

Integration of cultural factors into the behavioural models of virtual characters

Birgit Endrass, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Endrass, Birgit, and Elisabeth André. 2014. "Integration of cultural factors into the behavioural models of virtual characters." In *Natural language generation in interactive systems*, edited by Amanda Stent and Srinivas Bangalore, 227–51. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511844492.010>.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



1 Integration of Cultural Factors into the Behavioural Models of Virtual Characters

Birgit Endrass and Elisabeth André

Human Centered Multimedia, Augsburg University

1.1 INTRODUCTION

The design and implementation of **embodied virtual agents** that converse with human users and/or other agents using natural language is an important application area for natural language generation (NLG). While traditional NLG systems focus on information provision (the transformation of content from a knowledge base into natural language for providing information to readers), embodied conversational agents also need to exhibit human-like qualities. These qualities range from the integration of verbal and nonverbal communicative behaviours – such as facial expressions, gestures and speech – to the simulation of social and emotional intelligence. To be believable, embodied conversational agents should not all show the same behaviour. Rather, they should be realised as individuals that portray personality and culture in a convincing manner.

Whereas a number of approaches exist to tailor system behaviour to **personality** (for example, see the approach by (Mairesse & Walker, 2008) for NLG and the approach by (Hartmann et al., 2005) for gesture generation), few researchers have so far taken up the challenge of modelling the influences of **culture** on communicative behaviour. Even when communication partners speak the same language, irritations and misunderstandings can arise due to cultural differences in what people say and how they say it. For example, Germans tend to get straight to business and be rather formal, while casual small talk is more common in the USA. As a consequence, Americans might perceive Germans as rather reserved and distant, and Germans might feel rather uncomfortable about sharing private thoughts with Americans they have just met. Furthermore, cultural misconceptions can arise from differences in nonverbal communication; gestures or body postures that are common in one culture do not necessarily convey the same meaning in another culture. An example is the American *ok* gesture (bringing the thumb and the index finger together to form a circle). While it means *ok* in American culture, it is considered an insult in Italy and is a symbol for money in Japanese culture. The performance of gesture can also vary across cultures; for example, while gesturing expressively is considered a sign of engagement in some cultures, it is regarded as inappropriate in others.

Possible culture-related misunderstandings are sometimes not even recognised. If communication partners tend to take a common basis of cultural knowledge for

granted, they may interpret each other's behaviours in their own culture-specific ways. Behaviours may be decoded wrongly, ignored, or interpreted as deliberate misconduct. Consequently, people might be confronted with rejection without knowing why, which in turn can lead to frustration.

In a similar manner, computer-based systems can be misunderstood: the programmer of a software may use culture-specific indicators, which may be misperceived by the user. This could easily happen with a programmer and a user from different cultural backgrounds. For example, Marcus and Alexander (2007) described different perceptions of a homepage by users with different cultural backgrounds. Regarding virtual agents, these differences can play an even more crucial role. Since virtual agents try to simulate human behaviour in a natural manner, social background should be taken into account. If cultural factors are overlooked in the implementation of communicative behaviours for a virtual agent, this may affect user acceptance.

Iacobelli and Cassell (2007), for example, investigated the perceptions of ethnic differences in verbal and nonverbal behaviours by virtual agents. They changed the behaviours of the virtual agent to match behaviours of different ethnicities, but used a constant, racially-ambiguous appearance for the virtual agent. They tested children's perceptions of the ethnic identity of the virtual agent. Children were able to correctly assign the virtual agent to different ethnicities by behaviour, and they engaged with the virtual agent that behaviourally matched their own ethnicity in a promising way for educational applications¹.

If culture is not explicitly considered in the modelling process of a virtual agent, the character nevertheless will show a certain cultural background – usually that of the programmer since he or she is the one who judges the naturalness of the agent's behaviour. For example, if the programmer is a member of a Western culture where direct eye contact is considered to indicate honesty, the character will probably maintain a lot of eye contact. However, a user from a different cultural background, *e.g.* a member of an Asian culture, might judge this behaviour as impolite.

Integrating culture into the behavioural models of virtual agents is a challenging task. We need to adapt both the content and the form of an agent's utterances to particular cultures. For example, an agent might choose different topics in small talk and use different discourse markers to indicate politeness depending on the culture it represents. It might also vary the amount and quality of gestures depending on its assigned cultural background.

In this chapter, we investigate approaches to the generation of culture-specific behaviours for virtual agents. We first present an overview of culture models from the social sciences; these identify different levels of adaptation we will use as the basis for a computational approach to the generation of culture-specific

¹ This effect can be explained by the **similarity principle** (Byrne, 1971), which states that interaction partners who perceive themselves as being similar are more likely to like each other. This phenomena happens partly unconsciously (Chartrand & Bargh, 1999).

behaviours. Findings from the social sciences are often described in tendencies rather than as concrete rules of behaviour that could be implemented for a computational model, but data collections can augment findings from the research literature and help identify typical behaviour patterns that have not been described in sufficient detail. We introduce a hybrid (theory-based and corpus-driven) approach to integrating social factors into the behavioural models of virtual agents. We exemplify our approach for German and Japanese cultures.

1.2 CULTURE AND COMMUNICATIVE BEHAVIORS

Culture plays a crucial, though often unrecognised, role in the perception and selection of communicative behaviours. But what exactly is culture and how can we distinguish different cultures and their behaviours? Most people instinctively know what the term **culture** means; however, it is hard to formalise and explain what drives people to feel they belong to a certain culture. In this section, we survey different explanations of the notion of culture, including different levels of culture, dichotomies that distinguish cultures and dimensional models of culture.

1.2.1 LEVELS OF CULTURE

We first look at theories that use **layers** to describe the influence of culture on human behaviour. These layers highlight, among other things, that culture does not only determine behavioural differences on the surface but also works on the cognitive level.

Culture can be seen as one social factor that influences a whole group of people. It is, however, hard to formalise the extent to which each of these aspects determines an individual's behaviour. Hofstede (2001) referred to the collection of factors that influence human behaviour as a mental program unique to each person. This so-called “software of the mind” can be categorised into three layers: human nature, culture and personality (see Figure 1.1), where

- **Human nature** represents the universal level in an individual's mental program and contains physical and basic psychological functions.
- **Personality** is the level that is specific to the individual.
- **Culture** is the middle layer. Culture is specific to the group and the environment, ranging from the domestic circle, through the neighbourhood and workplace, to the country.

Enormous efforts have been made to incorporate two of these three layers into virtual agents. Human nature was included through **embodiment** and the simulation of credible verbal and nonverbal behaviours. Virtual characters simulate people's physical natures in more and more sophisticated ways, using natural-seeming speech and nonverbal behaviours such as gestures and body postures. Basic psychological functions have also been integrated into virtual agents; for

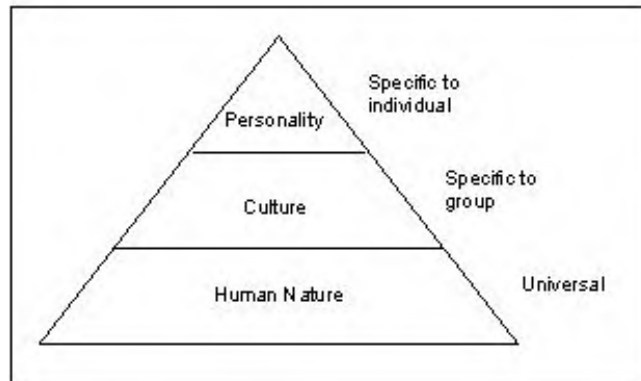


Figure 1.1 Hofstede's levels of uniqueness in an individual's mental program (Hofstede, 2001)

example, implementations of the ability to express **emotions** and act accordingly are described in (Aylett et al., 2005, Gratch et al., 2002). There has also been research on the expression of **personality** in virtual agents, for example, (Rist et al., 2003, Kang et al., 2008). However, culture has come into focus only recently.

