OLGA - a Multimodal Interactive Information Assistant

Olle Sundblad and Yngve Sundblad
CID - Center for User Oriented IT Design
KTH
S-100 44 Stockholm, Sweden
+46 8 790 9227, +46 8 790 7147
osu@nada.kth.se, yngve@nada.kth.se

ABSTRACT

This video is a description and use scenario of an interactive information assistant with both speech and visual direct manipulation interface. The 3D-animated assistant, nick-named OLGA, is intended to help in situations where people seek information. It is here demonstrated with consumer advice about microwave ovens. The demonstrator is modular and distributed, with separate modules from different partners and computers communicating via a server. The OLGA project is highly interdisciplinary, involving researchers from linguistics, speech technology, graphic illustration and computing science. Possible extensions and other uses of the assistant are demonstrated.

Keywords

Multimodal interaction, discourse modeling, speech/voice, animation, information assistant, software architecture

INTRODUCTION AND BACKGROUND

The aim of the OLGA project is to create and study the use of an interactive multimodal tool with voice and visual interaction for navigation with help of an animated person ("Olga") in an information space, representing possible services and products over the net, available publicly and privately, e.g. via interactive computer-TV.

Today the OLGA system can only answer questions about microwave ovens. We chose that area because consumer advice is an area where an interactive assistant can be of great help and it is feasible as a microworld where the vocabulary is limited.

The project is an interdisciplinary effort to use, integrate and extend knowledge and experience on

- modeling discourse and dialogue at the department for Linguistics, Stockholm University
- speech based dialogues at the department of Speech, Music and Hearing, KTH, [1]
- modeling visual (graphic) dialogues at NADA, the department for Numerical Analysis and Computing Science, KTH, [3]
- coordinating / synchronizing parallel programs at SICS, the Swedish Institute for Computer Science, [2]

PRIMARY GOALS AND RESEARCH QUESTIONS

- Development of a microworld for modeling discourse and dialogue for consumer advice on microwave ovens. How do current methods work in such an advice situation with speech and visual interaction and how can domain knowledge be represented for this purpose?
- Modeling and implementing speech for animated dialogues where at least trained users can have a spoken dialogue with the system on microwave ovens. How can visual couplings such as lip-synchronizing, facial expressions and gestures be made with animated figures?
- Modeling and animating the visual look of OLGA and the visual dialogue with the user. How can the visual look be coupled to the speech dialogue? What can be visualized and how (graphics, picture, table, film)?
- Handling coordination and synchronization situations in multimodal interaction. How can existing techniques be used and extended for multimodal interactive information assistance?

OLGA

DISCOURSE AND DIALOGUE

As input for modeling the discourse and dialogue consumer advisers were interviewed and advice sessions were video filmed and analyzed by the Linguistics department. Not only linguistic forms but also contents, structure, subject progression and communication activities were analyzed with implications for the advice situation.

Knowledge representation

This resulted in knowledge of typical questions in the advice situation, typical "turns" in the conversations etc., to be represented in a data base, as well as facts about microwave ovens (obvious and trivial) and at least some knowledge about the surrounding world (in order to make the dialogue with the human user reasonably natural, is a much more difficult problem). History of the current dialogue session must also be represented so that assistant does not unnecessary repeat itself of assumes knowledge the user does not have.

Scenarios

The data were used to simulate microwave oven advice dialogues and to create four scenarios used for the development of the demonstrator.

SPEECH AND VISUAL DIALOGUES

The speech analysis is based on test persons reading typical advice situation sentences, on which the demonstrator's recognition ability is based. The text recognized is analyzed syntactically and matched with domain specific templates. OLGA's speech is based on formant synthesis, with female tone parameter setting.

To be able to use different channels for different types of information gives a more natural interface. For example, if she is to present the a search result of ten microwave ovens with prices and test results she would choose to display it in graphics with a pointing gesture. If a search failed to return any result, she would give a verbal explanation with a disappointed gesture.

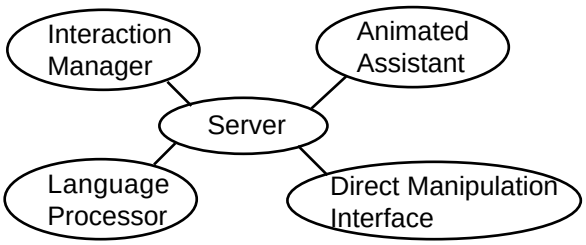
ANIMATION

The advantage of using a cartoon figure as Olga instead of a human model is that you do not have as high expectations of a figure. It is also possible to exaggerate gestures to increase understanding. The head is proportionally larger than the body. This combined with lip synchronized speech, makes it easier for users to understand what Olga says.

The animated figure also gives the system access to means of communication that humans normally use, for example gestures and facial expressions. All these things helps the dialog run smother. A nod or raise of eyebrows can show that Olga understood what the user said even though there is not yet any information to show. The antennas can add to emotions and normal gestures, for example, they can droop when she's sad, spark if she's angry etc. The 3D model of Olga was made with Alias Wavefront.

COORDINATION AND SYNCHRONIZATION

The OLGA system consists of four modules and a server:



When someone speaks to Olga, the Language Processor translates sound to text. The text is then sent to the Interaction Manager, which interprets the meaning of what the user said and decides how to react. If the response includes speech a message is sent to the Animated Assistant module, if graphical output is to be used a message is also sent to the Direct Manipulation Interface module. All messages are sent to the server which then sends the message to the target module. This allowed us to develop the modules independently from each other, even at different operating systems and at different locations. It also makes it easy to change a module or add a new one.

Here is a short example from the script used for the session on the video, where OLGA offers (says) a tip, which is accepted in voice from the user and displayed by OLGA.

```
im   aa   state(offer_tip)#
im   aa   say(you have chosen ovens without digital
          timing, do you want to know more about the
          advantages of digital timing?)#
aa   im   speech_done#
im   dmi  input_unblocked#
im   lp   input_unblocked#
lp   im   hear('okay, (what is it good for?)',
          [[type:accept]])#
im   dmi  input_blocked#
im   lp   input_blocked#
im   aa   state(understood)#
im   dmi  show(dmi6,explain,microwave,
          [digital_timing:_115410])#
```

CONCLUSIONS AND POSSIBLE EXTENSIONS

Future work in the OLGA project will include user studies, more complex settings, improving, extending and integrating the speech interface and the visual animations.

ACKNOWLEDGMENTS

This work has been made by the OLGA project group: Eva-Marie Wadman and Olle Sundblad (CID, KTH), Kari Fraurud, Christina Hellman and Nikolaj Lindberg (Linguistics, Stockholm University), Scott McGlashan (SICS) and Jonas Beskow (Speech Technology, KTH), who made the research and implementations, supervised by the project leader group: Björn Granström (Speech Technology), Gunnel Källgren (Linguistics), Yngve Sundblad (CID) and Henrik Wahlforss (Nordvis).

All mentioned, and also colleagues at our research laboratories contributed to the result in many fruitful interdisciplinary discussions.

REFERENCES

1. Bertenstam, J. & al.The Waxholm system - a progress report. *Proc. Spoken Dialogue Systems*, Vigsø 1995.
2. Linda Process Communication, *Sicstus Prolog Manual*, http://www.sics.se/isl/sicstus/sicstus_28.html, 1996.
3. K. Tollmar and Y. Sundblad. The Design and Building of the Graphic User Interface for the Collaborative Desktop. *Computers and Graphics*, 1994, 179-188.