Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 00 (2015) 000–000

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015

Cross-cultural training analysis via social science and computer vision methods

Peter Tu[a], Jixu Chen[a], Ming-Ching Chang[a], Ting Yu[a], Tai-Peng Tian[a], Gabriela Rubin[b], Julia Hockett[b], Aubrey Logan-Terry[b]

[a]*General Electric, 1 Research Circle, Niskayuna NY 12309, USA*
[b]*Georgetown University, 3700 O St. NW, Washington DC 20057, USA*

Abstract

Computer Vision technology is an invaluable addition to cross-cultural communication training for military personnel. It allows trainers to assess trainees in real time and provide feedback grounded in social science research. The present study reports on a joint analysis of military cross-cultural training data by Computer Vision specialists from GE Global Research as well as analyses from Georgetown University's Social Interaction Research Group (SIRG). Data for this study were collected over 10 days at the Army Infantry Basic Officer Leaders Course (IBOLC). 80 lieutenants participated in classroom role-play scenarios designed to assess their ability to communicate cross-culturally. GE and SIRG researchers video-recorded interactions among the role players and Soldiers and correlations were observed between these automatic interpretations and those gleaned by the SIRG analysis team in order to augment understanding of the efficacy of the cross-cultural training. For the Computer Vision methods, Each person was represented as a stream of visual cues which include: position, articulated motion, facial expressions and gaze directions. The social science researchers conducted multimodal (including embodied elements such as eye gaze, hand gestures, and body positioning), mixed method (qualitative and quantitative) discourse analyses of the data. SIRG researchers developed a coding scheme, marking specific human behavioral features within each interaction. From such coding, SIRG identified key skills in cross-cultural interaction, including observation and adaptation to unfamiliar communicative norms, rapport building, and trouble recovery (for details see Logan-Terry & Damari, forthcoming). Various correlations between raw computer vision measurements and the social science coding scheme was observed. Such results represent a significant step towards establishing the efficacy of the joint analysis of automated Computer Vision and established social science methods with regards to complex social interaction analysis.

Peer-review under responsibility of AHFE Conference.

Keywords: Computer vision; Social signal processing; Embodied communication; Sociolinguistics

Peer-review under responsibility of AHFE Conference.

1. Introduction

The relevance of non-verbal communication, such as eye gaze, body positioning, gestures, and facial expressions, has been widely studied in the fields of social science and computer vision. Social signal processing (SSP) [1] is a research field within the discipline of computer vision that focuses on enabling computers to interpret non-verbal cues for the purposes of interpreting human social interactions in an automatic fashion. Computer vision techniques such as facial expression analysis [2], gaze direction estimation [3,4] and gesture recognition [5] have been successfully applied to social interaction analysis. However, there exist two major challenges in the development of SSP. First, the automatic extraction of non-verbal cues from raw video data is a challenging task [6,7]. Second, studies in computer vision and social science often make use of different nomenclatures and modes of analysis. Within the field of social science, sociolinguistics researchers use the umbrella term 'embodied communication' to denote the study of non-verbal features of language, and their usage to show involvement [8], build rapport [9], complete an action [10] and send implicit messages, termed 'metamessages', that help the addressee interpret what was said [11,12]. For the purposes of consistency in this paper, we will be utilizing the social science term 'embodied communication' when discussing non-verbal features of language. In this paper, researchers from social science and computer vision fields jointly perform a case study on a military social interaction/cross-cultural communication training activity, in order to combine our research backgrounds and methodologies with the goal of bridging the gap between these two disciplines.

We investigate the use of a computer vision system in evaluating the efficacy of social interaction/cross-cultural communication in a military training course. The need of social intelligence in modern military operations is rising, as military personnel are often called to act as street-level diplomats and negotiators, where dynamic social proficiencies are the key for mission success. These high-stakes encounters can be shaped by the potential threat of violence on all sides, making interactional success an issue of life or death. These training courses serve as a low-stakes environment for the trainees to practice intercultural communication strategies, and learn from their errors in a safe, controlled environment. The goal of the evaluation software, developed as part of a larger Defense Advanced Research Projects Agency (DARPA) funded project, is to facilitate this kind of social interaction training by leveraging an automatic visual system to provide on-line social interaction evaluation to improve course design and feedback, and investigate the correlation between machine distilled visual cues and assessment obtained from social scientists.

