AI & Soc (2017) 32:25–35
DOI 10.1007/s00146-014-0573-4

ORIGINAL ARTICLE

Real-time gesture translation in intercultural communication

Béatrice S. Hasler · Oren Salomon · Peleg Tuchman · Amir Lev-Tov · Doron Friedman

Received: 16 January 2014 / Accepted: 9 October 2014 / Published online: 16 October 2014

Abstract Nonverbal behavior plays a crucial role in human communication and often leads to misunderstandings between people from different cultures, even if they speak the same language fluently. While translation systems are available for verbal communication, translators for nonverbal communication do not exist yet. We present the conceptual design and an early prototype of a real-time gesture translator using body tracking and gesture recognition in avatar-mediated intercultural interactions. It contributes to the ambitious goal of bridging between cultures by translating culture-specific gestures to enhance mutual understanding. Possible applications of the gesture translator are discussed as a facilitating tool for global business meetings and as a means of technology-enhanced conflict resolution and prevention.

Keywords Nonverbal communication · Intercultural communication · Translation systems · Body tracking · Gesture recognition · Avatars · Virtual environments

B. S. Hasler (✉) · O. Salomon · P. Tuchman · A. Lev-Tov · D. Friedman
Advanced Reality Lab, Sammy Ofer School of Communications, Interdisciplinary Center Herzliya, Kanfei Nesharim St., 46150 Herzliya, Israel
e-mail: hbeatrice@idc.ac.il

O. Salomon
e-mail: orensalo@gmail.com

P. Tuchman
e-mail: pel6413@gmail.com

A. Lev-Tov
e-mail: amir.levtov@gmail.com

D. Friedman
e-mail: doronf@idc.ac.il

1 Introduction

Globalization offers many opportunities for the establishment of international relations, multicultural networks, and workforces. While cultural diversity may enrich our social and work environment, failure to communicate effectively between cultures can become a stumbling block for intercultural meetings (Barna 1994). Although language is the most obvious barrier in intercultural encounters, nonverbal communication (NVC) is also known to vary across cultures (Andersen et al. 2003). Cultural differences in the use and interpretation of nonverbal signals provide an additional, often unconscious source of misunderstanding (Ting-Toomey 1999). Even if conversational partners speak the same language fluently, they tend to interpret nonverbal messages in a culture-specific way (Stening 1979); that is, according to the sociocultural conventions that each of them is familiar with.

When traveling abroad, we often mistakenly assume that we can circumvent language barriers by using simple hand gestures to communicate. However, as Archer (1997) stated, "just as there is no reason to expect an English word to be recognized internationally, there is no reason to expect an American hand gesture to be recognized" (p. 80). The popular fallacy of a 'universal language' of gestures is likely to result in misunderstandings. Two common types of mistakes include (1) using a gesture that has a different meaning abroad and (2) failing to interpret a foreign gesture correctly (Archer 1997). Every traveler has probably encountered such misunderstandings, which may have resulted in funny, embarrassing, or even threatening situations. Cultural misunderstandings can influence interpersonal perceptions in a negative way, and thus, hinder the establishment of positive intercultural relationships. For

instance, a person may be considered as rude or impolite without having the intention to appear in such a way.

Success or failure of intercultural contact depends on the extent to which each person understands the other's culture (Schneller 1989). The ability to use and interpret nonverbal signals correctly is of particular importance for mutual understanding. The increasing need for global understanding has led to a vast amount of cross-cultural training materials, ranging from field guides that illustrate the cultural variety in NVC (Armstrong and Wagner 2003; Axtell 1998) to video documentaries (Archer 1991) and interactive online role-playing games (Johnson et al. 2004; Maniar and Bennett 2007). Courses are being taught at various educational levels, as well as for professionals in order to prepare them for business travels and effective collaboration in multicultural work groups. Such trainings typically do not only include the acquisition of foreign language skills, but aim to increase the awareness of cultural differences in (nonverbal) communication styles, norms, and expectations.

While trainings can be an effective method to prepare for an intercultural encounter, they are an expensive and time-consuming option. We propose an alternative approach to facilitate understanding between cultures instantaneously during an intercultural contact. This novel approach utilizes a real-time translator for nonverbal signals, similar to translation systems that exist for verbal messages. While such a translator could potentially be built for any type of NVC, our work is targeted toward a translation of arm and hand gestures with their respective meanings across cultures.

The gesture translator has been conceptualized as part of an European Union Research Project, BEAMING[1] (Steed et al. 2012). The BEAMING project deals with the science and technology intended to give people a real sense of physically being in a remote location with other people, without actually traveling. It aims at developing the next generation of telepresence using breakthroughs in virtual reality, augmented reality, tele-robotics and haptic technologies. Once such a telepresence has been constructed, it is possible to envision a technology-enhanced experience that is actually better than being there physically. The system has a representation of both the destination and the remote visitor, and both can be transformed in order to achieve some desired effect on the locals or the visitors. Transformation of cultural factors has been considered as a potential enhancement of intercultural meetings due to their high practical relevance. The gesture translator described in the current paper is an example of such a cultural transformation.

[1] http://beaming-eu.org.

The gesture translator is still at an early stage of its development. The paper focuses on its conceptual design and discusses the technical implementation of a first prototype as a 'proof of concept.' We conclude with a description of the next steps in the implementation and evaluation of the gesture translator and outline possible use cases in global business meetings and intercultural conflict resolution and prevention.

