Evoking Gestures in SmartKom – Design of the Graphical User Interface

Nicole Beringer
Institute of Phonetics and Speech Communication,
Schellingstr. 3, 80799 Munich, Germany

ABSTRACT

In the SmartKom project we are developing an intelligent computer–user interface which can deal with various kinds of input, e. g. speech, gestures or facial expression of the user. This interface allows computer experts as well as novices to communicate quasi–naturally with an intelligent multimodal system.
To train a gesture analyzer, data is required, preferably realistic data. Our institute is responsible for the recording and annotation of such data.
Since the planned dialogue system does not already exist the data collection has been realized using the Wizard of Oz–techique: A simulation of the system by humans (the so–called wizards) while the subjects are influenced by instructions suggesting that they are interacting with an existing machine. Our subjects are recorded while solving some small tasks in the SmartKom domains, the recordings are annotated afterwards with regard to speech and gestures.
At this point we would like to present the design of the graphical user interface that we use in the Wizard of Oz recordings particularly with regard to evoking "deictic" and/or non–deictic gestures.
We start with a short outline of the SmartKom project including the data collection, the recruitment of the subjects and the task. Then we describe in detail the design of the graphical user interface which is used in the recording setup. The main focus lies on finding a layout which elicits gestures and on designing the graphical user interface appropriately in order to invite the subjects to use gestures.
Finally, results and future work are discussed in the last section.

1 THE SMARTKOM PROJECT[1]

The aim of the SmartKom project (started in January 2000) is the development of a multimodal computer–user interface which allows the user to communicate naturally with an adaptive and self–explanatory machine.

SmartKom is being developed for three application scenarios, each with different requirements:

1) SmartKom Home/Office to operate machines at home (e.g. TV, workstation, radio),
2) SmartKom Public to have a public access to the internet or other services, and
3) SmartKom Mobile as a mobile assistant.

The system understands input in the form of natural speech as well as in the form of gestures. In order to "react" properly to the intentions of the user the emotional status is analyzed via the facial expression and the prosody of speech.
The output of the system is presented with a graphical user interface – a computer screen projected onto a graphic tablet – and with synthesized language.

MULTIMODAL DATA COLLECTION

In the first phase of the project we are collecting data for the following three different main purposes:

1) The training of speech, gesture and emotion recognizers.
2) The development of user–, language–, dialogue– models etc. and of a speech synthesis module.
3) The general evaluation of the behavior of the subjects in the interaction with the machine.

In each session of a Wizard–of–Oz experiment the spontaneous speech, the facial expression and the gestures of the subjects are recorded.

The technical equipment is as follows:
Audio:
- a microphone array of 4 Sennheiser ME 104
- a directional microphone (Sennheiser ME66/KG) and
- (alternating) a headset or a clip–on–microphone (Sennheiser ME 104).

Video:

1 http://smartkom.dfki.de/index.html

- a fixed positioned digital camera captures the face of the subjects (*facial expression).*
- a second digital camera captures the gestures in a *side view of the subject* (upper body)
- and an infrared sensitive camera (from a *gesture recognizer*: SIVIT/Siemens) captures the hand gestures (2–dimensional) in the plane of the output projection.

Other:

- The coordinates of pointing gestures on the work space are recorded (SIVIT)
- as well as the inputs of a pen on the graph tablet.
- Graphical output as video stream

PREPARATIONS/TECHNICAL DETAILS

Each subject is recorded twice in sessions of about 4.5 minutes. length
The recordings are done directly onto harddisc. They are synchronized manuallyand all layers including the audio and the gesture annotations are aligned in a QuickTime frame [2] which is distributed via DVDs to the project partners for further investigation.

THE CORPUS

At the time this paper was written we had 78 sessions of about 4.5 min multimodal material recorded.

THE SUBJECTS

35 voluntary, naive subjects have been recorded until now and paid for their participation. The 18 females and 17 males were between the age of 16 and 60. Among them were 24 students (including 2 phd students, 1 pupil) and 11 employed persons (including 1 retired person).

