Factors that Influence the Perception of Feedback Delivered by a Pedagogical Agent

Kristen E. Link, Roger J. Kreuz, Arthur C. Graesser, and the Tutoring Research Group

The University of Memphis

Address correspondence to:

Kristen E. Link

Department of Psychology

The University of Memphis

Memphis, TN 38152-6400

(901) 678-2883 (Phone)

(901) 678-2579 (Fax)

klink@memphis.edu

Abstract

Three experiments were conducted to assess the relative importance of speech parameters and facial expressions in the delivery of feedback by a pedagogical agent. In Experiment 1, we manipulated linguistic form (i.e., positive, negative, or neutral terms), rate, pitch, pause, and emphasis. In Experiment 2, we manipulated eye size, mouth curve, brow height, and brow curve. In a third study, both speech parameters and facial expressions were manipulated. In all three experiments, the participants were asked to indicate how positive or negative the agent's feedback seemed to be. Across the studies, the variables collectively accounted for a significant amount of the variance. More specifically, the linguistic expressions and mouth curve emerged as significant predictors of the participants' ratings. This suggests that these two features should be implemented by developers wishing to provide appropriate feedback in their pedagogical agents.

Factors that Influence the Perception of Feedback Delivered by a Pedagogical Agent

There has been a great deal of interest in developing software systems with talking heads, avatars, and other forms of animated agents for educational purposes. Such systems typically have an animated tutor, and are often referred to as pedagogical agents (André, Rist, & Müller, 1998; Cassell et al., 1994; Cassell & Thórisson, 1999; Johnson, Rickel, & Lester, 2000; Lester, Towns, & FitzGerald, 1999; Paiva & Machado, 1998). The complexity of such systems vary considerably, from simple "talking heads" to full-blown embodied agents that use multiple modalities, such as facial expressions, gestures, intonation, and appropriate feedback, to interact with students.

The design of some of these pedagogical agents has been motivated by existing research that maps the features of each modality onto particular emotions. The impact of facial features on communicating emotions has been studied extensively (e.g., Ekman, 1993; Ekman & Friesen, 1984; Ellison & Massaro, 1997; Massaro & Egan, 1996). The results of this work have revealed that emotional states are diagnostically predicted by a relatively small number of facial features, such as eyebrow and mouth deflection, eyebrow height, and openness of the eyelids. In a separate body of research, vocal features that potentially express emotions have been explored (e.g., Crystal, 1975; Davitz, 1964; Frick, 1985; Murray & Arnott, 1993; Scherer, 1986; Scherer & Scherer, 1981). These include duration, pitch, intensity, voice quality, and speech rate.

There has been some controversy with regard to which of these modalities are important for effective communication (Cassell, Sullivan, Prevost, & Churchill, 2000; Cassell & Thórisson, 1999). Johnson, Emde, Scherer, and Klinnert (1986) reported that participants were able to recognize emotions almost perfectly from vocal cues alone. Ochsman and Chapanis (1974) studied communication through various modalities and claimed that the voice channel was the most important. However, Trower, Bryant, and Argyle (1978) suggest that the nonverbal channel is primarily used to communicate interpersonal attitudes. Unfortunately, there exist little data that address the relative contributions of the modalities used by animated agents, so the issue is unresolved. Moreover, none of the available research has investigated these modalities and parameters in the pedagogical context of tutoring.

In this paper, we will evaluate the effects of manipulating linguistic expressions, intonation, and facial cues on individuals' perceptions of the evaluative feedback from a pedagogical agent. When a learner answers a question or offers a solution to a problem, tutors give short feedback that evaluates the quality of the contribution. Feedback such as "yeah," "uh-huh," and "no" convey positive, neutral, or negative feedback and are typically delivered with appropriate intonational contours and facial expressions. This study reports three experiments that vary the features of short feedback and that assess their impact on the perception of feedback quality (i.e., positive, negative, versus neutral).

The present study was in part inspired during the development of a fully automated computer tutor called AutoTutor. The goal of that project was to construct a pedagogical agent that tutors students on the subject of computer literacy, and to do so in a conversationally appropriate manner (Graesser, Wiemer-Hastings, Wiemer-Hastings, Kreuz, & the Tutoring Research Group, 1999). The system is designed to handle a mixed initiative dialog by collaboratively fleshing out answers to questions of varying difficulty, including deep reasoning questions. AutoTutor uses latent semantic analysis (LSA; Landauer & Dumais, 1997), a statistical technique for representing world knowledge, to analyze student contributions. AutoTutor computes the similarity between a contribution and information about computer literacy represented in the system. AutoTutor then utilizes a dialog advancer network (Person, Bautista, Kreuz, Graesser, & the Tutoring Research Group, in press) to provide a pedagogically appropriate dialog move, such as a prompt, hint, or elaboration.