This interdisciplinary study resulted in both macro-level and micro-level analyses: the computer vision researchers conducted a macro-level analysis, focusing on aggregate levels of rapport and hostility simultaneously from all individuals in the scenario. The social scientists conducted a discourse analysis that focused on the individual interlocutors, analyzing both verbal and embodied communication to assess the trainee's usage of 'good stranger behavior' [13]. We found that rapport levels in the computer vision approach correspond to the social science rapport-building codes, particularly those for greetings and gift giving. In addition, we found that the discourse analytic approach is better equipped to differentiate between a source of trouble and a remedy to a source of trouble, often confused in the computer vision approach, while the computer vision approach succeeds at identifying occasions of non-explicit sources of trouble that are not referred to by the interlocutors and thus cannot be coded by the social scientists. These findings have implications for the improvement of social interaction/cross-cultural communication training in military settings, as well as improvements in joint methodologies for social science and computer vision researchers.

2. Methodology

2.1. Previous methods

There is a wide breadth of social science/sociolinguistic research on embodied communication. Across cross-cultural contexts, gestures can be used in order to aid or enhance understanding, indicate the topic of conversation, identify an addressee, and convey meaning. Gestures range from iconic 'emblems' to less regimented 'gesticulation' accompanying talk [14]. They are often conventionalized, and are linked not only to linguistic structure but also to other social and cultural aspects of speech. Research has shown that gestures can inform, add emphasis, and function as part of situationally appropriate rituals such as greeting or departing rituals [10,15,16,17,18].

Eye gaze is another feature of embodied communication that has been studied extensively in social science and linguistics research. Eye contact is key in determining addressivity and engagement [8,19]. It allows speakers to identify their addressee and gauge interest and involvement, while simultaneously allowing listeners to demonstrate their attentiveness to talk [20]. Backchannels, such as nodding or shaking one's head or saying "uh-huh", are verbal and embodied signals that can also indicate listenership and engagement in the conversation [21]. However, the same features can also be used to indicate sympathy, agreement or disagreement. This can lead to variation in how backchannels are interpreted.

Additionally, body positioning and movement can indicate degrees of intimacy and relative affiliations of participants [19,22] Coding variables for proxemics tend to include: distance, postural identifiers (i.e. sitting, standing), and orientation of frontal body plane (i.e. degree one faces another) [23,24]. Culturally appropriate distance has been shown to enhance persuasion and likability, while misunderstandings and withdrawals occur when the appropriate distances is misjudged [25]. Studies have also found a relationship between body movement and hostility or perceived threats. One study found that in situations of perceived threat, participants would increase the distance between themselves and the source of the perceived threat, creating a personal buffer zone [26]. Observing and mirroring these types of embodied communicative behaviors, especially in situations with varying norms, can be instrumental in the creation of rapport and interpersonal involvement, which contribute to successful interaction [9,27].

Computer vision methods for the purposes of interpreting human behaviors have been successfully applied to social interaction analysis. Visual cues that can be automatically harvested include: facial expressions [2,28], pupil motion/gaze direction [3,4] as well as body motions and gestures [5]. For the purposes of group level social interaction analysis, an opportunistic standoff multiple-camera sensing platform capable of capturing various visual cues from live video streams has been developed [29]. The system uses sets of fixed RGB+D cameras that allow for reliable person detection and tracking. In addition a ring of Pan-Tilt-Zoom (PTZ) cameras automatically target individuals so as to capture high-resolution facial shots for facial expression and gaze analysis.

2.2. Methodology

Our data collection took place at the Army Infantry Basic Officer Leaders Course (IBOLC) at Fort Benning. This data collection occurred as part of a larger, DARPA-funded interdisciplinary project called Strategic Social Interaction Modules (SSIM). The overall goal of the SSIM project is the development of methods for the teaching and subsequent measurement of interactional skills related to effective cross-cultural communication skills for military and law enforcement personnel.[1]

In the training course, uninstructed trainees participated in a set of challenging role-play scenarios, which are designed to assess their communication and decision-making skills [30]. There were approximately eighty lieutenant participants that took part in this social interaction/cross-cultural competence training. While this training took place in a classroom, both the trainee and the role-players could operate freely in the scenario. The two scenarios we analyzed for this paper were titled "Cafe Conundrum" and "La Comandanta". *Cafe Conundrum* required the trainee to pick up on implicit signals of distress from a civilian with whom they had already built rapport, and locate the hostile insurgent who is the source of that stress. *La Comandanta* challenged the trainees to build rapport with a local militia leader with whom there was minimal shared linguistic repertoire. Their mission was to disarm the militia without losing the militia leader's respect and support. Each iteration of the scenarios was recorded, both by the computer vision cameras and by the social science researchers.

