AUTOMATED CULTURE TRAINING SYSTEM

Lorenzo Scagnetti
Master of Science
Computer Science
January 2012
School of Computer Science
Reykjavík University

M.Sc. PROJECT REPORT

ISSN 1670-8539

Automated Culture Training System

by

Lorenzo Scagnetti

Project report submitted to the School of Computer Science
at Reykjavík University in partial fulfillment of
the requirements for the degree of
Master of Science in **Computer Science**

January 2012

Project Report Committee:

Hannes Högni Vilhjálmsson, Supervisor
Associate Professor, Reykjavik University

Kristinn R. Thórisson
Associate Professor, Reykjavik University

Emanuela Merelli
Associate Professor, Camerino University

Luca Tesei
Assistant Professor, Camerino University

Automated Culture Training System

Lorenzo Scagnetti

January 2012

Abstract

It is increasingly common that we engage in face-to-face conversations with people of different cultures as part of our daily lives. The results of such cross-cultural communication are sometimes affected by misunderstandings that arise from culturally different interpretations of the same message. This thesis focuses on the different interpretations of non-verbal behaviors, gestures in particular. The thesis proposes the Automated Culture Training (ACT) system that addresses the problem of misunderstanding in cross-cultural communication by providing an interactive 3D training environment, where people can quickly pick up cultural knowledge and apply it in a series of simulated social settings. Each lesson involves an interaction between the learner and an automatic character of a given culture. During the interaction, the learner chooses what gesture to use at any given moment, and the character gives immediate positive or negative feedback. The contribution of this thesis is a modular technical framework for the ACT system based on a clear abstraction between communicative behavior (the visible action) and the communicative function (the interpretation). Furthermore, the framework keeps a clear separation between data and its processing, for example by treating the cultural description and the description of each exercise purely as input data. The automatic character incorporates a complete perception-action loop, which allows it to dynamically react to learner input. The result is a fully functional prototype that demonstrates best-practice engineering principles and is ready for further development of content and testing with users.

Automated Culture Training System

Lorenzo Scagnetti

Janúar 2012

Útdráttur

It is increasingly common that we engage in face-to-face conversations with people of different cultures as part of our daily lives. The results of such cross-cultural communication are sometimes affected by misunderstandings that arise from culturally different interpretations of the same message. This thesis focuses on the different interpretations of non-verbal behaviors, gestures in particular. The thesis proposes the Automated Culture Training (ACT) system that addresses the problem of misunderstanding in cross-cultural communication by providing an interactive 3D training environment, where people can quickly pick up cultural knowledge and apply it in a series of simulated social settings. Each lesson involves an interaction between the learner and an automatic character of a given culture. During the interaction, the learner chooses what gesture to use at any given moment, and the character gives immediate positive or negative feedback. The contribution of this thesis is a modular technical framework for the ACT system based on a clear abstraction between communicative behavior (the visible action) and the communicative function (the interpretation). Furthermore, the framework keeps a clear separation between data and its processing, for example by treating the cultural description and the description of each exercise purely as input data. The automatic character incorporates a complete perception-action loop, which allows it to dynamically react to learner input. The result is a fully functional prototype that demonstrates best-practice engineering principles and is ready for further development of content and testing with users.

Dedicated to my Family and Friends.

Acknowledgements

First of all I would like to thank my parents for the support and love they give me.

Thanks also to my sister Laura and my little brother Leonardo and all my family for the precious suggestions they gave me all along.

Special thanks to Hannes, my supervisor, for the suggestion and help he gave me during this project but also during the entire academic year spent in Iceland.

Thanks to Ari Shapiro, he gave me a great help with my project.

Thanks to the Reykjavik University and the CADIA staff for their professionalism and great company during my work days there, in particular to Angelo Cafaro we spent a lot of time together talking about first impressions and different meaning of gestures across culture.

Thanks to Camerino University for giving me the possibility to participate to the Double Degree Program.

Thanks to all the Erasmus people I really enjoy the time spent with you guys, a special thanks to the people that lived with me in my guest house, this year spent with you was awesome.

Last but not least, I would like to say thank you to the Italians that study and works in the Reykjavik University.

Contents

List of Figures

Chapter 1

Introduction

In everyday life people are involved in a wide range of social situations, one of the most frequent is the communication between people, that allow them to exchange information among each others. There are several kinds of communications, in this thesis we want to focus our attention in Face to Face communication with the particular context of cross cultural communication. We have a Face to Face communication only if there is a possibility of direct eye contact between the speaker and the listener and we have a cross cultural communication when the participants belong to different cultures. Nowadays it is very common to participate in a cross cultural communications for example at university or at work. When people are engaged in cross cultural communication, they have to face the misunderstanding problem between people. A misunderstanding happens when the listener understand different message respect the original one of the speaker, this fact can influence the result of the conversation, for example someone can get upset and leave the conversation for a misunderstanding. So it is very important to be culturally aware to reduce or avoid misunderstandings. In this thesis we focus our attention on the misunderstandings generated by Non-Verbal Bheaviors with particular attention on gestures.

Computer Science and in particular the Artificial Intelligence (AI) field actually is able to reproduce the natural Face to Face communication by using Virtual Agents, they reach an impressive level of believability and in the last years some researchers start to study how to give to them a culture able to influence their behaviors. Collecting all this information we came up with our idea, we want to create a 3D application able to teach how to behave when you are in a conversation with people of different culture; the name of the system is Automated Culture Training(ACT) System. The main goal of the ACT System is to teach new cultural skill to the user. We want to reach several goals in its development: we want an automated system able to generate all the exercises by reading some input files, a cultural independent system able to work with any culture, a system with an architecture with a good level of modularity.

1.1 Thesis Overview

The thesis is divided in five more chapters:

- **Chapter 2:** we explain the theory behind our system and the motivations that support us in his creation.
- **Chapter 3:** we talk about the relevant works in the same application domain related to our work.
- **Chapter 4:** we introduce our approach of the system, our goals and we show the technologies used.
- **Chapter 5:** we describe how the ACT System works and is features.
- **Chapter 6:** we talk about the result achieved with this project and the future steps.

Chapter 2

Background

2.1 Verbal and Non-Verbal Communication

Communication takes place when two or more people exchange meaningful information between them. The communication process is composed by a sender, a message and a receiver, the process may end when the receiver understands the complete message from the sender. We can have different classifications of communications on the bases of several parameters, for our point of view the two most important attributes are: amount of time for message delivery and spatial distance between the sender and receiver. Considering this two parameters we can have the following classification:

- **Face to Face Communication:** the sender and the receiver are in the same place and the message is received in real-time.
- **Distance Communication:** the sender and the receiver are in different places, the message can be received in real-time or not. For example the telephone, sms or letter.

When we interact with someone else, we exchange information with the receiver by use two kind of communication devices:

- **Verbal Communication:** uses words, sentences and language to transmit the message.
- **Non-Verbal Communication:** uses body movements to transmit the message.

