Avatar Gestures

Thomas Rieger
Technische Universität Darmstadt,
Interactive Graphics Systems Group, 3D Graphics Computing
Fraunhoferstr 5, 64283 Darmstadt, Germany
Email: rieger@gris.informatik.tu-darmstadt.de
Internet: http://www.gris.informatik.tu-darmstadt.de/~rieger

ABSTRACT

This paper describes the concept and control of a 3d Avatar system with mimic and gestures as a conversational user interface. The Avatar system including gestures and mimic is based on morphing techniques. It allows to generate sign language and mouth motion at lip reading quality in real time. The concept of a Speech Act is introduced and a table is defined to classify conversation fragments. Furthermore, two rule-based systems have been implemented to control the Avatar's behavior and gestures. The system is shown to be useful in several application areas.

Keywords

Sign language, Lip Reading, Gesture, Morphing, Conversational Interfaces

1. INTRODUCTION

In Computer Graphics and in many other research areas 3d Avatars are utilized as conversational user interfaces with numerous applications. Some common examples are virtual hosts of the news [ABHR01] or assistants in software bundles [BrRi01][RiBF01]. In these cases Avatars typically show only a limited number of gestures, which have just a few meanings. So far, only motion capturing systems are able to show complex gestures or sign language in real time [SignLang]. In this work, Avatar mimic is described, and a novel set of rules and classifications is set to define Avatar gestures. Basic conversation scenarios and relative mimic are classified; based on this more complex situations are defined and the concept of Speech Acts is introduced in order to describe action/reaction situations. This leads to a better Human-System communication as described in [CSPC00].

For our system, we use of morphing (on the base of [AlBM00]) in order to provide high quality features such as gestures, sign language and lip motion; moreover every Avatar movement is calculated and rendered in real time in the system.

Journal of WSCG, Vol.11, No.1., ISSN 1213-6972
WSCG'2003, February 3-7, 2003, Plzen, Czech Republic.

2. MOTIVATION

The main target of this work is to increase the acceptance of Avatars in the Human-Computer-Communication. An important aspect is that Avatars could be very helpful as assistants for handicapped persons thanks to the features of lip reading and sign language

3. OVERVIEW

Man-Machine- Communication

In order to get the best result in Human-Computer-Communication (with an Avatar), it is necessary to give the system Meta information about the conversation and the conversation context according to the respective situation. This Meta information can be collected in different ways. One possibility is the dialog history. A conventional conversation begins with the greeting and ends with the leave taking. Another possibility is to get additional information from a GUI or an initialization. However, in a real conversation it is necessary to understand and to interpret the situation. A Dialog Manager handles this part in the EMBASSI project (Multimodal Assistance for Infotainment and Service Infrastructures) [EMBASSI].

Architecture

Our Avatar system is based on a system that has been published earlier [ABHR01].

The Avatar system is partitioned in several subsystems [ABHR01]. In this case the *Behavior Manager* and the *Gesture Manager* are of particular importance. Both of them use a complex rule-based system, which are designed to be easy to extend.

Through the Dialog Manger [see Figure 1] the Avatar System gets different information:

- **Speech Acts**
  [See chapter SPEECH ACTS]
  Not always available.
- **Text**
  Text to speak for the Avatar
  Not always available.
- **Basic Behavior**
  For example: *be happy*
  Always available.
- **Gesture Meta Data**
  For example: Avatar-Position on Display
  Always available.

The creation of behavior and gestures is based on this information. The *Behavior Manager* is responsible for the avatar mimic, speech, and head-movements. In addition, the *Behavior Manager* supports the *Gesture Manager* with information about the current activities. That is needed for a clocked and synchronized control of gesture and speech. The *Gesture Manager* calculates the hand-movements and hand morphing. The *Avatar Controller* executes and renders the calculated behavior of the *Behavior Manager* and the gesture of the *Gesture Manager*. The *Avatar Controller* also controls the *Text to Speech* device.

