Longitudinal Affective Computing
Virtual Agents that Respond to User Mood

Lazlo Ring, Timothy Bickmore, Daniel Schulman

College of Computer and Information Science, Northeastern University
360 Huntington Ave - WVH 202, Boston MA 02115
{lring,bickmore,schulman}@ccs.neu.edu

Abstract. We present two empirical studies which examine user mood in long-term interaction with virtual conversational agents. The first study finds evidence for mood as a longitudinal construct independent of momentary affect and demonstrates that mood can be reliably identified by human judges observing user-agent interactions. The second study demonstrates that mood is an important consideration for virtual agents designed to persuade users, by showing that favors are more persuasive than direct requests when users are in negative moods, while direct requests are more persuasive for users in positive moods.

Keywords: Affect, Agents, Mood, Longitudinal Study, Persuasive Technology

1 Introduction

As virtual agents spend more time with users—as educators, counselors, and companions—they will need to adapt to changes in users' beliefs, attitudes, and feelings over relatively long periods of time. This is particulary true for agents that enlist the user in behavior change through persuasion. To date, most virtual agent systems treat psychological construct as static, immobile trait confined to a single conversation. Virtual agents that have been developed to sense and respond to user affect also suffer a limitation, they only track and respond to user affective states that change within seconds or minutes [1][2]. In systems that respond to changes in affective state, most agents acknowledge the change and (if warranted) empathize with the user [3]. Expanding upon the idea of responding to changes in affective state, we look to explore other useful adaptations for agent behavior that could be made in response to user mood.

We are interested in the interaction between agent persuasion and the longitudinal affective construct, mood. For the purpose of this study, we define mood (following Larsen [4]) as an affective state that differs from emotion—the external expression of affect[5]—in two distinctive ways: duration and intensity. Whereas emotions last only a matter of seconds (from initial perception and reaction to decay), moods can last hours or days. Moods are also usually perceived to be less intense than emotions and are generally less-specific [6]. This lack of specificity, however, has led to difficulty in modeling mood, unlike emotion which has a variety of well-established models (e.g., basic emotions [7], categories of cognitive elicitors [8], etc.). Because of this, most researchers categorize mood using Russell's & Posner's circumplex model of affect [9], which models affective state (including mood) in terms of valence and arousal.

In this paper we describe a series of empirical studies exploring the detection and use of mood in long-term interactions with virtual agents. In section 3, we investigate whether the assessment of mood in user-agent interactions through verbal and nonverbal behavior is feasible. To answer this, we see if expert human judges can reliably assess the mood of users interacting with a virtual agent. We then investigate whether there is evidence of a mood construct in user-agent interactions, independent of momentary affect that can be assessed using the circumplex model. Then in section 4, we investigate the relationship between user mood and persuasion by an agent. In this study, we explore whether persuasive messages used by an agent are more effective if they are tailored in response to user mood.

2 Related Work

There have been many attempts to detect emotion and affect during agent based interactions. D'Mello investigated the automated detection of affect through various sensors [1]. Three studies were conducted to collect data on how affect expresses itself through body postures and eye movements, and how an automated affect detector could be created. Participants interacted with the AutoTutor pedagogical agent, and then were judged by one of three techniques: experts trained in Ekman's facial action coding system [7], self-report of their affective state, or judgement by their peers. Their results found that posture features and the tracking of eye movement could predict a participant's affective state with 70% accuracy.

The relationship between mood and persuasion has also been explored by multiple researchers. Aderman [10] found that the form of a request significantly impacted a participant's willingness to comply based upon their mood. Following a positive or negative mood induction, participants were asked to sort cards, with the request being phrased as either a study requirement or a favor to the experimenter. Participants in the negative mood condition were found to sort significantly more cards when the task was phrased as a requirement, where as those in the positive condition were found to sort significantly more cards in the favor condition.

3 User Mood Classification by Human Judges

To begin our exploration of user mood in long-term interactions with virtual agents we first wanted to determine whether human observers could reliably identify user moods in these interactions, based on their verbal and nonverbal conversational behavior.