2 Culture and communication

For the purpose of the current research, we define *culture* from an ethnographic point of view. Culture in the ethnographic perspective refers to a code or system of meanings, historically shared among a group of people (Hall 1992). Cultural codes originate in shared traditions, beliefs, norms, and values that provide us with a complex frame of reference through which we perceive our environment and structure interpersonal activities (D'Andrade 1984). Shared meanings are communicated by means of natural language and other symbol systems within a cultural community.

Ting-Toomey (1999) argues that we only see and hear culture through these verbal and nonverbal symbols, while their origin remains hidden from our view. The ethnographic approach focuses on the 'visible dimension' in which culture is manifested; for example, by observing how communication codes (i.e., ways of speaking) vary across cultures. Accordingly, we define *communication* as the process of producing (i.e., encoding) and interpreting (i.e., decoding) meaningful verbal and nonverbal messages (Hall 1992). We use the term *intercultural communication* to refer to the symbolic exchange process in which two or more people from different cultures negotiate meanings in an interactive situation (Ting-Toomey 1999).

2.1 A classification of gestures

Researchers have been debating about what is regarded as a gesture (McNeill 1992). As our focus lies on interpersonal communication, we only consider those body movements that are treated by interlocutors as an integral part of the conversation, or more specifically, "as part of what a person meant to say" (Kendon 1997, p. 110). Conversational gestures serve multiple communicative functions. The nonverbal messages conveyed by conversational gestures can repeat, accent, contradict, complement, or substitute for a verbal message (Liu et al. 2011). Although conversational gestures may involve any part of the body, including head and face movements (Ekman and Friesen 1969), we use a classification of gestures that focuses mainly on movements of the arms and hands (Kendon 1982).

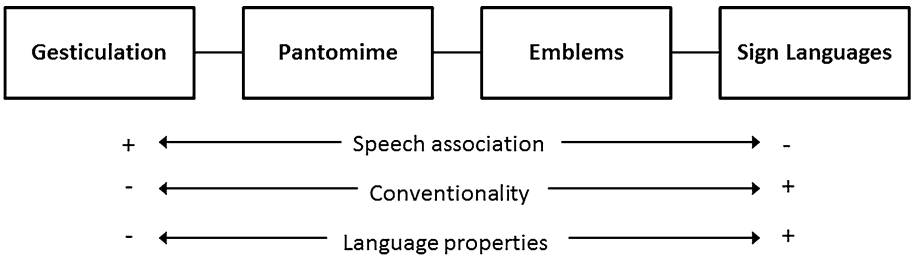

Fig. 1 Continuum of conversational gestures (based on McNeill 1992)

McNeill (1992) has introduced the notion of gestures as a continuum. He differentiates between four types of gestures in reference to Kendon's (1982) classification, ranging from *gesticulation* to *pantomime*, *emblems*, and *sign languages*. These types of gestures are placed on a continuum of different dimensions, which specify (1) the degree to which they accompany speech, (2) the extent to which they are conventionalized, and (3) the extent to which they possess linguistic and semiotic properties (see Fig. 1). As we move from the left to the right side of the continuum, association with speech declines, while social regulation and the presence of language properties increase.

Gesticulation refers to spontaneous hand and arm movements that are closely linked to the activity of speaking. These speech-associated movements are idiosyncratic and do not adhere to any standardized form. Although researchers disagree on its communicative value (see Goldin-Meadow 1999), gesticulation is a common phenomenon across cultures and comprises a significant part of our nonverbal expressions. Gesticulation has been found to make up about 90 % of human gestures (McNeill 1992). People even gesticulate when the listener is not in visible range (e.g., on the phone). It is also common for blind people to gesticulate when speaking to one another.

Pantomimic gestures depict objects, events or actions, with or without accompanying speech. As in gesticulation, linguistic properties are absent in pantomimic gestures. However, in pantomime, sequences of gestures form a unit, which increases their semiotic function. Pantomimic gestures are less spontaneous than gesticulation, but they are not as conventionalized as emblems and sign languages. The absence of standardized well-formedness leaves space for idiosyncratic expressions in pantomime. Examples of pantomime can be found in word guessing games in TV shows or played among friends, or as an artistic expression in theater performances.

Emblems are primarily hand gestures that are used in replacement of words or phrases. Emblems often have names or standard phrases to describe them, such as giving the "thumbs up", the O.K. sign, or waving good-bye. Each emblem has a direct literal translation to a corresponding word or phrase. Emblems have a shared (agreed-upon) meaning within each society, which makes them highly conventionalized. However, emblems are culturally specific. Their form and meaning may differ from one culture to another (Morris et al. 1979). Emblems are learned as specific symbols, but they are not structured like language. However, they have standards of well-formedness, which gives them a language-like property that is absent in gesticulation and pantomime.

Sign languages are fully developed linguistic systems used by deaf people to communicate, such as American Sign Language (ASL). Gestures used in sign languages are very different from conversational gestures used by the hearing population. These gestures are the direct equivalents of letters and words and function independently of speech. Conversational gestures used by hearing people are learned informally through observation. In contrast, sign language is taught explicitly. The gestures of a sign language are highly conventionalized with standardized ways of producing the movements. It is tempting to assume that sign languages would be understood internationally. However, they have enormous variations across cultures.

