Expressing social attitudes in virtual agents for social coaching

Hazaël Jones[1] Mathieu Chollet [2] Magalie Ochs[3]
Nicolas Sabouret[4] Catherine Pelachaud[3]

[1] Montpellier SupAgro; UMR ITAP
[2] Institut Mines-Telecom ; Telecom Paristech
[3] CNRS LTCI ; Telecom ParisTech
[4] LIMSI - CNRS ; Universit Paris-Sud

hazael.jones@supagro.inra.fr, nicolas.sabouret@limsi.fr,
{mathieu.chollet, magalie.ochs, catherine.pelachaud}@telecom-paristech.fr

Résumé

L'utilisation des agents virtuel pour le coaching social a connu une forte croissance ces dernires annes. Dans ce domaine, l'agent virtuel doit tre capable d'exprimer diffrentes attitudes sociales pour permettre l'utilisateur de s'entraner surmonter des situations de la vie relle. Dans cet article, nous proposons un modle d'attitudes sociales qui permet un agent virtuel de raisonner sur l'attitude sociale la plus approprie exprimer au cours d'une interaction avec l'utilisateur, en s'appuyant sur un modle d'agent affectif. L'expression de cette attitude se fait travers le comportement non-verbal.

Mots Clef

Attitudes sociales, motions, informatique affective, agent virtuel, comportement non-verbal.

Abstract

The use of virtual agents in social coaching has increased rapidly in the last decade. In social coaching, the virtual agent should be able to express different social attitudes to train the user in different situations than can occur in real life. In this paper, we propose a model of social attitudes that enables a virtual agent to reason on the appropriate social attitude to express during the interaction with a user given the course of the interaction, but also the emotions, mood and personality of the agent. Moreover, the model enables the virtual agent to display its social attitude through its non-verbal behaviour.

Keywords

Social Attitudes, Emotions, Affective computing, Virtual Agent, Non-verbal behaviour.

1 Introduction

Social coaching workshops constitute a common approach to help people in acquiring and improving their social competencies. The main difficulty with this approach is that it relies on the availability of trained practitioners as well as the willingness of the people to engage in exploring their social strengths and weaknesses in front of their peers and practitioners. For this reason, the use of virtual agents in social coaching has increased rapidly in the last decade [22, 2, 8]. However, most of the proposed models focus on the simulation of emotions and do not take into account the different social roles that the virtual agent may embody. Yet, given its role and the course of the interaction, the virtual agent should be able to express different social attitudes to train the user in different situations that can occur in real life. For this reason, one of the key elements of a virtual agent in the domain of social coaching is its ability to reason on social attitudes and to express them through its behaviour. This is why, in this research work, we propose a model of social attitudes for expressive virtual agents.

Social attitude can be defined as "*an affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons, colouring the interpersonal exchange in that situation (e.g. being polite, distant, cold, warm, supportive, contemptuous)*" [18]. As highlighted in [23], one's social attitude depends on one's personality but also one's moods that is directly influenced by the course of the interaction. One's social attitude is mainly conveyed by one's non-verbal behaviour [1].

Our aim is to develop a model of social attitudes that can be used by virtual agents to select determine expressive behaviours given its simulated affective state and the course of the interaction. The model we propose in this paper considers user's emotions, mood and personality and compute and display agent's appropriate social attitudes, based on classical work from the literature in affective computing [7, 14]. Moreover, it enables the virtual agent to display its social attitude through its non-verbal behaviour, based on models of social attitude expression for virtual agents such as [3]. Our model has been developed in the context of job interview (JI) simulation. This context is interesting for several reasons: 1) social attitude plays a key role in JI: the applicant tries to fit the social norm and 2) the recruiter is in a role-play game, which can be simulated with a virtual agent. Moreover, job interview is a type of social coaching situation with high social impact. The methodology used to develop such a model combined a theoretical and an empirical approach. Indeed, the model is based both on the literature in Human and Social Sciences on social attitudes but also on the analysis of an audiovisual corpus of job interviews and on post-hoc interviews with the recruiters on their expressed attitudes during the job interview.