Reliability Analysis: We used videotaped recordings of longitudinal user-agent conversations collected as part of a study of an eldercare companion agent[11] . Fifteen conversations conducted by three participants were selected for reliability analysis. Two minute video segments were extracted from the beginning, middle, and end of each conversation, resulting in a total of 41 video clips for analysis (4 conversations were too short to use all three time points). Three research assistants were asked to view each of the 41 video clips and rate each for arousal and valence using the Affect Grid[12], a self-report instrument that assess arousal and valence on a 2-dimensional grid, where arousal ranges from unpleasant (1) to pleasant (9) and valence ranges from sleepiness (1) to high arousal (9). Judges were also asked to specify a single English word that best described user mood. Video clips were provided for judges to view in any order, and as frequently as they liked.

Results: Arousal scores assigned by judges ranged from 3 to 9 (mean 6.57, SD 1.15) and valence scores ranged from 4 to 9 (mean 6.59, SD 1.17). Judges used 26 English words to describe the moods they observed. The most commonly used words were: "happy" (42 instances), "content" (16), "good" (12), "neutral" (9), and "calm" (8). Ratings of arousal and valence were significantly correlated among the three judges, with intraclass correlation coefficients of 0.662 for arousal ($p < .001$) and 0.646 for valence ($p < .001$). Of the 41 video clips, judges only agreed on English mood labels 12 times: 11 of these were pairs of judges, and only once did all 3 judges volunteer the same label (in all of these cases the label was "happy").

User Mood Variability Analysis: We next sought to characterize the amount of variance in user affect, both within and between conversations in order to determine the amount of variance that was due to momentary affect (within a conversation), the amount due to mood (between convresation), and the eamount due to personality (between subjects). 145 clips from 42 videotaped interactions (described above) was used in this study, with two minute video segments extracted from the beginning, middle, and end of each conversation as before.

For each of valence and arousal, we performed a restricted maximum likelihood fit (using lme4 [13] in R [14]) of a 3-level variance components model:

$$\begin{aligned} &y_{ijk} = \beta_0 + P_i + M_{ij} + \epsilon_{ijk} \\ &P_i \sim N(0, \tau_P^2), \quad M_{ij} \sim N(0, \tau_M^2), \quad \epsilon_{ijk} \sim N(0, \sigma^2) \end{aligned}$$

where y_{ijk} is the average of the judges' ratings for participant i, conversation j, and videotape segment k. We tested for significant intraclass correlation at the level of conversations with a restricted likelihood ratio test [15] that compared this model against a 2-level model which omitted the M_{ij} term.

Results: In the full corpus, valence was observed to range from 2 to 9 (overall mean 5.82, SD 1.32) and arousal ranged from 3 to 9 (mean 5.50, SD 1.40). The estimated variance of valence and arousal within and between conversations is shown in Table 1. For both arousal and valence, most variance was accounted for at the level of participants, followed by segments and conversations. There was significant intraclass correlation found at the level of conversations, both for valence ($RLR = 9.44$, $p = 0.001$) and for arousal ($RLR = 5.89$, $p = 0.007$).

Table 1: Variance of valence and arousal between and within conversations

	Valence		Arousal	
	Variance	% Total	Variance	% Total
Participants(trait) τ_P^2	0.55	0.50	0.82	0.70
Conversations(mood) τ_M^2	0.16	0.15	0.11	0.09
Segments(affect) σ^2	0.39	0.35	0.36	0.28

Discussion: We demonstrates that user affect can be reliably assessed by human judges using arousal and valence scores, on the basis of observed verbal and nonverbal conversational behavior during interactions with a virtual agent. The use of English words was not a reliable measure of affect as there was essentially no agreement among the judges on terms used. We also show that there is significant intraclass correlation in ratings at the level of conversations, while controlling for overall intraclass correlation. This demonstrates that these assessments partially captured mood: a phenomenon occurring on a larger time scale than a single conversation, yet distinct from an individual's overall baseline affective state. Thus, a longitudinal model of affective state should include both inter-subject (a subject specific baseline) and inter-conversation (mood) components.

