Gesture Activated Mobile Edutainment (GAME)

Intercultural Training of Nonverbal Behavior with Mobile Phones

Rehm, Matthias; Leichtenstern, Karin; Plomer, Joerg; Wiedemann, Christian

Published in:
9th International Conference on Mobile and Ubiquitous Multimedia

Publication date:
2010

Document Version
Accepted author manuscript, peer reviewed version

Gesture Activated Mobile Edutainment (GAME)

Intercultural Training of Nonverbal Behavior with Mobile Phones

Matthias Rehm
Department of Architecture, Design, and Media Technology (CREATE), Aalborg University
Niels Jernes Vej 14
9220 Aalborg, Denmark
matthias@create.aau.dk

Karin Leichtenstern, Jörg Plomer, Christian Wiedemann
Human Centered Multimedia
Universitätsstr. 6a
86150 Augsburg, Germany
leichtenstern@informatik.uni-augsburg.de

ABSTRACT

An approach to intercultural training of nonverbal behavior is presented that draws from research on role-plays with virtual agents and ideas from situated learning. To this end, a mobile serious game is realized where the user acquires knowledge about German emblematic gestures and tries them out in role-plays with virtual agents. Gesture performance is evaluated making use of build-in acceleration sensors of smart phones. After an account of the theoretical background covering diverse areas like virtual agents, situated learning and intercultural training, the paper presents the GAME approach along with details on the gesture recognition and content authoring. By its experience-based role-plays with virtual characters, GAME brings together ideas from situated learning and intercultural training in an integrated approach and paves the way for new m-learning concepts.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors, human information processing*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*artificial, augmented, and virtual realities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*haptic I/O, input devices and strategies, interaction styles*

General Terms

Human Factors

Keywords

Virtual Agents, Mobile Edutainment, Gesture Recognition, Intelligent Tutoring System

MUM '10, December 1-3, 2010, Limassol, Cyprus.

1. INTRODUCTION

Learning a foreign language often concentrates on the verbal aspect of the learning endeavor. But communication is not only concerned with verbal proficiency but it is inherently multimodal ranging from facial expressions over gestures to spatial behavior, which often follow culturally determined heuristics. As an example consider a dinner table discussion. The structure of such a multiparty conversation can vary from a very orderly turn after turn sequence to a very lively situation where several interactions and discussions take place at the same time between different participants. Often such nonverbal aspects of communication give rise to severe misunderstandings [25]. For instance, the first group in our example might classify the second one as chaotic and unfocused whereas the second group might think of the first one as restrained, distant and cold. Another well studied example is the use of space in interpersonal encounters [7]. While for instance in Northern Europe a certain distance between interlocutors is generally acceptable, in an Arabic context, this distance should not be to far in order to allow for touching between interlocutors. Again the interpretation of the other group's behavior is bound to differ, often resulting in the first group finding the second group invasive or pushy and the second thinking about the first as distant and cold. This is due to the fact that behavior is interpreted based on unconscious cultural heuristics that are formed by our personal interaction histories in the cultural groups to which we belong. The aim of intercultural training is thus two-fold [10]. First, the trainee has to realize that a communicative situation is ambiguous and can be interpreted in different ways (awareness) before he can try to adjust to different perspectives (knowledge and skills).

In this paper we present our ideas on assisting this endeavor by a technical solution. Thereby we draw motivation from two different areas. First, virtual agents, e.g. in the form of embodied conversational characters [3], offer natural interaction possiblities because of their potential to emulate verbal and nonverbal human behavior. In general, nonverbal interaction comprises facial expressions, gaze behavior, gestures, and body posture, which all play sometimes distinct, sometimes redundant roles in face to face communication. Virtual characters have also been show to be engaging tools for tutoring systems (see Section 2), and thus present a good starting point for exemplifying different perspectives in intercultural training and thus presents our first motivation in developing GAME.

Second, current trends in intercultural training emphasize the importance of a coaching approach [6], which is centered on the trainee's needs, goals, and especially on his agenda. This means that instead of following standard lessons in the classroom, small-scale experience-based learning sessions are delivered to the trainee anytime anywhere. Ideally, these learning sessions are tailored to the specific context and situation. For instance, being at the train station in Munich triggers a learning session on how to purchase a train ticket in Tokyo station. The coaching idea is the second motivation for our work.

With GAME (Gesture Activated Mobile Edutainment) we present our first step in this direction by realizing an experienced-based role-play with virtual characters on a smartphone for increasing intercultural awareness and training nonverbal behavior, in our case German emblematic gestures. In this paper, we first present the relevant ideas from diverse research directions like virtual characters and cultural training that have been integrated in the system (Section 2). Then, the GAME approach is presented, starting with the challenge of gesture recognition on the mobile phone over training and game mode of the application and ending with details on the authoring of learning scenarios (Section 3). Afterwards, an exploratory study is presented that took place on a public event (Section 4) before the paper concludes (Section 5).

2. THEORETICAL BACKGROUND

GAME brings together different research directions from cultural training over role-plays with virtual characters to mobile learning in a comprehensive edutainment scenario drawing heavily from previous work in these diverse areas. In the following, a short introduction is given to diverse backgrounds.

2.1 Enculturated Agent Systems

Culture itself is a multiply defined notion that gives rise to many misconceptions ranging from theater and art over language and national affiliation. Thus, it is necessary to specify exactly, what is meant by culture in the envisioned training system as this notion affects several levels of the system like the content of the learning scenarios or the behavior of the virtual characters. We claim that it is indespensable to base a system that integrates cultural aspects of interaction on a thorough theoretical foundation that allows for reliably predicting patterns of behavior that are influenced by cultural heuristics. Hofstede [10] presents a starting point with his theory of cultural dimensions that defines culture as a five-dimensional concept and relates positions on the dimensions to certain behavioral heuristics [11]. Thus, it becomes possible to predict behavioral tendencies based on the position of a culture in this five-dimensional space. Although Hofstede's work has been successfully adapted in the area of cultural usability (e.g. [19]; [20]), attempts for enculturating interactive systems like virtual agents have so far been mostly ad hoc and often without a thorough theoretical or empirical foundation.