2 General Architecture

Our interview simulation for social coaching involves two main actors, namely the participant (*i.e.* the person that is training on the system) and the interlocutor (*i.e.* person or virtual agent that respond to the trainee). In our platform, the interlocutor is replaced by a virtual agent. Although the model presented here is general and can be applied to different interaction situations, our corpus and the derived cognitive architecture and non-verbal behaviour are designed in the context of job interview simulations when the agent acts as recruiter.

Fig. 1 presents our general architecture.

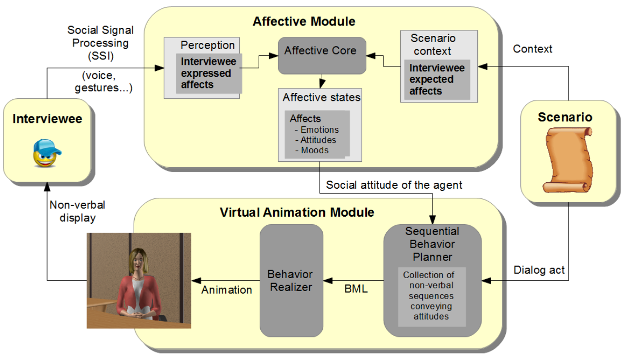

Figure 1: General architecture

This architecture is organized as follows. First, a *scenario* describes the context of the interaction and the expectations at the different stages of the interview in terms of affects expression[1]. For example, after a tough question, the interviewer will expect negative affects from the interviewee (distress, agitation). The *affective module* (described in ection 4) takes as inputs the scenario information and the detected social cues. It computes a social attitude for the virtual agent (the recruiter in our case). These attitudes are turned into non-verbal behaviours in the virtual agent *animation module* (Section 5) and used in the next interaction step. This module is based on the Greta platform [15] and is composed of an *intent planner* that generates the communicative intentions (what the agent intents to communicate) based on the scenario, and a *behaviour planner* that transforms the communicative intentions into a set of signals (e.g. speech, gestures, facial expressions) based on the agent's attitude and mental states.

In the following sections, we first present the corpus that was used to build our social attitude model. We explain how the corpus was collected, based on mock job-interviews, and it was annotated. We then give details of the Affective Module and the Animation Module and we show their interrelation.

3 Corpus

We have collected a corpus of real interpersonal interaction between professional recruiters and job seekers interviewees. The recordings consisted in creating a situation of job interviews between 5 recruiters and 9 interviewees. The setting was the same in all videos. The recruiter and the interviewee sat on each side of a table. A single camera embracing the whole scene recorded the dyad from the side. This resulted in a corpus of 9 videos of job interview lasting between 15 and 20 minutes each. We discarded 4 videos as the recruiter was not visible due to bad position of the camera. Out of the 5 remaining videos, we have so far annotated 3, for a total of 50 minutes and 10 seconds of video.

The non-verbal behaviours of the recruiters during the job interview have been annotated. We also annotate information on the interaction: the *turn taking* (i.e. who is speaking), the *topic of the discussion* (i.e. document related or general), the perceived affects of the interviewee (e.g. embarassed, relieved, bored...), and the attitude of the recruiter (i.e. the level of perceived dominance and friendliness of the recruiter). Different modalities of the recruiter's non-verbal behaviour have been annotated (e.g. the gaze behaviour, the gesture, the posture, the head movements, etc). The coding scheme, the resulting annotations, and the inter-annotators agreements are described in more details in [4]. These annotations have been used to construct the animation module for the virtual agent's expression of social attitudes (Section 4). Moreover, the annotation on the recruiter's social attitude has been used

[1]In our work, we use the SSI framework that recognizes the user's emotions expressed through his voice and his facial expressions.

for the evaluation of the affective model to check that it produces outputs that correspond to the human behaviour (Section 6).
In addition to these videos, post-hoc interviews with the recruiters were used to elicit knowledge about the expectations and mental states during the interview, following the methodology proposed by [16]. This knowledge was used in the affective module to select the relevant social attitudes and to set up the rules to give the capability to the virtual agent to select the appropriate social attitude to express.

4 Reasoning on social attitudes

The Affective Module is based on a set of rules that compute categories of emotions, moods and attitudes for the virtual recruiter, based on the contextual information given by the scenario and the detected affects (emotions, moods and attitudes) of the participant (Section 3). The computation of the virtual agent's emotions is based on the OCC model [14] and the computation of the agent's moods in the PAD space [13] is based on the ALMA model [7]. The personality is representy by a vector in the OCEAN space [5]. The details of the computation of emotions and moods will not be presented in this paper; it can be found in [10].