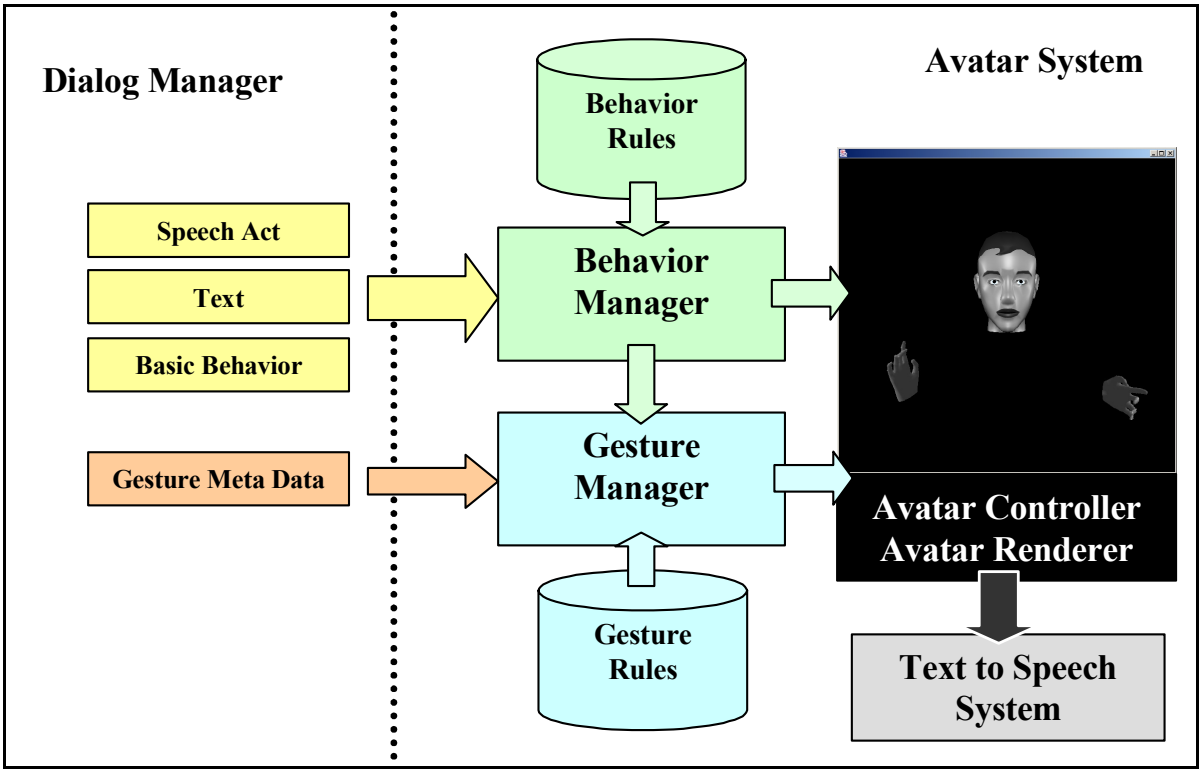

Figure 1: Abstract architecture illustration of the behavior manager and gesture manager

Weighting

Weighting is completely handled by the Behavior Manager. The *Speech Act* [see SPEECH ACTS] has the most important quota in the behavior creation. The *Speech Act* also partially overrules the *Basic Behavior* if there is a discrepancy between them. However, after the execution of the *Speech Act*, the Avatar falls back in the *Basic Behavior*. In the case of a neutral disposition of the *Speech Act*, the basic behavior is decisive. The *Text* has no influence on the Avatar behavior.

Lip sync

To achieve a realistic appearance, it is an important requirement for the Avatar to produce speech output with lip-sync. Since the dialogue with the user is not known in advance, prerecorded animation sequences with lip animation cannot serve as a solution. Instead, real-time lip sync with audio output generated by a *Text to Speech System* is implemented [ABHR01].

TECHNIQUES

Every motion in the Avatar System is calculated and rendered in real time. The animations are not prefabricated but based on complex rule systems for the behavior and the gesture [see Figure 1]. In contrast to a motion capturing system, it is possible to blend different animations. A set of rules defines what combinations of animations are allowed to be blended. To accomplish reasonable rules, several research projects including evaluations and tests are undertaken at the University of Cologne and at the Humbold University of Berlin in the context of the EMBASSI project [EMBASSI].