The commercially most successful intelligent tutoring system that employs virtual characters and focuses on cultural aspects is the tactical language training [14]. It is used as a training tool for soldiers that face ex-patriate missions. In the training sessions, the users have to solve tasks by employing their language knowledge in the given situation. The main interaction modality is speech. Additionally, users can select gestures to accompany their utterances that are then played as an animation of their avatar. Culture is equated in this case with the language that is trained and used as a back story for creating animations for the virtual characters. The training goal is language proficiency.

In [16], an intelligent tutoring system is described that is tailored at teaching business etiquette in intercultural encounters. Again, culture is used as a back story for the role-play with a virtual character that determines the "production design". The system aims at teaching (stereo-)typical rules of behavior like "do not bring alcohol as a present in Arabic countries", and allows the user to put his knowledge about such rules to a test in a kind of adventure game. The interaction is realized as a text input.

The above mentioned systems focus on language and knowledge about cultural rules. According to Ting-Toomey [25], the most severe misunderstandings in intercultural communication arise due to different perspectives on appropriate nonverbal behavior in communicative situations. A parameter-based model of culture is described in [13], where certain nonverbal behaviors (proxemics, gaze) of virtual agents are modified in a culture-specific way (US, Mexican, and Arabic) relying on the model parameters. The necessary data for this approach is drawn from a literature review. It turns out that the information from the literature is in most cases merely qualitative in nature, often gives only mean values or does not give information about a culture under investigation. A consequence of this is a mix of culture-specific behavior in the system, e.g. American turn-taking with Arabic proxemics and gaze, which makes it difficult to pinpoint effects found in preliminary perception studies to cultural variables.

A similar problem was encountered in [22], which led to a thorough empirical study in order to deal with the lack of reliable data. Based on the results and Hofstede's dimensional model [10], a probabilistic model of nonverbal behavior is derived, which is employed to categorize and interpret observed user behavior and to control the animations of virtual characters. The user actually performs nonverbal behaviors, e.g. by using a Wiimote, which allows for executing and analyzing gestures. Thus, it becomes possible to give the user a direct feedback on his performance. A prototype is described that gives feedback to users on their performance by adapting the nonverbal behavior of a group of agents. That the collected data presents a rich source of comparative data is exemplified in [5], where the data corpus is used to analyze and model cultural aspects of verbal interaction. A plan-based approach for realizing culture-specific small talk between virtual agents in first meetings is developed based on the empirical insights gained from the recordings.

That the neglect of a thorough cultural model can result in quite dubious systems is exemplified in [28] with a collaborative role-playing game. The approach is problematic because the game itself is culturally biased as it is a typical Western military action game. Moreover, the decisions players have to make in the game seem to be solely based on the developers intuition and thus their own cultural background. Thus, another bias is introduced on how to behave "correctly" in the game. Thus, showing different success rates when comparing US teams and multinational teams is not a suprising result.

2.2 Experience-Based Role-Plays

All of the above systems make use of virtual characters as a useful tool for training systems. Isbister [12] has convincingly argued for the use of agents to further cross-cultural communication skills between users. Compared to life role-playing games, learning with virtual agents adds new experiences to the learning process.

- Repeatability: The training scenario can be repeated as often as necessary without annoying a human training partner. Moreover, either one user can repeat a given lesson until he finishes successfully, or several users can train with the same agent successively.
- Emotional Distance: Because culture and cultural communication is a quite critical theme, people might easily get offended when treated (in their opinion) wrongly. Additionally, trainees are often hesitant in trying novel nonverbal behavioral styles. Interacting with an agent, the user does not have to be afraid of doing something wrong, or feel embarrassment.
- Intensity: With a virtual agent, special nonverbal features can be displayed in varying intensities, allowing to highlight even subtle differences in behavior. An added benefit is the possibility of isolating certain features allowing the user to concentrate only on those features.
- Generalization: The same agent and virtual scene can be used to simulate different cultures. Thus the same system can be reused and adopted for instace to contrast the behavior of two cultures and point out the differences.
- Feedback: If the user's behavior is logged during an interaction, the agent can be used to replay this behavior and exemplify/emphasize problems or progress and can contrast the behavior either with previous behavior of the user or with the target behavior.

Although it is often claimed that virtual agents have positive effects on the learning experience, there is only one reliable large-scale evaluation study so far that investigates the effects of experience-based role-plays with virtual characters in detail. FearNot!v2 is an anti-bullying learning software that is designed to exemplify coping strategies for bullying in school and to let children test different strategies in a safe environment. An evaluation study has been conducted in 30 classes in two countries to evaluate the effects of employing virtual agents in training systems [23], which shows that agents can be successfully employed for experience-based learning. Whereas interaction in FearNot was purely text driven, a follow-up system has been introduced, which makes use of the same agent architecture and integrates also some nonverbal behaviors [1].

The general idea behind employing experience-based role-plays is situated learning (e.g. [4]; [27]), which relates directly to the coaching idea. In this paradigm, learning has to take place in specific situations which provide rich contextual clues. Transfered to the language learning scenario for instance, instead of learning the dialogue for buying bread in class, you go to an actual bakery and buy bread there. This of course is not possible in most cases because often one starts learning a new language in one's home country. Thus, role-plays with virtual characters are argued to be a good substitute to experience and learn in specific situations.

2.3 Intercultural Training

With GAME, we aim at providing the means to train gestures anytime anywhere in role-plays following suggestions by Hofstede [9], who describes three steps of intercultural training (Figure 1):

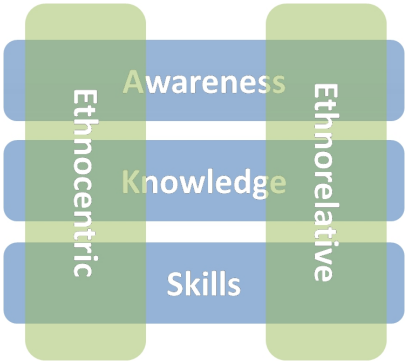

Figure 1: Two dimensional model of intercultural coaching.

