Real-time Translation of Nonverbal Communication in Cross-Cultural Online Encounters

Béatrice S. Hasler[1], Oren Salomon[2], Peleg Tuchman[1], Ady Nae O'Malley[2] and Doron A. Friedman[1]

Advanced Virtuality Lab, Sammy Ofer School of Communications, Interdisciplinary Center
Herzliya, Kanfei Nesharim St., P.O. Box 167,
46150 Herzliya, Israel
[1] {hbeatrice, ptuchman, doronf}@idc.ac.il
[2] {orensalo,adloyada}@gmail.com

Abstract. Nonverbal behavior plays a crucial role in human communication, and often leads to misinterpretations and misunderstandings between people from different cultural backgrounds; even if they speak the same language fluently. While translation systems are available for verbal communication, real-time translators for nonverbal communication (NVC) do not exist yet. We present the conceptual design and an early prototype of a real-time NVC translator using body tracking and gesture recognition systems for avatar-mediated interactions. It contributes to the ambitious goal of bridging between cultures by facilitating cross-cultural online meetings. Possible applications of the NVC translator are discussed as a facilitating tool for global teamwork and conflict resolution.

Keywords: Nonverbal communication, cross-cultural communication, translation systems, body tracking, gesture recognition, avatars, virtual environments

1 Introduction

Globalization offers many opportunities for the establishment of international relations, multicultural networks and workforces. However, culturally diverse groups often face challenges regarding effective communication and mutual understanding. Although language is the most obvious barrier in cross-cultural encounters, NVC is also known to vary across cultures [1], and provides an additional, mostly unconscious source of misunderstandings. Even if conversational partners speak the same language fluently, nonverbal messages tend to be interpreted in a culture-specific way (i.e., according to the sociocultural conventions that each of the interlocutors is familiar with). Different interpretations of culture-specific NVC are likely to result in misunderstandings, and hinder effective cross-cultural communication. Probably every traveler has encountered such misunderstandings due to cultural differences in NVC, which may have resulted in funny, embarrassing, or even threatening situations.

For example, hand gestures like the “O.K.” sign, “thumbs up”, and the “V” sign for victory, carry different meanings across cultures, and could be obnoxious or in some cases even insulting when interpreted wrongly. Thus, a person may be considered as impolite or rude without having the intention to appear in such a way. Cultural differences also exist in very basic social interactions, such as different greeting rituals (e.g., waving or shaking hands in Western cultures vs. bowing in some East-Asian cultures). Another example are different nonverbal indications of agreement and disagreement (e.g., nodding typically means “yes” in Western cultures, whereas in most Arab countries it actually means “no”; while Indians would move their head sideways to indicate “yes”).

There are numerous field guides, including books [2,3] and online resources [4,5] that provide examples of the different meanings assigned to gestures across different cultures. The increasing need for intercultural competences has also led to a vast amount of cross-cultural training materials, including interactive online role-playing games [6,7,8]. 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 not only include acquisition of foreign language skills, but also awareness of cultural differences in communication style and nonverbal behavior.

Cross-cultural trainings may be an effective but expensive and time consuming option to facilitate understanding between people of different cultures. We therefore propose a real-time translator for nonverbal messages, similar to translation systems that exist for verbal messages. In contrast to training systems, which are typically used in preparation for cross-cultural encounters, the NVC translator would have its application in real-time interactions between individuals with different cultural backgrounds.

The NVC translator is still at an early stage of its development. The current paper focuses on the conceptual design of the NVC translator, and discusses the technical implementation of a first prototype as proof-of-concept. We conclude with a description of the next steps in the implementation and evaluation of the NVC translator, and a discussion of possible application scenarios.

2 Conceptual Design

The basic idea of the NVC translator is that it translates the culture-specific meaning of a particular gesture by replacing it with an equivalent gesture used in the culture of the interlocutor. Instead of presenting the translated gesture as a static picture, which is most often the case in traditional training materials, it is performed by an animated avatar. If no equivalent gesture exists in the target culture, it would explain the meaning of the gesture with visual or verbal annotations.

