Rishabh Chatterjee

Literature Review

Non-Verbal Behaviors and Rapport

October 8, 2017

The nature of rapport can be described as a dynamic structure of three interrelating components: mutual attentiveness, positivity, and coordination (Tickle-Degnen & Rosenthal, 2017). The rapport levels between students and tutors, and the learning outcomes of the student are positively related (Perkins, Schenk & Stephan, 1995). Keeping this in mind, the goal for Virtual Agent Tutoring Systems (VATS) today should include increasing the rapport level between itself and the student along with teaching.

Rapport may be measured using verbal (tonal quality, pitch, etc.) and non-verbal (facial gestures, posture, etc.) behaviors. This paper will focus on the latter, trying to find patterns with rapport levels and different non-verbal behaviors, and ultimately suggesting a set of non-verbal behavior rules which VATS can follow to increase rapport levels, thereby having a positive effect on the learning outcomes of the students. In order to design such tools, it is pivotal to first understand the problem space of non-verbal behaviors and rapport and identify the gaps in the current space.

The context of the non-verbal behavior and the duration of the same, play a deciding role on how the behavior is perceived. For example, smiling, a generally well-meaning gesture can be interpreted as mocking if used by the VATS when the students is unable to solve a question. Similarly, gazing at a student for too long may be perceived as staring and may make the situation rather uncomfortable for the student. In all the non-verbal behaviors being explored in the paper, I will address the question of when, and for how long to carry out the gesture in order to positively affect the rapport level.

As Admoni & Scassellati (2017) observed, gaze can be divided into four different categories: mutual, referential, joint, and aversion. In the context of this paper, mutual gaze is eye gaze that is directed from the VATS to the student’s eyes or face, and vice versa. The essential component of this type of gaze is reciprocity. The reciprocity makes the mutual gaze a two tier process – generation of eye contact from the VATS, and detection of eye contact from the student. Referential gaze, often referred to as deictic gaze, is gaze that is directed at an object or location in space. This type of gaze may be observed when the object is spoken about (verbal cue), but may also occur independent of any verbal reference. Joint attention involves sharing focus on a common object (Moore & Dunham, 2014). It may be a result of a sequence of events, like mutual gaze followed by referential gaze, followed by mutual gaze again to ensure that the experience is shared. Gaze aversions are shifts of gaze away from the main direction of gaze (the partner’s face). Gaze aversion is generally used to avoid eye contact, perhaps because the partner is feeling a bit uncomfortable in the current conversation. It may be perceived as a sign of low rapport levels. Note that if the student is gazing elsewhere while thinking about how to solve a question, but the VATS is looking at the student, it is not gaze aversion. Hence, when I talk of gaze aversion here it refers to intentional gaze aversion rather than coincidental / unintentional gaze aversions. I will also include another gaze type, gaze stare, to the aforementioned types. A gaze at the student’s face for more than a threshold period of time, and vice versa, will be categorized as a gaze stare.

A meta-analysis indicated that a participants’ evaluation of their partner’s level of positivity (indicative of rapport) was positively associated with smiling, nodding and forward posture (Schulman & Stephan, 2012). Nodding can be seen as a sign of coordination (Cassell, Gill, & Tepper) and approval of an action, perhaps a correct answer by the student. It can also be perceived as a sign of reassurance by the VATS of the student enquires whether he / she is solving the problem correctly. However, Schulman & Stephan also observed that stronger self-reported rapport is associated with less nodding in earlier conversations. Keeping this in mind, it must be determined when and to what extend the VAS should nod in order to increase rapport levels in the long run. Having somewhat similar connotations, forward posture represents and interest in the current conversation by the student. Since the student may have a higher interest in the current conversation, it is a good opportunity for the VATS to ask some questions or hold some conversation (educational or small talk) to increase rapport level.

Ochs, Niewiadomski & Pelachaud (2010) identified three different types of smiles by virtual agents in a context free scenario – amused, polite, and embarrassed. The most common smile a person exhibits is the amused, or Duchenne, smile. It signifies a “genuine” smile, a sign of enjoyment. Somewhat the opposite of the amused smile is the polite smile, also called the non-Duchenne smile. It is a social, or controlled smile used to mask actual feelings. Smiling in a negative situation is categorized as an embarrassed smile. Perceptual studies have shown that people unconsciously distinguish between amused and polite smiles (Frank, Ekman & Friesen, 1993). Keeping this in mind, it is essential to determine when the VATS should generate either type of smile since the student may feel undermined in situations like the VATS generating a non-Duchenne smile for approval of a question correctly solved, etc. The three types of smiles have some distinguishing morphological features that are mentioned in the paper by Ochs, Niewiadomski & Pelachaud (2010).