1. Awareness: The first step of gaining intercultural competence is being aware and accepting that there are differences in behavior. The hardest part of this learning step is to accept that there are no better or worse ways of behaving and especially that one's own behavior routines are not superior to others. To realize this step in a learning system with embodied conversational agents, the trainee is confronted with a group of characters displaying the behavior routines of the target culture. With the knowledge of the trainee's cultural background, the agents could also contrast the behavior of the target culture with the behavior of the trainee's culture. Comparing the behavior patterns the trainee recognizes that there are differences but might not be able to name them.
2. Knowledge: In the second step, the trainee's knowledge of what exactly is different in the behavior is increased, which can be interpreted as getting an intellectual grasp on where and how one's own behavior differs. For instance the trainee might have felt a little bit uncomfortable in step one due to a different pattern of gaze behavior. In step two, he will gain the knowledge on how his behavior patterns differ from the patterns of the target culture and what the consequences are. In the learning system, the user is confronted with reactions to his behavior by his interlocutors. For instance, the agents could move away if the user comes too close. Moreover, the agents could replay specific behavior routines of the user and contrast them to the behavior routines of the target culture, pointing out where exactly the user's behavior deviates from the target culture.
3. Skills: Hofstede argues that the first two steps are sufficient to avoid most of the obvious blunders in intercultural communication. If the trainee has the ambition to blend into the target culture and adapt his own

behavior, a third step is necessary, the training of specific nonverbal communication skills. If e.g. avoiding eye contact in negotiations is interpreted as a sign of disinterest in the target culture, it might be a good idea to train sustained eye contact for such scenarios. Again, virtual characters can play a vital role in this learning step due to the above mentioned features (see Section 2.2).

Apart from the three steps introduced by Hofstede, Bennett [2] argues concisely that the success of a learning session is tightly related to the user's stage of intercultural awareness. He establishes a succession of six stages from ethnocentrism to ethnorelativism that the trainee passes through and that differ in applicable teaching methods. This means, a full-blown contextual coaching application for cultural awareness will have to take all these dimensions into account by integrating the two-dimensional model depicted in Figure 1.

2.4 Summary

With GAME, we present a first step in this direction. The system integrates interactive role-plays with virtual characters with knowledge and skills training described by Hofstede. To this end, a mobile serious game is realized where the user acquires knowledge about German emblematic gestures and then trains to perform these gestures in role-plays with virtual agents. Gesture performance is evaluated making use of build-in acceleration sensors. Currently, Bennett's ideas on the transition from ethnocentric to ethnorelative perspectives have not been integrated. But in Section 3.3, possibilities for content authoring are presented that can be used to define different learning scenarios taking Bennett's stages into account.

As we have seen in this section, mobile edutainment scenarios have the potential of relating a number of theoretical concepts on innovative learning. By its experience-based role-plays with virtual characters, GAME brings together ideas from situated learning and intercultural training in an integrated approach and paves the way for new m-learning concepts.

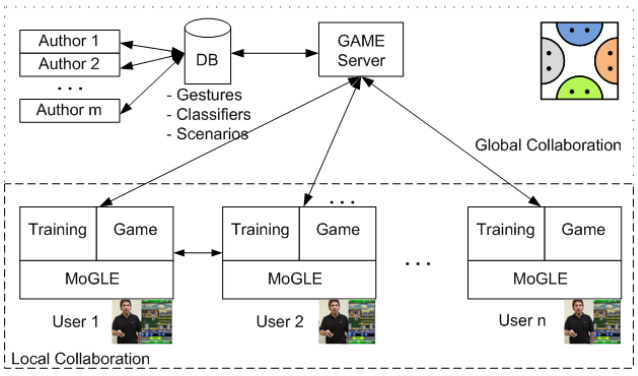

Figure 2: The GAME architecture.

3. THE GAME APPROACH

Figure 2 gives an overview of the whole GAME architecture. GAME has been realized as a collaborative mobile environment. The user can choose to either run GAME in single user mode or in competitive mode. Collaboration can be implemented locally by one user becoming the master, the others the slaves or remotely by connecting to the GAME server. Here in this account we focus on the single user scenario and disregard the collaborative aspect of the system. The user can load new scenarios as well as gestures along with classifiers from the server. Content is authored by a XML-based authoring tool allowing for specifying narrative structure, cut scenes as well as gesture informaton, and can be done by expert community members from the target culture.

Relating to the three steps of intercultural training by Hofstede (see Section 2.3), GAME focuses on the second and third step assuming that the user already has a certain level of cultural awareness. Thus, by playing with the system, the user acquires knowledge and skills of culture-specific behavior, in our example about German emblematic gestures. To this end, the system features two modes, one dedicated to training specific skills (training mode, Section 3.2), the other allowing the user putting his new knowledge and skills to a test in specific situations, a visit to a beergarden (game mode, Section 3.3). Both modes require analyzing the user's gestures. Details on this process are given in the next section.

3.1 MoGLE – Mobile Gesture Learning Environment

GAME aims at training German emblematic gestures. Thus, the user's gestural input has to be classified. Current smartphones offer acceleration sensors, which can be utilized to this end. Accelerometer based gesture recognition has been shown to work at a high level of accuracy (e.g. [18], [21], [24], [26]). Based on previous work on gesture recognition with Nintendo's Wiimote controller presented in [21], we aimed a utilizing the acceleration sensors of handhelds for the same end. Thus, the general ideas from [21] have been adapted. In order to become leaner and faster to operate on the restricted environment of a mobile phone, MoGLE (Mobile Gesture Learning Environment) restricts the number of available features and offers only a Naïve Bayes classifier in order to minimize calculation efforts on the mobile device.

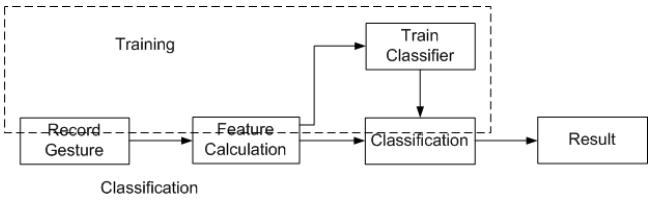

Figure 3: The standard classification pipeline has been integrated in MoGLE.