2.2 Cross-cultural differences in gestures

A person's cultural background influences how gestures are displayed and understood. Emblems in particular have been widely studied because of their heterogeneity across cultures (Morris et al. 1979). We differentiate between four cases of relations between emblematic gestures and their meanings across cultures. These cases are schematically illustrated in Table 1.

In the first case ("Equivalence"), a gesture that is used to communicate a specific meaning in one culture is different from the gesture used in another culture to convey the same meaning. Examples are cross-cultural differences in greeting gestures (e.g., waving or shaking hands in Western cultures vs. bowing in some East-Asian cultures). Cross-cultural differences also exist in head gestures used to indicate agreement or disagreement. For instance, a head shake typically means "no" in most European countries, with some exceptions, such as in Greece, where a single upwards nod is used to indicate "no". Other examples are pointing gestures, which carry the same meaning across

Table 1 Mapping matrix of gestures and meanings

Case	Gesture	Culture *X* (sender)	Culture *Y* (receiver)	Potential outcome	Action
1. Equivalence	G_a	M_1	–	Non-understanding	Translation (visual animation)
	G_b	–	M_1	Misunderstanding	
2. Equality	G	M_1	M_1	Positive understanding	No translation required
3. Confusion	G	M_1	M_2	Misunderstanding	Deletion of gesture, verbal description
4. Absence	G	M_1	–	Non-understanding	Verbal annotation

M meaning, *G* gesture, – no meaning, identical indices represent identical meaning or gesture, different indices represent different meaning or gesture

cultures but are performed differently (e.g., using the index finger or the entire hand in order to refer to an object or a person).

In the second case ("Equality"), an identical gesture is used in two cultures and carries the same meaning in both cultures. Despite cultural variations in emblems, some are used in identical form and meaning across different cultures. For example, a circle formed by the thumb and index finger means "O.K." in the USA and many European countries. It has been found that some nonverbal behaviors are rooted in biological processes and thus, even have universal meanings. A prominent example is the facial display of basic emotions, such as fear and anger, which is constant across cultures (Ekman and Friesen 1986). However, culture determines whether or not an emotion will be displayed or suppressed in a given occasion and to what degree.

In the third case ("Confusion"), an identical gesture used in two cultures has a different meaning in each culture. For example, hand gestures like the "O.K." sign and "thumbs up" carry different meanings across cultures and could be obnoxious or even insulting when interpreted wrongly. The American "O.K." gesture actually means "money" in Japan, "zero" in France, and is an obscene gesture in Brazil. Similarly, the American "thumbs up" gesture for "good luck" has a vulgar meaning in Southern Italy and Iran. Likewise, the Israeli gesture for "wait a second" would be understood in Italy as "What do you want?" It is crucial to pay attention to the subtle differences regarding how seemingly identical gestures are performed. A gesture may look similar at first sight, but a slight difference in position or motion can change its meaning dramatically. For example, in England, the "V" sign performed with the palm front means "Victory!", while the palm-back "V" sign is a sexual insult. In Germany, the gesture for "stupid" is performed by placing the index finger on the forehead. The American gesture for "smart" is nearly identical, but the finger is held further to the side, at the temple.

In the fourth case ("Absence"), a gesture that is common and has a clear meaning in one culture is meaningless in another culture. In contrast to the first case, the other culture does not have an equivalent gesture to express the respective meaning. This case points to an interesting observation by Archer (1997) that gestural categories of meaning are not universal. Thus, the assumption that there is at least one gesture for each category of meaning in every society but only the way these gestures are performed would vary within each culture is not true. Not every culture has an obscene gesture, a gesture for "beautiful woman," a "shame on you" gesture, a gesture for "well done," "crazy," "smart," etc. For example, France is one of the only cultures that have a gesture for "I'm bored." A gesture for "two people are in love" exists in Thailand, whereas most other cultures have no such gesture. While some countries in Latin America (e.g., Mexico) have several obscene gestures, some northern European countries, including Switzerland, Norway, and the Netherlands, have no native obscene gesture. Thus, culture not only determines what a specific gesture means, it also determines whether a gesture is necessary in a society to fulfill a certain communication need.

It is interesting to note that the cultural diversity in gestures persists despite potential homogenization effects of global mass media. Although the Western cultural imperialism has clearly influenced the world in many aspects, it does not seem to have erased all cultural differences. While we all may be drinking Coca-Cola and playing basketball, gestures seem less likely to homogenize (Archer 1997).

2.3 Quality of intercultural communication

Communication is generally considered as successful if the sender's intention matches the receiver's attribution of the meaning (Salomon 1981). Schneller (1989) presents a model of 'communication quality' defined as a function of decoding accuracy and interpretation certainty. This model is particularly suitable as a measure of NVC quality in cross-cultural settings. It distinguishes between correct and incorrect decoding of nonverbal signals, as well as different levels of certainty about whether the meaning of a

nonverbal message has been understood. Both decoding accuracy and interpretation certainty largely depend on the level of prior knowledge about an interlocutor's culture.