The Verbal Communication can be used in every kind of communication, instead the Non-Verbal Communication can be used only when there is a direct eye contact between the sender and the receiver. The body language includes:

- Facial Expression;
- Gaze and Pupil dilatation;
- Gesture and other body movements;
- Posture;
- Bodily Contact;
- Spatial Behaviour;
- Appearance;
- Non-Verbal Vocalization;
- Smell.

One of the most interesting area of the body language are the gestures, they are integral part of language as much as words, phrases, and sentences in fact gesture and language are one system [10]. Figure 2.1 shows one kind of gesture: "The Thumbs Up Gesture".

Figure 2.1: The Thumbs Up Gesture.

In this thesis we are going to analyse Non-Verbal Behaviors in Face to Face Communication with particular attention to the gestures performed during the conversation.

2.2 Gestures and Culture

We are mainly focus in gestures performed during the interaction between people. When we communicate, we use an infinite variety of gestures in [10] is presented the following categorization:

- **Iconic Gestures:** are strictly related to concrete semantic content of the speech. Example display an action while you are talking about it.

- **Metaphoric Gestures:** draw an image that represent or stand for some abstract concepts used during the conversation. Example draw a square for ring to constraints.
- **Deictic Gestures:** are pointing movement to persons or objects in the scene related to the context of the speech. Example point to an object when you introduce it in the conversation.
- **Beat Gestures:** defined as movements that haven't any specific meaning they are used for emphasize some concepts during the conversation. Example flick the wrist when you say "gone" as in "but he was gone!".
- **Emblem Gestures:** are fully lexicalized gestures with a conventionalized form-meaning relation, that we can use without speech. Example the Thumbs Up gesture Figure 2.1 means agreement in certain cultures.
- **Ritual Gestures:** gesture used for specific purpose with specific meaning, they can be used without speech. Example greeting gesture are in this category.

From this categorization we can point out that some kind of gestures Iconic, Metaphoric, Deictic and Beat need to be related to speech for have a meaning and at the same time help the delivery of the message, instead the Emblems and Ritual Gestures have a proper meaning even if they are not used whit speech.

For this project we decided to focus only on the Emblem and Ritual Gestures, because the meaning of those is strictly constructed within a Culture.
Culture is the way of life, especially the general norms and beliefs, of a particular country or a group of people at a particular time. Culture has a great influence over the body language used in communication, that goes from meaning to shape of the Non-Verbal Behaviors. Several study have been made about culture's behaviors differences; a briefly overview on the most important studies is in [11]. The studies presented are different comparative researches on gesture and speech between two specific cultures like Spanish and German. This researches are used by the author to point out that gesture and speech vary along the dimensions: gesture-space, amount and degree of differentiation of conventionalized gestures, amount and kind of gestures used in relation to speech-phrases, semantic content and sequential position if referential gestures, in relation to cross-linguistic differences.

Emblem and Ritual gestures are really interesting in cultural context, because the meaning of this type of gestures can change across cultures. An example of an Emblem Gesture is presented in Figure 2.2 "The Hand Purse Gesture".

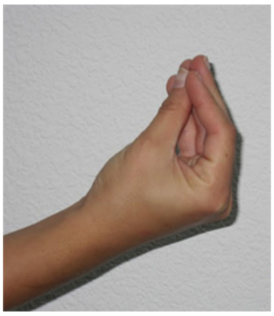


Figure 2.2: The Hand Purse Gesture.

The Hand Purse Gesture has the thumb and fingers are in contact at the tip, the facing upward. This emblem gesture means "query" in Italy, "good" in Portugal, Greece and Turkey; "fear" in Belgium and France [10].

The Hand Purse has different meaning across culture and they are very different, it changes from "query" something in Italy to show "fear" in France. This huge difference of meaning can create problems of misunderstanding when we are in a situation of face to face communication in a cross-cultural context. This kind of communication is the one that we want to address with our system. In the next section we are going to show what is cross-cultural communication and the motivations that guide us for develop ACT system.

2.3 Cross-Cultural Communication

We are in a situation of Cross-Cultural Communication, when two or more people of different culture communicate among each other face to face. As we said in the previous section culture influences the Verbal and Non-Verbal communication of a group of people, such as nations. This kind of communication is not so easy to carry on, because sometimes the result of the interaction is not what we expected, due to some behaviors that have different meaning across culture, if this happen means that there was a misunderstanding during the conversation.

For example the Bulgarian Culture, for them the "Head Nod Gesture" means "No" and "Head Shake Gesture" means "Yes" [9], now imagine a conversation between a Bulgarian and an Italian, in Italy the same gestures have the opposite meaning, in this condition it is very easy to have a misunderstanding between them.

A misunderstanding is the failure to understand something correctly, so if we think in terms of communication process the message send from the sender to the receiver is understood in a different way, this is very common when we use gesture for communicate something because each culture can assign a different meaning for the same gesture. This can create a lot of problems in several situations like global marketing, military operations but also in our daily life, as consequence of the globalization is normal nowadays to find little community of different cultures inside our nations, so to be aware of other cultures is very important for all of us for carry on conversation with them without problems. Our main idea is to create a system able to teach the users the correct non-verbal behaviors to use when they are talking to a person of different culture to attempt to avoid misunderstanding between them.

Figure 2.3: Cross-Culture Communication.

Chapter 3

Related Work

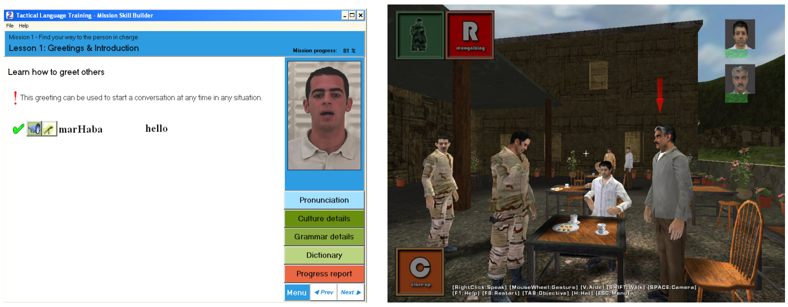
3.1 Tactical Language Training System

The Tactical Language Training System(TLTS)[7] gives learners basic training in foreign language and culture, focusing on communicative skills (verbal and non-verbal). The TLTS is used by United States soldiers to acquire enough communicative skills to carry out civil affair missions. The training is divided into two parts:

- Theoretical Lesson: the student learns new skills, a virtual agent teaches him.
- Interactive Mission Simulation: the student attempts to use the new skills in a simulated mission in a 3D environment.

During all the course the student is followed by an agent that helps him while training, such us giving feedback on pronunciation or giving suggestions in the interactive simulation to achieve the goal. To make the whole experience more compelling and interesting, the application was developed as a game. The user can interact with the system by using his voice, keyboard and mouse. Figure 3.1 shows some snapshots of the TLTS.