Figure 3 illustrates the standard classification process that has been integrated in MoGLE. To train the classifier, a training set is recorded for each gesture class preferably by different users. Features are calculated on the raw signals and the resulting feature vector along with the information about the gesture class is used to train the Naïve Bayes classifier. For realtime classification, features are calculated for each gesture and the classifier calculates the most likely class for the feature vector. Currently, MoGLE is running under WindowsMobile on an HTC Touch Diamond[1]. The

[1]Please contact the authors to receive a copy of the software.

acceleration sensors are working with a framerate of 60 Hz for each axis. On the raw data, standardized statististical features are calculated for each axis: minimum, maximum, length, mean, median, and gradient.

Different evaluations were run to ensure that performance is comparable to the results presented earlier. In [21], we have shown that accelerometer-based gesture and expressivity recognition is robust and reliable.

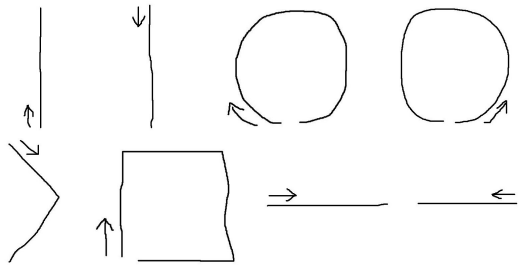

Figure 4: VCR control gestures: Top row from left to right: gestures for play, stop, next, previous. Bottom row from left to right: gestures for increase, decrease, fast forward, fast rewind.

To evaluate MoGLE, we replicated one of the experiments done with the Wiimote. The gesture set used as our benchmark are control gestures for a video recording device, which were first introduced by Mäntijärvi and colleagues ([15]; [18]). Thus, using this gesture set allows us to evaluate MoGLE against two reference applications. In the original approach by [18], the raw acceleration data is quantified and then used for training HMM models, i.e. no higher level feature calculation is done on the gestures. In principle HMMs could be used for continous gesture recogniton but the test set for the VCR control does not take this advantage into account rendering the original classification problem easily solvable by classification methods that require less computing power like Naïve Bayes. The VCR control gesture set is given in Figure 4.

Table 1: Results for first evaluation: Video control gestures.

	MoGLE Naïve Bayes	Wiimote Naïve Bayes	Mäntijärvi et al. HMM
Result	95.8%	99.6%	97.2%

In [18], different training procedures have been tested in order to increase the recognition rate of the classifier. The best result that was achieved is 97.2% accuracy. This is taken as the benchmark to compare MoGLE against. Gestures were recorded under the same conditions. One user did 30 gestures per class, which were recorded in two sessions. In each session, 15 gestures per class were performed. Recognition rates were calculated by a 14-fold cross-validation. The experiment was replicated for the Wiimote and showed that the faster, computationally less complex Naïve Bayes classifier is sufficient to solve the recognition task for a given user with a recognition rate of 99.6% for the eight class problem. Results are given in Table 1 and show that changing the gesturing device from the Wiimote in [21] to a mobile device and running the classification process on the device

Table 3: Recognition results for the fifteen emblematic gestures.

Gesture	Rec. Rate	Gesture	Rec. Rate
Come Here	0.74	Yummy	0.97
Go Away	0.90	Idiot	0.92
Handshake	0.93	Stupid	0.95
Go On	0.98	Threat	0.98
Unsure	0.95	Me	0.97
Get Up	0.97	No	0.95
Eating	0.95	Time	0.95
Drinking	0.98	**Average**	**0.94**

itself produces comparable results with a recogntion rate of 95.8%.

Having shown that the type of a gesture is reliably recognizable, we aimed next at evaluating the performance of MoGLE for our task of German emblematic gestures. Fifteen emblematic gestures have been selected that are partly derived from the Berlin dictionary of German everyday gestures (Berliner Lexikon der Alltagsgesten, BLAG[2]) and partly based on their usefulness in the selected training scenarios (see Section 3.3). Table 2 gives an overview of the selected gestures along with their index in the BLAG (given in parantheses if applicable) and a short description of their meaning.

Figure 5: Snapshot from three users performing the "Go On" gesture with the mobile phone.

Performing gestures with the mobile phone might differ from a hands-free performance of the same gesture. To get insights into how users handle the device when performing each gesture, data was collected from a focus group of eight persons. Each person was asked to take the mobile phone and perform the gesture several times. Figure 5 gives some snapshots of the recordings for gesture "Go On". The information gathered from these tests was used to create the database of training samples for the classifier. To train the classifier, three trainers provided 10 training samples for each gesture resulting in a database of 450 gestures. Table 3 gives an overview of the results of a 10-fold cross validation on this training database. The mean recognition result for the 15-class problem is 93.8%, which is a reasonable result

[2]http://www.ims.uni-stuttgart.de/projekte/nite/BLAG/ (28 October 2010)

Table 2: German emblems selected for GAME. BLAG index in parantheses if applicable.

Name	Gesture	Description
Come Here	Waving a hand rhythmically towards the body	Signaling a person to come closer
Go Away	Waving a hand rhythmically away from the body	Signaling a person to go away
Handshake	Moving right hand rhythmically up and down	Greeting someone
Go On	Rotating hand in front of body	Signaling a person to come to a conclusion
Unsure	Rotating one's hand back and forth (A23)	Signaling not being sure about a topic
Get Up	Raising upwards-pointing flat hands (A26)	Signaling a person to stand up
Eating	Putting hand to mouth	Asking for/Offering something to eat
Drinking	Drinking from a container (A05)	Asking for/Offering something to drink
Yummy	Rubbing splayed hand in circle across tummy	Signaling that food was good
Idiot	Pointing with index finger to forehead	Reproaching someone for being an idiot
Stupid	Waving a hand in front of onet's eyes (A01)	Reproaching someone for being stupid
Threat	Cutting the throat (A21)	Threatening someone
Me	Pointing with index finger to own chest	Selecting oneself
No	Moving hand horizontally back and forth (A04)	Signaling disagreement
Time	Indicating to one's wrist (A02))	Indicating that time is running out, somebody is late