According to Schneller (1989), three types of understanding may occur in intercultural communication: *positive understanding*, *misunderstanding*, and *non-understanding*. Positive understanding is given if a nonverbal signal has been correctly decoded by the receiver; that is, when the sender's intention is highly similar to the attributed meaning by the receiver. In the case of emblems, positive understanding is most likely if the same gesture is used to communicate a specific meaning in two cultures (Case 2: "Equality" in Table 1). Misunderstanding is likely to occur if a nonverbal signal is interpreted by the receiver according to a different culture-specific meaning. This is particularly the case in emblematic gestures that are identical or similar in two cultures but carry different meanings in each culture (Case 3: "Confusion" in Table 1). Non-understanding occurs in situations when a transmitted nonverbal signal (e.g., gesture) does not exist in the receiver's culture and can therefore not be interpreted (Case 4: "Absence" in Table 1).

Each of these types of understanding can either be conscious or unconscious. Most of the times we are aware if non-understanding occurs, but we may not always be conscious about the reason behind it. We might intuitively understand a person from another culture without being consciously aware of our differences (unconscious positive understanding). Conscious positive understanding, on the other hand, is likely if we are aware of the cross-cultural differences but possess the necessary skills in managing them effectively. Some misunderstandings may reach a conscious level, especially if they lead to funny or threatening situations (e.g., in the case of misinterpretations of emblems). The most severe case, however, is when misunderstandings remain undetected. We may wrongly attribute an interlocutor's communication missteps to personal factors (e.g., personality flaws) instead of cultural factors. Such attribution errors due to unconscious misunderstandings may spiral into major escalatory conflicts (Ting-Toomey 1999).

3 Conceptual design

3.1 Idea and intended purpose

The basic idea of the gesture translator is that it translates the culture-specific meaning of gestures in an intercultural communication process. Its ultimate goal is to enable people from different cultures to understand each other's nonverbal messages immediately, without having to know how to interpret the culture-specific gestures used by their interlocutor. In contrast to training systems that would be applied prior to an intercultural contact, the gesture translator has its application as a real-time mediator between individuals with different cultural backgrounds. Thus, it provides instantaneous assistance and omits the need for preparation and training. Instead, learning of cultural differences (and similarities) in gestures and their associated meanings may occur as a side effect during the mediated intercultural encounter.

We make use of ubiquitous tracking technologies in order to let participants express themselves and interact in the most natural way (i.e., as if they were meeting face-to-face). While earlier motion tracking systems required tracking suits and markers, depth cameras, such as *Kinect*, make it possible to track a user's body movements without having to wear special clothing. In recent years, interactive gesture-based applications have been developed, such as computer games, in which a user can control an avatar (i.e., his digital representation in a virtual environment) through natural body movements. Gestures performed by a user may be used to trigger predefined movements of his avatar or to manipulate objects within the virtual environment. In order to activate such avatar animations or game actions by human gesture inputs, gesture recognition algorithms are applied to the motion tracking data. It is also possible to map a user's body movements continuously onto an avatar so that the avatar appears to move in the same way as its user.

The idea of applying gesture-based interaction and avatars to facilitate intercultural communication is new. However, there are similar approaches aimed to improve communication between deaf and hearing people using automatic sign language recognition (Cooper et al. 2011; Ong and Ranganath 2005; Zieren et al. 2006). The goal of sign language translators is to read a deaf person's gestures and facial mimicry and translate them into spoken language, and vice versa.

3.2 Scenarios

The first version of the gesture translator is designed for emblems since they are well defined regarding their form and culture-specific meaning. The scenario involves two participants, each of which has a different cultural background. Although the system would eventually work bidirectionally, we currently focus on a unidirectional scenario. It considers one participant to be the sender (i.e., encoder) and the other participant to be the receiver (i.e., decoder). In a bidirectional scenario, both participants would have their body movements tracked and mapped onto an avatar over the course of their interaction. The unidirectional scenario only requires tracking of one participant.

Each of the cases illustrated in Table 1 has a different potential outcome and requires different kind of actions,

depending on how the sender's gesture is decoded by the receiver. Thus, whenever the system recognizes a culture-specific gesture, it takes the corresponding actions.

In the first case, the sender's gesture is not understood by the receiver, but an equivalent gesture exists in the receiver's culture to express the same meaning. This case requires a translation, which can be implemented by replacing the sender's gesture by the equivalent gesture used in the receiver's culture. The translation output in this case may be a prerecorded avatar animation depicting the equivalent gesture that overrides the sender's original gesture.

In the second case where the same gesture is used in both cultures and carries the same meaning, no translation is required. In the third case, the gesture used by the sender also exists in the receiver's culture but has a different meaning. This is the most critical case as it may lead to severe misunderstandings. A possible solution is to delete the gesture performed by the sender and replace it by a verbal description of the intended meaning in order to prevent a potential misunderstanding. This solution is particularly suitable if no equivalent gesture exists in the receiver's culture. However, if there is an equivalent gesture available, the same solution could be used as in the first case; that is, replacing the gesture by an alternative, prerecorded avatar gesture. In the fourth case where the gesture performed by the sender is meaningless to the receiver, translation is not possible. Instead, verbal annotations may be displayed in order to explain the intended meaning of the respective gesture.