Culture itself can also be described in terms of layers. For example, Trompenaars and Hampden-Turner (1997) distinguished between implicit and explicit layers of culture (see Figure 1.2). The explicit layer contains observable realities such as language, clothes, buildings or food. The middle layer consists of norms and values that are reflected in the explicit layer. While norms are related to a group's sense of right and wrong, values are associated to the sense of good and bad. Thus norms determine how people think they *should* behave and values determine the way people *wish* to behave. The innermost layer of culture contains basic assumptions that have vanished from conscious questioning and become self-evident. When basic assumptions are questioned, the result is confusion or even annoyance, as for example when one asks an American or Dutch person why he or she thinks that all people are equal.

In sum, implicit and explicit layers of culture can be distinguished, with explicit layers consisting of things that can be observed in reality and implicit layers containing internal values and basic assumptions. For the simulation of cultural behaviours in virtual agents, the explicit layer is thus of special interest since this layer holds observable differences in verbal and nonverbal behaviour. Implicit layers of culture could also be taken into account, for example when building a cognitive model to control the behaviour of virtual agents.

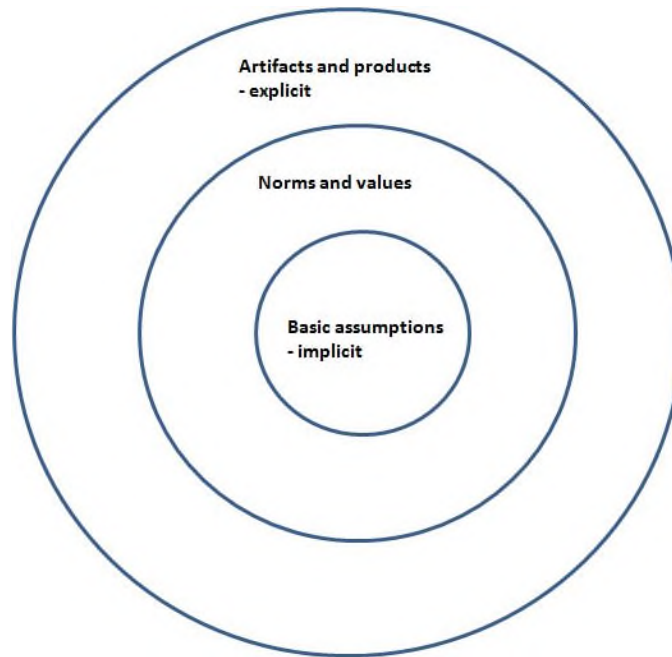


Figure 1.2 Implicit and explicit layers of culture (Trompenaars & Hampden-Turner, 1997)

1.2.2 CULTURAL DICHOTOMIES

Cultural **dichotomies** categorise different cultural groups. Most dichotomies focus on one aspect of culture, for example different perceptions of time, and describe prototypical behaviours for the groups that are being distinguished. Cultural dichotomies seem well suited for computational modelling since concrete distinctions between cultures are clearly stated, and are described in terms of behavioural differences.

Hall (1966), for example, distinguished so-called **high-contact** and **low-contact** cultures, that present behavioural differences in proxemics (the use of space) and haptics (the use of gesture). Ting-Toomey (1999) characterised the varieties of these cultural groups in more detail. According to her, features of high-contact cultures include direct facing, frequent direct eye contact, close interaction and a rather loud voice, whereas features of low-contact cultures include indirect facing, greater interpersonal distance, little or no touching, indirect glances and a soft or moderate voice.

Hall (1966) also distinguished between **high-context** and **low-context** cultures. In high-context cultures little information is explicitly encoded in verbal messages and communication relies heavily on physical context and nonverbal clues. Thus, interlocutors are expected to “read between the lines” in order to decode the whole meaning of a verbal message. In contrast, in low-context cul-

tures the content of most communications is explicitly encoded. The speaker is thus expected to construct clear messages that can be understood easily without reference to context and nonverbal cues. Most Western cultures are low-context cultures whereas most Asian cultures are high-context cultures.

Another dichotomy analysed by Ting-Toomey (1999) is the distinction between **monochronic** and **polychronic** cultures. One behavioural pattern described for monochronic cultures is that members tend to do one thing at a time. Most Western cultures are in the monochronic group. Most Asian cultures belong to the polychronic group; members of these cultures prototypically tend to do several things at a time. Generalising these behavioural patterns, members of Western cultures tend to finish one thing before starting another, while it is more common in Asian cultures to switch back and forth between tasks.

1.2.3 HOFSTEDE'S DIMENSIONAL MODEL AND SYNTHETIC CULTURES

Although culture is often described in terms of abstract behavioural tendencies, there are approaches that define cultures as points in multi-dimensional attribute spaces. These **dimensional models** constitute an excellent starting point for building behaviour models for virtual agents, since they lend themselves to implementation.

An example of a dimensional model for culture is given by Hofstede (2001, 2011). His model is based on a broad empirical survey, covering more than 70 countries (however, only the largest 40 countries have been analysed in detail). The culture of each country is captured as a set of scores along five dimensions², power distance, individualism, masculinity, uncertainty avoidance and long term orientation, where

- **Power distance** captures the extent to which an unequal distribution of power is accepted by the less powerful members of a culture. Scoring high on this dimension indicates a high level of inequality of power and wealth within the society. A low score on the other hand indicates greater equality within social groups, including government, organisations, and families.
- **Individualism** captures the degree to which individuals are integrated into groups. On the individualist end ties between individuals are loose, and everybody is expected to take care of him- or herself. On the collectivist end, people form strong, cohesive in-groups.
- **Masculinity** captures the distribution of roles between the genders. The two extremes are masculine and feminine; masculine values include assertiveness and competitiveness, while feminine values include empathy and negotiation.

² Originally, Hofstede used a four-dimensional model. The fifth dimension, long term orientation, was added later in order to better model Asian cultures. So far this dimension has been applied to 23 countries.

- **Uncertainty avoidance** captures tolerance for uncertainty and ambiguity. The extent to which a member of the culture feels uncomfortable or comfortable in an unknown situation is a key factor of this dimension. Uncertainty avoiding cultures try to minimise the possibility of unknown situations and stick to laws and rules, whereas uncertainty accepting cultures are more tolerant of different opinions.
- **Long term orientation:** long term orientation is associated with thrift and perseverance whereas short term oriented cultures show respect for tradition, fulfilling of social obligations, and saving “face”.

In addition to positioning each culture in this five-dimensional space, Hofstede explains how the dimensional scores of a culture impact the behaviour of its members (Hofstede, 2011).