The ultimate goal of the NVC translator is to enable interlocutors to immediately understand each other (including the subtleties in their nonverbal messages), without having to take any trainings in the use and interpretations of nonverbal behaviors of the target culture. An immediate (nonverbal) understanding between members of different cultures would enable them to focus on their actual collaboration task or conversational topic without having to worry about how to express themselves or how to read the other's messages.

In order to let participants express themselves and interact in the most natural way (i.e., as they would when meeting face-to-face), we use ubiquitous tracking technologies. These novel human-computer interfaces make it possible to map a user's body movement in real-time onto an avatar. In a scenario with two participants, both would have their body movements tracked and mapped onto an avatar over the course of their interaction. Whenever the system recognizes a gesture performed by one participant that has a different meaning in the other participant's culture, it translates (or annotates) it accordingly. The NVC translator would typically provide bi-directional gesture translations providing both interlocutors with translated output. However, different translation modes may be chosen depending on the purpose of the interaction, or individual preferences.

We built a prototype of the NVC translator that maps a participant's body movements onto an avatar. It has the ability to detect a specific motion, and replaces it with another (prerecorded) animation. As a proof-of-concept, we implemented a translation of greeting gestures typically used in Western cultures (i.e., wave) and some East-Asian cultures (i.e., bow) (see Fig. 1).

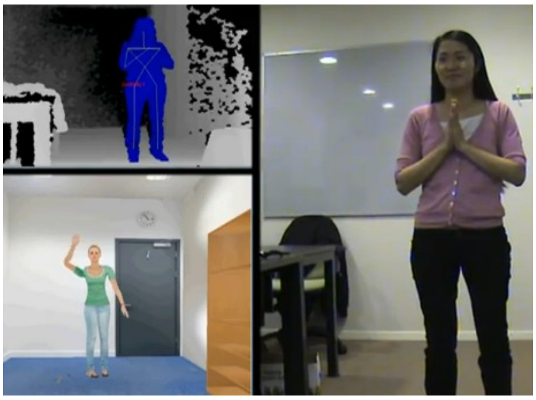

Fig. 1. Prototype of the NVC translator for Western and Asian greeting gestures

2 Prototype

The first version of the prototype was developed using the *Kinect* depth camera. In comparison to other natural human-computer interfaces, 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 torso, neck, head, shoulder, elbow, hip, and knee positions and orientations.

Gesture Recording and Preprocessing. The method includes a training phase and a real-time component. In the training phase, we asked participants to perform specific gestures of the target culture. In order to make sure that the system will be able to recognize these particular gestures even if they are performed with individual variations (both between and within participants), we recorded each gesture multiple times.

Kinect tracks 15 joint positions at 30 frames per second. We treat each coordinate separately, and for each movement we process a vector of length 44 per frame. We take the orientation data (x,y,z,w) of 11 joints. The absolute position is not included in the vector because we do not want the participant's position in the room to affect the recognition. A typical gesture lasts 1-2 seconds, so for each motion we get 44 vectors of approximately length 30-60. For each of these values we train a Markov Model (MM) that identifies the probability of a specific movement with specific orientation. The training estimates the transition and emission probabilities for a Hidden Markov Model (HMM) with 3 states using the Baum-Welch algorithm.

After we obtain 44 MMs of a specific gesture, we test the contribution of each model to recognition. We compare the recorded data with the target motion and without it. We use the Viterbi algorithm [9] in order to find the probability that this motion data includes our target gesture. For each model we get a probability vector in length of the number of testing sequences. Now we compare the probabilities of the two motion data sets (see Figure 2). We recognize models that can serve as good predictors; each such model is related with one orientation coordinate (such as model 33 in Figure 3). For each model we assign a score in [0,1] based on the distance between the probability vectors of the two datasets. For each model, where p_m is the mean probability of the target motion and p_o is the mean probability of the other motions, the distance is:

$$\frac{\left||\mathbf{p}_o| - |\mathbf{p}_m|\right|}{\max(d_M)};$$

$i = 1 \dots 44$ i is the model number

$dm_i = |po_i - pm_i|$ dm_i is the distance in model i

$W_i = \frac{dm_i}{\max(dm)}$ W_i is the weight vector

Real-time Gesture Recognition. For real-time gesture recognition, we use a variable time window, which is the maximum length of the movements that we look at. Test operations are carried out using the Viterbi algorithm for each of the 44 MMs. The results are multiplied by the weight vector that we prepared for each model in the training phase.