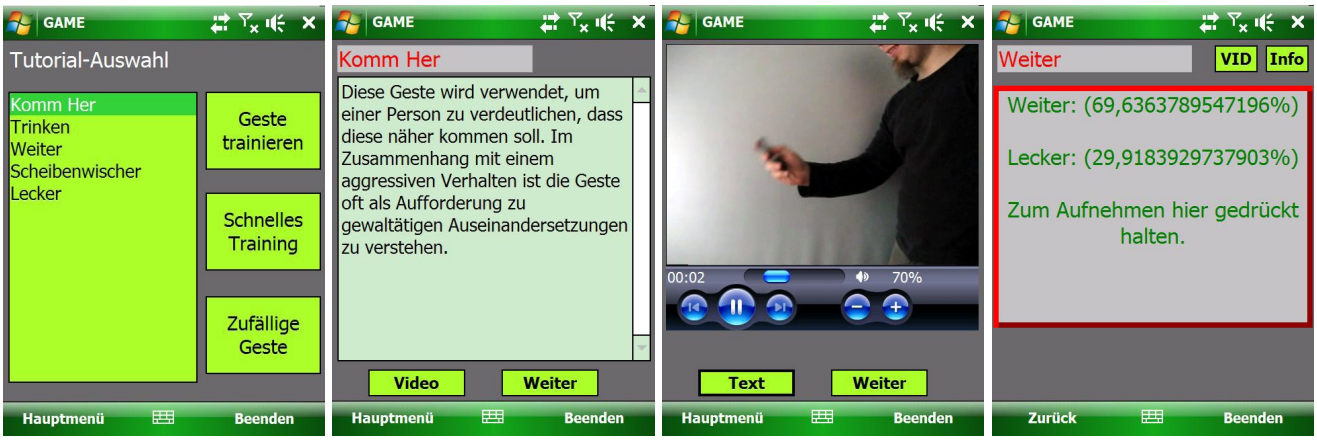

Figure 6: Training sequence for gestures in the "Greeting" scenario: gesture selection, information text, video sequence, gesture execution.

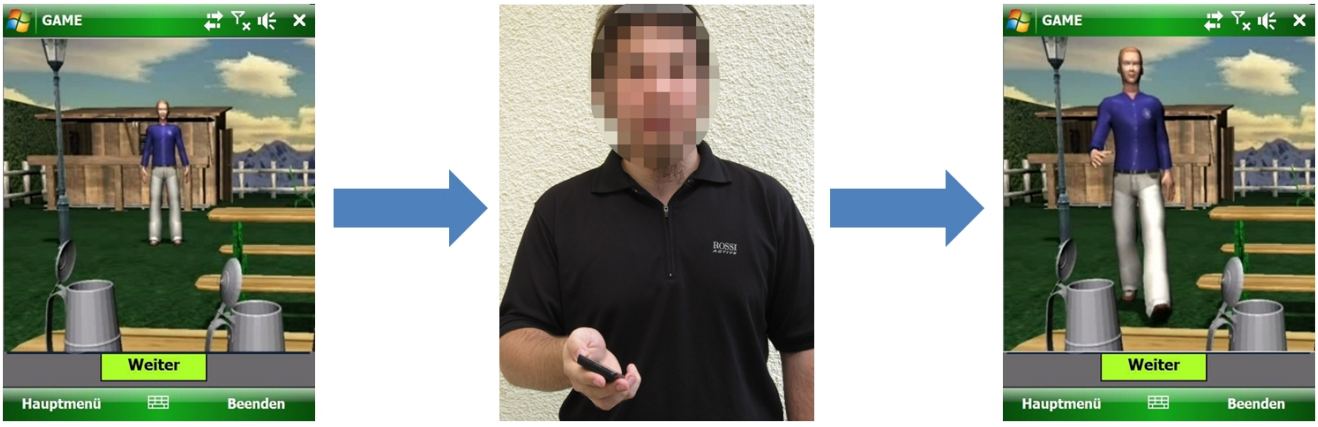

Figure 7: A short game sequence with the user reacting to a waiting agent that moves closer if the gesture is performed correctly.

for employing the classifier in the game and comparable to the results obtained earlier.

3.2 Training Mode

The application consists of two modes, a training mode to concentrate on specific gestures and the game mode for the experience-based role-play. If the user chooses the training mode, he is able to acquire in-depth knowledge about single gestures as well as to practise his skills by doing the gestures. Thus, the training mode allows training and rehearsing in isolation without having to concentrate on the contextual factors for gesture use.

Figure 6 gives an overview of the training cycle. The start screen (Figure 6 left) offers three options to the user: (i) gesture training, (ii) quick training, and (iii) random training. The standard option is (i) – gesture training. By selecting this option, an information text about the gesture is presented next, giving details about the meaning and usage of the gesture (Figure 6 second from left). The user can now choose to directly try out the gesture (Button "Weiter"), or to see a small video of how the gesture is performed (Button "Video"). A snapshot from such a video is given in Figure 6 (second from right). Having seen the video, the user now performs the gesture and gets the feedback on his performance in auditory and textual form. The recognition result is shown in Figure 6 (right). The gesture is performed by pressing on the grey area (e.g. with the thumb) and releasing this press after the gesture has been performed. After each gesture, the recognition results are given in textual form in the grey area and are accompanied by an auditory feedback signal for good, medium, and bad performance. This can be repeated until the user is satified with the result.

If the user chose option (ii) at the beginning (quick training instead of the standard gesture training), he jumps directly to the gesture execution without information on the gesture and how it is performed. If necessary the information text as well as the video can be requested at any time by pressing the "Info" and "VID" buttons respectively (Figure 6 right).

The last option, (iii) random training, allows the user to rehearse what he has trained before by presenting a random gesture from the list of available gestures, which the user has to perform. This mode was integrated for motivational reasons to keep the training session more engaging.

3.3 Game Mode

The game mode realizes the experience-based role-play and is based on standard techniques for intercultural training [17]. Two scenarios have been integrated so far: "The Greeting" and "The Visit". The greeting allows to rehearse greeting rituals in the target culture, whereas the visit represents a less formal interaction during dinner with a family in the target culture. In GAME, both scenarios take place in a beergarden (typical German meeting place) and differ in length and number of gestures that are performed (5 during the greeting, 15 during the visit). Both scenarios are technically realized as interactive narratives. A short video is presented that triggers a reaction of the user in the form of a gesture. Depending on the gesture and its performance a cut scene is played, which in turn leads to another trigger video. To give a short example (Figure 7), the greeting scenario starts with the user entering the beergarden and noticing an agent that is apparently waiting for someone. The user's reaction should now be to either wave hello or signal the agent to come closer. The latter will for instance result in a video showing the agent moving closer to the user.