4 The Effect of Form of Request and Mood on Persuasion

Following Aderman's work (described in section 2), we decided to adapt his methodology to investigate the effects of mood and persuasive request phrasing on exercise motivation. This specific area was chosen due to previous literature showing that agents are effective exercise counselors, and that they elicit similar effects from dialogue phrasing as found in human-human interactions [16][17].

The study was conducted in the context of the "Virtual Laboratory" system [18], in which a standing group of participants interact with a virtual exercise promotion agent up to once a day from their home computers. The agent encourages participants to walk every day and tracks their progress through a supplied pedometer that the agent discusses with them.

Our manipulation consists of the agent asking participants to exercise, phrased as either a favor to the agent or direct request. Our hypothesis was that participants will walk significantly more steps when they are in a negative mood and are told to walk using a favor dialogue, and when they are in a positive mood and are told to walk using a request dialogue.

Measures: The Affect Grid (Section 3) was used by participants to rate their mood. Finally participants upload the amount of steps they walked since their last session via a pedometer provided at the beginning of each session with the system.

Experimental Protocol: This study was divided into two separate interaction phases: a desensitization phase, and a collection phase. In the desensitization phase (5 days), participants did not interact with the agent, but instead were given an Affect Grid each session for five sessions. This was done to both reduce habituation effects from prior interactions with the agent, and to collect baseline valence and arousal measurements for each participant. This data was used to calculate the change in valence and arousal each day in the following phase.

In the collection phase (2 months), participants first filled out the Affect Grid at the beginning of each session, then conducted their usual counseling conversation with the agent but with the following change: instead of negotiating daily pedometer step count goals the agent asks participants to walk as either a favor or as a request. The exact language used was:

Favor: *I was wondering if you'd mind doing me a favor and take a walk before our next session.*

Request: *Would you take a walk before our next session.*

The manipulation was randomly selected every day for every participant (within-subjects).

Results: Twenty-one participants (mean age 61.5) interacted with the system over two months, resulting in 696 unique interactions (mean=33.1 per participant, SD = 16.2) with the agent, with one participant dropping out of the study. For each interaction, the number of steps the participant had walked since their last session along with their valence and arousal were recorded.

A linear mixed-effects regression model was used to fit the data. This model is an extension of linear regression models that allows for the linear predictors to contain both random and fixed effects. This model used the study condition of favor (Coded as 0) versus request (Coded as 1), the number of interactions, and the difference in participant's valence and arousal from their baseline to estimate the number of steps they walked since their last interaction. Baseline arousal and valence was estimated for each participant using their average valence and arousal recorded via the Affect Grid during the desensitization phase of the study. The average of these scores were used to model the participant specific baseline affect found in study 1. Steps were put on a logarithmic scale to restrict the range of outcomes to greater than 0 steps, and to account for the right tail skew of the measure. Since exact p values and confidence intervals cannot be calculated for mixed effect models analytically, a semi-parametric bootstrap was used, as described by Carpenter, et al [19]. All statistics were calculated using R-2.14.1 and the lme4 package [14][13].

As shown in Table 2, if the agent used the request dialogue while the participant was in a positive mood they walked significantly more steps, and if they agent used the favor dialogue while the participant was in a negative mood

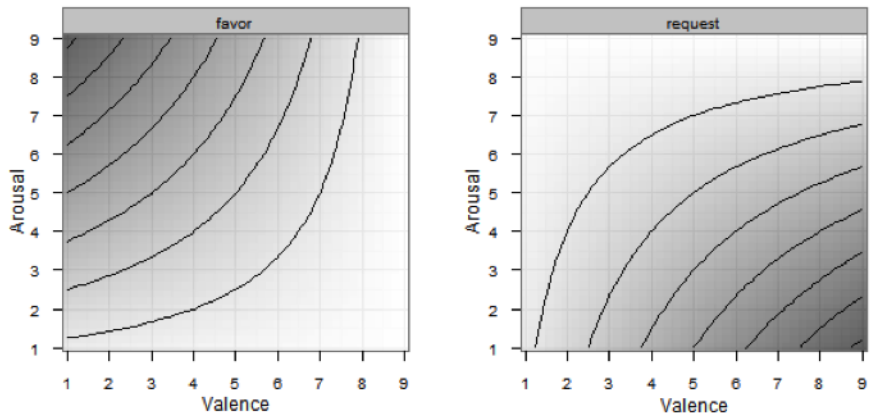

Fig. 1: Change in the number of steps walked based on mood and dialogue manipulation. Darker areas represent where each dialogue had the most positive effect on step count.