If the value exceeds some threshold, there is high probability that the performed movement matches a gesture to be translated, and the translation action can be deployed. In our proof-of-concept implementation this results in the 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 participant's gestures onto the avatar in real-time without translations (see Fig. 4).

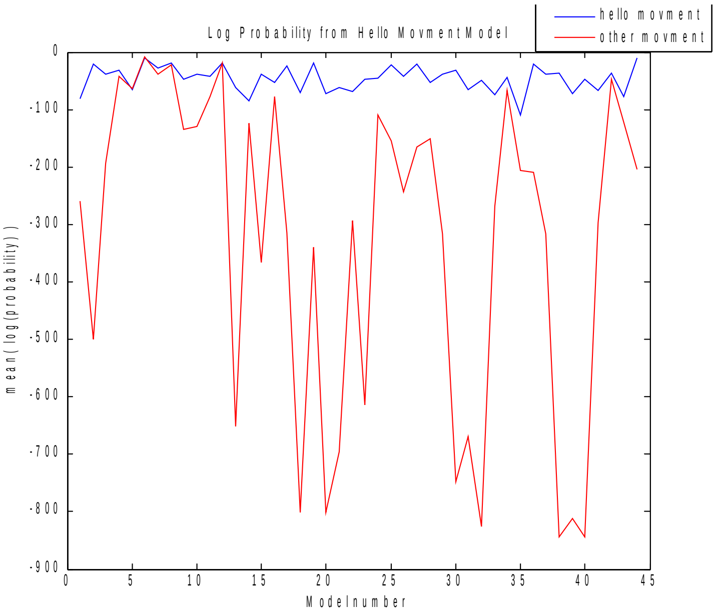

Fig. 2. Average logs of motion recognition, for a HMM with 3 motions. Upper line: clean record of motion; lower line: records of other movement