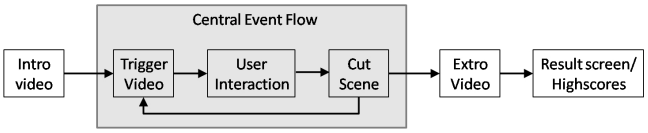

Figure 8: Overview of general game flow with central interaction sequence highlighted.

3.4 Authoring

Figure 2 depicts the possibilities of training gestures and classifiers (Section 3.1) as well as authoring the content of the learning scenarios by expert community members. The "Greeting" scenario will serve as the example for detailing the authoring process. Figure 8 introduces the general game flow with the central interaction loop highlighted and Figure 7 gives one example for the central interaction loop with a trigger video showing an agent waiting in the beergarden (right), the user performing the "Come Here" gesture (middle) resulting in a cut scene, where the agent moves towards the user (right).

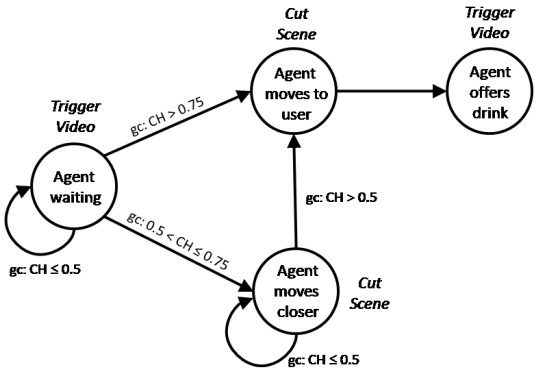

Figure 9: Finite state machine for the example sequence of the "Greeting" scenario (gc: gesture check, CH: "Come Here" gesture).

The game flow can be expressed as a finite state machine with conditional transitions that evaluate the user's performance. Figure 9 gives a detail of the state machine for the "Greeting" scenario that deals with the sequence depicted in Figure 7. Each video from the central event flow constitutes one state, the transitions correpond to the user interactions. In the example, the trigger video is the first state and shows an agent waiting in the beergarden. The user now performs a gesture and depending on his performance one of three successor states is activated. If the performance was really bad, i.e. the "Come Here" gesture was recognized with a probability of less than 0.5, the system remains in the state "Agent waiting". If the performance was good, i.e. recognition probability greater than 0.75, the system moves into the state "Agent moves to user" and the correponding video of this cut scene is played. After that there is an unconditioned transition to the next trigger video that corresponds to state "Agent offers drink". If the user's performance is

less than optimal but still acceptable, i.e. recognition probability between 0.5 and 0.75, the system moves to the state "Agent moves closer" and the corresponding video of this cut scene is played. This cut scene then serves also as the next trigger video, as the user has not yet succeeded in his task. In order to not frustrate the user by repeated failures, the thresholds for the evaluation of the next user gesture are relaxed somewhat in that a recognition probability of over 0.5 will be counted as a success.

The FSM translates into a corresponding XML-structure. Along with the resources needed for the scenario like gestures and video files, the XML-structure specifies the flow of the interaction as well as the conditions for the transitions between states.

Apart from authoring the content of the system, it is possible to localize the interface because the idea is that the system should be used by learners from a variety of other cultures. Localizing the interface is straighforward and takes into account the texts used in the interface. All textual information in the system like button and menu labels as well as instruction texts are fully configurable without resorting to the source code. Labels and texts are read from external files during the startup phase and can be edited with any text editor.

4. EVALUATION

In order to see if the resulting interface and the game play are attractive to users, a first exploratory evaluation was conducted on a public event for the German year of science in 2009 that took place in the city center of Augsburg. For this event, the Department of Computer Science presented a number of interactive demos along with information on the study programmes. During this event, participants were recruited on site.

4.1 Design

20 participants could be won (15 male, 5 female) for the study, which consisted of a training phase followed by a single player role-play with the greeting scenario. Afterwards participants filled out an AttrakDiff questionnaire [8], which is used to measure if the system is perceived as usable, innovative and motivating for the user, which is measured in so-called hedonistic and pragmatic qualities of the system (see Section 4.2.2). Additionally, participants were asked to give their subjective impressions about the input possibilities and the game play.

Thus, three different sources of information are available for the evaluation. (i) Log data: All user actions have been logged during training phase and role-play allowing to analyze the success of gesture executions. (ii) Hedonistic and Pragmatic Quality: By requesting a graded response to adjective pairs like "complex – simple", the AttrakDiff questionnaire results in a rating of the system's usability as well as its ability to engage the user in the interaction. (iii) Subjective Impressions: Participants have been asked to write down their subjective impressions about the game play and the gestural input possibilities.

4.2 Results and Discussion

4.2.1 Log Data

In our first explorative analysis we wanted to find out if users are able to handle the device and successfully play the game by performing gestures and if the training mode has an effect on the gesture performance in the game. For the analysis we divided the users into low performers with mean success rates below 0.5 and high performers with mean success rates above 0.5. The log data revealed that 7 of the 20 participants were low performers. Next, we compared the number of training rounds low and high performers did and saw that the low performers either directly started with the game or did on average less training rounds than the high performers. Figure 10 (left) gives the box plot for this relation. What is apparent from the plot is that users with high success rates had on average more training rounds than users with low success rates. A correlation analysis (Pearson) showed a significant positive correlation (0.509, $p<0.05$) between training and the success rates in the game.

Thus, the log data analysis highlights that although we designed the gestures and classifiers based on user observations, there is still a need for getting acquainted with handling the device to perform conversational gestures. This does not come as a complete surprise as users have never done this before. The amount of trying out the device is not overly large because on average users need 7 training units for 5 gestures to become a high performer, i.e. basically they have to try out each gestures ones. On the other hand this result raises the question if training with the device will carry over to performing the gestures without the device. This is the topic of a follow-up study that is currently run with newly arrived Erasmus students that have no prior knowledge about the German language nor any experience with the German culture.