A so-called **synthetic culture** (Hofstede et al., 2002) is a thought experiment in the characteristics of a hypothetical culture at the extreme of a single cultural dimension. For each synthetic culture Hofstede et al. (2002) defined a profile that contains the culture’s values, core distinction, and key elements as well as words with a positive or negative connotation. The individualistic synthetic culture, for example, has the core value “individual freedom” and the core distinction is the distinction between the self and others. Key elements are statements such as “Honest people speak their mind”, “Laws and rights are the same for all” or “Everyone is supposed to have a personal opinion on any topic”. These key elements are golden rules for appropriate behaviour in this culture and explain the way in which members of this culture would think. Words with a positive connotation in the extreme individualistic culture include self, “do your own thing,” self-respect, dignity, I, me, pleasure, adventure, guilt, and privacy. Words with a negative connotation include harmony, obligation, sacrifice, tradition, decency, honour, loyalty and shame. For the collectivistic synthetic culture the connotations of these words would be the other way round. Hofstede et al. (2002) also defined stereotypical behaviours for members of synthetic cultures. Extreme individualists for example are described as verbal, self-centred, defensive, tending to be loners and running from one appointment to the next.

Although synthetic cultures are a valuable tool, no existing culture is exclusively influenced by one dimension. For example, the U.S. American culture scores high on the individualist dimension, but also scores high on the masculinity dimension. A combination of these two dimensions explains the culture better than either in isolation.

1.3 LEVELS OF CULTURAL ADAPTATION

Culture determines not only *what* we communicate, but also *how* we communicate it. In addition, different aspects of human communicative behaviour are influenced by people’s cultural background, including verbal communications,

posture, eye gaze and gesture. Culturally adaptable embodied virtual agents should consequently include culture in models for every aspect of communicative behaviour. In this Section, we review computational approaches to culture-specific adaptation of the behaviour of virtual agents in terms of both the content and form of communication, and of communication management.

1.3.1 CULTURE-SPECIFIC ADAPTATION OF CONTENT

Language, as the main medium of human communication, is the most obvious barrier to intercultural communication. However, even for speakers of the same language, verbal behaviour can vary vastly across cultures. For example, members of different cultures will tend to choose culturally relevant topics and use culture-specific scripts when discussing those topics.

Isbister et al. (2000) describe a culturally adaptable conversational agent for small talk conversations. The agent, a so-called “Helper Agent”, in the appearance of a dog, virtually joins users of a chat room. The agent plays the role of party host, interacting with the human interlocutors when their conversation stagnates by introducing prototypical small talk topics. The agent distinguishes between **safe** topics (*e.g.* weather, music) and **unsafe** topics (*e.g.* religion, money). According to (Isbister et al., 2000), the categorisation of a topic as safe or unsafe varies by culture. However, their agent does not consider the cultural background of the human interlocutors when introducing conversational topics.

Yin et al. (2010) present conversational agents that tailor their conversational scripts based on the assigned cultural background. Two different virtual agents were designed, one representing a member of an Anglo-American culture and the other resembling a member of a Hispanic culture. The cultural background of each virtual agent was expressed in their physical appearance and physical context (the appearance of the flat in the background of the agent). In addition, each agent used different conversational scripts. While the Anglo-American agent focuses on the interlocutor’s well-being, the Hispanic agent shows interest in the interlocutor’s family and friends. Thus the content of conversations varies with simulated cultural background.

Another example of a culturally adaptable conversational agent is the Tactical Language Training System (TLTS) presented in (Johnson et al., 2004). This system is designed to help soldiers acquire cross-cultural skills through interaction in a game-like virtual environment while learning a foreign language. The soldier can interact with the system using speech, and through the selection of culture-specific gestures. This system is interesting because the content of both verbal and nonverbal communicative behaviours are culturally specific.

1.3.2 CULTURE-SPECIFIC ADAPTATION OF FORM

Another important aspect of culture-specific communication is the form and performance of conversational behaviours.

GENERATION OF CULTURE-SPECIFIC WORDING

Personality has been considered as a parameter of linguistic variation (see, for example, (Mairesse & Walker, 2008)), but only a small amount of research has considered culture, and that mainly for **politeness behaviours**. Most noteworthy is the work of House, who has performed a series of contrastive German-English discourse analyses over the past twenty years (*e.g.* House (2006)). Among other things, she observed that Germans tend to be more direct, and more self-referential, and resort less frequently to using verbal routines such as “how-are-you’s” (and consequently tend to interpret such phrases literally).

Alexandris & Fotinea (2004) investigated the role of discourse particles as politeness markers to inform the design of a Greek speech technology application for the tourist domain. They performed a study in which evaluators had to rank different variations of written dialogues according to their perceived degree of naturalness and acceptability. The study revealed that dialogues in Modern Greek with discourse particles emphasising approval of the user are perceived as friendlier and more natural, while dialogues without any discourse particles, or containing discourse particles that perform other functions, were perceived as unnatural. The authors regard these findings as culture-specific elements of the Greek language.

Johnson et al. (2005) investigated the potential benefits of politeness in a tutoring system. By examining interactions between a real tutor and his students, they identified a set of politeness behaviour templates, which they implemented in their system. Their system automatically selects different tutorial strategies, which use different politeness behaviours, depending on the type of expected face threat (Brown & Levinson, 1987) to the student.

Wu & Miller (2010) present another example of a tutoring system that incorporates culture-specific politeness behaviours. In their system, a user can interact with the virtual agent by selecting phrases and/or gestures from an interactive phrasebook. The system then calculates the appropriateness of the selected action(s) from the perspective of a member of a Middle Eastern culture based on the social relationship between the interlocutors, and categorises the action(s) as polite, nominal, or rude.

GENERATION OF CULTURE-SPECIFIC GESTURES AND POSTURES

An interesting aspect of culture-specific behaviour is the **expressivity** of non-verbal behaviours. How someone performs a gesture can sometimes be as crucial for the observer’s perception as the gesture itself. For example, a hand gesture may be small (involving only the fingers, hand and wrist) and performed near the torso; or it may be large (involving the whole arm) and performed near the face. These types of difference in dynamic variation can be described according to expressivity parameters, including spatial extent, speed, power, fluidity, repetitivity, and overall activation (see (Pelachaud, 2005)). The spatial extent parameter, for example, describes the arm’s extent toward the torso. The fluidity param-

eter captures the degree of continuity between consecutive gestures, while the repetitivity parameter holds information about the repetition of the stroke. The overall activation parameter captures the number of gestures performed.

Expressivity is a function of both individual and social factors, including personality, emotional state and culture. Lipi et al. (2009) present a cross-culture corpus study of the expressivity of body postures using Hofstede's model of cultural dimensions. Each body posture was analysed in terms of several behavioural parameters, including spatial extent, rigidity, mirroring, frequency and duration. Then, a Bayesian network model was trained on this data and used to predict nonverbal expressions appropriate for specific cultures.

1.3.3 CULTURE-SPECIFIC COMMUNICATION MANAGEMENT

When people communicate, they do not typically need to think about the management of their conversation; **turn taking** and **grounding** behaviours are produced automatically. However, these communication management behaviours are also realised in culturally specific ways. According to Ting-Toomey (1999), several types of culture-specific **regulators**, including vocalics, kinesics and oculesics, control the flow of a conversation. Vocalics are verbal feedback signals, such as the English *uh-huh*, silences, and interruptions. Kinesics include hand gestures and body postures, while oculesics include eye gaze and face orientations. Regulators are learned at a very young age and are thus used at a very low level of awareness. Ting-Toomey (1999) states that the indiscriminate use of regulators in cross-cultural communication often causes misunderstandings and frustration. However, people from contrasting cultures may not be able to name the reason for their distress since regulators are used subconsciously.