TLTS supports new languages and cultures through a set of authoring tools, the system relies on a lot of manual work and could exploit more automation. Actually the system exist for Arabic, Pashto, French, Cherokee and other. Instead with our approach, used in the ACT System, we want to create a cultural independent system able to switch between cultures and exercises just by changing the initialization file of the application with the goal to have an automated and a cultural independent system.



(a) TLTS Theoretical Lesson (b) TLTS Interactive Mission Simulation

Figure 3.1: Screenshots from TLTS.

3.2 ELECT BiLAT

ELECT BiLAT[6] provides soldiers students an engaging and compelling training environment to practise their skills in conducting meetings and negotiations in a specific cultural context. The ELECT BiLAT architecture is based on a commercial game engine integrated with research technologies to enable the use of virtual human characters, scenario customization, as well as coaching, feedback and tutoring.

In this system is integrated the Culturally Affected Behavior (CAB) Model [14] a recent work on cultural behaviors that focuses on cultural norms, used for modelling cultures. The CAB model is based over two main social theory: the Theory of Mind [12] defined as the human ability of give mental states such as intentions, beliefs, and values, not only to oneself but to other as well, the second theory is the Schema Theory [3] which says that a culture can be represented as a shared organization of schema. Graphs called Sociocultural Networks are used to modelling the culture, Figure 3.2 shows an example of this graph for Iraqi Sunni Culture.

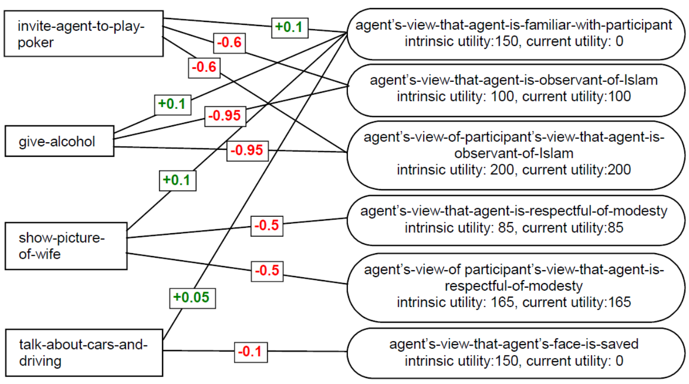


Figure 3.2: Sociocultural Networks Iraqi Sunni Culture[14]

The nodes of this graph represent tasks (rectangular nodes) or cultural norms (rounded nodes), each cultural norms has stored in it an intrinsic utility value showing how much important is for the culture and a current utility value indicating how the simulation is going. The arcs connect each task(action) to one or more cultural norms, a prefixed value is assigned to them indicating if the action is good or bad for the current culture. The CAB Model allows the ELECT BiLAT to be cultural independent because keeping the same task model, what the user have to do in the simulation, you need to change the sociocultural network over the task model for switch the training into another culture. Figure 3.3 shows some snapshots of the training system.

(a) Farid Iraqi Police Office (b) Fritz German Police Office

Figure 3.3: Screenshots from ELECT BiLAT.

Our approach is not so different to CAB Model, we have two model the task model in which is specified what the user has to do and a culture model in which we define the culture. The main difference between our approach and the one used in the ELECT BiLAT is how we represent the culture. We use a set of rules, each rule is composed by a condition that is the action performed in a specific context and the result is the evaluation of the action in the culture, what the receiver understands. With our representation we are focusing on the meaning of gestures and not how specific actions affect the state of interaction. Another difference is the data used in the two representation in the CAB Model they used values representing the influence of an action on the interaction, in our case we works with several values that represent the meaning(intent), the result of the rule, of gesture in a specific culture. This approach assures also to us the culture independences of the system.

Chapter 4

Approach

With our Automated Culture Trainer (ACT) system we want to achieve several goals that are:

- A complete automated culture trainer system, in a 3D environment, able to teach users how to behave when interact with people of different culture.
- A system able to handle different kind of exercises and cultures by reading corresponding input files.
- A good level of modularity to have a system easy to manage.

In the following sections we are going to introduce the framework and a specific language that we used inside our system.

4.1 Behavior Markup Language(BML)

The Behavior Markup Language(BML)[8] is a core part of the SAIBA framework [8]. The goal of BML is to provide a new standard language for the generation of multi-modal behavior for communication. BML is an XML based language that can be embedded in a large XML message or document, the tags that specify the beginning and the end of BML code are: <bml>....</bml>. In this block is stored the description of all the behaviors, verbal or non-verbal, which the animated character has to render. Each BML behavior is considered as a single elements called "Behavior Element" describing the general type of behavior, where for each of them it is possible to specify some optional parameters that define a more accurate behavior, the level of accuracy description of the behaviors depends on the animation engine that is used. An important feature of BML is the specification of temporal constraints, that allow you to synchronize different behavior elements using synchronization points. We can assign to each behavior element an ID when is declared, so the other can reference to this ID from within the BML block and use its synchronization point to time its own animations accordingly. Figure 4.1 shows the main types of synchronization points that can be reference for each behavior and how they represent key points in time during the life time of the behavior. Actually it is also possible to schedule behaviors or wait for an opportune event. Listing 4.1 show an example of a BML block whit three behaviour elements synchronized by using sync point, this block tells to a character to perform the gesture "beat", when the gesture "beat" is in "Ready sync point" the character has to look "object1" till the gesture "beat" arrives to the "Relax sync point",and when the gesture "beat" is in "Stroke sync point" the character has to perform an head nod. The BML lenguage is still in active development the new feature and challenge are available in [16] or on its main website [1]. The ACT system used the BML for define the verbal and non-verbal behaviors that the animated characters have to perform.

Listing 4.1: BML Example.[8]

```
<bml>
  <gesture id="g1" type="beat" />
  <head type="nod" stroke="g1:stroke" />
  <gaze target="object1" start="g1:ready" end="g1:relax"/>
</bml>
```