4 Proof of concept

We implemented an example of the first case ("Equivalence") of the mapping matrix as shown in Table 1. In this first case, an equivalent gesture exists in the receiver's culture to express the same meaning conveyed by the sender's gesture. As a proof of concept, we implemented a translation of greeting gestures typically used in Western cultures (i.e., waving) and some East-Asian cultures (i.e., bowing) (see Fig. 2).[2]

We track a participant's body movements and map them continuously onto an avatar. Thus, the avatar appears to move in the same way as the participant. The first prototype of the gesture translator has the ability to detect a specific motion and replace it with another (prerecorded) animation.

This prototype was developed using the *Kinect* depth camera. Compared to other gesture-based interaction devices, such as *Wiimote*, which is based on acceleration data in three spatial directions, *Kinect* provides a larger set of tracking points (i.e., 15 joints). The *Kinect* skeleton includes positions and orientations of the torso, neck, head, shoulders, elbows, hips, and knees.

4.1 Gesture recording and preprocessing

The method includes learning a model in a training phase and using it for recognition in real time. In order to collect data for the training, we asked participants to perform the two greeting gestures of interest as well as some other random movements. Although emblems are performed according to conventionalized movements, they vary among individuals (de-Graft Aikins 2011). Physical characteristics, such as body shape and height, influence limb movements and stride. Gestures also vary for a given individual from instance to instance. Situational factors, such as mood or fatigue, are likely to affect the way an individual performs a gesture. Thus, in order to train the model and make it robust for both inter- and intra-individual variations and enable the system to recognize them, we recorded each gesture for each participant multiple times.

Kinect tracks 15 joints and returns information regarding their position and orientation in the three-dimensional space. For our feature set, we took only the orientation data and excluded the absolute position because we do not want the participant's position in the room to affect the recognition. We chose 11 joints and used their orientation data in quaternion format (x, y, z, w), resulting in 44 scalar values per frame. Taking a series of frames of a certain movement, we represent the data of an observed gesture, by a matrix M of the size $n * m$, where $n = 44$ and m is the number of frames. *Kinect* tracks at a rate of 30 frames per second, a typical gesture lasts 1–2 s, so for each joint, we get a vector of the length $m = 30$–60. Instead of directly modeling all orientation features of all joints, we look at a given orientation feature of a given joint and check its variation over time. In terms of the matrix M, we train a model from a single row vector r from each movement associated with a given gesture.

Early research on sign language recognition has been done by Assan and Grobel (1997), and Starner and Pentland (1995). They use Hidden Markov Models (HMM) to recognize ASL and the Sign Language of the Netherlands, respectively. A more recent effort which is inspired by speech recognition techniques is described in Dreuw et al. (2007). We chose to implement a simpler algorithm, which still uses HMM as a training model. We chose HMM since we want to capture the temporal relations between basic elements of a gesture (hidden states) as well as the observed density of each one of

[2] A demo video can be found at http://www.youtube.com/watch?v=Wp6VPb2EaFU.

Fig. 2 Prototype of the gesture translator for Western and Asian greeting gestures. On the right, a participant's bowing gesture, on the upper left, the recognized participant and her joints in a typical frame, and on the bottom left, the translated gesture onto her avatar

them. This is done in a similar way that is being used in speech recognition while training HMM for words' acoustics or by training HMM for phonemes' acoustics given words/sentences. The basic elements of a gesture correspond to hidden states. The relations between them correspond to transition probabilities, and the observed density in a given state corresponds to the emission probabilities. The transition and emission probabilities are being estimated in a training phase using the Baum–Welch algorithm, and the computation of our target gesture's likelihood, given gesture observations, is performed using the Viterbi algorithm (Rabiner 1989). The structure we chose for the HMM is three states left-to-right with reflexive transitions for each state.

Each of the HMM models one orientation coordinate. After we obtain 44 HMM of a specific gesture, we want to assign them with weights according to their contribution to the prediction. We test the contribution of each model to recognition accuracy by comparing likelihoods of target gestures against likelihoods of nontarget ones. For each model, we get a probability vector in length of the number of testing sequences and compare the probabilities of the two data sets (see Fig. 3).

For each model, we assign a weight in the range [0, 1] based on the distance between the probability vectors of the two datasets. The weight W_i of a model $i(1 \le i \le 44)$ is given by

$$W_i = \frac{\left|P_{\mathrm{T}}^i - P_{\mathrm{O}}^i\right|}{N} \tag{1}$$

where P_{T}^i and P_{O}^i are the mean log likelihood of the target gesture and the other gestures, respectively, and N is a normalization factor given by

$$N = \max_{1 \le i \le 44} \left|P_{\mathrm{T}}^i - P_{\mathrm{O}}^i\right| \tag{2}$$

In our training phase, we got $P_{\mathrm{T}}^i \ge P_{\mathrm{O}}^i$, otherwise the model i is ignored.

4.2 Real-time gesture recognition

For real-time gesture recognition, we use a variable time window, which is the maximum length of a gesture in the training data. In each window, the extracted feature vectors are used as an observation input for the 44 HMM. Running the Viterbi algorithm on each of the HMM, we get a vector P of 44 probability values. Then, the weighted average S of the values is taken using the weight vector W computed for the models:

$$S = \mathrm{PW}^{\mathrm{T}} \tag{3}$$

If S exceeds a threshold (tuned in the training stage), there is high probability that the performed movement matches the gesture, and a translation action can be deployed. An example of such a model is model number 37 (see Fig. 4) that has the highest value.