There are a number of computational approaches to modelling verbal regulators for turn taking, *e.g.* (Jonsdottir et al., 2008, Sato et al., 2002, Sidner, 2004). However, hardly any effort has been made so far to integrate culture as a parameter in these models. An exception is the work of Endrass et al. (2009, 2010) who focus on culture-specific aspects of communication management for dialogues with culturally adaptable virtual agents.

1.4 APPROACHES TO CULTURE-SPECIFIC MODELLING FOR EMBODIED VIRTUAL AGENTS

In principle, there are two types of approach that might be taken to integrate aspects of culture into the behavioural models of virtual agents: **bottom-up** and **top-down**. The top-down approach is model-driven: descriptions of culture-related differences in behaviour are extracted from the research literature and transformed into computational models that can be incorporated into the virtual agents' behaviour repertoires. By contrast, the bottom-up approach is data-driven: corpora of human communicative behaviours are analysed to identify



Figure 1.3 User, represented by an avatar, interacting with a member of a different culture in the Tactical Language Training System (from (Johnson et al., 2004); used with permission)

culture-specific behavioural tendencies, and these are then integrated into the behavioural models of the virtual agents.

1.4.1 TOP-DOWN APPROACHES

The most well-known system that aims at simulating culture-specific behaviours in virtual agents is the Tactical Language Training System (TLTS) (Johnson et al., 2004, Johnson & Valente, 2008). The TLTS is used to teach functional verbal communication skills in foreign languages and cultures, including Iraqi, Dari, Pashto, and French. The TLTS virtual environment is a virtual village, where users can gain communicative skills by interacting with virtual agents to perform everyday activities, such as asking for directions or buying food. The user's goal is to learn how to communicate in culturally appropriate ways. The user is represented by an avatar in the virtual world; the user speaks for the avatar and can choose gestures by selecting them from a menu. The virtual agents represent different cultural backgrounds, and incorporate culture-specific communication behaviours that are derived from the research literature; that is, the TLTS uses a top-down approach to simulating culture-specific communication. Figure 1.3 shows a screenshot of the TLTS in an Iraqi environment.

Mascarenhas et al. (2009) present another top-down approach to simulating culture-specific communication. Instead of simulating existing national cultures, they simulate Hofstede's synthetic cultures (Hofstede et al., 2002). This work focuses on culture-specific **rituals**, symbolic social activities carried out in a predetermined fashion. Groups of virtual agents representing different synthetic cultures perform rituals in a culturally inflected manner. Figure 1.4 shows the virtual agents in a dinner ritual. Characters in the left picture, representing a low power culture, rush to the table, while characters in the right picture wait for the elder to sit first, as they represent a high power culture. In an evaluation, participants were asked to observe the agents' performances of culture-specific

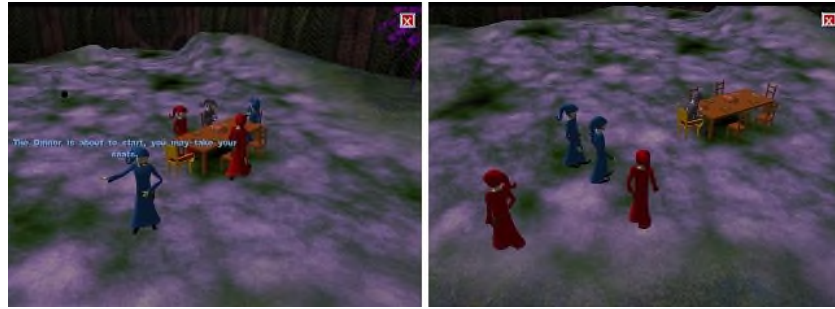


Figure 1.4 Group of virtual agents interacting in different culture-specific rituals (from (Mascarenhas et al., 2009); used with permission)

rituals and then categorise each group of agents using a set of adjectives. Participants identified significant behavioural differences between the groups of agents, and were able to relate these differences to culture.

1.4.2 BOTTOM-UP APPROACHES

Abstract theoretical models of culture-specific behaviours are useful for understanding cultural differences and implementing heuristics for virtual agents. However, more accurate and natural simulations of culture-specific behaviours may be obtained using data-driven approaches. Kipp et al. (2007a,b) describe a data-driven approach to simulating nonverbal behaviours that are performed by a virtual agent to resemble a certain speaker. The approach requires an annotated **multimodal corpus**, from which gestural profiles and animation lexicons are extracted for each speaker. Conversational gestures are synthesised for the virtual agent at runtime by an animation engine for any given input text in the style of the particular speaker. Kipp et al. illustrated their approach using data from two talk show hosts, Jay Leno and Marcel Reich-Ranicki. In an evaluation, human observers were able to correctly assign the virtual agent's behaviour to the speaker whose data was used to train the model.

Martin et al. (2005) use a similar data-driven approach to simulate emotional behaviours. Video recordings of human speakers are manually annotated to identify multimodal emotional behaviours and capture their expressivity parameters. The annotations for each individual human speaker form a behavioural profile, which is used by an animation engine to resynthesise emotional behaviours for performance by a virtual agent. Figure 1.5 left shows a frame from a video recording where a woman talks about a recent trial in which her father was kept in jail. Figure 1.5 right shows the corresponding virtual agent, displaying anger fading into despair. Martin et al. (2005) state that they focus on modelling visible behaviours, rather than internal models that describe the motivations for those behaviours.



Figure 1.5 Frame from video corpus displaying the emotions anger and despair (left) and simulation with a virtual agent (right), (from (Martin et al., 2005); used with permission)

1.5 A HYBRID APPROACH TO INTEGRATING CULTURE-SPECIFIC BEHAVIOURS INTO VIRTUAL AGENTS

In this section, we present a **hybrid approach** to simulating culture-specific communication behaviours. Like (Martin et al., 2005) and (Kipp et al., 2007b), we aim to model visible behaviours – the explicit layer of culture, rather than the implicit layers. However, we focus on behaviours that are generalizable across a culture, rather than individual behavioural differences. We use an iterative, top-down then bottom-up, approach. First, we explore the research literature to find descriptions of stereotypical behaviours for different cultures. Then, we validate and ground these descriptions in empirical data through annotation of multimodal corpora. Finally, we extract computational rules from the annotated data. To summarise this combination of top-down and bottom-up approaches, the social sciences tells us *what* aspects of communication behaviours might be relevant to culture-specific models, and the empirical data tells us *how* differences in these aspects manifest themselves. We exemplify our approach for German (Western) and Japanese (Asian) cultures, since the differences between Western and Asian cultures are supposed to be large.

1.5.1 IDENTIFYING CULTURE-SPECIFIC COMMUNICATION BEHAVIOURS

1.5.1.1 CULTURAL PROFILES FOR GERMANY AND JAPAN

The two cultures of Germany and Japan are very different. Table 1.1 summarises some of these differences as identified in the research literature. Germany is a Western culture, while Japan is an Asian culture. In terms of cultural dichotomies (see Section 1.2.2) the following distinctions can be made:

- Germany is a mediate-contact culture while Japan is a low-contact culture (Ting-Toomey, 1999). This means that standing close to or touching the con-

Classification	Germany	Japan
regional	Western	Asian
contact dichotomy	mediate-contact	low-contact
context dichotomy	low-context	high-context
time dichotomy	monochronic	polychronic
power distance	low power distance	high power distance
individualism	individualistic	collectivistic
masculinity	masculine	masculine
uncertainty avoidance	avoiding	avoiding
long term orientation	long term	short term

Table 1.1. Summary of culture profiles for Germany and Japan

versation partner should not be very common in either culture, but would be more acceptable in Germany than in Japan.