they walked significantly more steps ($p < .01$). Additionally, it was found that participants walked significantly more steps when their valence and arousal scores were opposite in sign ($p < .01$). Thus, when a participant is in a high arousal, low valence state, a favor message predicts more walking, whereas when a participant is in a low arousal, high valence state, request predicts more walking (Figure 1).

However, the effect of the manipulation decreased over time, as shown by the quadratic session terms in Table 2, such that it was no longer significant after

Table 2: A Linear Mixed-Effect Regression Model Predicting Participant's Step Count (log-transformed). Inter-subject Variance: (Estimate: .258, 95% CI [.142, .343]), Residual Variance: (Estimate: .623, 95% CI [.566, .678]). Legend: V = Valence, A = Arousal, C = Condition (Favor coded as 0, Request coded as 1), S = Session, S^2 = Sessions×Sessions (To model habituation of affect over time)

Parameter	Est.	SE	P	Parameter	Est.	SE	P
$Intercept$	6.19e-03	1.96e-03	0.93	$V \times S^2$	2.62e-06	-2.99e-06	1.00
V	2.06e-01	1.79e-03	0.39	$A \times S^2$	-1.20e-05	6.54e-06	0.99
C	-2.08e-02	-1.15e-03	0.87	$V \times A \times C$	4.59e-01	1.55e-02	0.52
S	5.50e-03	-1.95e-04	0.57	$\boldsymbol{V \times A \times S}$	**1.02e-01**	**4.45e-04**	**0.03**
S^2	-1.04e-04	3.25e-06	0.59	$\boldsymbol{V \times C \times S}$	**-8.53e-02**	**-2.42e-06**	**0.03**
$\boldsymbol{V \times A}$	**-1.45**	**-4.22e-03**	**0.01**	$A \times C \times S$	2.73e-02	-1.17e-04	0.42
$\boldsymbol{V \times C}$	**1.07**	**-1.977e-03**	**0.01**	$V \times A \times S^2$	-1.65e-03	-1.13e-05	0.11
$A \times C$	-1.22e-01	2.60e-03	0.74	$V \times C \times S^2$	1.41e-03	4.13e-06	0.09
$V \times S$	-8.66e-03	1.29e-05	0.72	$A \times C \times S^2$	-8.29e-04	-7.96e-08	0.25
$A \times S$	-1.06e-03	-2.10e-04	0.96	$V \times A \times C \times S$	-2.44e-02	-1.23e-03	0.73
$C \times S$	1.27e-03	1.30e-04	0.91	$V \times A \times C \times S^2$	-3.98e-04	2.47e-05	0.81

a month. This habituation effect is consistent with previous research on affect [20][21], showing the decay of the manipulation through the course of the study.

Discussion: We found that the form of a persuasive message should be tailored based on user mood in order to be maximally effective. These results are contrary to our hypotheses and findings in the previous literature, but our experiment differs from the earlier work in three key aspects. In Aderman's original work, participants were asked to do a favor or request for the experimenter, whereas in our experiment the participant is doing a favor for the agent. However, due to the virtual nature of the agent, the agent cannot benefit from this request; therefore the participants are indirectly doing a favor for themselves. This change in perspective could account for the reversal of the observed trend since the persuasive outcome of interest is self-efficacy instead of altruistic behavior. Additionally, the majority of studies on mood observed only a single session of affect while disregarding the longitudinal property of mood in the process.