THE TASK (SMARTKOM PUBLIC)

The subjects were carefully instructed how to test a new prototype of a dialogue system. It was suggested that the system could understand spoken language and gestures without giving a detailed explanation either on the functioning or on the gestures the subjects could use. The aim was to encourage the subjects to experiment with their gestures as well as with the system itself.
While working with the system for the first – and in the second session for the second – time they had to imagine being in Heidelberg (a German town). Their job was to solve a task like planning a trip to the cinema / restaurant in the evening, finding a hotel, programming their VCR at home or arranging a sightseeing tour. They also were encouraged to try different things. After the instruction two sessions were recorded (with a short break in between). Afterwards, in a different room, the subjects were questioned if problems had appeared, which aspects of the system they approved of and which they resented, and if the system all in all gave more the impression of being a human or a machine.
For results please refer to [3].

2 DESIGN OF THE WIZARD OF OZ USER INTERFACE

It is known from compatibility experiments [4, 5] that people not only imitate the movements of their dialog partner [6] but also tend to adjust their movements to the location and speed of objects [7].
This fact can be of use for the designing of the Wizard of Oz User Interface.

IDEA

The main purpose of creating the Wizard of Oz User Interface is to simulate a working system that differs

2 http://www.apple.com/quicktime/
3 S. Steininger: Labeling of Gestures in SmartKom – Concept of the Coding System, GW2001
4 Stins,–John–F.; Michaels,–Claire–F.: Strategy differences in oscillatory tracking: Stimulus–hand versus stimulus–manipulandum coupling. Journal–of–Experimental–Psychology:–Human–Perception–and–Performance. 1999 Dec; Vol 25(6): 1793–1812.
5 Stins,–John–F.; Michaels,–Claire–F.: Stimulus–response compatibility is information–action compatibility. Ecological–Psychology.1997; Vol 9(1): 25–45.
6 Bandura: Social learning theory of identificatory processes, in Goslin, Handbook of socialization theory and research. Chicago, 1969
7 Worringham,–Charles–J.; Beringer,–Dennis–B.: Directional stimulus–response compatibility: A test of three alternative principles. Ergonomics.1998 Jun; Vol 41(6): 864–880.

from the usual computer interfaces, especially from internet applications. Moreover the graphical design is mainly based on series of webpages that correspond to the most likely way subjects are expected to solve the problem given to them. Care was taken to give them a different look and feel to usual webpages. All of the design was explicitly different from that of usual webpages (and thus their look and feel). The reason for this was that we wanted the subjects not only to use simple "click" gestures but above all new kinds of gestural interaction. Apart from that the design had to be easy to handle for by the wizards. Together with the demand to present a completely new working system, the graphical user interface had to provide the possibility to affect gestures while the subjects used the Wizard of Oz system.

The goals have been achieved by
1) avoiding visible weblinks
2) creating "unusual" buttons
3) providing elements which are likely to be moved in some way by the subject
4) showing self–moving elements
5) showing pages which require the same sort of affirmation or negation

all leading to deictic or non–deictic gestures of the subjects as you can see in the results given below.

Deictic gestures are here defined as gestures which are actively used by the subjects in order to interact with the system instead of or in combination with speech, e.g. pointing to graphical elements. All these gestures can be learned while interacting with the system.

Non–deictic gestures on the other hand refer to speech underlining gestures or nonintended reactions towards the system, e.g. waving of the hands to say "no".

REALISATION

Showing the demonstration of the User Interface in combination with some characteristic interactions of the subjects it can be well seen how the different gestures could be evoked. Of course, especially deictic gestures can be evoked in different degrees of peculiarity. non–deictic gestures are not that easy to evoke by graphical design. They are more or less side effects.
Although it is not easy to imagine the dynamic User Interface I try to give an idea of how deictic gestures can be evoked by describing the following (static) figures.

DEICTIC GESTURES

To train the infra–red gesture recognizer which the final system will use, we have to provide deictic gestures like
- pointing with one or more fingers
- pointing with the whole hand
- using both hands
- circling objects or regions
- etc.

in the Wizard of Oz recordings. Therefore we have realized the ideas 1) to 4) of the previous section:

Unvisible weblinks:
As you can see in figures 1 and 2 the graphical user interface provides general information about cinemas, restaurants as well as hotels, pca, weather forecast (last three not illustrated here). All display pages are presented without visible links. The wizards operate the system by using the frame on the right which is not visible to the subjects.
The visible user interface is devided into four frames:

SmartKom Assistant	***Bubbles for Orientation***
Help	***Task Information***

The SmartKom Assistant in the left upper corner gives information about the current state of the system: inactive (sleeping), active (eyes open), working (thinking), error (dead). The Help–Frame provides advice for the use of the system in general or in particular situations. The information presented in the Task–Information–Frame appears in form of headlines in the frame above and these again in form of balloons.