- Germany is a very low-context culture, while Japan is a very high-context culture (Ting-Toomey, 1999). This means that German speakers will tend to be very direct and explicit, while Japanese speakers will rely more on context and nonverbal behaviours.
- Germany is a monochronic culture, while Asian cultures, including that of Japan, are polychronic (Hall & Hall, 1987). This means that in Germany clock time is important, and tasks are solved sequentially. In Japan, on the other hand, the notion of time is more relative, and people may be involved in several tasks simultaneously.

Table 1.1 also summarises Hofstede’s five-dimension analysis of German and Japanese cultures. Germany is a low power distance (egalitarian), individualistic and long term oriented culture, while Japan is a high power distance, collectivistic and short term oriented culture.

1.5.1.2 BEHAVIORAL EXPECTATIONS FOR GERMANY AND JAPAN

The top-down categorisation of German and Japanese cultures from the research literature can be used as a basis to state expectations for behavioural differences. Following the discussion in Section 1.3, we will list expectations for cultural-specific adaptations to: (1) communication content, (2) the form of communication behaviours, and (3) communication management. We do not attempt to cover every culture-specific variation in these three aspects, but focus on one or two adaptations for each one that are likely to highlight the contrasts between German and Japanese cultures. Table 1.2 summarises our expectations for behavioural differences between German and Japanese cultures based on their behavioural profiles.

Behavioural Aspect	Germany	Japan
topic selection	more private	less private
pauses	avoided	consciously used
overlaps	uncommon	common during feedback
nonverbal expressivity	more expressive	less expressive

Table 1.2. Summary of behavioural expectations for Germany and Japan

COMMUNICATION CONTENT

For communication content, we focus on casual small talk such as one might make when first meeting someone. We chose this domain, since meeting someone for the first time is a fundamental interaction in everyday communications, occurring in every culture as well as in cross-cultural encounters – so much so, that it is the topic of the first chapter of most language learning books. Small talk is often thought of as neutral, non-task-oriented conversation about safe topics, where no specific goals need to be achieved. But small talk serves important social purposes. Schneider (1988) categorises topics that might occur in small talk conversations into three groups, with the choice of topic depending on the social context:

- The **immediate situation** group holds topics that are elements of the so-called *frame* of the situation. In order to explain the idea of a frame, Schneider (1988) uses a small talk situation that takes place at a party. Possible topics within a party frame could be the atmosphere, drinks, music, participants or food.
- The **external situation** or “supersituation” group holds topics in the larger context of the immediate situation. This category is the least limited of the three. Topics within this category could be the latest news, politics, sports, movies or celebrities.
- The **communication situation** group holds topics pertaining to the conversation participants themselves. Topics in this category could include hobbies, family and career.

However, Schneider (1988) only considered Western cultures in his studies and did not look at topic selection in other cultures. Thus, these groups do not necessarily hold true for other cultures. According to (Isbister et al., 2000), the categorisation of small talk topics as **safe** or **unsafe** varies with cultural background. If the distinction into safe and unsafe topics varies with culture, we expect that the overall choice of topic categories is also dependent on culture. For example, in Table 1.1, Germany was categorised as a low-context culture, while Japan was categorised as a high-context culture. Ting-Toomey (1999) describes people belonging to high-context cultures as having a lower “public self” than people belonging to low-context cultures. A typical behavioural pattern for members of high-context cultures is not to reveal too much personal information during a first-time meeting. Consequently, we might expect topics from the communi-

cation situation group to be more common in German small talk conversations than in Japanese small talk conversations.

COMMUNICATION FORM

The degree of **expressivity** of nonverbal behaviours is known to vary by culture (see Subsection 1.3.2). In particular, the individualism dimension in Hofstede's model is related to the expression of emotions and acceptable emotional displays in a culture (Hofstede et al., 2010). For communication form, we focus on the expression of emotions. In individualistic cultures such as Germany, it is more acceptable to publicly display emotions than in collectivistic cultures such as Japan (Ekman, 2001). This suggests that nonverbal behaviours may be performed more expressively in German conversations than in Japanese ones. We expect the display of emotions by German speakers to more obviously affect the expressivity of gestures in a way that increases parameters such as speed, power or spatial extent. This should also be the case for body postures as the whole body can be used to express emotions.

COMMUNICATION MANAGEMENT

In Subsection 1.3.3, we discussed **regulators** (such as verbal feedback, silences, gestures, posture and eye gaze) are used to manage the flow of a conversation. For communication management, we focus on the use of **silence** and on **overlapping speech**. As summarized in Table 1.1, Japan is a collectivistic culture. In these cultures silence can occur in conversations without creating tension. In addition, pauses can be a crucial feature of conversations in collectivistic cultures. These observations do not hold true for individualistic cultures such as Germany's (Hofstede et al., 2010). Rather, in German culture silences are likely to create tension and thus pauses in speech are avoided if possible.

The power dimension in the model of Hofstede et al. (2010) may also influence the flow of communication. High power distance cultures are described as verbal, soft-spoken and polite, and in these cultures interpersonal synchrony is much more important than in low power distance cultures, whose members tend to talk freely in any social context (Ting-Toomey, 1999). One way to achieve interpersonal synchrony in a conversation is by giving verbal feedback. This **feedback** often occurs while the conversational partner has the floor, creating overlapping speech segments. We expect this to occur more often in the Japanese culture due to their higher score on the power distance dimension.

1.5.2 FORMALISATION OF CULTURE-SPECIFIC BEHAVIOURAL DIFFERENCES

As we discussed previously, culture-specific behavioural tendencies identified from the research literature can be hard to formalise for use in computational models. The hypothesis that there should be more pauses in Japanese conversations than in German ones, for example, is not quantified: it does not indicate

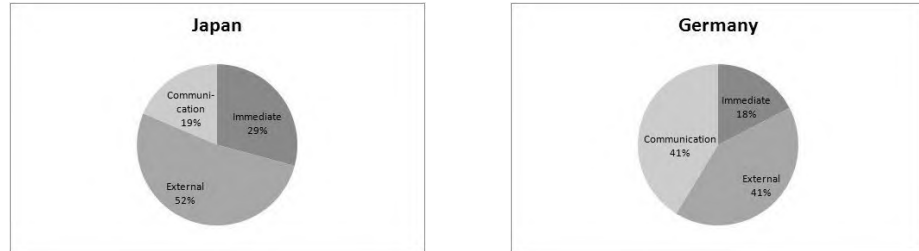


Figure 1.6 Average distribution of topic categories during small talk conversations recorded in Japan and Germany (Endrass et al., 2011b)

about how many pauses we should include in the virtual agent’s behaviour or how long these pauses should last. To get a deeper insight into these issues and to obtain some quantitative data, we analysed the video corpus recorded for the Cube-G project (Rehm et al., 2007). This corpus was recorded in Germany and Japan and includes three prototypical interaction scenarios (a first-time meeting, a negotiation and a conversation with status differences). Altogether, more than twenty subjects from each of the two cultures participated, and about twenty hours of video material were collected. Participants interacted with actors to ensure that they had not met before and that all scenarios lasted for about the same time. For the first-time meeting, participants were asked to get acquainted with their conversational partner in order to be better able to solve later tasks together. The analysis described in this section focuses on this first-time meeting scenario, which lasted for around five minutes for each subject. The analysis includes all the German videos and about half of the Japanese videos due to translation issues.