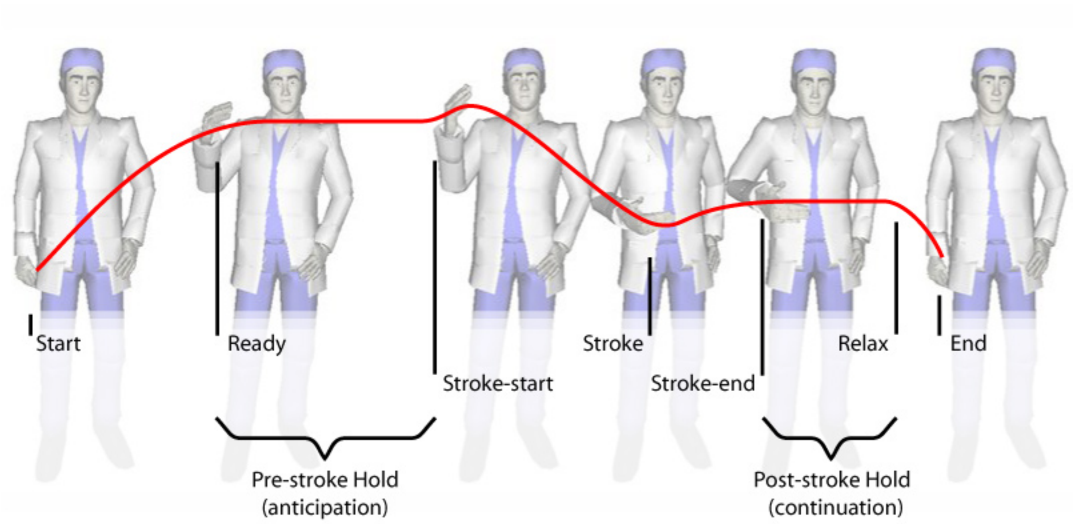


Figure 4.1: The synchronization points of a communicative behavior [8]

[1] http://bml-initiative.org

4.2 Functional Markup Language (FML)

The Functional Markup Language (FML), is XML based language, that aims to represent an higher level of description up from BML, one that can describe the functional level of a multimodal agent. FML represents the goals that the agent wants to achieve by using some behaviors. A brief history of the evolution of FML is here in [18]. Actually FML is not yet a standard, the first attempt to propose a specification was in 2005 by Hannes Vilhjálmsson and Stacy Marsella [17]. They say that a rich enough description at this level is very important because it is the basis for the generation of verbal and non-verbal behavior generation. FML specification is divided into two part, the former define certain basic semantic units associated with the communicative event , the latter can further annotate these units with various communicative or expressive functions. Listing 4.2 shows an example of this specification. In this example the units are: Participant, Turn, Topic, Performative, Content. While the elements that can go to give a more accurate description to the units are in Listing 4.2 are : Emphasis, Affect, Social. The example doesn't show all the elements the other are: Contrast, Coping, Cognitive, Certainty, Illustration. The successive important date for the FML was in 2008 at the first Workshop on FML the main goal of this event was to get together all the researchers on the FML for share among them their founding and the needs about it and for plan the next step towards the Functional Markup Language[5]. The ACT system used the FML for define the goal of the action that the animated character has to achieve, but also for interpreter the meaning of the non-verbal behavior performed.

Listing 4.2: FML Example.[17]

```
<participant id="ali" role="speaker"/>
<participant id="trainee" role="addressee"/>
<turn id="turn1" start="take" end="give">
  <topic id="topic1" type="new">
    <performative id="perform1" type="enquiry">
      <content>goal trainee ? here</content>
  </performative>
  </topic>
</turn>
<emphasis type="new">perform1:here</emphasis>
<affect type="fear">perform1:goal</social>
<social type="maintain_distance">trainee</social>
```

4.3 CADIA Populus

CADIA Populus[13] is part of the Humanoid Agents in Social Game Environments(HASGE) research project at the Center for Analysis and Design of Intelligent Agents(CADIA) at Reykjavik University in collaboration with CCP Games. CADIA Populus is an interactive 3D environment for social behavior simulation, it is geared toward the development of autonomous behaviors for agents and avatars in multyplayer game environments.The main goal of this project is to invent some new methods to autonomously generate believable human behaviors in animated characters. The main feature of this framework are:

- The characters are either fully controlled by the game AI or they are avatars under the direction of human players.
- A particular emphasis is places on the automatic generation of non-verbal behavior that supports communication and socialization.
- The feet, torso, head and eyes of each character is affected by forces that adjust orientation or location. These forces react to the social situation surrounding the character according to a set of behavior rules.
- Provides a conversation system that manage the formation of conversation between characters.

This framework is still in development new behaviors especially social behaviors are continuously added. Thanks to CADIA Populus constructing environments and manipulating the social situation is made very easy for the developer. Screenshot from CADIA Populus environment is shown in Figure 4.2 in which there are some characters interacting. The ACT system used the CADIA Populus for its pathfinding system that allow us to move the character in the scene and for its conversation system that mange the formation of the interaction between two or more characters.

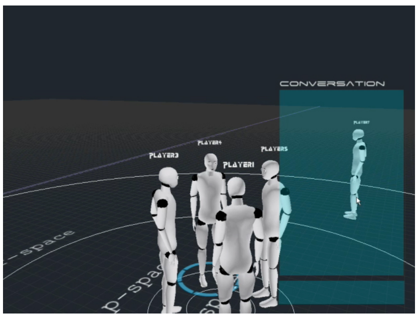


Figure 4.2: Screenshot from CADIA Populus.

4.4 SmartBody

SmartBody[15] is developed at University of Souther California's Information Science Institute(USC/ISI). It is part of the VHuman project at USC that aims to simulate virtual humans that are responsive and life-like in their behaviors. The main feature of SmartBody is its ability to transform BML code into prefixed character animation. For achieve responsive and life-like behavior the developers propose a combination of animation approaches. The architecture of the motion engine is based on motion controllers that can be hierarchically interconnected in real-time in order to achieve continuous motion. This kind architecture allow us to don't care about how to handle continuous sequence of animations, SmartBody do that for us. As we said before SmartBody takes in input BML, so for handle it on top of the motion engine SmartBody needs a behavior and a schedule manager for parse the incoming BML and map each BML request to a set of skeleton-driven motion controllers, than the motion engine takes care of realizing with the appropriate timing specified in BML with the time constraints. The rendering engine is not directly connected to the motion engine in fact they communicate with a network protocol; this solution allow SmartBody to work independent of the rendering engine and this gives to developers the possibility to include SmartBody in different rendering engine without modify the SmartBody code. The ACT system used SmartBody characters in the 3D environment because they have an high level of believability and for render animations defined with BML code.

Figure 4.3 shows the graphics user interface of SmartBody.

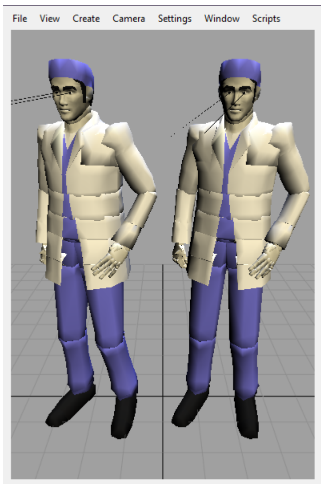


Figure 4.3: SmartBody GUI.

4.5 CADIA BML Realizer

CADIA BML Realizer(BMLR)[1] is an open source animation toolkit for visualizing virtual humans in rich 3D environments developed at Reykjavik University. BML language introduced in section 4.1 is the input of this animation system. As consequence of use BML the developers had the idea of use USC's SmarBody character animation engine introduced in section 4.4 as core of toolkit, but for use it they had to apply some modification to make completely independent, open and very easy to use. The most significant modification was the adoption of Panda3D rendering engine, they rebuilt the communication layer of SmartBody to make it usable in a simple single-animation environment. Figure 4.4 shows some screenshots from CADIA BML Realizer. The ACT system used the BML Realizer, because it allowed us to include SmartBody inside Cadia Populus.