For the computer vision methods, the training facility was instrumented with a collection of fixed and PTZ cameras. Tracking was performed so as to compute the location of both the training subject and all of the different role players. PTZ cameras were then tasked with collecting high-resolution facial imagery of all individuals. Each

[1] DARPA SSIM website: http://www.darpa.mil/Our_Work/BTO/Programs/Strategic_Social_Interaction_Modules_SSIM.aspx

person was then represented as a stream of visual cues, which include: position, articulated motion, facial expressions and gaze directions. PTZ cameras characterize each individual's gaze directions and facial expressions including anger, fear, joy, surprise, and frustration on a per-frame basis in real-time.[2] Based on these measurements, various aggregate statistics were computed on a frame by frame basis resulting in measures of four visual analytics, namely affect, proximity, engagement and body motion. These visual analytics are normalized to the scale of [0,1]. (I) Emotional affect is a pooling of expressions (smile and frustration) extracted from the participants of the group. (II) Proximity is calculated using the average distance of each participant to the group center. (III) Engagement captures whether the group is sufficiently engaged in the social interaction. We use the gaze direction of each participant as an indication of engagement. If most of the participants are looking roughly in the direction of the group center then we deem that engagement level to be high. (IV) Activity/Motion is an expressive cue for social interaction, e.g., gesturing and uneasy shifting. It is designed to estimate the number of individuals that are highly "animated".

Graphical models for instantaneous group level concepts such as rapport and hostility were applied to these visual analytics in a continuous fashion. We model rapport and hostility as two latent variables in a Hidden Markov Model (HMM). At each time step the probabilities of hostility and rapport are estimated from observed visual analytics.

Social scientists at the Social Interaction Research Group (SIRG) in Georgetown University's Department of Linguistics conducted multi-modal, mixed method discourse analyses of the data, using both audio and video data to gain a deeper understanding of the interaction. The general framework for the qualitative analysis is grounded in Interactional Sociolinguistics [12,27,31,32,33], Conversation Analysis [34,35], Ethnomethodology [36], and Pragmatics [37,38,39]. For the larger DARPA-funded project, SIRG researchers had developed a coding scheme, marking specific human behavioral features within each interaction. From such coding, SIRG identified three overarching key skills in cross-cultural communication: observation and adaptation to unfamiliar communicative norms, rapport building, and trouble recovery.[3] Within those codes, the videos were flagged for a number of specific features that supported that skill, such as 'greeting', 'gift giving', or 'use of local language'. Cohen's Kappa was performed in order to establish intercoder reliability, and descriptive statistics were run using SPSS. The coding, completed using the software ELAN, was then sent to the computer vision researchers for comparison.

Initially comparison of the two methodologies was achieved via the construction of a software package that allows for visualization of both computer vision measurements along with social interaction codes. As shown in Figure 1, this system replays video and shows rapport and hostility levels (as a probability) and visual cues (affect, proximity, etc.) in parallel with social interaction codes (e.g. "gift offering"). This was followed by analyses of the correlations between these observations both at micro and macro levels.

---

[2] See Chang et al. (forthcoming) for further details on this system [29].
[3] See Logan-Terry & Damari (2015) for further details on these key skills [13].

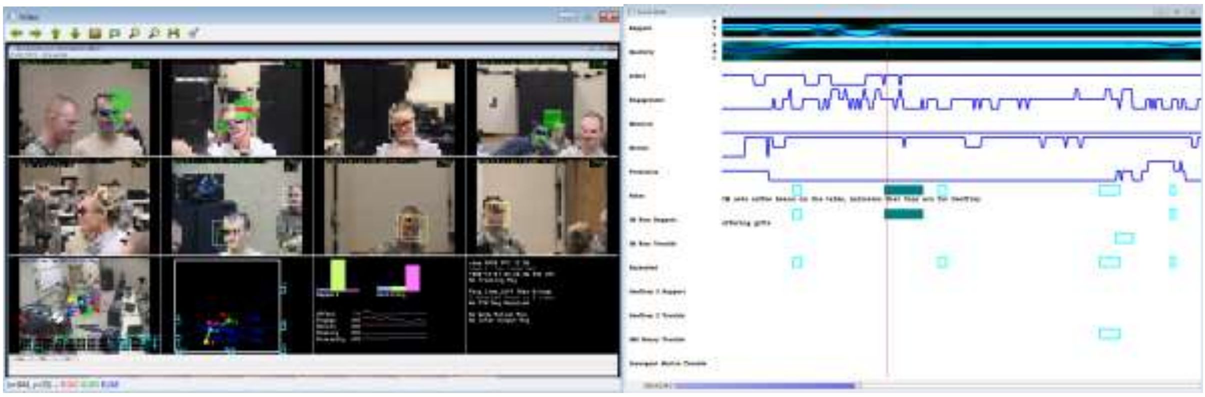

Fig. 1. Social Interaction analysis software for comparison of visual cues and social interaction codes.