The video corpus was annotated using the Anvil tool (Kipp, 2001) for the behavioural tendencies summarised in Table 1.2. In particular:

- Topics that occurred in the conversations were categorised as pertaining to immediate, external or communication situations, as described in Section 1.5.1.
- Expressivity was annotated for every gesture, taking into account the parameters power, speed, spatial extent, repetition and fluidity (see Section 1.3.2). We also labeled body postures.
- Pauses and overlaps were calculated as segments on the time line where either neither of the participants spoke, or both participants spoke at the same time.

ANALYSIS OF TOPIC SELECTION

We compared the frequencies of the three topic categories within first-time meeting conversations involving German and Japanese participants. Topics pertaining to immediate and external communication situations occurred significantly more often in the Japanese conversations than in the German ones ($p = 0.014$ for immediate situations and $p = 0.036$ for external situations), while topics cov-

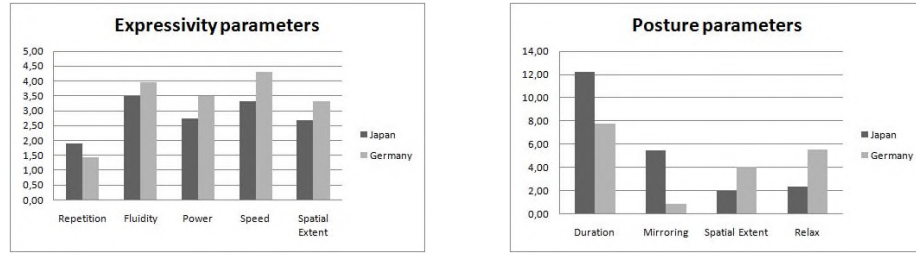


Figure 1.7 Ratings of expressivity (left) and posture parameters (right) averaged over participants (Endrass et al., 2011a)

ering the communication situation occurred significantly more often in the German conversations ($p = 0.035$) (Endrass et al., 2011b). We also calculated the average percentage distribution of topic categories in the German and Japanese conversations, shown in Figure 1.6. This distribution is in line with our top-down expectations and provides a possible target topic distribution for a computational model of culture-specific topic choice.

ANALYSIS OF NONVERBAL EXPRESSIVITY

Using a seven-point scale for each parameter, we annotated the expressivity of nonverbal behaviours in each conversation for each of the parameters power, speed, spatial extent, repetition and fluidity. We found significant differences between the two cultures for all parameters (ANOVA, $p < 0.01$). Figure 1.7 (left) shows the average ratings of the expressivity parameters for German and Japanese conversational participants. Gestures were performed faster and more powerfully in the German conversations than in the Japanese ones, and German participants used more space for their gestures compared to Japanese participants. Gestures were also performed more fluently in the German conversations, and the stroke of a gesture was repeated more in the Japanese conversations (Endrass et al., 2011a).

We used the coding scheme described in (Bull, 1987) to label the type/shape of body postures. Postures most frequently observed in the German conversations were folding arms in front of the trunk and putting hands in the pockets of trousers, while the most frequent postures in the Japanese conversations were joining both hands in front of the body and putting hands behind the back. It is notable that ratings for postures frequently observed in the German conversations were rated higher in spatial extent and lower in rigidity (on a seven-point scale), compared to postures frequently observed in the Japanese conversations. In addition, our analysis indicates that the Japanese conversational participants remained in the same posture longer, engaged in more frequent mirroring, took up less space, and displayed a more rigid posture in comparison to the German participants (see Figure 1.7 (right)).

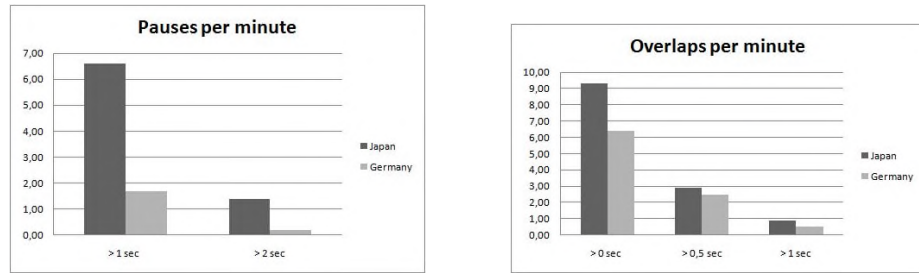


Figure 1.8 Pauses (left) and overlaps in speech (right) per minute, averaged over participants (Endrass et al., 2011a)

ANALYSIS OF PAUSES AND OVERLAPS

We counted pauses in conversations that lasted for more than one second (short pauses) and pauses that lasted for more than two seconds (longer pauses); we discarded very brief pauses, such as those used for breathing. We found more pauses in conversations involving Japanese participants than in those involving German participants; these differences were significant both for short pauses (two-tailed t-test, $p < 0.001$) and for longer pauses ($p < 0.001$). In conversations involving German participants, there were on average 7.1 short pauses and 1.3 longer pauses. By contrast, in conversations with Japanese participants, there were on average 31 short pauses and 8.4 long pauses. Figure 1.8 (left) shows the distribution of short and longer pauses that were found on average per minute in each video.

We counted, for each conversation, the total amount of overlapping speech, as well as the number of overlaps lasting for more than 0.5 seconds (short overlaps) and the number lasting for more than 1 second (longer overlaps). In line with our predictions, there was significantly more overlapping speech in Japanese conversations ($p = 0.04$); however, the differences in numbers of short overlaps and numbers of long overlaps were not significant. The average occurrences of overlapping speech per subject per minute for the two cultures are shown in Figure 1.8 (right). We observed an average of 6 overlaps per minute in conversations with German participants, and of 9 overlaps per minute in conversations with Japanese participants (Endrass et al., 2011a).

1.5.3 COMPUTATIONAL MODELS FOR CULTURE-SPECIFIC CONVERSATIONAL BEHAVIOURS

For the generation of culture-specific dialogues, a distributed planning system was implemented based on the SHOP planner (University of Maryland, 2005). This distributed approach allows us to implement autonomous agents that can act proactively on their internal goals, as well as react to the ongoing conversation. Dialogue planning consists of triggering appropriate speech acts given

the dialogue history, the agent's goals and other agent-specific features, here including cultural features.

COMPUTATIONAL MODEL FOR TOPIC SELECTION

In each conversational agent's knowledge base a personal motivation is provided for each of available topic. This motivation represents the agent's internal drive to talk about the topic, *e.g.* because of an increased personal interest in the topic or a general desire to talk a lot. In our analysis of human-human conversations, we found that Japanese participants prefer topics from the immediate and external situation, while German participants will also select topics from the communication situation (see Figure 1.6). To implement a culture-specific computational model for topic selection, we also set cultural thresholds for each topic based on the topic category and the selected culture for each agent. So, a more "Japanese" agent might have a high personal motivation to talk about children, but a high cultural threshold against discussing communication situation topics, while a more "German" agent, similarly motivated to discuss children, would have a lower cultural threshold for the topic. In order for an agent to initiate discussion of a topic, its personal motivation would have to exceed its cultural threshold.