In our proof-of-concept implementation, this results in an avatar playing a prerecorded animation of the equivalent gesture in the target culture. This gesture overrides the tracking data, which is resumed a few seconds after playback of the prerecorded animation. Otherwise, we continue mapping the

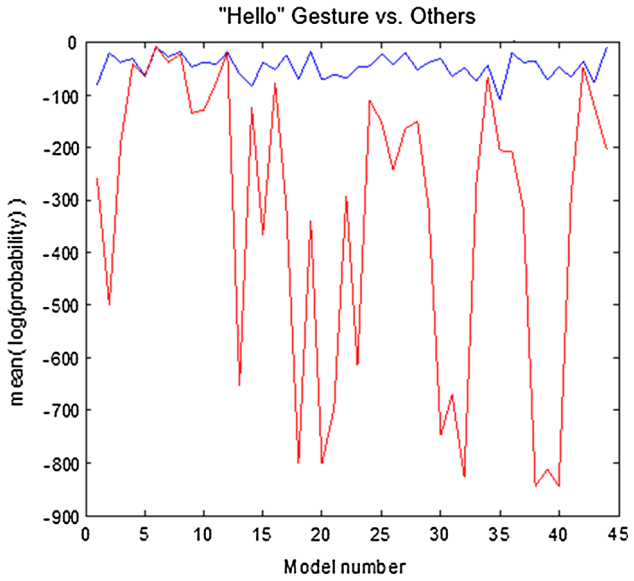

Fig. 3 Gesture recognition mean log likelihoods. The upper (*blue*) line indicates our target "Hello" gesture and the lower (*red*) line indicates the other gestures (color figure online)

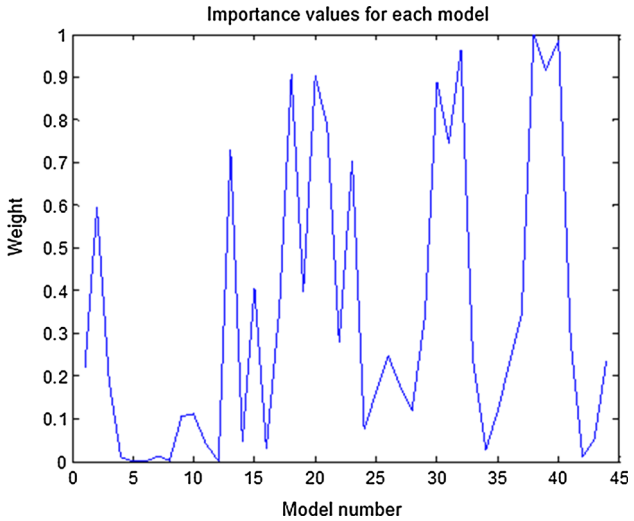

Fig. 4 Weight values for each model, based on the distance between "Hello" gestures and the other gestures

participant's gestures onto the avatar in real time without translations (see Fig. 5a, b). The beginning and the end of the animation sequence is blended with the live animation stream from the participant in order to avoid a "jump cut".

5 Future work

5.1 Lessons learned from the prototype

We implemented a prototype based on a simple intercultural greeting scenario between East-Asian and Western counterparts as a proof of concept. Greeting gestures were chosen because they were relatively easy to capture with the data provided by the *Kinect* skeleton. However, many emblematic gestures involve finger movements, which require additional devices (e.g., data gloves or very high resolution cameras) for accurate tracking.

One of the main lessons learned from the proof-of-concept implementation is concerned with the animated translation output. We chose to override the original motion by a prerecorded animation if it matches a gesture that is known to the system as being culture-specific. However, even though the system operates in real time, it requires a significant percentage of a gesture's frames to appear before the target gesture is detected. The system displays the original movements on the participant's avatar until it is being recognized as a gesture that needs to be translated and only then plays the alternative prerecorded gesture. The result is that a significant part of the gesture to be replaced will be shown before the replacement gesture is displayed. This procedure might lead to an interruption of the flow of communication. Further evaluation of the delay in recognition is necessary.

For future developments of the gesture translator, we suggest considering alternative solutions without having to interrupt and override the participant's original movements. It may be possible to indicate a delay of nonverbal transmission due to automatic translation in ways that users could adapt to. Alternatively, a separate window may be used to display the translation output. This would look similar to the display of a human sign language translator that is often used on TV to translate a news reporter's speech for deaf viewers.

While this defies the goals of BEAMING in providing a seamless communication medium, we know that people can get used to such augmented displays on TV and in using augmented reality systems. The goal of telepresence systems, such as BEAMING, is to replace face-to-face interaction while retaining all the subtleties of such interactions. Thus, ideally the augmented content would be superimposed on the participants' avatars in order to maintain the visual attention on a seemingly coherent flow of the interaction.