4.1 Virtual recruiter's social attitudes

Several research has shown that one's social attitude is influenced by one's affective state (e.g. his emotions and moods [6]) and the course of the interaction (e.g. the affective reaction of the other [23]). For instance in [23], Wegener et al. show that a positive mood can help influence a change of attitude in the interlocutor, and that people tend to feel a higher likelyhood toward interlocutors that are in a positive mood.
Relations between attitudes and personality and moods and attitudes [23] has been exhibit in literature. Although we cannot give all the details of these papers here, their results show that one's mood has an influence not only on the interlocutor's attitude, but also on one's own reaction to events. This knowledge has been turned into expert rules that compute values for the attitude of the virtual agent.

Computation of attitudes. The way we compute attitudes follow this principle: an agent can adopt an attitude according to its personality [20] and to its actual mood [23]. For example, an agent with a non-aggressive personality may still show an aggressive attitude if its mood becomes very hostile. The mood compensates the personality and vice versa. For this reason, we use a logical-OR as condition on these two dimensions. As a consequence, in our model, the attitude can be triggered by one of these two dimensions. Then, the maximum value (mood or personality) is kept to compute the corresponding attitude, as is classically done in Fuzzy logics. We also use a threshold θ to define the minimum value for a trait to have an influence on the attitude.
As an example, the definition of the attitude "friendly" is based on Agreeableness (as was shown by Costa [5]) and positive and arroused moods (as proposed by Isbister [9]), i.e. exhuberance in the PAD space: If $(P_a > \theta) \vee (M_p > \theta \wedge M_a > \theta)$, then:

$$val(friendly) = max(\frac{M_p + M_a}{2}, P_a)$$

with P_a the value of the Agreeable personality trait in the OCEAN model, M_p and M_a the values of pleasure and arrousal in the PAD space.
Using similar rules, we compute 7 categories of attitudes: friendly, aggressive, dominant, supportive, inattentive, attentive and gossip. Details on that computation can be found in [17].

4.2 Interpersonal circumplex

The non-verbal behaviour model of our agent, presented in the next section, does not work directly with the categories that were identified in the Knowledge elicitation phase of the corpus collection. It makes use of continuous values that relies on the annotation of corpus which uses the Friendly and Dominant dimensions of the interpersonal circumplex [9]. For example, the *aggressive* attitude is defined on the circumplex as a vector $(-0, 5, 0, 5)$ (friendly at -0.5 and dominant at 0.5).
To convert the attitudes represented by categories into continuous values of dimensions, we rely on the work by Isbister [9].When several attitudes are triggered at the same time, we compute the global attitude (that is the attitude that emerges) as the average of the associated vectors of these attitudes. The vectors' magnitudes influence this average giving more importance to an attitude with a large magnitude (*i.e.* intensity)). For n attitudes $a_i \in a_1, ..., a_n$:

$$\overrightarrow{attitude} = \frac{1}{n} \sum_{i=1...n} \overrightarrow{a_i}$$

with a_i the attitudes in the circumplex.

4.3 Evaluation of the affective model

To evaluate the affective module, we compare the affects that are computed from the affective module with the manual annotation of emotions from the job interview corpus. To perform such an evaluation, we chose arbitrarily one video of this corpus. The affective states of the interviewee and the social attitudes of the recruiter have been manually annotated. Our evaluation consists in comparing if, given the affective states of the interviewee, the affective module computes social attitudes for the recruiter that are identical (i.e. similar in terms of dimensions representation) as the manually annotated social attitude of the recruiter in the video.