Morphing

The used morphing technique is based on the work of Alexa et al. [AlBM00]. The Avatar consists of several morph targets for the head and for the hands. Thereby, the head, the left hand and the right hand are independent from each other, and can be controlled discretely.

A Morph Target displays a state of an expression [AlBM00]. Every morph target is weighted between '0.0d' and '1.0d'. Higher values result in more influence of a specific morph target in the morph result.

In order to morph over several targets (see, e.g., Figure 2) we use linear interpolation. For the example discretized into five steps we, thus, obtain the following values:

1. Morph Target	1.0	0.5	0.0	0.0	0.0
2. Morph Target	0.0	0.5	1.0	0.5	0.0
3. Morph Target	0.0	0.0	0.0	0.5	1.0

The number of steps that the system calculates is adjustable and depends on the hardware. In general, the quality rises with the number of steps, but also the system requirement is increasing.

More morph targets typically result in higher quality and abilities of the Avatar [CBBC99]. To be able to show gesture in sign-language quality it is necessary to design one or more morph targets per sign. However, already a small count of morph targets delivers good results.

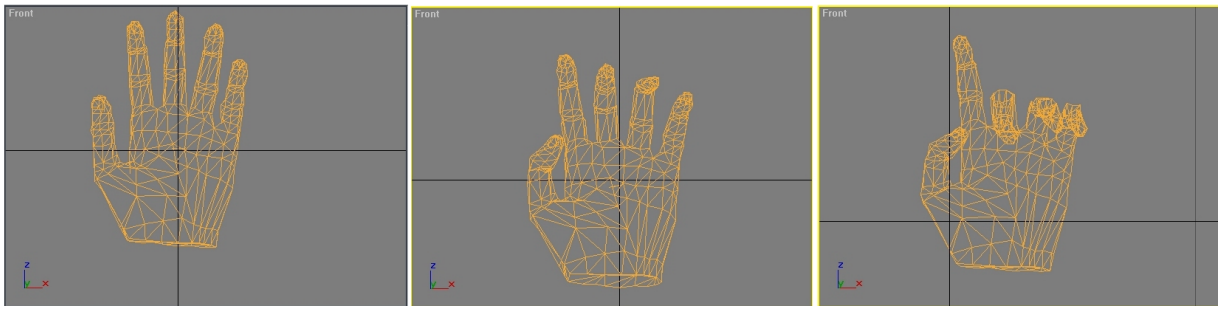

Figure 2: Some Morph Targets of the hand models

SPEECH ACTS

Discourse

To get a highly realistic appearance in Human-Computer-Communication it is necessary to analyze human conversations. A human unknowingly expects some actions and reactions from the conversational partner in a discourse. These actions and reactions can be verbal and/or nonverbal. The nonverbal actions and reactions will be expressed with the mimic and the gesture of the Avatar. One problem of the use of a nonverbal activity is to find the right point of time in a discourse to show a special gesture. A discourse can be subdivided in fragments. Thereby we have to distinguish between monologues and dialogues. In order to create a general definition of patterns of behavior for an Avatar system, we need to look for

- Extensibility
- Simplicity
- Parameterizing

Monologue

The handling of monologues is very simple, because the evolution of the discourse is known from the beginning. A simple example of fragments of monologues:

GREETING +
INFORMATION +
LEAVE TAKING

Scenario: Weather Report.

Every fragment of Monologue is mapped on exactly one Speech Act [see Table 1].

FRAGMENT 1:

GREETING = message_greeting

FRAGMENT 2:

INFORMATION = message_inform

FRAGMENT 3:

LEAVE TAKING= message_closing

Dialogue

In a dialogue the evolution of the discourse depends on all the conversational partners. Here, a good timing for a natural appearance of the Avatar is very important. An example of fragments of dialogues:

GREETING +
QUESTION / ANSWER +
LEAVE TAKING

Scenario: Directory Assistance.

Every fragment of dialogue is mapped on not less than one Speech Act [see Table 1].