COMPUTATIONAL MODEL FOR NONVERBAL EXPRESSIVITY

We used a Bayesian network model for culture-specific nonverbal expressivity. Using this approach, culture as a nondeterministic concept can be modelled in an intuitive manner and without giving up a certain amount of variability that is necessary to ensure that an agent is perceived as an individual. In the network, the causality of cultural background and corresponding most probable behaviour is modelled according to Hofstede's dimensional model (see Section 1.2.3). The Bayesian network was implemented using the GeNie modelling environment (Decision Systems Laboratory, 2007), and incorporates nodes reflecting the culture, the dimensions of the culture, and the aspects of gestural expressivity that are influenced by each dimension. Each gesture is divided into three phases: preparation, stroke and retraction. The preparation and retraction phases are used for animation blending, while in the stroke phase, the actual gesture is performed. The stroke phase of a gesture is customised for different parameters and degrees of gestural expressivity. For example, the repetition parameter can be varied by playing the stroke phase several times, while the speed parameter controls the speed of performance of each repetition. To vary spatial extent, the stroke phase is blended with a neutral hand position close to the trunk of the virtual agent.

In some cases, culture-specific behaviours cannot be generated by customising culture-neutral behaviours because culture is reflected by specific (usually emblematic) gestures and postures that need to be accurately executed in order not to be misunderstood. To account for this fact, the library of nonverbal behaviours was extended with behaviours extracted directly from the corpus.



Figure 1.9 Prototypical male (left) and female characters (right) resembling members of German and Japanese cultures

For example, a prototypical posture for a German agent would be putting their hands in the pockets of their trousers, or folding their arms in front of the trunk, while a prototypical posture for a Japanese agent would be folding both hands in front of the body.

1.5.4 SIMULATION

The environment for our virtual agents is the Virtual Beergarden scenario from <http://mm-werkstatt.informatik.uni-augsburg.de/projects/AAA>. We designed culture-specific virtual agents that had a prototypical German or Japanese appearance, and spoke with a German or Japanese synthesised voice. Figure 1.9 shows examples of prototypical characters in the Virtual Beergarden scenario.

Each virtual agent has a dialogue planner, and access to a verbal knowledge base containing small talk sentences relating to different topics and speech acts that can be triggered by the planner. Each agent also has access to over forty different animations of nonverbal behaviour, including gestures and body postures. Some behaviours are culture-specific (e.g. a bow for a Japanese greeting, a hand wave for a German greeting). Other behaviours can be exhibited by an agent of any cultural background, but are realised in a culture-specific way using our Bayesian model for nonverbal expressivity. For behavioural examples of our virtual agents, please see our website at <http://www.hcm-lab.de/projects/animations/>.

1.5.5 EVALUATION

We conducted two evaluations of the perception of our virtual agents, one focusing on verbal behaviour (Endrass et al., 2011b) and the other on communication management and nonverbal behaviour (Endrass et al., 2011a). We did this to reduce dependencies between the different types of behaviour that might influence participants. The evaluation of verbal behaviour did not include variations

in nonverbal behaviours such as gestures, while in the evaluation of nonverbal behaviour the agents spoke gibberish. In each evaluation, we asked participants to observe and rate generated dialogues. In the studies, the virtual agents' appearance was left constant to resemble the observers' cultural background and changed only the culture-specific behavioral model. Thus, German participants observed Western-looking characters and Japanese participants watched Asian-looking characters with varying behavior.

In line with previous work (*e.g.* (Byrne, 1971)), we hypothesised that our participants would prefer agents whose behaviour reflected their own cultural background. That is, we expected German participants to prefer German-acting agents, while we expected Japanese participants to prefer Japanese-acting agents.

First we look at the results of the verbal behaviour study. German participants found the German version of the small talk dialogues significantly more appropriate and interesting than the Japanese version, thought they would quite like to join the conversation, and thought that the agents got along with each other better. Japanese participants found the Japanese version of the small talk dialogues significantly more appropriate and interesting than the German version, and thought that the agents got along with each other better.

In the second study, we measured differences in the usage of pauses, in instances of overlapping speech, in gestural speed, in the spatial extent of gestures and in postures. Each aspect was tested in isolation. For German participants, we found statistically significant preferences for the German version of the dialogues for overlapping speech and spatial extent of gestures. For all other aspects, we observed trends towards a preference for the more German behaviours, although none were statistically significant. For Japanese participants, we found statistically significant preferences for the Japanese version of the dialogues for postures. For most other aspects, we observed trends towards a preference for the more Japanese behaviours. However, for pauses in speech and overlapping speech we observed trends towards a preference for the less Japanese behaviours. One reason for this outcome might be that the dialogues in this evaluation involved gibberish rather than actual speech. Since the Japanese version of the dialogues contained both more pauses and more overlaps in speech, but lacked linguistic content, participants might have chosen the "safe" solution.

In sum, our evaluation studies suggest that human observers of different cultural backgrounds seem to prefer agent behaviours designed to reflect their own cultural background in most cases. We conclude that the integration of culture-specific features into the behavioural models of virtual agents can make them more believable and acceptable to users.

1.6 CONCLUSIONS

In this chapter, we examined the role of culture in the expression of conversational behaviours. Through an analysis of the research literature, we demon-

strated that culture can influence the expression of conversational behaviours at the content, form and communication management levels. We then presented a hybrid approach to the generation of culture-specific behaviours for embodied virtual agents. This approach is both model-driven (informed by qualitative models of culture from the social sciences) and data-driven (able to capture and encode quantitative information from corpora of labeled culture-specific conversational behaviours). It is capable of modelling culture-specific variations in the selection of communication content, in the form of conversational behaviours and in the management of the conversation. Our approach goes beyond what many think of as traditional NLG by: (1) considering both verbal and nonverbal behaviours, (2) focusing on social and personal behaviours, and (3) being adaptive to particular cultural backgrounds. In two evaluations, we demonstrated that observers of virtual agents are able to distinguish between culture-specific versions of both verbal and nonverbal conversational behaviours, and that they prefer those that reflect their own culture. This means that by incorporating culture-specific models into virtual agents, we can increase their believability and acceptability to users.

Although we demonstrated the feasibility and utility of modelling culture for several conversational behaviours, richer and more complete models may lead to further improvements in the generality and performance of culture-specific virtual agents. There is scope here for a variety of research projects including empirical ones that look at particular behaviours and/or particular cultures, and more engineering ones that concern themselves with statistical modelling and feature extraction for conversational behaviours. In particular, although our approach is partly data-driven, the collection and annotation of the corpus we used was time consuming and expensive. More research on the automatic analysis of conversational behaviours from video is also needed if we are to produce more humanlike embodied conversational agents.

Acknowledgments.

This work was funded by the European Commission under grant agreement eCute (FP7-ICT-2009-5).