(a) CADIA Lab Environment (b) Falcon Environment

Figure 4.4: Screenshots from CADIA BML Realizer.

Chapter 5

The ACT System

The ACT System goal is to create an application able to help people to avoid non-verbal behaviors misunderstanding in cross cultural communication. As we mentioned in Section 2.2 the same gesture can be interpreted in different ways across cultures and sometimes a polite gesture for a culture can be a rude one in others, with the consequence to cause problems between people during a conversation. Take as an example two business managers of different cultures that are having a meeting for define the last detail of an important contract, in this situation is very important to know the correct non-verbal behavior to use, if a misunderstanding will happen between them there is possibility that the deal's off, that's why avoid cross-cultural misunderstanding is important. In this system we focus our attention only on the non-verbal communication without take into account the verbal one. The ACT System gives the possibility to face in first person the situation of communicate with people of different cultures and in different situation and context. The ACT System explains the situation that the user has been brought into, then runs the user through several tests to check how much the user know about the other culture, the user has to choose the right non-verbal behavior to answer the test. After that the system gives to the user a positive or negative feedback about his answers and if it is wrong it gives a little explanation. We develop a virtual environment where the interaction between the characters take place, Figure 5.1 shows a screenshot from the ACT System. In the environment there are two characters facing each other in an office, this is the case where the system asks the user to choose between wave hand or handshake gesture for greeting the other avatar representing the other culture.

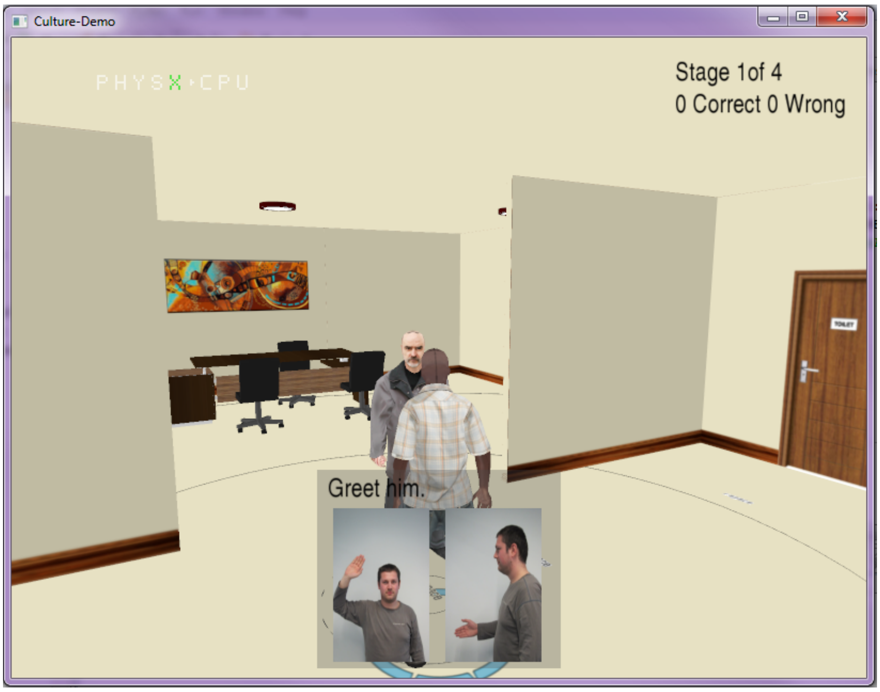


Figure 5.1: Screenshot ACT System

5.1 The Architecture

Figure 5.2 shows the architecture of the ACT System, now we are going to describe each components in detail.

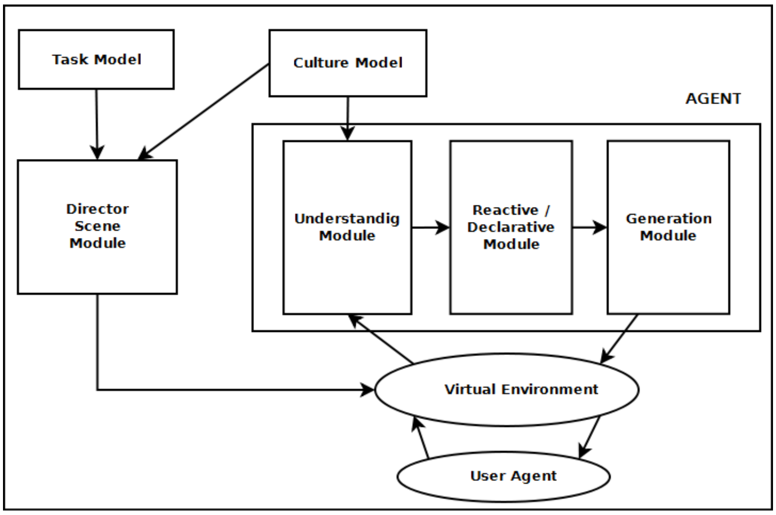


Figure 5.2: System Architecture

Task Model.

The Task Model is an XML file where are stored all the necessary information for describe the tasks that the user have to perform during the training. Some line of code from a Task Model file is showed in Listing 5.1, the XML schema is divided into two parts:

- Initial Set-Up;
- Test Description.

In the former we have the general information useful for the initial set-up of the training environment like the description of the situation and the initial positions of the characters inside the scene. In the latter we are going to define each single scene presented in the simulation; actually a task model file is composed by four scenes used for training the following non-verbal behaviors:

- Greet Someone;
- Express Agreement;
- Express Disagreement;
- Close the conversation.

For each scene we have to specify a question to submit to the user, the situational context, the general intent expected in the test, the positions of the characters in the environment, and the FML expected for this test. We use the coordinates of CADIA Popolus for specify the position of the characters in the scenes. In Listing 5.1 is reported some line of code from a task model file for Italian culture which describe the general context, talk to the boss, and the first scene where the task for the students is to greet the boss.

Listing 5.1: Task Model Example.

```
<Description>
        You are meeting your Italian Boss.
        Press "Enter" to start.
</Description>
<InitialPosAvatar>0,-3</InitialPosAvatar>
<InitialPosNpc>6,3</InitialPosNpc>
<Test>
  <Prompt>Greet him.</Prompt>
  <Context>1</Context>
  <Intent>greet</Intent>
  <PosAvatar>1,0</PosAvatar>
```

```
  <PosNpc>2,0</PosNpc>
  <Task>
    <Fml>
      <greet>
        <SocialDistance>0.9</SocialDistance>
        <Power>0.95</Power>
        <Respect>0.9</Respect>
        <Warmth>0.1</Warmth>
      </greet>
    </Fml>
  </Task>
</Test>
```

Culture Model.