The interview in the selected video is composed of 8 speaker turns. In the first 4 questions/answers turns, the interviewee is confident and gives good answers to the recruiter (positive detected affects). The interviewer expectations has been annotated positively during this first sequence. Then, the 4 following questions appear more difficult for the interviewee; he shows expression of negative affects. On the other hand, expectations of the recruiter were annotated positive for questions 5 and 6 and negative for questions 7 and 8.
Fig. 2 shows the recruiter affects. Fig. 2(up) shows the output for the recruiter's affects after each of its question. We can notice that positive affects (triangles) are positively correlated with supportive attitude while negative affects (squares) are positively correlated with aggressive attitude. Fig. 2(bottom) displays the annotated social attitudes of the recruiter as it evolves during the interaction. Positive values mean that the recruiter is perceived supportive by the user while negative values as aggressive. We also indicate where the 8 questions happen in the course of the interaction.
Comparing both data, the outputs of the affective module and the manual annotation, we can remark that the results are not really comparable in term of intensity of social attitudes. However the attitudes computed by the affective module coincide with the manually annotated attitudes. The variation of attitudes from supportive to aggressive happens in both cases. This example shows that our affective module computes, for a given input (the affect states of the interviewee), similar attitudes for the recruiter of those that are perceived in real human-human job interview.

5 Expression of attitudes

Research has shown that interpersonal attitude are conveyed through non-verbal behaviours (see Section 1). However it is insufficient to look at signals independently of the other surrounding signals: a smile is a sign of friendliness, but a smile preceded by head and gaze aversion conveys submissiveness [11]. In our work, to give the capability to a virtual agent to convey attitudes, we choose a corpus-based approach to find how attitudes are expressed through signal sequences. We then use this knowledge to generate new sequences of non-verbal behavior for animating our ECA. In the following section, we present an algorithm that extract sequences of non-verbal signal.

5.1 Frequent sequences extraction

In order to extract significant sequences of non-verbal signals conveying interpersonal attitudes from our corpus, we used a *sequence mining* technique. Using the attitude annotation files described in section 3, we segmented our corpus in time intervals preceding attitude variations. With a clustering technique, we regrouped together these attitude variations, such as small (resp. large) increases (resp. decreases) of friendliness (resp. dominance). The next step consisted of applying the Generalized Sequence Pattern (GSP) frequent sequence mining algorithm described in [21]. This nets us a set of frequent sequences for each type of attitude variation, we can characterize each of these sequences with several *quality measures*: *Support*, *i.e.* how many times the sequence appears in the data ($[0;\infty] \in \mathbb{N}$) ; *Confidence*, *i.e.* the proportion of a sequence's occurrences happening before one type of attitude variation ($[0;1] \in \mathbb{R}$). Keeping sequences with a support of at least 10, we extracted a set of 879 sequences for dominance variations and 329 for friendliness variations. For instance, the $HeadNod \rightarrow Smile$ sequence was found frequently before large friendliness increases ($Support = 32$, $Confidence = 0.59$). In the following section, we describe the algorithm for generating non-verbal signals sequences conveying attitudes.

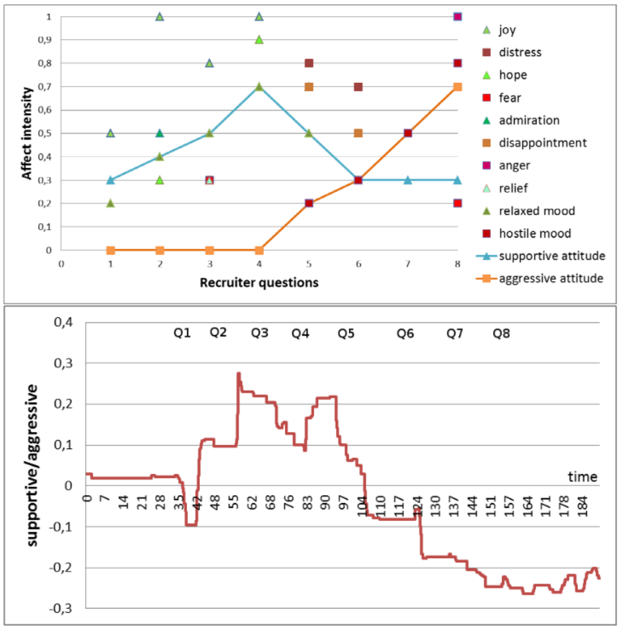

Figure 2: Computed recruiter affects after each question (up). Annotated recruiter state in real time (bottom)