4.2.2 Hedonistic and Pragmatic Quality

The AttrakDiff questionnaire asked the participants to select a graded response (seven point scale, -3 to +3) to adjective pairs that fall into four different categories. Participants had to rate 28 pairs in all, i.e. 7 pairs for each category.

- Pragmatic Quality (PQ): Describes the usability of the product and clarifies if the user can reach his goals with the system. An example pair for this category is "complex – simple".
- Hedonistic Quality - Stimulation (HQ-S): Describes if the product is stimulating in presenting new, innovative and motivating ways of interaction and content presentation. An example pair for this category is "conservative – innovative".
- Hedonistic Quality - Identity (HQ-I): Describes if the user is drawn into the interaction and can identify with the system. An example pair for this category is "amateurish – professional".
- Attractivity (ATT): Describes a global rating based on perceived quality of the product. An example pair for this category is "discouraging – encouraging".

Figure 10 (middle and right) gives the result of the AttrakDiff analysis. An overview for the hedonistic and pragmatic quality of the system is given in the middle. It shows that users reacted positively towards the system on both dimensions, rating it as attractive to use and self-oriented, which means that the training was perceived as a positive experience for personal development. This result is compatible with the goals we had for the system because it was designed

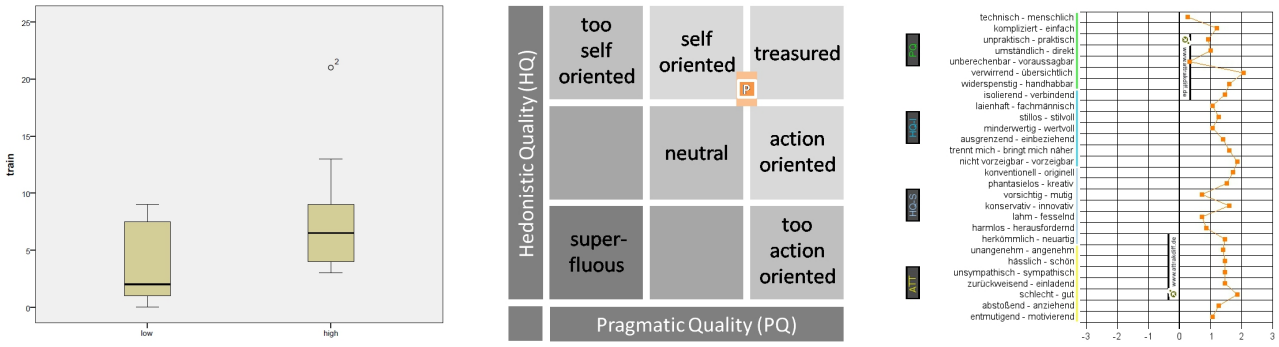

Figure 10: Left: Relation between number of training rounds and success rate. Middle: Overview of AttrakDiff evaluation. Right: Details of AttrakDiff evaluation.

to support the user in his self-directed study of knowledge and skills of nonverbal behavior. The detailed analysis (Figure 10 right) gives the mean ratings of all adjective pairs and corroborates the first impression. For nearly all pairs, the ratings are on the positive side. For two pairs ("technical – human", "unpredicable – predictable") results are rather neutral instead.

Concerning the "unpredictable" vs. "predictable" dimension, we observed that for the low performers it was not always clear why the system did not register their gestures as correct resulting in low ratings for this dimension because for them the system seemed to recognize their gestures on a random basis. A reason for the low score on the dimension "technical" vs. "human" could be that conversational gestures are generally done without technical requisites. Thus, the gestural interaction becomes suddenly mediated by the mobile device, which introduces a technical layer to the interaction. For other gesture types this might not pose a problem, e.g. conducting an orchestra, which is often mediated by a baton. Moreover, the advent of game consoles that make use of acceleration sensing to introduce embodiment into the game play might also have an influcence on this rating when users get more acquainted with gesture recogniton devices.

4.2.3 *Subjective Impressions*

Consistent with the AttrakDiff results, users were quite positive about the interaction possibilities offered by the system and the game play. Two comments recurrently came up that should be considered during the further development. Some of the buttons were perceived as being to small especially if the user did not use a stylus but operated the system solely with his fingers. The second comment concerned the event flow during the training mode. To select a new training gesture, the user always has to go back to the main menu (see Figure 6 left). Several users requested a possibility to change the training gesture directly from the result screen (Figure 6 right), for instance by introducing a next button or a drop-down menu.

4.2.4 *Discussion*

This first evaluation revealed the positive potential of our approach. Participants were able to handle the device and interact with the application successfully by performing gestures. The analysis of the hedonistic and pragmatic qualities showed that the system is perceived as motivating and innovative by the users. A follow-up study is currently evaluating if the experience-based training has an effect on the user apart from being motivating.

5. CONCLUSION

The work presented in this paper is based on the idea of marrying mobile technology with the possibilities of experience-based role-plays. It draws its motivation from two sources. First, virtual characters have been shown to be a successful tool for intelligent tutoring systems. Second, intercultural training is facing a shift towards coaching endeavors that require to deliver training units anytime and anywhere. With GAME we present a first step in this direction. A mobile edutainment platform has been developed that challenges the user with active tasks where he has to put his knowledge and skills about nonverbal behavior to a test in interactions with virtual characters. To this end, the GAME platform offers gesture recognition and authoring possiblities. Scenarios are defined as finite state machines with conditioned transitions between states.

So far, the experience-based role-plays with virtual characters have been brought to the mobile device, freeing the user from desktop based stationary interactions. The aim is to realize a coaching approach that takes the user's context (location, agenda, etc.) into account for suggesting a learning session. Thus, a proactive system is envisioned as the next step that decides on scenarios based on contextual clues like location of the user or the user's agenda and the user's stage of intercultural development (ethnocentric to ethnorelative).

Acknowledgments

The work work described in this paper was funded by the German Research Foundation (DFG) under research grant RE 2619/2-1 (CUBE-G).