As it can be seen in figures 1 and 2 the subject gets information to read which can be illustrated by photos as can be found in movie advertisements, TV guides, touring guides or menu cards.
The subject's task is to choose the preferred information from the ones presented.

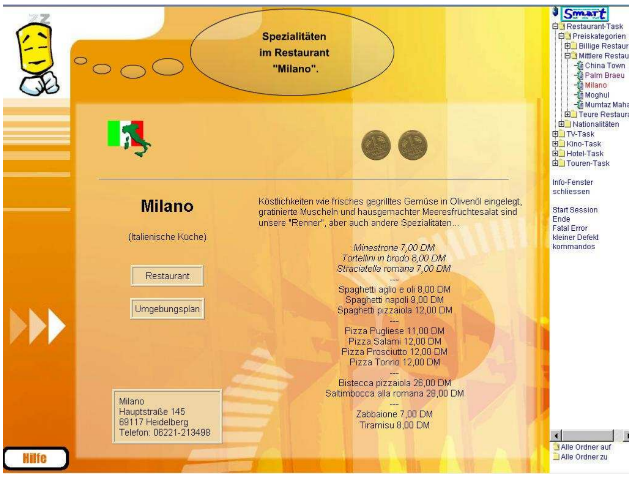


Figure 1: General overview of the movies in Heidelberg

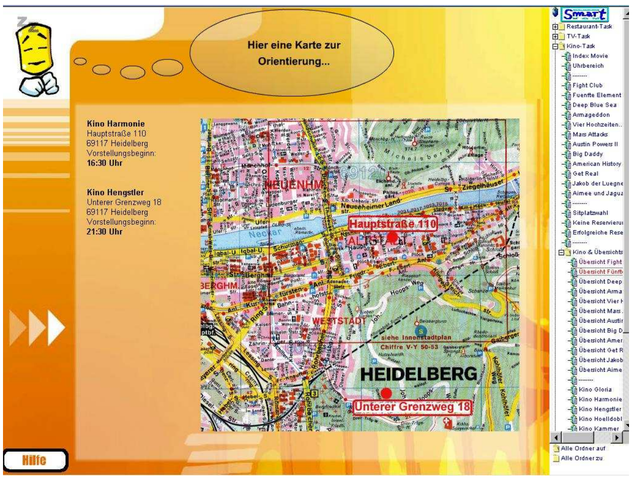
Figure 2: Menu of the restaurant Milano

Without weblinks the subjects have two possibilities to get further information:
- they use speech
- they indicate text or images

Therefore, we created

Unusual buttons:
By "unusual" buttons we understand any form of icons which lead to further information. These can be
- maps
- menu cards
- units of time
- images
- Flash applications

We realized unusual buttons – among other – as can be seen in figures 3 to 5.



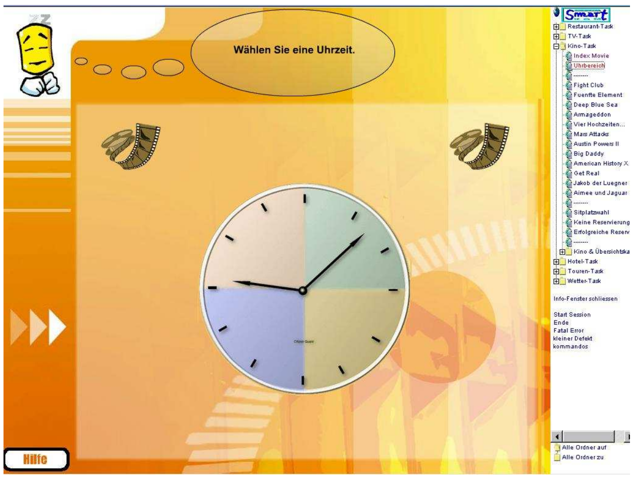


Figure 3: How to get to the cinemas in question

Figure 4: Time scheduler

Maps invite to zoom in or to circle the interesting region (Fig. 3). This was also foreseen in our application as will be shown in the demonstration. To zoom in, people could touch or surround areas. To get more information about the places in question they could either point at the names on the left handside or the corresponding spot on the map.