AutoTutor has been designed to mimic a human tutor, and is based on an analysis of 66 hours of transcripts of naturalistic tutoring (Graesser, Person, & Magliano, 1995). The system provides feedback when the student makes a contribution. In real-world tutoring, feedback can be given in a variety of ways. For example, the feedback may be the words that are spoken (e.g., "right"), the way it is spoken (e.g., quickly or slowly), a facial expression (e.g., raised eyebrows), or gestures (e.g., head nod). These feedback terms are followed by appropriate continuations or corrections to continue the dialog with the student. The accuracy of the system is dependent on the sensitivity of the latent semantic analysis computation, but research suggests that this mechanism can discriminate the quality of student contributions (Graesser, Wiemer-Hastings, Wiemer-Hastings, Harter, Person, & the Tutoring Research Group, 2000; Wiemer-Hasting, Wiemer-Hastings, & Graesser, 1999).

In naturalistic tutoring, it has been shown that human tutors tend to provide positive feedback more frequently than negative feedback, because of pragmatic phenomena like politeness considerations (e.g., Person, Kreuz, Zwaan, & Graesser, 1995). Therefore, AutoTutor might say "good" in response to a contribution that has a high match to the system's list of correct answers for a particular question. Conversely, in response to a student's erroneous contribution (which might have a high match to a list of incorrect answers), AutoTutor will provide a response such as "no." In cases of an intermediate match with the correct and incorrect answer lists, AutoTutor will offer a neutral response, such as "okay."

Fox (1993), in her research on human tutoring with expert tutors, has found that tutors paused before providing negative feedback to students' error-ridden contributions. In contrast, Graesser, Person, and Magliano (1995) reported that novice tutors provided negative feedback quickly and corrected the error. Therefore, the expertise of the tutor may affect the onset time of negative feedback. In addition, while providing feedback, there may be a discrepancy between the feedback term used by tutors and their facial expressions and intonation. For example, the linguistic expression might be "yeah" while the intonation and facial expression signals a negative response. The linguistic expression may be used to satisfy politeness goals (e.g., Brown & Levinson, 1987) whereas the tutor's intonation and facial expression may satisfy pedagogical goals, such as giving discriminating feedback.

In addition, it is not entirely clear how neutral feedback terms like "okay" are interpreted by the student. Some individuals might perceive "okay" as relatively positive, whereas others may construe "okay" as being somewhat negative. As mentioned above, other modalities, such as intonation cues and facial expressions, might also affect the perception of such feedback.

We designed a series of experiments in which features of these modalities were systematically manipulated in order to determine their relative importance. Specifically, in Experiment 1, we manipulated verbal features of the feedback: linguistic expression, rate, pitch, pause, and emphasis. In Experiment 2, we manipulated facial features: mouth curve, eyebrow height, eyebrow curve, and eye size. Finally, in Experiment 3, we manipulated verbal and facial features at the same time: linguistic expression, mouth curve, rate, and pitch.

Experiment 1

Method

Participants. Participants were 30 undergraduate introductory psychology students from The University of Memphis, participating for course credit.

Materials and procedure. Student contributions were created to precede AutoTutor's feedback. Sixteen advanced medical facts (e.g., "Six active cortical hormones are characterized by a double bond in the steroid skeleton.") were read by a male student and digitally recorded. Advanced facts were chosen so that the participants would not know the correct answer and whether the feedback should be positive or negative.

There were three categories of linguistic expressions, with 3 feedback terms in each category. Positive feedback terms included "yes," "good," and "right." Neutral feedback terms included "uh-huh," "okay," and "well." Finally, negative feedback terms included "wrong," "no," and "what?" These feedback terms were generated by AutoTutor with the Microsoft Agent speech engine (TruVoice, 1997). The engine has a mark-up language that allowed us to manipulate the following intonation variables: rate (75 words per minute (wpm) vs. 150 wpm), pitch (50 Hz vs. 150 Hz), pause between the feedback term and the previous student contribution (0 msec vs. 2000 msec), and emphasis of a particular word (on or off). When considering all four manipulated parameters, there were 16 combinations. Each feedback expression was presented in all 16 possible combinations of the intonation variables, resulting in 144 feedback items. Each item was randomly paired with one of the 16 student contributions.