FRAGMENT 1:

GREETING = message_greeting

FRAGMENT 2:

QUESTION = query_input

ANSWER = message_inform

FRAGMENT 3:

LEAVE TAKING= message_closing

Table of Speech Acts

In the context of the EMBASSI Project [EMBASSI] in cooperation with the University of Cologne [Kraemer] and the ZGDV Darmstadt [ZGDV], a list of Speech Acts has been compiled. In these Speech Acts, definitions are set to manage the Avatar behavior. The Avatar Behavior is expressed with the Avatar mimic and the Avatar gestures. Both of them, gestures and mimic, have to be considered for the definition of Speech Acts. The base of the definition is the observation of patterns of Human behavior.

message_greeting	friendly greeting
message_closing	friendly closing
message_inform	
status=warning	show importance
status=busy	show regret
status=error	how sorry
status=ok	positive neutral
status=failed	show sorry
status=offer	friendly offer
message_accept	show agreement
message_reject	show rejection
message_command	show importance
message_cancel	show attention
message_acknowledge	show happiness
message_correct	neutral
query_input	show interest
query_selection	show alternatives
query_request_knowledge	show interest
query_request_repair	show interest

Table 1: Speech Act Overview

Mapping

Almost every Speech Act can be accompanied by a gesture. The gesture is often helpful beside the mimic to amplify the mood and to aid or to intensify information. To correctly employ the gesture, the Speech Acts have to be defined as accurately as possible. However, it is mostly sufficient to describe gestures in an abstract way. In the implementation, the Gesture Manager decides which type of signs, of the described category, will be shown.

Speech Act	message_greeting
Global	friendly greeting
Concrete	
Friendly voice	Begin until end
Eyebrows up	On begin
Eyes wide open	On begin 1 sec.
Smile	Before begin speaking
Beck	Before begin speaking
Smile	After speaking
Head angular 10 degree	After speaking 4 sec.

Table 2: Example of Speech Act definition

Incidentally it is not needed to have a text for every Speech Act. Many Speech Acts can be executed without the Avatar speaking. The decision about the type of the rendering, for example: with or without Speech or Gesture, is made by the *Behavior Manger* (See Figure 1). The actions shown (see Table 2), act as orientation for the *Behavior Manager*. Like the *Gesture Manager* the *Behavior Manager* has a choice of different expressions.

GESTURE

Each gesture is a combination of several transformations and morphs. Translations and rotations of the hands follow calculated timestamps. Timestamps of the morphs are also calculated in real time. The timestamps for rigid body motion and morphing are interdependent.

Rule system

Many gestures can be blended and/or combined with others. In this way, a smooth transition from one gesture to another gesture is possible. The transition depends on the rules and on the speed of the animations. Again, the speed of the animations depends on the mood settings (Global Behavior) and the hardware.

The rules are implemented in LISP.