To get more information about the timetable of movies or tv the subjects had to touch one of the time units marked on

the page.(Fig. 4).
Alternatively of course, they may circle a time slot to indicate the chosen period.

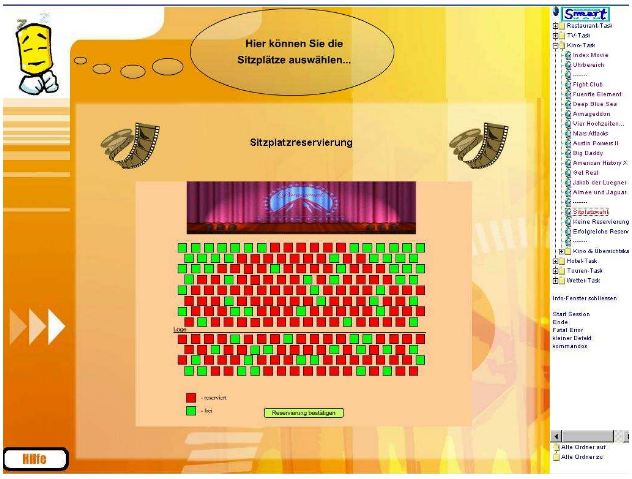


Figure 5: Make a reservation

Figure 6: Choose your preferred actor or programme

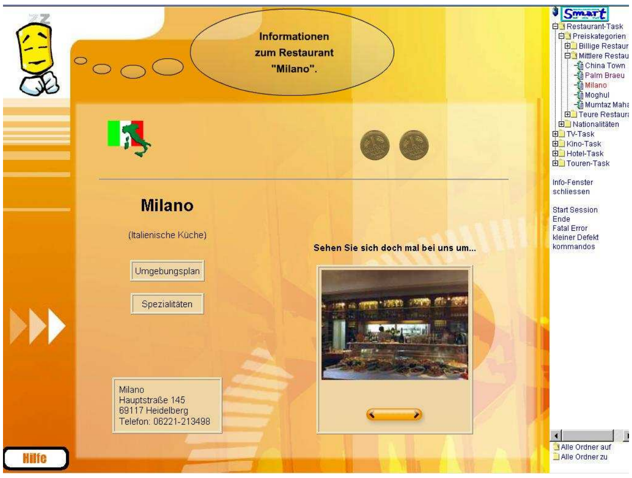
In order to make a reservation for a movie in one of the cinemas provided the subjects had to choose seats in the Flash application in Figure 5. (They had the possibilities to press each place separately or simultaneously or to circle the chosen places).
Pressing one of the actors or programme in Figure 6 leads to more information about the films the chosen actor played in for the week in question or about the contents of the chosen programme including the starting times.

Figure 7: System Front Page

Figure 7 shows a possibility of an "unusual" button we have not foreseen. The triangles on the left encouraged some subjects to press (which of course didn't lead to any success!).
Because of this observation we decided to intersperse graphical elements without any functions to see if the subjects react to them. If so, the relevant elements will be included in future tasks.
Another possibility to evoke deictic gestures is to provide

Movable elements.
We used two different methods
- QuickTime applications to provide the opportunity to look around the place in question (Figure 8)
- Moving objects at the display from one point to another (Figures 9 and 10)

Both methods are intended to evoke "drag (and drop)" gestures.
The QuickTime application in figure 8 is activated if the subject touches the photo on the right. Then the room shown is moved to the left or right depending on where the picture was touched by the subject.
Figure 8: Have a look at the chosen restaurant



If the subject drags and drops one of the transparent balls representing the TV channels into the monitor, this channel is chosen. It is also possible to select more than one ball by circling.