5.2 Extensions of the gesture translator

The most obvious next step in the extension of the gesture translator is to add more gestures. We will set up a database of culture-specific gestures and record the various gestures, which will ideally be performed by members of the respective cultures. However, creating and updating a gesture database for gesture recognition systems is not trivial (Turk 2002). The system needs to be trained for each gesture that has been added, and the more gestures a

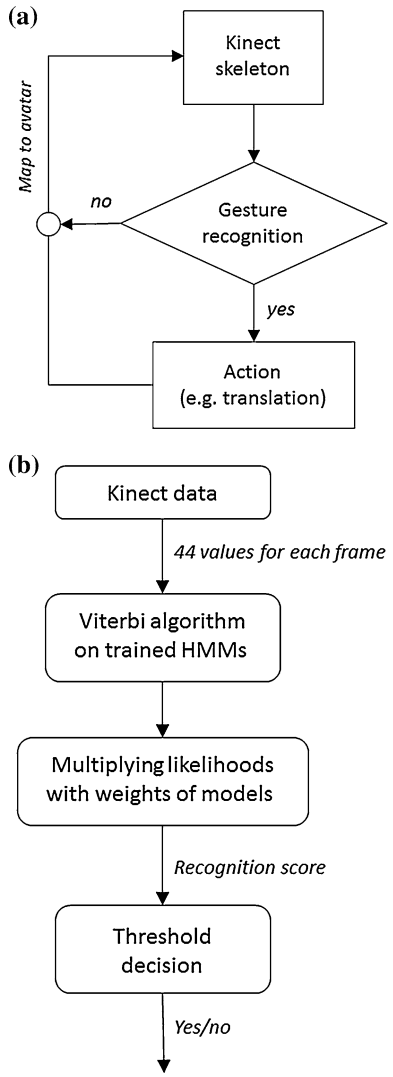

Fig. 5 **a** A flow diagram of the overall system. **b** A flow diagram of the gesture recognition algorithm

database contains, the more sophisticated the recognition algorithm needs to be in order to be able to differentiate correctly between similar gestures. Furthermore, it needs to be taken into consideration that the more gestures we add, the more training is required in order to assure recognition accuracy.

While in our current scenario, the translation is directed from Culture *X* to Culture *Y*, a bidirectional translator would monitor the movements performed by both participants and provide them both with translated output if required. Different translation modes may be chosen depending on the purpose of the interaction or individual preferences. For example, a variable component is whether or not the participants know that their gestures have been translated in the receiver's view. In order to increase learning effects, feedback could be given to the sender if she used a culture-specific gesture that is meaningless or has a different meaning in the receiver's culture.

The first prototype of the gesture translator has been conceptualized for two participants from different cultures, which would eventually provide bidirectional translation outputs. The translator may be extended to support multicultural gatherings involving multiple participants with various cultural backgrounds. A multi-user setting would require the system to synchronize multiple tracking inputs and run multiple comparisons of gestures and meanings at the same time in order to generate the right translation output for each participant. The complexity of a multi-user scenario not only brings about technical constraints, but also requires careful design considerations. The translator interface would need to be designed in a way that is intuitive to use and comprehend and does not disturb participants' focus of attention on the content of the group interaction.

An alternative scenario is a single-user application, in which the gesture translator is used as an interactive gesture dictionary. Given that an extensive cross-cultural gesture database exists, a user could perform a specific gesture, choose the target culture, and would receive the respective translation output. The output could either be an animation of an equivalent gesture that carries the intended meaning or an alert if the performed gesture has a different meaning or is meaningless in the target culture. An interactive gesture dictionary has several advantages compared to verbal descriptions of gestures or still images commonly used in field guides. As gestures are dynamic by nature, an avatar-based animated gesture output is clearly more informative, and the nuanced differences in the way gestures are performed may be easier to comprehend. The interactive feature also has advantages over video tutorials as it makes it possible to "look up" specific gestures without having to know how they are called. This single-user application could also be used as a training tool. For instance, a salesman who is preparing for a business travel to Japan may practice his sales talk while being connected to the translator. He would receive immediate feedback if he uses a gesture that could be misinterpreted in Japan.

5.3 Evaluation of the translator's performance and social impact

It has yet to be evaluated how scalable our proposed method for gesture recognition and translation will be when more gestures, cultures, and simultaneous participants are added. Recognition accuracy needs to be evaluated with a larger set of gestures. User evaluation studies are required in order to examine the translator's usability. Different types of translation and annotation outputs may be appropriate for different kinds of participants and

application scenarios. Especially participants' level of prior knowledge of the target culture should be considered as a factor in generating user-specific translation outputs.

Besides these technical and design evaluations, empirical studies need to be conducted in order to test the translator's social impact in intercultural settings. It has yet to be evaluated whether the gesture translator would indeed improve mutual understanding between members of different cultures. We can further test implicit and explicit learning effects that may occur intentionally or unintentionally during its use. Intentional learning, for example, may be supported by integrating feedback features. Research questions of interest could be: Does the gesture translator increase users' awareness of cross-cultural differences? Does it make people more sensitive for potential misunderstandings in intercultural communication? Does it improve users' intercultural communication skills? Are these hypothesized learning effects limited to the specific situation in which the translator has been applied or would they generalize to other contexts?

6 Conclusions

A majority of cross-cultural conflicts can be traced to miscommunication or ignorance (Ting-Toomey 1999). Many of these conflicts are rooted in cross-cultural differences in NVC. The gesture translator conceptualized in the current paper sets the ambitious goal to enhance mutual understanding in intercultural interactions. It is intended to bridge between cultures in the role of a facilitator or mediator by adding an interpretation layer to the NVC channel.