5.2 Sequence generation

Given an input attitude that an ECA should express and an input utterance tagged with communicative intentions defined in the Functional Markup Language (FML) [12], the objective of our model is to generate a sequence of non-verbal signals that conveys the appropriate attitude. Our algorithm follows three steps:
Building minimal sequences - In a conversation, communicative intentions can be expressed through non-verbal behavior as well as through speech. For

instance, in Western culture, it is possible to convey uncertainty by performing a particular hesitation gesture. We specified behavior sets for every possible input communicative intention, *i.e.* the different non-verbal signals that can be displayed by an ECA to express the intention. The first step in our algorithm builds non-verbal signals sequences expressing an input message by selecting one signal in the behavior set of each communicative intention of the input message. Such a resulting sequence is called a *minimal sequence.*

Generating candidate sequences - For each minimal sequence obtained in the previous step, we retrieve all the time intervals where it is possible to insert other signals. For instance, if there is enough time between two head signals, we might insert a head nod or a head shake. For this purpose, we represent the extracted frequent sequences (Section 5.1) with a Bayesian Network (BN). This enables us to represent the causal and non deterministic relation of the attitudes on the signals (*e.g.* there might be more smiles for friendliness increases) and the sequences of signals (*e.g.* hands rest pose changes appear after gestures). An interesting feature of this model is that non-verbal signals sequences that did not occur in our data can still be generated, and their likelihood can be evaluated. Starting with the minimal sequences obtained after the previous step, we use the BN to add new signals in the available intervals, pruning out sequences that are too unlikely. The remaining sequences are called *candidate sequences.*

Selecting the final sequence - For selecting the final sequence, we defined a score variable of a candidate sequence s as $Sc(s) = P(s) * Conf(s)$, where $P(s)$ is the probability of s computed by the appropriate Bayesian Network (BN for dominance or friendliness depending on the input attitude), and $Conf(s)$ is equal to the sequence's confidence. The sequence with the highest score Sc is selected. In the next section, we present an study we realized to evaluate our model.

5.3 Evaluation

We evaluated our model with an online study realized using Adobe Flash technology. Participants were asked to compare 8 pairs of videos of a virtual character acting as a job recruiter expressing non-verbal signals when speaking (see Figure 3). For every pair of videos, the virtual recruiter said a different job interview question of the scenario . The character's speech was identical in both videos, however, the non-verbal behavior of the recruiter was different. The left video was generated with Greta's existing *Behavior Planner*, which does not consider attitudes and therefore considered neutral, while the right video was generated with our model with one of the 8 following attitudes *high dominance, low dominance, high low submissiveness, high submissiveness, high friendliness, low friendliness, low hostility, high hostility.*

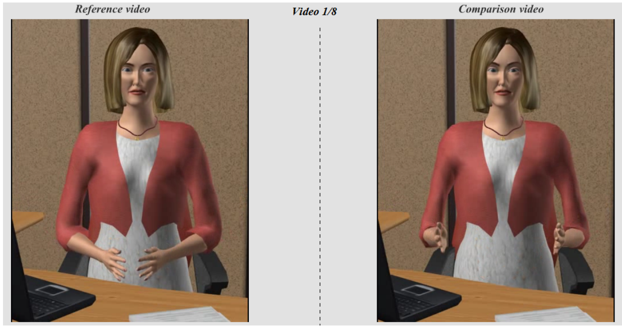

Figure 3: The main screen of the online study.

For every pair of videos, the participants answered were asked (*Q1*) how the right video compared to the left in terms of attitude (*e.g. much more dominant*) and how they would rate the attitude's intensity (*Q2*). Eighty-one participants took part in our study. The results of the study validate partially our model: the results of *Q1* showed that the expression of dominance, friendliness and hostility were recognized, however submissiveness was not recognized. All four results were statistically significant. For *Q2*, the only significant difference was found between intensity of large and small decreases in friendliness, however the participants identified large decreases as small decreases, and vice-versa. Therefore, it seems that our model cannot simulate attitudes of different intensities.

6 Concluding remarks

This paper proposes a new architecture for expressive agents that can reason about and display social behaviours. Unlike classical reactive models, our approach combines an affective reasoner that generates affects for the virtual character with a sequence selection mechanism based on a domain corpus annotated on two dimensions: dominance and friendliness. The methodology we propose contains several stages, from corpus collection and annotation, knowledge elicitation with experts for the definition of rules, implementation of behaviours corresponding to sequences and sequence selection based on the generated internal affects.