Figure 9: Choose your favourite channel(s)

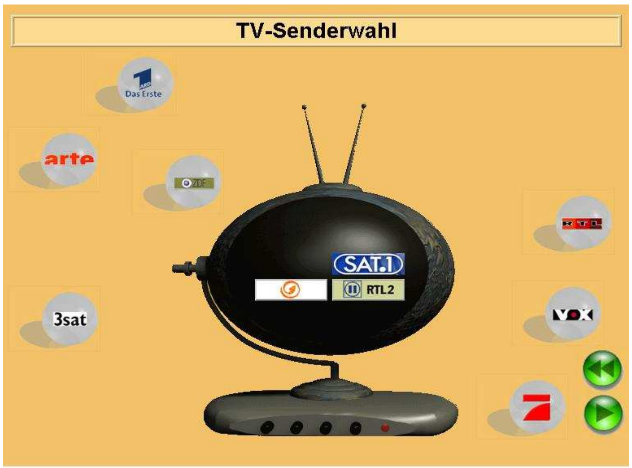


Figure 10: Chosen channels

The last tested method to create deictic gestures is to present

Self–moving elements

By this method the subject is encouraged to "imitate" the movement in general. From experiments in psychology it is known that subjects tend to adjust the direction and speed of their own movements to observed movements. Therefore, we provided moving images, as can be seen in figures 11 to 13. The idea is to imitate the movement in either an deictic or a non–deictic gesture. Figures 11 to 13 give an idea of how the dynamic website looks like in the several stages of the call before it ends in the static version.

Figure 11: Touring guide before the moving element

Figure 12: Touring guide during motion

Figure 13: Touring guide after the moving element

Non–Deictic Gestures

Although our Wizard of Oz user interface is primarily designed to evoke deictic gestures, it is also possible to evoke non–deictic gestures.

Unfortunately, these cannot be planned and are thus a side product of the User Interface **and** the wizards' reactions. Some examples are given in the demonstration video.

3 Results

Frequency of Gestures

We evaluated 52 out of 70 recordings according to the gestures. We found

- 278 deictic gestures in 52 dialogues
- 123 non–deictic gestures in 52 dialogues

by counting the gestures manually and analysed in what situations they occured.

The **deictic** gestures were evoked by:

- avoiding visible weblinks in 113 cases
- creating "unusual" buttons in 115 cases (143 cases including the background triangles and the help buttons)
- providing elements which can be moved in some way by the subject was not used at the time of the evaluation although the application was offered
- showing self–moving elements is included in the recording task at the moment.

In 18 dialogues no unusual buttons were pressed.

The **non–deictic** gestures were mostly a reaction to the wizards' output as assumed.

Evaluation of the Subjects' Intention towards Gesticulating with a Machine

Finally, while evaluating the interviews, it has to be mentioned that most of the subjects tend to interact with the system via speech. For them it seems to be unusual to gesture without a human dialog partner and not necessary if the system understands them via voice input. Without a clear demonstration of the recognizable gestures during the instruction though, some of the subjects found it difficult to use gestures. Finally, some wished to have a clear emphasis on the elements which can be activated by gestures, what of course contradicts our intention to find out what gestures are produced spontaneously and in which context.

4 Conclusion and Further Work

The Wizard of Oz User Interface is still under construction for further tasks.

We have already shown some examples of what kind of gestures showed up in such a situation and we are currently searching for new ways of evoking gestures by the design.

The more recordings we have the easier we probably will find new objects for the display to animate the subjects to produce deictic gestures.
We plan to examine this question systematically in order to improve and enlarge our knowledge of how to evoke different gestures by the design of the graphical user interface.
Momentaneously, we search for alternative graphical design apart from "unusual" buttons and animated websites as there are interactive task changes, faked telephone connections etc.
These are included in the new tasks (e.g. PCA, webbrowser) we start the data collection for soon.
We hope to be able to present the results in the near future.

5 References

1. http://smartkom.dfki.de/index.html
2. http://www.apple.com/quicktime/
3. S. Steininger: Labeling of Gestures in SmartKom – Concept of the Coding System, GW2001
4. Stins,–John–F.; Michaels,–Claire–F.: Strategy differences in oscillatory tracking: Stimulus–hand versus stimulus–manipulandum coupling. Journal–of–Experimental–Psychology:–Human–Perception–and–Performance. 1999 Dec; Vol 25(6): 1793–1812.
5. Stins,–John–F.; Michaels,–Claire–F.: Stimulus–response compatibility is information–action compatibility. Ecological–Psychology.1997; Vol 9(1): 25–45.
6. Bandura: Social learning theory of identificatory processes, in Goslin, Handbook of socialization theory and research. Chicago, 1969
7. Worringham,–Charles–J.; Beringer,–Dennis–B.: Directional stimulus–response compatibility: A test of three alternative principles. Ergonomics.1998 Jun; Vol 41(6): 864–880.