References

- Alexandris, C. & Fotinea, S.-E. (2004). Discourse particles: Indicators of positive and non-positive politeness in the discourse structure of dialog systems for modern Greek. *Sprache und Datenverarbeitung: International Journal for Language Data Processing*, **28**(1-2), 19-29.
- Aylett, R., Louchart, S., Dias, J., Paiva, A., & Vala, M. (2005). FearNot! - an experiment in emergent narrative. In T. Panayiotopoulos et al., editor, *Proceedings of the International Conference on Intelligent Virtual Agents (henceforth 'IVA')*.
- Brown, P. & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press, Cambridge, UK.
- Bull, P. (1987). *Posture and Gesture*. Pergamon Press, Oxford, UK.
- Byrne, D. E. (1971). *The attraction paradigm*. Academic Press, New York, NY.
- Chartrand, T. L. & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, **76**(6), 893-910.
- Decision Systems Laboratory (2005-2007). GeNIe and SMILE. <http://genie.sis.pitt.edu/>. last viewed: 26.02.2012.
- Ekman, P. (2001). *Telling Lies - Clues to Deceit in the Marketplace, Politics, and Marriage*. W. W. Norton & Co., New York, NY, 3rd edition.
- Endrass, B., Rehm, M., & André, E. (2009). Culture-specific communication management for virtual agents. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (henceforth 'AAMAS')*.
- Endrass, B., Huang, L., Gratch, J., & André, E. (2010). A data-driven approach to model culture-specific communication management styles for virtual agents. In van der Hoek, Kaminka, Lespérance, Luck, and Sen, editors, *Proceedings of AAMAS*.
- Endrass, B., André, E., Rehm, M., Lipi, A. A., & Nakano, Y. (2011a). Culture-related differences in aspects of behavior for virtual characters across Germany and Japan. In *Proceedings of AAMAS*.
- Endrass, B., Nakano, Y., Lipi, A. A., Rehm, M., & André, E. (2011b). Culture-related topic selection in small talk conversations across Germany and Japan. In *Proceedings of IVA*.

- Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., & Petajan, E. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, **17**(4), 54–63.
- Hall, E. T. (1966). *The Hidden Dimension*. Doubleday, Garden City, NY.
- Hall, E. T. & Hall, M. R. (1987). *Hidden differences: Doing business with the Japanese*. Anchor Books / Doubleday, New York, NY.
- Hartmann, B., Mancini, M., Buisine, S., & Pelachaud, C. (2005). Design and evaluation of expressive gesture synthesis for embodied conversational agents. In *Proceedings of AAMAS*.
- Hofstede, G. (2001). *Culture's Consequences - Comparing Values, Behaviours, Institutions, and Organizations Across Nations*. Sage Publications, Thousand Oaks, CA, 2nd edition.
- Hofstede, G. (2011). <http://www.geert-hofstede.com/>.
- Hofstede, G., Hofstede, G.-J., & Minkov, M. (2010). *Cultures and Organizations. Software of the Mind. Intercultural Cooperation and its Importance for Survival*. McGraw-Hill USA, 3rd edition.
- Hofstede, G. J., Pedersen, P. B., & Hofstede, G. (2002). *Exploring Culture - Exercises, Stories and Synthetic Cultures*. Intercultural Press, Yarmouth, United States.
- House, J. (2006). Communicative styles in English and German. *European Journal of English Studies*, **10**(3), 249–267.
- Iacobelli, F. & Cassell, J. (2007). Ethnic identity and engagement in embodied conversational agents. In *Proceedings of IVA*.
- Isbister, K., Nakanishi, H., Ishida, T., & Nass, C. (2000). Helper agent: Designing an assistant for human-human interaction in a virtual meeting space. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (henceforth 'CHI')*.
- Johnson, W. L. & Valente, A. (2008). Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. In *Proceedings of the National Conference on Innovative Applications of Artificial Intelligence*.
- Johnson, W. L., Marsella, S., & Vilhjálmsdóttir, H. (2004). The DARWARS Tactical Language Training System. In *Interservice / Industry Training, Simulation, and Education Conference*.
- Johnson, W. L., Mayer, R. E., André, E., & Rehm, M. (2005). Cross-cultural evaluation of politeness in tactics for pedagogical agents. In *Proceedings of the Conference on Artificial Intelligence in Education*.
- Jonsdóttir, G. R., Thorisson, K. R., & Nivel, E. (2008). Learning smooth, human-like turntaking in realtime dialogue. In *Proceedings of IVA*.
- Kang, S.-H., Gratch, J., Wang, N., & Watt, J. H. (2008). Agreeable people like agreeable virtual humans. In *Proceedings of IVA*.
- Kipp, M. (2001). ANVIL - a generic annotation tool for multimodal dialogues. In *Proceedings of the European Conference on Speech Communication and*

- Technology* (henceforth 'Eurospeech').
- Kipp, M., Neff, M., & Albrecht, I. (2007a). An annotation scheme for conversational gestures: How to economically capture timing and form. *Language Resources and Evaluation*, **41**(3–4), 325–339.
- Kipp, M., Neff, M., Kipp, K. H., & Albrecht, I. (2007b). Towards natural gesture synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In *Proceedings of IVA*.
- Lipi, A., Nakano, Y., & Rehm, M. (2009). A parameter-based model for generating culturally adaptive nonverbal behaviors in embodied conversational agents. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction*.
- Mairesse, F. & Walker, M. A. (2008). Trainable generation of big-five personality styles through data-driven parameter estimation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (henceforth 'ACL').
- Marcus, A. & Alexander, C. (2007). User validation of cultural dimensions of a website design. In N. Aykin, editor, *Proceedings of the international conference on Usability and internationalization*.
- Martin, J.-C., Abrilian, S., Devillers, L., Lamolle, M., Mancini, M., & Pelachaud, C. (2005). Levels of representation in the annotation of emotion for the specification of expressivity in ECAs. In *Proceedings of IVA*.
- Mascarenhas, S., Dias, J., Afonso, N., Enz, S., & Paiva, A. (2009). Using rituals to express cultural differences in synthetic characters. In *Proceedings of AAMAS*.
- Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. In *Proceedings of the ACM international conference on Multimedia*.
- Rehm, M., André, E., Nakano, Y., Nishida, T., Bee, N., Endrass, B., Huan, H.-H., & Wissner, M. (2007). The CUBE-G approach - coaching culture-specific nonverbal behavior by virtual agents. In *Proceedings of the International Simulation and Gaming Association Conference*.
- Rist, T., André, E., & Baldes, S. (2003). A flexible platform for building applications with life-like characters. In *Proceedings of the International Conference on Intelligent User Interfaces* (henceforth 'IUI').
- Sato, R., Higashinaka, R., Tamoto, M., Nakano, M., & Aikawa, K. (2002). Learning decision trees to determine turn-taking by spoken dialogue systems. In *Proceedings of the International Conference on Spoken Language Processing* (henceforth 'ICSLP').
- Schneider, K. P. (1988). *Small Talk: Analysing Phatic Discourse*. Hitzeroth, Marburg, Germany.
- Sidner, C. L. (2004). Building spoken-language collaborative interface agents. In D. Dahl, editor, *Practical Spoken Dialog Systems*, pages 197–226. Kluwer, Dordrecht, The Netherlands.
- Ting-Toomey, S. (1999). *Communicating across cultures*. The Guilford Press, New York, NY.

- Trompenaars, F. & Hampden-Turner, C. (1997). *Riding the waves of culture - Understanding Cultural Diversity in Business*. Nicholas Brealey Publishing, London, UK.
- University of Maryland (2005). SHOP: Simple Hierarchical Ordered Planner. <http://www.cs.umd.edu/projects/shop/>. last viewed: 26.02.2012.
- Wu, P. & Miller, C. (2010). Interactive PhrasebookTM – conveying culture through etiquette. In *Proceedings of the International Workshop on Culturally-Aware Tutoring Systems*.
- Yin, L., Bickmore, T., & Cortés, D.-E. (2010). The impact of linguistic and cultural congruity on persuasion by conversational agents. In *Proceedings of IVA*.