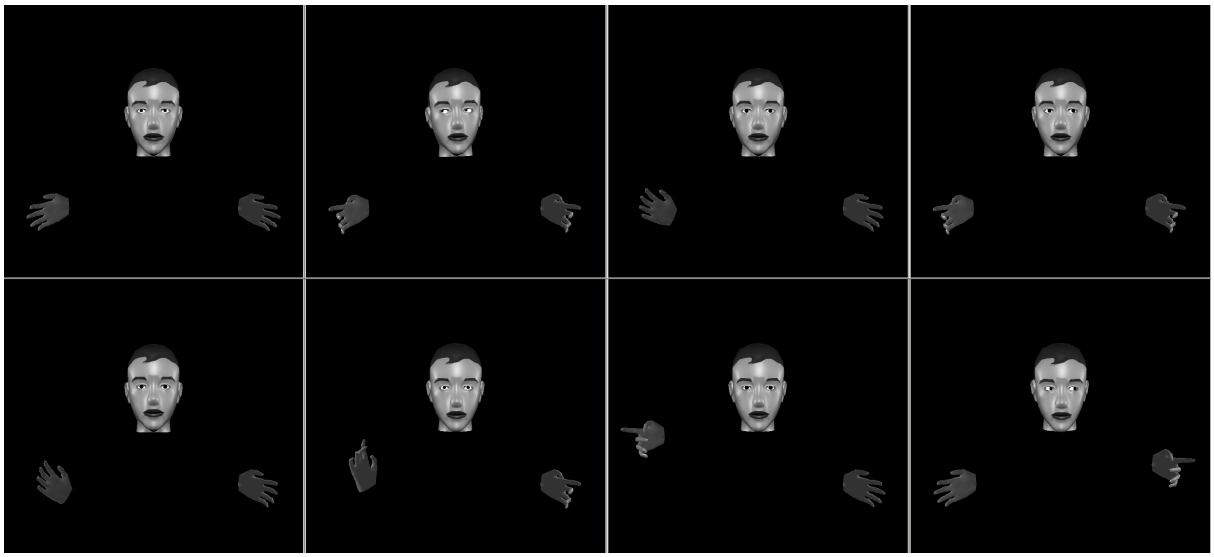

Figure 3: Simple Avatar gesture

Gesture Manager

The *Gesture Manager* handles the complete control of the gesture. The count of available gestures depends on the number of available morph targets of the hands and on the number of rules defined within the *Gesture Manager*. The rules define what gesture will be executed and combined with the next gesture. That depends on the *Gesture Meta Data* (See Figure 1) and on the information of the *Behavior Manager*.

Gesture Meta Data:

- *Avatar Position*
- *Avatar Size*
- *Position of other Windows and Objects on the Screen*

Behavior Manager Information:

- *Speech Act*
- *Mood*
- *Time Stamps*

If, for example, the avatar is in a small window outside the center of the screen, then the gestures become more distinct expression.

SIGN LANGUAGE

ASL, the American Sign Language [ASLBr](See Figure 5), is the most commonly used sign language; nevertheless, many international sign languages are available [WorldSig]. For example the German Sign Language [See Figure 4] is very interesting for use with the morphing technique, as the alphabet is a "finger alphabet" and several target signs are simple to model and morph.

Numerous online descriptions and links are available in the Internet [ASLBr]. The basic alphabet is shown in Figure 5, it is further necessary to consider that numbers, words, and in particular sentences have to be taken in analysis.

For instance, [ASLBr] shows an online dictionary, and it is clear how complex and long a complete sign language description can be. This results in a difficult work in defining a set of sign targets, which need to be modeled and which have to be morphed as a sequence of steps in the language description.

It is clear how a high quality avatar gesture could be used in various applications. The morphing technique produces good results: [ABHR01][BrRi01][RiBF01] are just some examples, but the field of applications is promising and future directions are possible. For instance Vanderheiden [Vanheid] proposes the introduction of sign language in cellular technologies in order to extend the accessibility of Telecommunications.

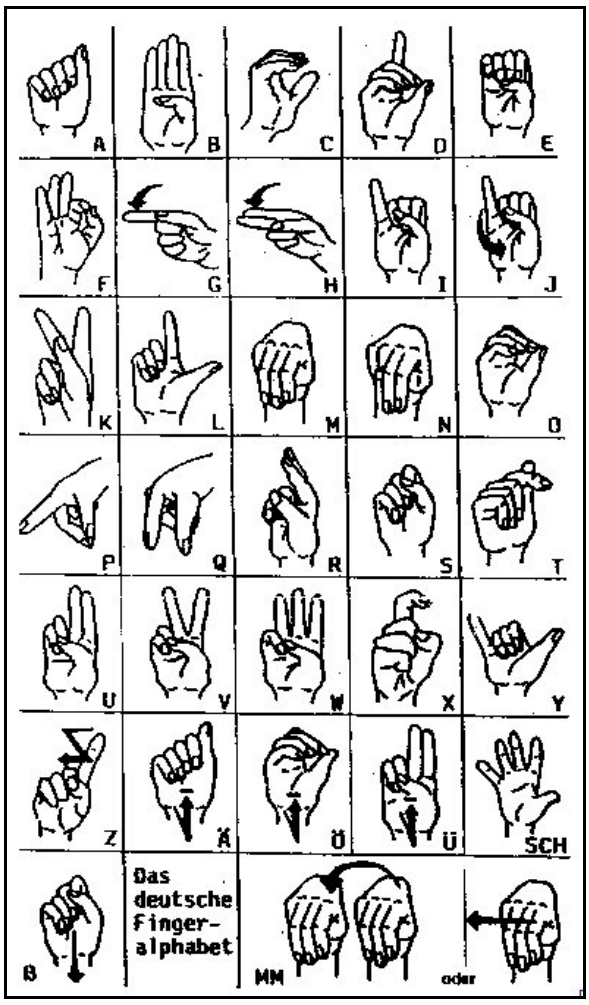

Figure 4: German Sign Language alphabet.