3. Results and discussion

During the analysis of the results of the two studies, researchers found that rapport levels in the computer vision (CV) approach are related to social science rapport-building codes, particularly those for greetings and gift giving. In an example of this correlation, the *La Comandanta* scenario presents the trainee with an offer of food. In an iteration in which the trainee gracefully accepts the gift, the CV rapport levels increase while the social science analysis marks the offer of the gift as a rapport-building action. In order to characterize this observation numerically, we compared the CV rapport levels and visual cues before and after typical rapport-building actions, namely “gift giving”, “name giving” and “asking for chair”. We observed that, after these actions, the average rapport level increases from 0.41 to 0.54, while the average hostility level decreases from 0.81 to 0.77. We also observed increases in the average value of several visual analytics (affect increases from 0.53 to 0.68, proximity increases from 0.47 to 0.65, engagement increases from 0.33 to 0.35).

In the area of problematic elements in the scenarios, the hostility levels found in the CV approach correlate with the social science coding for interactional trouble; in particular, our findings show decreases in CV hostility generally occur after a trainee engages in what was coded by social science researchers as an ‘interactional trouble remedy’. This is displayed in an iteration of the *La Comandanta* scenario, as the trainee receives a radio call from his commander during the interaction. This is categorized in the social science analysis as a source of interactional trouble, as the civilian is visibly displeased by the action, while the CV analysis identifies this action as an increase in the hostility in the room. The trainee’s recovery, in the form of a rapid ending of the call and return to the interaction in response to the civilian’s displeasure, is also marked in both analyses; the social science coding marks the recovery action as a resolution of the previous trouble source, and the CV analysis finds a decrease in the room-wide hostility level.

In short, CV approaches the scenario looking to calculate the aggregate levels of rapport and hostility in the room, as participants who are not engaged in the interaction can affect it through embodied communication. For example, the *Cafe Conundrum* scenario is designed to have a hostile insurgent observing the scene upon the trainee’s entrance. As such, the positioning and embodied cues from the insurgent have an increasing effect on the aggregate level of hostility in the scenario. The social science approach balances this macro-analysis with a micro-level analysis; it focuses on specific actions and engagements on the part of interlocutors. In the above example, the social science analyses focus on the specific actions of the trainee in response to the increased hostility in the room, rather than the general atmosphere.

The two approaches are complementary; each notes certain interactional and physical cues that the other misses. Specifically, the CV approach succeeds at identifying occasions of non-explicit sources of trouble that escape the social science analyses, while the social science approach is better equipped to differentiate between a source of trouble and a remedy to a source of trouble, often confused in the CV approach. In cases of the former, the social science approach does not identify sources of interactional trouble unless an interlocutor explicitly brings up the issue, thus creating a sense of moments of trouble that are allowed to pass without comment. An instance of this is in the *La Comandanta* scenario, requiring the trainee to perform a formal embrace as a culturally appropriate greeting. The trainee attempts to perform a handshake, which is culturally inappropriate greeting. Yet, as no interlocutor attempts to correct the greeting, the social science analysis merely takes note of the misstep. The CV approach, here, is able to provide a broader analysis of the room at large, thus capturing such moments in the broader measures of rapport and hostility. In cases of the latter, the social science approach is able to make distinctions based on knowledge of the interaction and the outcome. For example, the *Cafe Conundrum* scenario involves a physical altercation between two role players, requiring the trainee to diffuse the situation and, ideally, recover the interaction. In one iteration of this training scenario, a trainee uses both palms to gesture downwards in an attempt to lessen interactional tension. This tactic is successful, de-escalating the situation and regaining some of the rapport lost in the altercation. In this, the social science research approach was able to express the recovery action and the subsequent recovery of rapport, whereas the CV approach merely saw continued high levels of hostility due to the aggressive positioning of the interlocutors.

