Investigating non-verbal behaviors conveying interpersonal stances

Mathieu Chollet[1] and Magalie Ochs[2] and Catherine Pelachaud[2]

[1]Institut Mines-Telecom ; Telecom Paristech , mchollet@telecom-paristech.fr
[2]Institut Mines-Telecom ; Telecom Paristech ; CNRS-LTCI {mochs, cpelachau}@telecom-paristech.fr

Abstract. The study of the expression of affects and their expression by Embodied Conversational Agents is complex. This is because affects are expressed by non-verbal behaviors on a variety of different modalities, and that these behaviors are influenced by the context of the interaction and other interactants' behaviors. To overcome these challenges, we present a multi-layer framework and apply it to the study interpersonal stance dynamics. For this purpose, we built a corpus of non-verbal behavior annotations and interpersonal stance traces for job interviews.

Introduction - Embodied Conversational Agents (ECAs) are increasingly used in training and serious games. In the TARDIS project[1], we aim to develop an ECA that acts as a virtual recruiter to train youngsters to improve their social skills. Such a virtual recruiter should be able to convey different interpersonal stances, "*spontaneous or strategically employed affective styles that colour interpersonal exchanges* [1]": our goal is to find out how interpersonal stance is expressed through non-verbal behavior, and to implement the expression of interpersonal stance in an ECA. As a representation for interpersonal stance, we use Argyle's attitude dimensions [2], friendliness (also called warmth or affiliation) and dominance (also called agency).

Challenges in non-verbal behavior interpretation - A challenge when interpreting non-verbal behavior is that every non-verbal signal can be interpreted with different perspectives: for instance, a smile is a sign of friendliness [3]; however, a smile followed by a gaze and head aversion conveys embarassment [4]. Non-verbal signals of a person in an interaction should also be put in perspective to non-verbal signals of the other participants of the interaction: an example is posture mimicry, which can convey friendliness [5]. Finally, the global behavior tendencies of a person are important when interpreting their stance [6]. Another challenge is that, as Scherer points out [1], all kinds of affect dont happen in the same span of time. For instance, emotions have a very short duration, and to assess a persons emotion, one should only look at this persons very recent displays of emotion in their non verbal behavior. For moods, one has to look at a persons non-verbal behavior on a longer time span. It might get even longer to get a good sense of someones interpersonal stance. These different perspectives and time spans have seldom been studied together, and this motivates the use of multimodal corpora in order to analyze their different impact in a systematic fashion. To handle these challenges, we introduced a multi-layer model and a multimodal corpus we collected to study it. This model will be used to drive a computational model of interpersonal stance expression for an ECA.

Interpretation of signals using a multi-layer approach - In [7], we defined a multi-layer model to encompass the different non-verbal behavior interpretation perspectives. The *Signal* layer looks at one signal, its characteristics and their immediate interpretation. In the *Sentence* layer, we analyze the sequence of signals happening in a dialogue turn. The *Topic* layer focuses on the

[1] http://tardis.lip6.fr/

behavior patterns and tendencies happening in parts of the interactions where a topic is discussed (e.g. greetings, discussing job offer, saying goodbye...). Finally, the *Interaction* layer encompasses the whole interaction and looks at global behavior tendencies. These different layers allow to interpret interactants' interpersonal stances at every instant of the interaction, taking into account their behavior, their reactions to other interactants' behaviors, and their global behavior tendencies.

Multimodal corpus of interpersonal stance expression - In order to study how recruiters express interpersonal stance, we annotated three videos of job interview enactments, for a total of slighty more than 50 minutes. We consider full body non-verbal behavior, turn-taking, task and interpersonal stance [8]. Non-verbal behavior and interactional state annotations consist of labeled time intervals, while interpersonal stance was annotated as a trace. Trace data is prone to certain issues, such as scaling and reliability issues, thus we chose to follow Cowie's approach [9] by switching from an absolute perspective (i.e. what is the value of interpersonal stance at time t) to a relative perspective (i.e. is the interpersonal stance staying stable or is it changing at time t). This process allows to segment where the rater's perception of interpersonal stance varies or stays stable.

Investigating the perception of non-verbal behavior produced by an ECA - Our future work consists of investigating how behaviors contribute to the expression of interpersonal stance at every layer level, and how the four layers contribute to the perception of the interpersonal stance at every instant. For this, we intend to use our corpus as training data to learn the parameters of the model. The multi-layer model will then be implemented into our ECA as a behavior planning module which will receive target interpersonal stances as an input, and then compute which behaviors to express in order to reach these targets.

Conclusion - The complexity of non-verbal behavior expression and interpersonal stance perception in specific contexts motivates the use of a framework that considers all perspectives of behavior interpretation, and of a multimodal corpus as ground truth. We have proposed a multi-layer framework to handle the complexity of interpersonal stance expression, and we annotated videos of job interview enactments. Our future work consists of tuning our model using the multimodal corpus, and to then implement it as a behavior planner in our ECA platform before validation.

Acknowledgment - This research has been partially supported by the European Community Seventh Framework Program (FP7/2007-2013), under grant agreement no. 288578 (TARDIS).

References

1. Scherer, K.R.: What are emotions? and how can they be measured? Social Science Information **44** (2005) 695–729
2. Argyle, M.: Bodily Communication. University paperbacks. Methuen (1988)
3. Burgoon, J.K., Buller, D.B., Hale, J.L., de Turck, M.A.: Relational Messages Associated with Nonverbal Behaviors. Human Communication Research **10**(3) (1984) 351–378
4. Keltner, D.: Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. Journal of Personality and Social Psychology **68** (1995) 441–454
5. LaFrance, M.: Posture mirroring and rapport. In Davis, M., ed.: Interaction Rhythms: Periodicity in Communicative Behavior, New York: Human Sciences Press (1982) 279–299
6. Escalera, S., Pujol, O., Radeva, P., Vitria, J., Anguera, M.: Automatic detection of dominance and expected interest. EURASIP Journal on Advances in Signal Processing **2010**(1) (2010) 12
7. Chollet, M., Ochs, M., Pelachaud, C.: Interpersonal stance recognition using non-verbal signals on several time windows. In: Workshop Affect, Compagnon Artificiel, Interaction. (November 2012)
8. Chollet, M., Ochs, M., Pelachaud, C.: A multimodal corpus approach to the design of virtual recruiters. In: Workshop Multimodal Corpora, Intelligent Virtual Agents. (2013) 36–41
9. Cowie, R., Cox, C., Martin, J.C., Batliner, A., Heylen, D., Karpouzis, K.: Issues in Data Labelling. Springer-Verlag Berlin Heidelberg (2011)