Lip reading

Besides gestures, the quality of an Avatar may also offer the possibility to interpret the facial mimic, providing a Lip Reading feature. This could be of particular importance in applications where fewer place is available for the avatar displacement, so that just the head and the mouth movements are visible instead of the whole body.

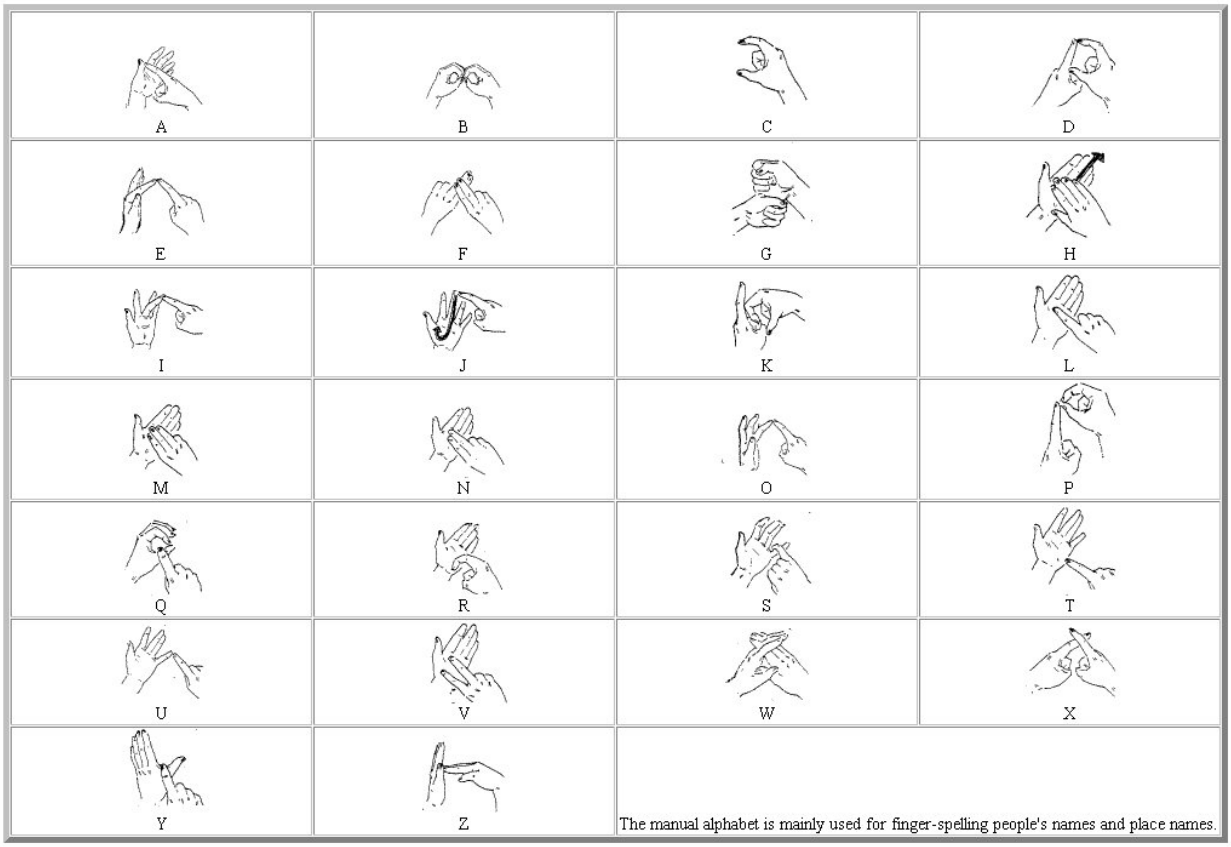

Figure 5: ASL (American Sign Language).