Going beyond the correlation between temporally local CV and social science observations, we now consider the ability of a CV system to predict an overall assessment of a given engagement as made by social scientists. For the social interaction/cross-cultural communication training course, we focus on the behavior of the trainee instead of the whole group. A regression model between raw computer vision measurements and the social science coding scheme focus on trainee's performance is derived. Two social scientists independently rate the social interaction for each trainee with respect to overall "rapport" level (from 1 to 5) by walking through the videos. Judgments are made according to the validated non-verbal cues of Bernieri et.al. [7]. The CV measurements (Table 1) are extracted from trainee and role-players [30] respectively.

Table 1. Computer vision measurement of trainee and role-players.

S_T	Time percentage of trainee with smile expression.
FT	Time percentage of trainee with frustration expression.
S_R	Average time percentage of role-players with smile expression.
F_R	Average time percentage of role-players with frustration expression.
D_L	Time percentage when subject is very close (< 0.6meter) to role-players.
D_H	Time percentage when subject is far (> 1.8 meter) from role-players.

All measurements are normalized to the range of (0,1]. Thus, the visual cues of a training session are characterized by a 6-dimensional vector $= [S_T, F_T, S_R, F_R, D_L, D_H]$. We use half of the data (40 subjects) to train a simple linear regressor $F(V) = W \cdot V$ to predict the rapport level (R) labeled by social scientists. The linear coefficient W can be obtained at the learning stage, by minimizing the prediction error in the training data: $W = argmin_W \sum_{i=1}^{N} \|F(V_i) - R_i\|$, where N is the number of training subjects. The minimization is solved using a standard SVD approach.

We applied this regression model to predict the rapport level of the remaining dataset. A positive correlation (0.35, $p < 0.05$) was observed between the predicted and reported rapport level. The CV results found that positive emotion emanating from the role players (smiling) is the strongest indicator of rapport, while frustration of both trainee and role-players is negatively correlated with rapport. However, smiling of the trainee is negatively correlated with rapport, which is counter-intuitive. However, the analysis found that low rapport measures resulted when trainees continued to smile even though the role-players were visibly frustrated.

Such results represent a significant step towards establishing the efficacy of the joint analysis of automated computer vision and established social science methods with regards to complex social interaction/cross-cultural communication analysis. In terms of social interaction/cross-cultural communication training for military personnel, these types of analyses can be used to improve scenario design to optimize the trainee's ability. These types of analyses illustrate potential for a shift toward real-time feedback on specific elements of the training interaction, allowing for immediate evaluation of the relative success of the trainee's cross-cultural performance.

4. Conclusion

Via the joint analysis of over 80 role-playing engagements using state of the art computer vision algorithms and well-established social science methodologies, the first steps towards understanding the efficacy of a combined approach have been made in this paper. Correlations between specific CV and social science observations were both observed and measured. While the semantic precision associated with direct human observation is still beyond the capabilities of automated methods, CV's ability to simultaneously observe all individuals at all times results in a form of real-time continuous measurement that for many applications would not be practical via manual methods.

Going forward, social scientists armed with these new forms of automatic measurement capabilities will be able to develop and verify new models of human interaction and cross-cultural communication skills which in turn will enable machines to provide real time feedback to participants and other stake holders. This form of automation will go beyond training and may one day result in a new form of situational awareness that can be used to monitor and facilitate a great variety of social interactions and cross-cultural communications that take place in our day to day lives. In this way, the present paper points to important methodological as well as real world implications of these types of state-of-the art joint research efforts.

Acknowledgements

This research was sponsored by the Defense Advanced Research Projects Agency and the US Army Research Office, through contracts to Aptima, Inc. (Contract W911NF-11-C-0266), via subcontract to Georgetown University as well as contracts with General Electric (Contract D13PC00002). The information presented here does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