The Culture Model is an XML file where is stored the description of a culture. In this file is stored the FML intent of a non-verbal behavior in a particular context, it is possible that a gesture appear more than once in the file, that's because the same gesture can have different meaning in different context. So for each gesture we are going to create one or more rules in the XML file and we are going to group them on the basis of their general intent, for have a more readable list. In Listing 5.2 there is an example of a XML culture model file describing the Italian Culture, the rules describe a non-verbal behavior used for greet people, we have the same gesture with different meaning in different context.

A rule is divided into two parts, the former is the condition that have to be satisfied for apply the rule and the latter is the result of the rule itself. In the condition we have:

- The non-verbal behavior specified with the Behavioural Markup Language (BML).
- The context in which the gesture is performed.

While the in the result is stored:

- The meaning of the non-verbal behavior in that particular context specified by using the Functional Markup Language (FML).

Actually in the culture model is possible to store only gestures that can be used in the scenes present in the task model but new non-verbal behaviors can be easily added if a new scene is added to the task model.

Listing 5.2: Culture Model Example.

```
<Rules type="greet">
  <Rule>
    <Condition>
      <Bml>
        <gesture type="wave"/>
      </Bml>
      <Context>0</Context>
    </Condition>
    <Result>
      <Fml>
        <greet>
          <SocialDistance>0.27</SocialDistance>
          <Power>0.0</Power>
          <Respect>0.5</Respect>
          <Warmth>0.95</Warmth>
        </greet>
      </Fml>
    </Result>
  </Rule>
  <!-- Second Rule for Greet -->
  <Rule>
    <Condition>
      <Bml>
        <gesture type="wave"/>
      </Bml>
      <Context>1</Context>
    </Condition>
    <Result>
      <Fml>
        <greet>
          <SocialDistance>0.27</SocialDistance>
          <Power>0.0</Power>
          <Respect>0.5</Respect>
          <Warmth>0.1</Warmth>
        </greet>
      </Fml>
    </Result>
```

```
  </Rule>
</Rules>
```

Director Scene Module.

The Director Scene Module is responsible to drive the ACT System.

It takes in input the Task Model and the Culture Model files, for initialize the system preparing the test to submit to the user and the scene to render into the virtual environment. Than its task is to update the virtual environment according to the scenes stored in the Task Model file, drives the characters inside the environment, shows and hides the Graphics User Interface.

Virtual Environment.

The 3D Virtual Environment is build over the CADIA Populus engine integrated with SmartBody animation toolkit by using the CADIA BML Realizer as middlewere between them. This integration permits us to use SmartBody characters inside CADIA Populus, in this way we can keep the advantages of the two systems. The Virtual Environment receives input from the Director Scene Module and the User. The inputs from the first one are used to change the scene of the environment like position of the avatars and so on, instead the User requests to render non-verbal behavior by his avatar. Figure 5.1 shows the current environment available in the ACT System, a modern office. The rendering engine is Panda3D[1] engine, because CADIA Populus uses it.

User Agent.

The User Agent is a Virtual Human representing the user inside the environment. Technically consist of a SmartBody character highlighted with a sky-blue circle around it. The User Avatar is driven by the Director Scene Module around the environment. Its main task is to render the behaviors selected by the user as answer of the exercise.

[1] http://www.panda3d.org

Agent.

This module is the Artificial Intelligence behind the NPC[2] represented by a SmartBody Character and its main task is to handle the Agent Perception Loop. The goal of the Agent Perception Loop is to generate a reaction to the User Agent actions. We can divide the Agent Perception Loop into three step:

1. Understand the behaviour performed by the User Agent, it consists to retrieve the FML intent of the gesture in the Agent culture and this is the task of the **Understanding Module**;
2. Analyse the FML obtained from the previous step and evaluate it, this is the task of the **Reactive/Declarative Module**;
3. Generate and perform the reaction of the Agent on the basis of the FML evaluation from the previous step, this is the task of the **Generation Module**

In Section 5.2 we are going to explain in detail how the understanding process works.

5.2 The Agent Perception Loop

Figure 5.3 shows the schema of the Agent Perception Loop, its goal is to build the NPC reactions to user agent actions. This process is activated when the user answers to the test and it produces the feedback about the chosen answer. The Agent Perception Loop is divided into three step.

The first step is performed by the Understanding Module, it takes as input the BML specification of the User Agent gesture and the situational context, then looks inside the Culture Module, associated to the NPC, for a rule with its condition satisfied by the input of the Understanding Module. Once the rule is found, we have the FML intent understood by the Agent, the result of the rule. With this module the NPC is able to interpret the gesture of the User Agent according to its culture.

The second step is handled by the Reactive/Declarative Module, its goal is to compare the FML understood in the previous step with the one expected for the exercise stored in the Task Module file, if all the FML value are in the range of the expected value, we give a fixed threshold to each of them, the answer is correct otherwise no. Now we have an evaluation of the FML understood by the Agent and we are able to say if there is a misunderstanding or not, if some FML values are out of range.

[2] NPC (Non Player Character)

The third and last step is performed by the Generation Module, its goal is to prepare the feedback to the user taking in input the FML evaluation from the previous step, it gives a positive feedback if all the FML values are in the expected range, otherwise is going to underline to the user which FML values are not satisfied by preparing a negative feedback. Two example of possible feedback in the ACT System are:

- **Positive:** "Well Done! Let's move to the next scene!"
- **Negative:** "Why don't you show me the correct Social Distance, Respect?"

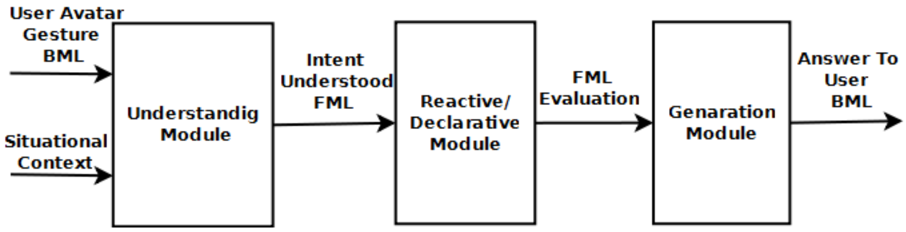


Figure 5.3: Perception Loop Schema

5.3 FML Specification

In this section we are going to talk in detail about the FML specification crated for the ACT system, Listing 5.3 shows some FML line of code that express the intent of Greet Respectful someone, taken from a Cultural Model. The FML specification used in the Task Model file and Culture Model is the same. Actually the system can handle four kind of general intents:

- Greet;
- Agreement;
- Disagreement;
- Salutation.