The gesture translator has several potential use cases. It may be used as a facilitating tool for global business meetings. An immediate nonverbal understanding would enable international business partners to focus on their actual task or conversational topic without having to worry about how to express themselves or how to read the other's nonverbal messages. Another possible application of the gesture translator is in the area of conflict resolution and prevention. A translator technology that reduces conflict potential due to misinterpretations of nonverbal messages may ease negotiations between members of rival groups. Although the gesture translator is still at an early prototype stage, these possible application scenarios illustrate its great potential and immediate importance.

Acknowledgments This research was funded by the European Commission under the ICT work program 2009 (Grant No. FP7-ICT-248620). The first author was supported by a Marie Curie Fellowship from the European Commission (Grant No. FP7-PEOPLE-2009-IEF-254277).

References

Andersen PA, Hecht ML, Hoobler GD, Smallwood M (2003) Nonverbal communication across cultures. Cross-cultural and intercultural communication. Sage Publications, Thousand Oaks, pp 73–90

Archer D (1991) A world of gestures: culture and nonverbal communication (video). Berkley Media LLC, Berkley

Archer D (1997) Unspoken diversity: cultural differences in gestures. Qual Sociol 20(1):79–105

Armstrong N, Wagner M (2003) Field guide to gestures. How to identify and interpret virtually every gesture known to man. Quirk Books, Philadelphia

Assan M, Grobel K (1997) Video-based sign language recognition using hidden Markov models. In: Gesture and sign language in human-computer interaction. Proceedings of international gesture workshop, Bielefeld, Germany

Axtell RE (1998) Gestures: the do's and taboos of body language around the world. Wiley, New York

Barna LM (1994) Stumbling blocks in intercultural communication. In: Samovar LA, Porter RE (eds) Intercultural communication: a reader, 7th ed. Wadsworth Publishing Company, Belmont, pp 337–346

Cooper H, Holt B, Bowden R (2011) Sign language recognition. In: Moeslund TB, Hilton A, Krüger V, Sigal L (eds) Visual Analysis of Humans: Looking at People. Springer, London, pp 539–562

D'Andrade R (1984) Cultural meaning systems. In: Shweder RA, LeVine RA (eds) Culture theory: essays on mind, self, and emotion. Cambridge University Press, Cambridge

de-Graft Aikins A (2011) Nonverbal communication in everyday multicultural life. In: Hook D, Franks B, Bauer MW (eds) The social psychology of communication. Palgrave Macmillan, New York, pp 67–86

Dreuw P, Rybach D, Deselaers T, Zahedi M, Ney H (2007). Speech recognition techniques for a sign language recognition system. In: Proceedings of Interspeech 2007, Antwerp, Belgium, pp 2513–2516

Ekman P, Friesen WV (1969) The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica 1(1):49–98

Ekman P, Friesen WV (1986) A new pan cultural expression of emotion. Motivation Emotion 10:159–168

Goldin-Meadow S (1999) The role of gesture in communication and thinking. Trends Cognit Sci 3(11):419–429

Hall BJ (1992) Theories of culture and communication. Commun Theory 2(1):50–70

Johnson WL, Marsella S, Mote N, Viljalmsson H, Narayanan S, Choi S (2004) Tactical language training system: supporting the rapid acquisition of foreign language and cultural skills. In: InSTIL/ICALL symposium: NLP and speech technologies in advanced language learning systems, Venice

Kendon A (1982) The study of gesture: some remarks on its history. Semiotic Inquiry 2:45–62

Kendon A (1997) Gesture. Annu Rev Anthropol 26:109–128

Liu S, Volcic Z, Gallois C (2011) Introducing intercultural communication. Global cultures and contexts. Sage, Los Angeles

Maniar N, Bennett E (2007) Designing a mobile game to reduce culture shock. In: International conference on advances in computer games technology, ACM Press, Lisbon, pp 252–253

McNeill D (1992) Hand and mind: what gestures reveal about thought. University of Chicago Press, Chicago

Morris D, Collett P, Marsh P, O'Shaughnessy M (1979) Gestures, their origins and distribution. Stein and Day, New York

Ong SCW, Ranganath S (2005) Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Trans Pattern Anal Mach Intell 27(6):873–891

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–285

Salomon G (1981) Communication and education: social and psychological interaction. Sage, Beverly Hills

Schneller R (1989) Intercultural and intrapersonal processes and factors of misunderstanding: Implications for multicultural training. Int J Intercult Relat 13:465–484

Starner T, Pentland A (1995) Visual recognition of American Sign Language using hidden Markov models. In: Proceedings of international workshop on automatic face- and gesture-recognition, Zurich

Steed A, Steptoe W, Oyekoya W, Pece F, Weyrich T et al (2012) Beaming: an asymmetric telepresence system. In: Spatial interfaces. IEEE Computer Graphics and Applications, pp 10–17

Stening BW (1979) Problems in cross-cultural contact: a literature review. Internat J Intercult Relat 3:269–313

Ting-Toomey S (1999) Communicating across cultures. Guilford Press, New York

Turk M (2002) Gesture recognition. In: Hale KS, Stanney KM (eds) Handbook of virtual environment technology: design, implementation, and applications. Lawrence Erlbaum, Mahwah, pp 223–238

Zieren J, Canzler U, Bauer B, Kraiss KF (2006) Sign language recognition. In: Kraiss KF (ed) Advanced man-machine interaction. Signals and communication technology. Springer, Berlin, pp 95–139