4. CONCLUSION

This work describes an Avatar System, which allows real time gesture in sign language quality.

The system is based on morphing technology and it is controlled by different rules for Avatar gestures and behavior. This Human-Computer-Communication system achieves high quality and provides a good solution for sign languages' new applications.

An important contribution of this work is the definition of a set of rules for Avatar mimic and gestures, and their classification. The idea of Speech Acts is described and a common conversation is analyzed: Discourse and fragments: Monologue and Dialogue are defined; actions and reactions, verbal and non-verbal mimic are illustrated. In the Context of the EMBASSI project [EMBASSI] a table of Speech Acts has been defined.

In addition, new perspectives are possible with the use of this system, i.e. future applications for sign language and lip reading.

5. FUTURE WORK

The most important future work is to extend the database of the morph targets to advance the gestures and its expressions. It is also necessary to extend the system of rules of the *Behavior Manager* and the *Gesture Manager*.

Furthermore, the Avatar System is still under evaluation at the University of Cologne and at the Humbold-University of Berlin in the EMBASSI context.

6. ACKNOWLEDGEMENTS

This work was partially funded by the German "Bundesministerium für Bildung und Forschung, BMB+F through the Focus Project EMBASSI (BMB+F-No. FKZ 01 IL 904 U8) [EMBASSI]. I would like to thank all EMBASSI project partners for their helpful ideas and support.

7. REFERENCES

Publications

[ABHR01] M. Alexa, U. Berner, M. Hellenschmidt, T. Rieger An Animation System for User Interface Agents Proc. Of WSCG 2001, Pilzen, Czech. Republic

[AlBM00] Alexa, M., Behr, J., Müller, W. (2000). The Morph Node. Proc. Web3d/VRML 2000, Monterey, CA., pp. 29-34

[BuKL02] Kerstin Bücher, Michael Knorr, Bernd Ludwig, Anything to Clarify? Report your Parsing Ambiguities?, Proceedings of the 15th European Conference on Artifical Intelligence, Lyon,2002, pp. 465--469

[BrRi01] N. Braun, T. Rieger, Group Conversation within a Internet TV Community through a Lifelike Avatar Proc. of Media Futures 2001, Florence, Italy

[CBBC99] J. Cassel, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, H. Yan. Embodiment in conversational interfaces: rea. Conference Proceedings of ACM CHI `99, 1999. p. 520-7.

[CSPC00] J. Cassel, J. Sulivan, S. Prevost, E. Churchill, editors. Embodied conversational Agents. Cambridge, MA: MIT Press, 2000

[Ludw01] Bernd Ludwig, Dialogue Understanding in Dynamic Domains, in: Proceedings of the 5th Workshop on Formal Semantics and Pragmatics of Dialogue (BI-DIALOG 2001), Bielefeld, Germany 2001

[MSAR01] W. Müller, U. Spierling, M. Alexa, Th. Rieger Face-To-Face With Your Assistant Computers & Graphics 25 (2001) 593-600

[RiBF01] T. Rieger, N. Braun, M. Finke, Community TV: An Approach to Interaction for Groups and Single Users on Internet Video Proc. of WSES (MIV 2001), Qawra, Malta

[Vanheid] Vanderheiden Gregg, Telecommunications - Accessibility and Future Directions, Proc of Technology And Persons With Disabilities Conference 2001, Northridg USA

Internet Links

[ASLBr] American Sign Language Browser http://commtechlab.msu.edu/sites/aslweb/

[EMBASSI] EMBASSI BMB+F-Project http://www.embassi.de

[Kraemer] Nicole Krämer, University of Cologne http://www.uni-koeln.de/phil-fak/psych/diff/darkraem.htm

[SignLang] Sign Language Animations http://www.signingbooks.org/animations/sign_language_animations.htm

[WorldSig] World Sign Language Library http://www.deafblind.com/worldsig.html).

[ZGDV] Zentrum der Grafischen Datenverarbeitung e. V. Darmstadt. http://www.zgdv.de/