Listing 5.3: FML Example.

```
<Fml>
  <greet>
    <SocialDistance>0.9</SocialDistance>
    <Power>0.95</Power>
    <Respect>0.9</Respect>
    <Warmth>0.1</Warmth>
```

```
  </greet>
</Fml>
```

The FML specification is surrounded by the tag <FML> and </FML>, the second tag specify the general intent that the character wants to achieve and in this case is <greet> someone, then are specified four values used for augmenting the description level of the general intent.

Using this specification we can define several specific intent from the general one, just by changing the values of the parameters, this structure help us to represent what happen in real life where we have different ways to do something, for example we have different way to greet someone, we can greet him friendly or in a respectful way, with our specification it is possible to specify this two goals. When we define the FML for describe an intent we have to specified all the values for the four variables and they can assume values in the range between 0 and 1 included. The meanings of those four variables are:

- **Social Distance:** indicates the relation that exist between the speaker and the receiver. A value equal 0 means Familiar Relation (Father-Son) instead 1 means completely stranger (two people never meet before).
- **Power:** indicates how much difference exist in the social scale between the speaker and the receiver. A value equal 0 means same social level(conversation Student-Student), 1 means a huge difference in social scale (conversation Prime Minister-Student).
- **Respect:** indicates how much respect the speaker wants to show to the receiver. A value equal 0 means low respect(rude person), equal 1 high respect(polite person).
- **Warmth:** is related to the level of friendliness,helpfulness,sincerity,trustworthiness and morality [4],so how much warmth do you want to show to the receiver. A value equal 0 means no warmth at all(you wants to keep the distance with the receiver), equal 1 means warmth (you want to be very close to the receiver).

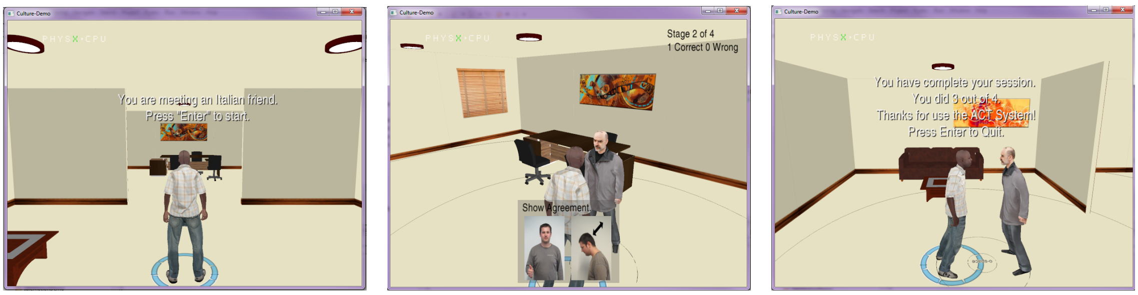
The Social Distance and Power are considered Sociological Variables and they are part of the Politeness Theory[2].

5.4 Graphic User Interface

The Graphic User Interface(GUI) allows the user to interact with the system. The GUI created for the ACT System is very simple and user friendly. Actually the student can interact only by using the mouse:

- Left Click over gesture's picture for select an answer.
- Move up or down the mouse wheel for control the camera's zoom.
- Press the mouse wheel and move the mouse for rotate the camera around the user avatar.

Figure 5.4 shows three screenshots from the ACT System in different moments of the training. The first picture, Figure 5.4a, is the beginning of the training when the initial message is showed to the user for let him know the situational context, in this case is "You are meeting an Italian friend. Press Enter to Start". In Figure 5.4b are presented two components of our GUI, one is the answer's frame that contains the question prompt and the possible answers. The gesture's pictures, inside the answer's frame, are buttons so for select your answer just a left mouse click over one of them. The second component is in the top right of Figure 5.4b, a frame box containing the current situation of the test: the actual stage and global score of the current simulation. In this case the student is in stage 2 with a global score of 1 correct and 0 wrong. Figure 5.4c, is an example of possible end of a training where the final message is showed to the user, the message contains the global feedback that consist of summary of the student score, like in the example below "You have complete your session. You did 3 out of 4.".



(a) Initial Message. (b) Stage and Score of simulation, Answer's Frame. (c) Final Message.

Figure 5.4: GUI Components.

5.5 Animations

The Animations available in the ACT System are in part provided by SmartBody itself and the other are developed by us. We use Blender 2.49b[3] a 3D free graphics application, it can export animations in collada[4] format accepted as input by SmartBody. This solution is not the best one, because Blender and SmarBody work with a different coordinates system, the first one works with Z-Up instead the second one with Y-Up, so create new animations is not so easy and require a lot of time. The best solution for the creation of the animations is to use Maya[5], that works with the same coordinate system of SmartBody, but we haven't the licence available. Figure 5.5 shows the Blender 2.49b window containing the skeleton of a SmartBody Characters during the creation of a wave gesture.

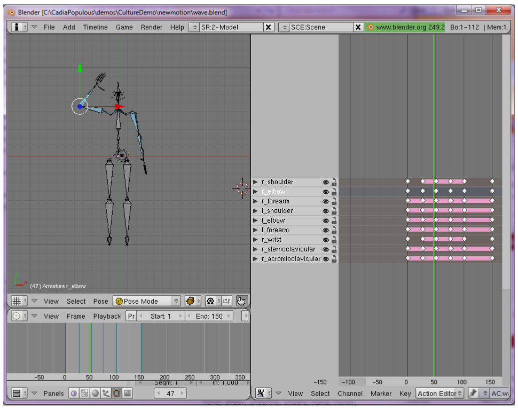


Figure 5.5: Blender 2.49b Screenshot

[3] "http://www.blender.org"
[4] "http://www.collada.org"
[5] "http://usa.autodesk.com/maya/"

Chapter 6

Conclusion

6.1 Result

Actually the ACT System is just a prototype, a lot of work still remain for have a system able to compete with the others like the TLTS introduced in Section 3.1. Although the system is at the beginning of its development we achieve all the goals prefixed for the project.

First of all the ACT System is a complete system that works properly, the student can carry on all the training from the beginning to the end of the simulation, learning new skills to use in his real life. The system is also cultural independent, it can work with all the cultures, you need only to create the culture model and the opportune gesture if not yet available. The system is automated it loads exercises from an external file, the Task Model, in this way it's more easy to add new exercises because they are not hard coded in the system. The architecture developed for the ACT System has a good level of modularity, it is easy to modify or add new components so we can gain time in the future upgrade of the system.