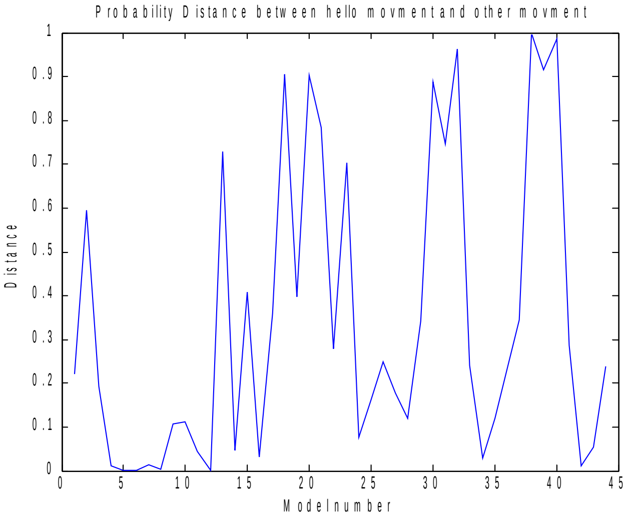

Fig. 3. A plot of average weight values for each model

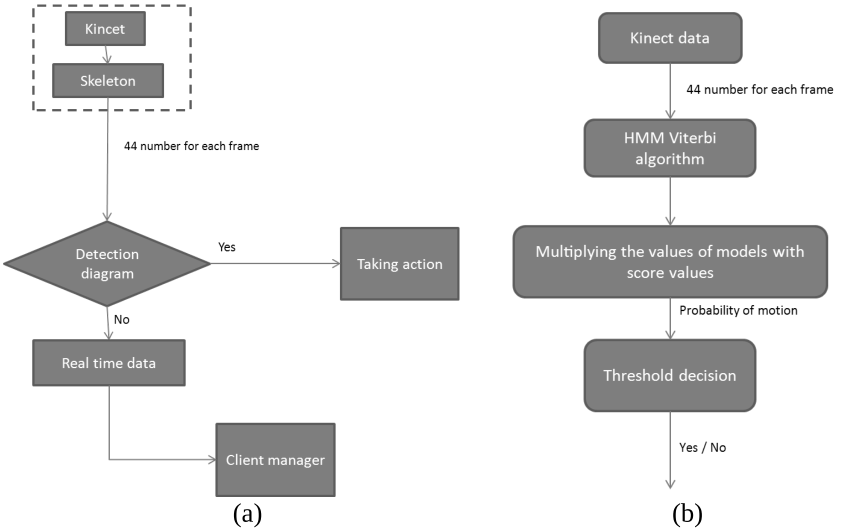

Fig. 4. A flow diagram of the real-time motion detection algorithm: (a) overall algorithm, (b) zoom in on the detection step ("taking action")

3 Future Work and Application Scenarios

We implemented a simple greeting ritual of East-Asian and Western counterparts as proof-of-concept. These gestures were chosen because they were easy to capture with the data provided by the *Kinect* skeleton. More sophisticated tracking systems are required in order to capture the subtleties in body postures, hand gestures, eye gaze, and facial expressions. The implementation of the actual NVC translator thus requires additional tracking devices. We will also set up a database of cultural differences in NVC, and how the various gestures are performed by individuals of the respective cultures. It has yet to be evaluated how scalable our proposed method for gesture recognition is (i.e., when more gestures and participants are added).

Furthermore, we need to take the context of a cross-cultural interaction into account, and to specify the requirements for different application scenarios. For example, depending on the purpose for which the NVC translator is being used, different translation modes and output modalities may be required. If the system was used to facilitate online meetings of multicultural work groups, a multi-user setting would be required with multiple translation mechanisms running at the same time considering each group member's culture. The NVC translator may also be used as a facilitating tool in conflict resolution between members rival (cultural) groups. As technology-enhanced conflict resolution is still very new, the potential benefits and pitfalls need to be considered carefully, and evaluated in empirical studies.

Although the NVC translator is still at an early prototype stage, these example application scenarios illustrate its great potential and immediate importance in bridging between cultures.

Acknowledgments. This research was funded by the European Commission under the ICT work programme 2009 (Grant No. FP7-ICT-248620). The first author was supported by an IEF grant from the European Commission (Grant No. PIEF-GA-2009-254277).

References

1. Andersen, P.A., Hecht, M.L., Hoobler, G.D., Smallwood, M.: Nonverbal Communication Across Cultures. In: W.B. Gudykunst (Ed.), Cross-cultural and Intercultural Communication, pp. 73--90. Sage Publications, Thousand Oaks, CA (2003)
2. Axtell, R.E.: Gestures: The Do's and Taboos of Body Language Around the World. John Wiley & Sons, New York, NY (1998)
3. Armstrong, N., Wagner, M.: Field Guide to Gestures. How to Identify and Interpret Virtually Every Gesture Known to Man. Quirk Books, Philadelphia, PA (2003)
4. http://soc302.tripod.com/soc_302rocks/id6.html
5. http://westsidetoastmasters.com/resources/book_of_body_language/toc.html
6. Maniar, N., Bennett, E.: Designing a Mobile Game to Reduce Culture Shock. In: International Conference on Advances in Computer Games Technology, pp. 252--253. ACM Press, Lisbon, Portugal (2007)
7. Johnson, W.L., Marsella, S., Mote, N., Viljalmsson, H., Narayanan, S., Choi, S.: 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, Italy (2004)
8. Warren, R., Diller, D.E., Leung, A., Ferguson, W., Sutton, J.L.: Simulating Scenarios for Research on Culture and Cognition Using a Commercial Role-play Game. In: Winter Simulation Conference, pp. 1109--1117, Orlando, FL (2005)
9. Veterbi, A.J.: Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. IEEE Transactions on Information Technolgy 13, pp. 260--269 (1967)