5 Conclusion

We found that inter-conversation mood is a significant component of user affect, and that mood can be reliably assessed on the basis of user verbal and nonverbal behavior during interactions with a virtual agent. We also found that mood should be taken into account when selecting persuasive messages in order to maximize compliance, although the effectiveness of a simple (non-varying) mood-based manipulations decays over time. In future studies, we plan to explore the use of non-invasive sensors to automatically detect user mood and investigate the application of the results found in this paper to develop affectively tailored dialogue systems. We also plan to explore how well our findings on mood hold up during much longer time periods, such as year-long interactions with an agent.

Acknowledgments. We thank Connor Westfall, Colin Shaughnessy, Bridgette Collado, and Ashley Kline for their assistance with this study. This material is based upon work supported by the National Science Foundation under Grant No. 1012086 and ISS-0545932. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. D'Mello, S., Graesser, A., Picard, R.: Toward an affect-sensitive autotutor. Intelligent Systems, IEEE **22**(4) (July-Aug. 2007) 53 –61
2. Burleson, W., Picard, R., Perlin, K., Lippincott, J.: A platform for affective agent research. In: 3rd International Conference on Autonomous Agents and Multi Agent Systems, ACM Press (2004)
3. Klein, J., Moon, Y., Picard, R.W.: This computer responds to user frustration. In: CHI '99 extended abstracts on Human factors in computing systems. CHI EA '99, New York, NY, USA, ACM (1999) 242–243

4. Larsen, R.J.: Toward a science of mood regulation. Psychological Inquiry **11**(3) (2000) pp. 129–141
5. Batson, C.D., Shaw, L.L., Oleson, K.C. Emotion. In: Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions. Thousand Oaks, CA, US: Sage Publications, Inc (1992) 294–326
6. Thayer, R.E.: The biopsychology of mood and arousal. Oxford University Press, New York (1989)
7. Ekman, P., Friesen, W.V.: Facial Action Coding System. Consulting Psychologists Press, Stanford University, Palo Alto (1977)
8. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge [England]; New York (1988)
9. Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology **17**(3) (2005) 715–734
10. Aderman, D.: Elation, depression, and helping behavior. Journal of personality and social psychology **24**(1) (1972) 91–101
11. Vardoulakis, L., Ring, L., Barry, B., Sidner, C., Bickmore, T.: Designing relational agents as long term social companions for older adults. In: Intelligent Virtual Agents. (2012)
12. Russell, J.A., Weiss, A., Mendelsohn, G.A.: Affect grid: A single-item scale of pleasure and arousal. Journal of Personality and Social Psychology **57**(3) (1989) 493–502
13. Bates, D., Maechler, M., Bolker, B.: lme4: Linear mixed-effects models using s4 classes (2011)
14. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (2008) ISBN 3-900051-07-0.
15. Scheipl, F., Greven, S., Küchenhoff, H.: Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. Computational Statistics & Data Analysis **52**(7) (March 2008) 3283–3299
16. de Melo, C.M., Carnevale, P., Gratch, J.: The effect of expression of anger and happiness in computer agents on negotiations with humans. In: The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3. AAMAS '11, Richland, SC, International Foundation for Autonomous Agents and Multiagent Systems (2011) 937–944
17. Schulman, D., Bickmore, T.: Persuading users through counseling dialogue with a conversational agent. In: Proceedings of the 4th International Conference on Persuasive Technology. Persuasive '09, New York, NY, USA, ACM (2009) 25:1–25:8
18. Bickmore, T., Schulman, D.: A virtual laboratory for studying long-term relationships between humans and virtual agents. In: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1. AAMAS '09, Richland, SC, International Foundation for Autonomous Agents and Multiagent Systems (2009) 297–304
19. Carpenter, J.R., Goldstein, H., Rasbash, J.: A novel bootstrap procedure for assessing the relationship between class size and achievement. **52**(4) (2003) 431–443
20. Titchener, E.B.: Lectures on the elementary psychology of feeling and attention. Arno Press, New York (1973)
21. Dijksterhuis, A., Smith, P.K.: Affective habituation: Subliminal exposure to extreme stimuli decreases their extremity. Emotion **2(3)** (2002) 203–214