6. REFERENCES

[1] Ruth Aylett, Ana Paiva, Natalie Vannini, Sibylle Enz, Elisabeth André, and Lynne Hall. But that was in another country: agents and intercultural empathy. In *Proceedings of 8th International Conferecne on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 329–336, 2009.

[2] Milton J. Bennett. A developmental approach to training for intercultural sensitivity. *International*

Journal for Intercultural Relations, 10(2):179–195, 1986.

[3] Justine Cassell, Timothy Bickmore, Lee Campbell, Hannes Vilhjalmsson, and Hao Yan. Designing embodied conversational agents. In Justine Cassell, Joseph Sullivan, Scott Prevost, and Elisabeth Churchill, editors, *Embodied conversational agents*, pages 29–63. MIT Press, Cambridge, MA, 2000.

[4] Willian James Clancey. A tutorial on situated learning. In *Proceedings of Computers and Education*, 1995.

[5] Birgit Endrass, Matthias Rehm, and Elisabeth André. Culture-specific communication management for virtual agents. In Keith Decker, Jaime Sichman, Carles Sierra, and Cristiano Castelfranchi, editors, *Proceedings of 8th International Conferecne on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 281–288, 2009.

[6] Sandra M. Fowler and Judith M. Blohm. An analysis of methods for intercultural training. In Dan Landis, Janet M. Bennett, and Milton J. Bennett, editors, *Handbook of Intercultural Training*, pages 37–84. Sage Publications Inc., 3rd edition, 2004.

[7] Edward T. Hall. *The Hidden Dimension*. Doubleday, 1966.

[8] Marc Hassenzahl. The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, 13(4):481–499, 2001.

[9] Geert Hofstede. *Cultures and Organisations — Intercultural Cooperation and its Importance for Survival, Software of the Mind*. Profile Books, 1991.

[10] Geert Hofstede. *Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Sage Publications, Thousand Oaks, London, 2001.

[11] Gert J. Hofstede, Paul B. Pedersen, and Geert Hofstede. *Exploring Culture: Exercises, Stories, and Synthetic Cultures*. Intercultural Press, Yarmouth, 2002.

[12] Katherine Isbister. Building Bridges Through the Unspoken: Embodied Agents to Facilitate Intercultural Communication. In Sabine Payr and Robert Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 233–244. Lawrence Erlbaum Associates, London, 2004.

[13] Dusan Jan, David Herrera, Bilyana Martinovski, David Novick, and David Traum. A Computational Model of Culture-Specific Conversational Behavior. In Catherine Pelachaud et al., editors, *Intelligent Virtual Agents (IVA'07)*, pages 45–56, Berlin, Heidelberg, 2007. Springer.

[14] W. Lewis Johnson and A. Valente. Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. In *Proceedings of IAAI*, 2008.

[15] Juha Kela, Panu Korpipaää, Jani Mäntyjärvi, Sanna Kallio, Giuseppe Savino, Luca Jozzo, and Sergio Di Marca. Accelerometer-based gesture control for a design environment. *Pers. Ubiquitous Computing*, 10:285–299, 2006.

[16] H. Chad Lane and Matthew J. Hays. Getting down to business: Teaching cross-cultural social interaction skills in a serious game. In *Workshop on Culturally Aware Tutoring Systems (CATS)*, pages 35–46, 2008.

[17] H. Losche. *Interkulturelle Kommunikation. Sammlung praktischer Spiele und Übungen*. Ziel, 2005.

[18] Jani Mäntyjärvi, Juha Kela, Panu Korpipää, and Sanna Kallio. Enabling fast and effortless customisation in accelerometer based gesture interaction. In *Proceedings of MUM'04*, pages 25–31, 2004.

[19] Aaron Marcus and Valentina-Johanna Baumgartner. A practical set of culture dimensions for global user-interface development. In M. Masoodian et al., editor, *Proceedings of APCHI*, pages 252–261. Springer, 2004.

[20] Aaron Marcus and Emily W. Gould. Crosscurrents: Cultural Dimensions and Global Web-User Interface Design. *ACM Interactions*, 7(4):32–46, 2000.

[21] Matthias Rehm, Nikolaus Bee, and Elisabeth André. Wave like an Egyptian — Acceleration based gesture recognition for culture-specific interactions. In *Proceedings of HCI 2008 Culture, Creativity, Interaction*, pages 13–22, 2008.

[22] Matthias Rehm, Yukiko Nakano, Elisabeth André, Toyoaki Nishida, Nikolaus Bee, Birgit Endrass, Michael Wissner, Afia Akhter Lipi, and Hung-Hsuan Huang. From Observation to Simulation — Generating Culture Specific Behavior for Interactive Systems. *AI & Society*, 24:267–280, 2009.

[23] Maria Sapouna, Dieter Wolke, Natalie Vannini, Scott Watson, Sarah Woods, Wolfgang Schneider, Sibylle Enz, Lynne Hall, Ana Paiva, Elizabeth André, Kerstin Dautenhahn, and Ruth Aylett. Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial. *The Jounal of Child Psychology and Psychiatry*, 2009.

[24] Thomas Schlömer, Benjamin Poppinga, Niels Henze, and Susanne Boll. Gesture recognition with a wii controller. In *Proceedings of Tangible and Embedded Interaction (TEI)*, 2008.

[25] Stella Ting-Toomey. *Communicating Across Cultures*. The Guilford Press, New York, 1999.

[26] Martin Urban, Peter Bajcsy, Rob Kooper, and Jean-Christophe Lementec. Recognition of arm gestures using multiple orientation sensors: Repeatability assessment. In *IEEE Intelligent Transportation Systems Conference*, pages 553–558, 2004.

[27] L. S. Vygotsky. Thinking and speech. In *The collected works of L. S. Vygotsky, vol. 1: Problems of general psychology*, pages 39–285. Plenum Press, New York, 1987.

[28] Rik Warren, David E. Diller, Alice Leung, Willian Ferguson, and Janet L. Sutton. Simulating scenarios for research on culture and cognition using a commercial role-play game. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, editors, *Proceedings of the 2005 Winter Simulation Conference*, 2005.