[1] Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and Vision Computing, 27, 1743–1759.
[2] Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions On, 31(1), 39–58.
[3] Vertegaal, R., Slagter, R., van der Veer, G., & Nijholt, A. (2001). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 301–308.
[4] Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. Pattern Analysis and Machine Intelligence, IEEE Transactions On, 32(3), 487–500.
[5] Metaxas, D., & Zhang, S. (2013). A review of motion analysis methods for human nonverbal communication computing. Image and Vision Computing, 31(67), 421–433.
[6] Grahe, J. E., & Bernieri, F. J. (1999). The Importance of Nonverbal Cues in Judging Rapport. Journal of Nonverbal Behavior, 23(4), 253–269.
[7] Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. Journal of Personality and Social Psychology, 71(1), 110.
[8] Goodwin, C. (1981). Conversational Organization: Interaction between Speakers and Hearers. New York: Academic Press.
[9] Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? Journal of Memory and Language, 66(1), 249–264.
[10] Kendon, A. (2004). Gesture: Visible Action as Utterance. Cambridge University Press.
[11] Bateson, G. (1955). A theory of play and fantasy. Psychiatric Research Reports, 2, 39–51.
[12] Tannen, D. (1984). Conversational Style: Analyzing Talk among Friends. Oxford University Press.
[13] Logan-Terry, A., & Damari, R. (2015). Key culture-general interactional skills for military personnel. This volume.
[14] Haviland, J. B. (2005). Gesture as cultural and linguistic practice. In Linguistic Anthropology. Oxford: EOLSS Publishers.
[15] Knight, N. (2009). Roots of human sociality: culture, cognition and interaction – Edited by N.J. Enfield & Stephen C. Levinson. Journal of the Royal Anthropological Institute, 15(4), 896–897.
[16] Goodwin, M. H. & Goodwin, C. (1986) Gesture and coparticipation in the activity of searching for a word. Semiotica, 62, 51-75.
[17] Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. Psychological Bulletin, 137(2), 297–315.
[18] McNeill, D. (1992). Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press.
[19] Kendon, A. (1990). Conducting Interaction: Patterns of Behavior in Focused Encounters. CUP Archive.
[20] Kidwell, M. (2009). What Happened? : An Epistemics of Before and After in "At-the-Scene" Police Questioning. Research on Language & Social Interaction, 42(1), 20–41.

[21] Yngve, V. (1970). On getting a Word in Edgewise. In Papers from the 6th Regional Meeting (pp. 567–578).
[22] Lebaron, C. D., & Streeck, J. (1997). Built Space and the Interactional Framing of Experience During a Murder Interrogation. Human Studies, 20(1), 1–25.
[23] Hall, E. T. (1963). A System for the Notation of Proxemic Behavior. American Anthropologist, 65(5), 1003–1026.
[24] Harrigan, J. A., Rosenthal, R. E., & Scherer, K. R. (2005) The new handbook of methods in nonverbal behavior research (Vol. viii). New York, NY, US: Oxford University Press.
[25] Hayduk, L. A. (1978). Personal space: An evaluative and orienting overview. Psychological Bulletin, 85(1), 117–134.
[26] Albas, D. C., & Albas, C. A. (1989). Meaning in Context: The Impact of Eye Contact and Perception of Threat on Proximity. The Journal of Social Psychology, 129(4), 525–531.
[27] Tannen, D. (2007). Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse. Cambridge University Press.
[28] Chen, J., Chang, Ming-Ching, Tian, T.-P., Yu, T., & Tu, P. (Forthcoming). Bridging computer vision and social science; a multi-camera vision system for social interaction training analysis. Manuscript Submitted to ICIP.
[29] Chang, M.-C., Chen, J., Tian, T.-P., Yu, T., & Tu, P. (Forthcoming). A live video social interaction analysis system. Manuscript Submitted to ICIP.
[30] Rubin, G., Logan-Terry, A., Damari, R., & Hockett, J. (2015). Practicing high-stakes communication: Role-play as a tool for military cross-cultural training. Poster presented at the Cross-Cultural Decision Making Conference, Las Vegas, NV.
[31] Gumperz, J. J. (1978). The conversational analysis of interethnic communication. In E. L. Ross (Ed.), Interethnic communication. Athens, Ga.: University of Georgia Press.
[32] Gumperz, J. J. (1999). On interactional sociolinguistic method. In S. Sarangi & C. Roberts (Eds.), Talk, work and institutional order: Discourse in medical, mediation and management settings (pp. 453–471). Berlin: Mouton de Gruyter.
[33] Tannen, D. (1981). New York Jewish Conversational Style. International Journal of the Sociology of Language, 30, 133–149.
[34] Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), 696–735.
[35] Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. Language, 53(2), 361–382.
[36] Garfinkel, H. (1984). Studies in ethnomethodology. Malden, MA: Polity Press/Blackwell Publishing.
[37] Grice, H. P. (1989). Studies in the way of words. Cambridge, Mass.: Harvard University Press.
[38] Searle, J. R. (1969). Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press.
[39] Austin, J. L. (1962). How to do things with words. Oxford: Clarendon Press.