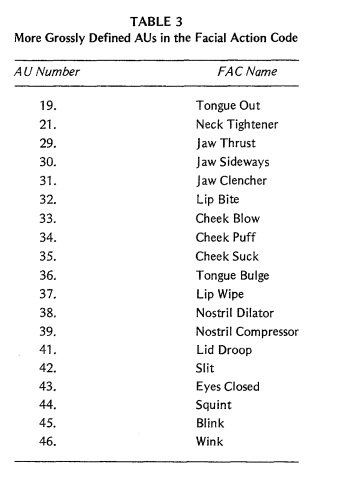
Ekman & Friesen (1976) worked on identifying facial action units (FAC) to attempt an approximation of the total repertoire of facial action including, but not limited to, the ones in Table 1. Of the FACs described, I will be looking into the correlation between eyebrow movement, nostril dilation, nose crinkling, and lip bite and rapport levels. More concretely, I will look into when does the student carry out these gestures, and when to generate them by the VATS to increase rapport levels.

Table : Ekman & Friesen (1976) FAC table 3

Zoric & Forchheimer (2011) worked on the generation of non-verbal behaviors form verbal cues. The main idea behind this model is that every verbal behavior, or utterance, has a corresponding non-verbal gesture associated with it. I will explore the possibility of the reverse direction between utterances and gestures – along with agreeing with the previous claim, I feel every non-verbal gesture has a corresponding verbal counterpart to it. For example, if the VATS is generating a Duchenne smile, the speech generated at that time must have a “soft” tonal quality, one that seems encouraging and supportive rather than an assertive verbal behavior. Similarly if the VATS is raising eyebrows to show doubt, the tonal quality and amplitude of the speech it generates must be different from those in the previous context.

The research by Cassell, Gill & Tepper (unknown year) has suggested that rapport levels are higher in friends than in strangers. The main difference between friends and strangers is that the former “knows” the person. Hence, if the VATS remembers the student with whom it has interacted before, it can perhaps try to start the next session at a higher rapport level than the previous one, thereby gradually increasing rapport level in the long run. The important question which needs to be answered in what about the student should the VATS remember in order to attain this gradually increasing rapport level at time 0 in the different sessions.

Drawing on all the readings about the problem space, I have the following questions that I wish to address in the paper.

* Which non-verbal behaviors have most positive impact on rapport levels and when and to what extent should the virtual agent use them?
* How to pair up non-verbal behaviors with verbal behaviors in order to get tuples of reactions that increase rapport level?
* What should the VATS “remember” and how to use that memory to positively affect rapport levels with the student in the long run?

While prior studies have partially worked on the automatic detection and generation of the some of the aforementioned non-verbal behaviors (Schulman & Bickmore, 2012; Zoric & Forchheimer, 2011), there is still no well defined set of rules or automatic process to determine when and how a VATS should use non-verbal behaviors to increase rapport levels between itself and the student. Along with this, what to remember about the student and how to use that in subsequent sessions is a field that has still not been explored. I intend to address these gaps in my research.

**References:**

1. Friesen, W. V. (1976). Measuring Facial Movement \*, *1*, 56–75.
2. Frank, M. G., Ekman, P., & Friesen, W. V. (1993). INTERPERSONAL RELATIONS AND GROUP Behavioral Markers and Recognizability of the Smile of Enjoyment, *64*(1), 83–93.
3. Ochs, M., Niewiadomski, R., & Pelachaud, C. (2010). How a virtual agent should smile? *Intelligent Virtual Agents*, 427–440. https://doi.org/10.1007/978-3-642-15892-6\_47
4. Perkins, D., Schenk, T. A., & Stephan, L. (1995). Learning O N Students ’ Perceived Ratings of College Instructors ’, 627–635.
5. Cassell, J., Geraghty, K., Gonzalez, B., & Borland, J. (2009). Modeling culturally authentic style shifting with virtual peers. In *Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI ’09*. https://doi.org/10.1145/1647314.1647338
6. Schulman, D., & Bickmore, T. (2012). Changes in verbal and nonverbal conversational behavior in long-term interaction. In *Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI ’12*. https://doi.org/10.1145/2388676.2388681
7. Admoni, H., & Scassellati, B. (2017). Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction*. https://doi.org/10.5898/JHRI.6.1.Admoni
8. Pejsa, T., Andrist, S., Gleicher, M., & Mutlu, B. (2015). Gaze and Attention Management for Embodied Conversational Agents. *ACM Transactions on Interactive Intelligent Systems*. https://doi.org/10.1145/2724731
9. Linda Tickle-Degnen, Robert Rosenthal, The Nature of Rapport and Its Nonverbal Correlates. (2017). https://doi.org/10.1207/s15327965pli0104\_1
10. Cassell, J., Gill, A. J., & Tepper, P. A. (n.d.). Coordination in Conversation and Rapport, 40–50.
11. Basic-Emotions\_Ekman. (n.d.).
12. Zoric, G., Forchheimer, R., & Pandzic, I. S. (2011). On creating multimodal virtual humans-real time speech driven facial gesturing. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-010-0526-y