This architecture, whose components have been tested separately, has been integrated using the SEMAINE platform [19] and is currently being tested with real users. This will allow us to validate the global behaviour of our platform in the context of social coaching. However, several components can still be improved. One first limit of our model is that we assume exact inputs from the perception module. In addition, we intend to provide the affective reasoner with a representation of the interaction from the recruiter's point of view. We believe that allowing the recruiter to reason about the actual and potential behaviour of the applicant, following a Theory of Mind paradigm, will allow a more credible decision process. We will also take the sequence extraction procedure further to take into account the user's non-verbal signals. This will require the sequence selection model to plan for and react to user signals.

References

[1] M. Argyle. *Bodily Communication.* University paperbacks. Methuen, 1988.

[2] R. Aylett, A. Paiva, J. Dias, L. Hall, and S. Woods. Affective agents for education against bullying. In *Affective Information Processing*, pages 75–90. Springer, 2009.

[3] D. Ballin, M. Gillies, and B. Crabtree. A framework for interpersonal attitude and non-verbal communication in improvisational visual media production. In *Proc. CVMP*, 2004.

[4] M. Chollet, M. Ochs, and C. Pelachaud. A multimodal corpus for the study of non-verbal behavior expressing interpersonal stances. In *Proc. IVA'13 Workshop Multimodal Corpora*, 2013.

[5] P. T. Costa and R. R. MacCrae. *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual.* Psychological Assessment Resources, 1992.

[6] Joseph P. F., Gordon H. B., and Susan E. K. The influence of mood on perceptions of social interactions. *Journal of Experimental Social Psychology*, 20(6):497–513, 1984.

[7] P. Gebhard. ALMA - A Layered Model of Affect. In *Proc. AAMAS*, pages 29–36, 2005.

[8] M. Hoque, M. Courgeon, J.-C. Martin, Bilge M., and Rosalind P. MACH: My Automated Conversation coacH. In *Proc. UbiComp.* ACM Press, 2013.

[9] K. Isbister. *Better Game Characters by Design: A Psychological Approach.* Morgan Kaufmann Publishers Inc., 2006.

[10] H. Jones and N. Sabouret. TARDIS - A simulation platform with an affective virtual recruiter for job interviews. In *Proc. IDGEI*, 2013.

[11] D. Keltner. Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, 68:441–454, 1995.

[12] M. Mancini and C. Pelachaud. The FML - APML language. In *Proc. AAMAS'08 FML Workshop*, Estoril, Portugal, May 2008.

[13] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual Differences in Temperament. *Current Psychology*, 14(4):261, 1996.

[14] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, July 1988.

[15] I. Poggi, C. Pelachaud, F. de Rosis, V. Carofiglio, and B. De Carolis. GRETA: a believable embodied conversational agent. In *Multimodal intelligent information presentation*, pages 3–25. Springer, 2005.

[16] K. Porayska-Pomsta, M. Mavrikis, S. D'Mello, C. Conati, and R. Baker. Knowledge elicitation methods for affect modelling in education. *International Journal of Artificial Intelligence in Education*, 2013.

[17] N. Sabouret, H. Jones, M. Ochs, M. Chollet, and C. Pelachaud. Expressing social attitudes in virtual agents for social training games. *Proc. IDGEI*, 2014.

[18] K. R Scherer. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, pages 92–120, 2001.

[19] M. Schröder, E. Bevacqua, R. Cowie, F. Eyben, G. Gunes, D. Heylen, M. ter Maat, G. McKeown, S. Pammi, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. F. Valstar, and M. Wöllmer. Building autonomous sensitive artificial listeners. *Trans. Affective Computing*, 3(2):165–183, 2012.

[20] M. Snyder. The influence of individuals on situations: Implications for understanding the links between personality and social behavior. *Journal of Personality*, 51(3):497–516, 1983.

[21] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. In P. Apers, M. Bouzeghoub, and G. Gardarin, editors, *Advances in Database Technology*, volume 1057 of *LNCS*, pages 1–17. Springer Berlin Heidelberg, 1996.

[22] A. Tartaro and J. Cassell. Playing with virtual peers: bootstrapping contingent discourse in children with autism. In *Proc. ICLS*, pages 382–389, 2008.

[23] D. T. Wegener, R. E. Petty, and D. J. Klein. Effects of mood on high elaboration attitude change: The mediating role of likelihood judgments. *European Journal of Social Psychology*, 24(1):25–43, 1994.