface**: Content is proactively pushed on user’s devices via subjective filtering
  - The new concept is: **To be informed** instead of relentless searching for needles in data haystacks
  - Is it the infobesity cure?
Abstract

- All-in-one mobile phones have changed our social communication behaviors and infotainment habits. For people on the move, to be informed about relevant and new content represents a challenge and an opportunity: on the one hand, user interaction is restricted by user’s situation contexts and phone size; on the other hand, the reachability is increased, mobile phones are always on and carried around. In this context, a unified user interface is proposed in the form of a ranked stack that combines pulled and pushed items; pulled results are typically provided by a multimodal search engine whereas pushed items are proactively included via subjective filtering of new available content. The user interaction concept will be illustrated through an interactive media application tailored towards mobile user experience.
Learning performance and validation
(data set) http://www.daviddlewis.com/resources/testcollections/reuters21578/

- The task is to learn which Reuters articles are about "corporate acquisitions" given that this is of user’s interest

- In the training set, there are 1000 positive and 1000 negative examples

- The test set contains 600 test samples (300 positive and 300 negative samples).
Support Vector Machine (SVM) approach

- The number of user’s inputs needed (relevant, not relevant)
Support Vector Machine approach suite

- Using unlabelled data as well
- With 10 (5 positive and 5 negative) user’s inputs, the relevance of information is improved 10 times (from 50% to 5% of errors reduction)

<table>
<thead>
<tr>
<th>Correct classification</th>
<th>Number of positive and negative labelled training articles out of 2000 that are used in the transductive SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>100,00%</td>
<td>2</td>
</tr>
<tr>
<td>95,00%</td>
<td>5</td>
</tr>
<tr>
<td>90,00%</td>
<td>10</td>
</tr>
<tr>
<td>85,00%</td>
<td>20</td>
</tr>
<tr>
<td>80,00%</td>
<td>40</td>
</tr>
<tr>
<td>75,00%</td>
<td>80</td>
</tr>
<tr>
<td>70,00%</td>
<td>160</td>
</tr>
<tr>
<td>65,00%</td>
<td>320</td>
</tr>
<tr>
<td>60,00%</td>
<td>640</td>
</tr>
<tr>
<td>55,00%</td>
<td></td>
</tr>
<tr>
<td>50,00%</td>
<td></td>
</tr>
</tbody>
</table>