

Gestures for an interactive animated agent

Magnus Nordstrand

Centre for Speech Technology, Department of Speech, Music and Hearing, KTH
magnusn@speech.kth.se

Abstract

This paper presents the architecture and implementation for an easily extendable gesture library to be used by a dialogue system. Main focus has been on creating and implementing gestures to signal turn taking etc. in order to facilitate for users to interact with a dialogue system. The idea was to make possible for the dialogue system to randomly select among a given set of gestures, depending what dialogue state it is in, and in that way make it more natural and non-repetative.

1 Introduction

1.1 The importance of gestures

When people communicate they often use both verbal and visual signals to in order to supply information. Normal speech communication is thus often multimodal. The visual modality can qualify the auditory information providing segmental cues on place of articulation, prosodic information concerning prominence and phrasing and extralinguistic information such as signals for turn-taking, emotions and attitude. It is evident that the use of such information in a dialogue system would be of great help in the communication between the user and the agent. The use of turn taking in such a system could prevent the user from talking when the agent is preparing to make a response and thereby make the dialogue more fluent. At CTT there is earlier work like Waxholm (Carlson & Granström 1996), Olga (Beskow, Elenius & McGlashan 1997) and August (Gustafson, Lindberg & Lundeberg 1999) as well as ongoing work like AdApt (Gustafson, Bell, et al. 2000) that has implemented facial gestures as a part of a multimodal dialogue system. The construction of a dialogue system that can handle fragmented user utterances, such as AdApt, would need a method to show the user when it gets a closing utterance (more input is unlikely to come) or a non-closing utterance (more input is likely to come) from the user. The use of an animated agent to perform gestures to signal that is therefore of great help.

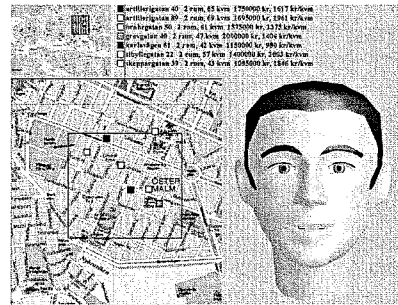


Figure 1. The AdApt user interface with the agent Urban.

2 Method

2.1 Dialogue states and gestures

The speakers in a dialogue, in this case the user and the agent, can be said to be in different dialogue states during the discourse. When looking at AdApt we can distinguish a number of states that the agent can be in. We decided to implement the following states in AdApt:

- *Idle* – The agent is waiting for the user to begin the interaction.
- *Busy* – When the agent gets a closing utterance from the user and prepares to make an appropriate answer.
- *Continued attentive* – This state is divided into three subgroups, *positive*, *neutral* and *negative* but they are still separate states when handled by the agent. The *negative* state is used when the agent gets a non-closing utterance from the user to signal that he hasn't understood, and wants more information to be able to carry out the request. The agent uses the gestures in the *positive* state when its answer to the user is what the user expected.
- *Attentive* – Is to show the user that the agent is listening and ready for input.

Each state has been divided into different transition states:

- *enter gestures* – These gestures are performed when the agent enters any given state.
- *sustain gestures* – These gestures are carried out between the enter and exit gestures. It contain gestures to make the agent look more natural such as blinking etc.
- *exit gestures* – The exit gesture corresponds to the enter gesture and is performed when the agent leaves a state to enter a new one.

The agent will always be in a given state and doesn't leave it until he is told to enter a new state. The agent randomly selects one of the gestures in the state it is about to enter as an attempt to make the agent seem more human. The probability for a gesture to be picked can be controlled by increasing the probability value connected to that specific gesture. Each state can consist of an infinitely number of gestures and the library can be extended with as many states you need.

The gestures used by the parametrized 3D-model (Beskow 1995) are created using the graphical interface in Wavesurfer by Sjölander and Beskow (2000). It gives you the option to change different facial parameters such as: brow raise, brow frown, smile, horizontal and vertical head and eye movements, eye closure and head tilt. The gesture file to be read by the agent contains only the parameter values that change during the gesture. Making the gestures look natural and relevant has mainly been done by introspection and trial and error. So far the gesture library contains about 40 gestures.

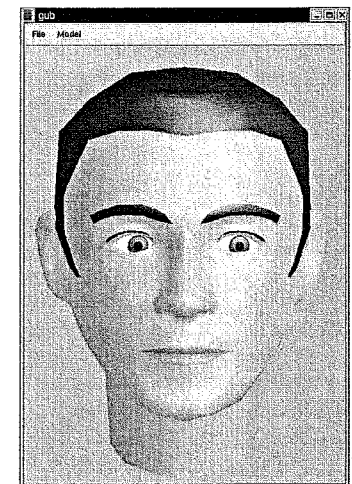


Figure 2. The agent performing a busy gesture.

2.2 State and gesture handling with XML

The dialogue system uses XML to encode states and gestures. Since we are using empty elements the agent will always be in a given state and doesn't leave it until it is about to enter a new one. This is an example of how an utterance can look like when it has been equipped with XML-tags:

```
<state name="contatt_negative">Jag förstår inte
```

The agent will perform a *continued attentive negative* gesture while it is saying *Jag förstår inte* ('I don't understand'). Exactly which one of the gestures in that state it will execute depends on the probability for a gesture to occur.

3 Results

3.1 Improvement or confusion

No user test has been performed as the implementation of gestures in the AdApt system isn't finished at this point. It is clear though that the timing of gestures is difficult, since the agent can execute any of the gestures in a state when making his utterance and the gestures are of different length and appearance. The general impression though is that even if a gesture does not completely correspond to an utterance, it is often an improvement to having no gestures at all.

4 Discussion

The work on creating a gesture library for an interactive animated agent is still in progress. More states and gestures will be made to extend the gesture library.

Acknowledgments

The work is done within the Centre for Speech Technology (CTT) at KTH. I would like to thank Jonas Beskow who came up with the idea and made most of the necessary programming. I also thank Jens Edlund for his help in implementing the agent in AdApt.

References

- Gustafson, Lindberg & Lundeberg. 1999. 'The August spoken dialogue system'. In *Proceedings of EUROSPEECH'99*, Budapest, Hungary, September 1999.
- Beskow, Elenius & McGlashan. 1997. 'OLGA - A dialogue system with an animated talking agent'. In *Proceedings of EUROSPEECH'97*, Rhodes, Greece, September 1997.
- Carlson R. & Granström B. 1996. 'The Waxholm spoken dialogue system'. In Palková Z., ed. *Phonetica Praensia IX. Charisteria viro doctissimo Premysl Janota oblata. Acta Universitatis Carolinae Philologica 1*, 1996. Prague: Charles University; 39-52.
- Gustafson, J., Bell, L., Beskow, J., Boye, J., Carlson, R., Edlund, J., Granström, B., House, D. and Wirén, M. 2000. 'AdApt - a multimodal conversational dialogue system in an apartment domain'. In *Proceedings of ICSLP*.