6.2 Demonstration

In this section we are going to present a demonstration of what happen when the user runs the system from beginning to end. Figure 6.1 shows the first scene in an example of simulation. At this point the user is made aware about the general context of the simulation, in this case is: "You are meeting your Italian Boss. Press Enter to Start.". When the user is ready he presses the Enter button and the simulation starts. Now the Director Scene

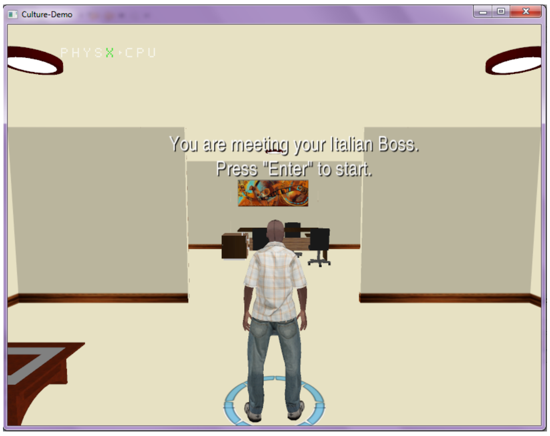


Figure 6.1: Initial Message to the User

Module takes the control of the system and update the current scene, moves the character in the scene and shows the question, on the basis of what is stored in the Task Model. When this process is over the user is in the situation presented in Figure 6.2. The user is now in a conversation with another character, his Italian Boss, and he has to choose the right gesture for greet him. The user can choose between two gesture: "Wave Gesture" or "Hand Shake Gesture", the gestures pictures in the frame shows the possible answers. For select the gesture the user has to click on its picture. Right after the selection his avatar performs the gesture, this event activates the Agent Perception Loop, Section 5.2, of the other character that takes in input the current context and the BML command of the gesture selected.

Figure 6.2: Test to the User

Now the user is waiting for the feedback response from the second character, the Italian Boss, to know if his answer is correct or not. The feedback message is the output of the Agent Perception Loop, its task is to understand the meaning of the user gesture in the culture of the Boss. Figure 6.3 shows how the user receives the feedback, in this case

Figure 6.3: Feedback to the User

is a negative one. The massage says: "Why do you not show me the correct respect, social distance, warmth and power.". When a gesture is not correct the system tells to the user which variable of the FML specification expected in the Task Model are out of range respect to the FML understood from the NPC character, the Italian Boss. In case of correct answer the feedback to the user is: "Well Done!!Let's move to the next stage!". The last thing that the system has to do after the feedback message is to update the current status of the simulation, in the top right of the window in Figure 6.3 there is the state of the simulation: Current Stage and Score. All this process is repeated depending on how many scene or exercise are stored in the Task Model, in this simulation the user has to answer to other three exercise that are: Show Agreement, Show Disagreement, Close the Conversation. The change of scene is always handle by the Director Scene Module. After the last exercise the user receives the result of the simulation, how many answers were correct. Figure 6.4 shows the end of the simulation for the user, where he did 2 out of 4. This is the end of the execution of our system.

Figure 6.4: Final Message to the User

6.3 Future Step

There are many thing we would like to do and possible future directions for the ACT System. The next step of the project should be an evaluation of it, made from some users, for receive some feedback and know if we are in the right direction or not, changing as consequence our plan for the future development of the system. The actual version is a prototype it can be improve under several aspects for example we want to give to the simulation a more game style, creating a more compelling and interesting story where the user has to achieve a goal instead to answer to some questions. We would like to create new scenarios for handle new situation,as example public space or open space. Actually a menu for allow the user to choose the avatar, the exercise and the culture to training is missing so we need to implement one. Add to the ACT System the possibility to interact with the NPC using voice for start to take into account the Verbal Behavior in the Cross-Cultural Communication. We would like also to improve the FML specification by adding new variables to increase the level of description, it might be possible to join in our work people that study misunderstanding in cross cultural communication that have more knowledge and insights about the problems we are attempting to address.

Bibliography

[1] Bjarni Thor Arnason and Aegir Thorsteinsson. *BML Realizer: An Open Source BML Animation Toolkit*. Reykjavik University, 2008.

[2] Penelope Brown and Stephen C. Levinson. *Politeness: Some universals in language usage*. Routledge, 1999.

[3] R. D'Andrade. *Human Motives and Cultural Models*, chapter Schemas amd motivation, pages 23–44. Cambrige University Press., 1992.

[4] Susan T. Fiske, Amy J.C. Cuddy, and Peter Glick. Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, 11(2):77–83, 2007.

[5] Dirk Heylen, Stefan Kopp, Stacy C. Marsella, Catherine Pelachaud, and Hannes Vilhjálmsson. The next step towards a function markup language. pages 270–280. Springer-Verlag, 2008.

[6] Randall W. Hill, James Belanich, H. Chad Lane, and Mark Core. Pedagogically structured game-based training: Development of the elect bilat simulation. In *In Proceedings of the 25th Army Science Conference*, 2006.

[7] W. Lewis Johnson, Stacy Marsella, and Hannes Vilhjalmsson. The darwars tactical language training system. SSA, 2004.

[8] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn Thorisson, and Hannes Vilhjalmsson. Towards a common framework for multimodal generation in ecas: The behavior markup language. In *Proceedings of Intelligent Virtual Agents 2006*, volume 4133, Berlin Heidelberg. Springer.

[9] Evelyn Z. McClave. Head movements in arabic, bulgarian, korean and african american english: What's cognitive and what's cultural. In Christian Cave, Isabelle

Guaitella, and Serge Santi, editors, *Oralité et Gesturalité: Interactions et comportements multimodaux*, volume Aix-en-Provence, France, pages 560–564. L'Harmattan.

[10] David McNeill. *Hand and Mind.* Chicago, 1992.

[11] Cornelia Muller. *Beyond Misunderstandings*, chapter Gesture and Speech Cross-Culturally. Benjamins, Amsterdam, 2005.

[12] S. Nichols and S.P. Stich. *Mindreading: an integrated account of pretence, self-awareness, and understanding other minds*. Clarendon, 2003.

[13] Claudio Pedica and Hannes Vilhjálmsson. *Social Perception and Steering for Online Avatars*, pages 104–116. Springer-Verlag, 2008.

[14] S. Solomon, M. van Lent, M. Core, P. Carpenter, and M. Rosenberg. A language for modeling cultural norms, biased and stereotypes for human behavior models. *BRIMS*, 2008.

[15] Marcus Thiebaux, Stacy Marsella, Andrew N. Marshall, and Marcelo Kallmann. *SmartBody: behavior realization for embodied conversational agents*, pages 151–158. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[16] Hannes Vilhjálmsson, Nathan Cantelmo, Justine Cassell, Nicolas E. Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Zsofi Ruttkay, Kristinn R. Thórisson, Herwin Welbergen, and Rick J. Werf. *The Behavior Markup Language: Recent Developments and Challenges*, pages 99–111. Springer-Verlag, 2007.

[17] Hannes Vilhjálmsson and Marsella Stacy. Social performance framework. volume Pittsburgh, PA. AAAI.

[18] Hannes Vilhjálmsson and Kristinn R. Thórisson. A brief history of function representation from gandalf to saiba. In *1st Function Markup Language Workshop at AAMAS*, 2008.

School of Computer Science
Reykjavík University
Menntavegi 1
101 Reykjavík, Iceland
Tel. +354 599 6200
Fax +354 599 6201
www.reykjavikuniversity.is
ISSN 1670-8539