Experimental sessions were run in small groups of approximately 4 individuals. Each participant was seated at a computer terminal. The experimenter, seated in the middle of the room, played each of the 144 student contribution-feedback pairs from a laptop computer equipped with external speakers. After each item was presented, participants rated the feedback on a six-point scale (1 = very negative, 2 = negative, 3 = not sure, but guess negative, 4 = not sure, but guess positive, 5 = positive, and 6 = very positive) by entering on the keyboard the number corresponding to their rating. The scale on which they made their ratings remained on the screen throughout the experiment.

Results and Discussion

The overall mean rating for all stimuli was 3.52 (SD = 1.58). We conducted a multiple regression analysis that included rate, pitch, pause, emphasis, and linguistic expression category as the predictor variables, and subjects' ratings as the criterion variable. The five predictor variables were entered into the regression equation as a single step (i.e., forced entry). These five variables accounted for a significant amount of the variance, F (5, 138) = 13.57, $p < .01$, $R^2 = .33$. However, the linguistic expression category was the only significant predictor ($R^2 = .32$). The beta-weights, t-scores, and R^2 for each of the variables appear in Table 1. The R^2 values were change scores, starting with the most robust predictor and then adding on the statistical contributions of each incremental predictor. We also assessed the possibility of interactions among the variables. Given that there were five predictor variables, there were ten 2-way interaction terms, ten 3-way interactions, five 4-way interactions, and one 5-way interaction. One of these interactions was significant (viz., linguistic expression x pitch x pause, accounting for an additional 6% of the variance). However, given the large number of effects analyzed, there is a high likelihood that this result is spurious. Therefore, this interaction will not be analyzed further.

Mean ratings were calculated for the three levels of linguistic expression category. The mean for the three positive terms was 4.53 ($\underline{SD}$ = 1.21). For the neutral terms, the mean was 3.61 ($\underline{SD}$ = 1.45), and for the negative terms, the mean was 2.44 ($\underline{SD}$ = 1.34). Therefore, the positive, neutral, and negative feedback terms were rated as predicted: the positive terms were rated on the high end of the scale, the negative on the low end, and the neutral terms were rated near the midpoint of the scale.

These results can be summarized simply. The participants relied entirely on the linguistic expression when they made their feedback judgments. They were not affected by the intonation variables that were manipulated.

Experiment 2

Animated agents can provide feedback not only verbally, but nonverbally as well, including the use of facial expressions and gestures (Cassell, Sullivan, Prevost, & Churchill, 2000; Johnson, Rickel, & Lester, 2000). In Experiment 2, we manipulated a variety of facial parameters and examined how these affected the perception of feedback. This type of nonverbal feedback was examined in isolation in order to prevent the possibility that these factors might be overshadowed by verbal cues.

Method

Participants. Participants were 30 undergraduate introductory psychology students from The University of Memphis, participating for course credit. Participants had not participated in Experiment 1.

Materials and procedure. The 16 digitally recorded medical facts used in Experiment 1 served as the student contributions. The feedback items, in the form of facial expressions, were generated using the Microsoft Agent software. Four facial features were manipulated: eye size (small, medium, or large), mouth curve (down, partially up, or up), eyebrow curve (low, medium, or high), and eyebrow height (low, medium, or high), yielding 81 possible facial expressions. The faces of AutoTutor shown in Figure 1 illustrate each value of these facial features. The facial expressions were projected on the center of a 69 in. wide x 76 in. high screen via an LCD projector.

Experimental sessions were run in large groups of approximately 15 individuals. The experimenter played each of the 81 student contribution-facial feedback pairs from a laptop computer equipped with external speakers and projected on the screen. After a student contribution was played, AutoTutor's facial expression changed from a default position (i.e., medium eye size and eyebrow height, low eyebrow curve, and partially up mouth curve) to one of the 81 facial expressions. The student contributions were randomly paired with the facial expressions, which were presented in a random order. Participants were given packets on which feedback scales were printed for each item. Participants rated the feedback on a six-point scale that was the same as in Experiment 1. They circled the number on the scale corresponding to their rating.

<u>Results and Discussion</u>

The overall mean rating for all facial expression combinations was 3.68 (SD = 1.56). A multiple regression analysis that included eye size, mouth curve, brow height, and brow curve as the predictor variables, and subjects' ratings as the criterion variable, was performed. The four predictor variables were entered into the regression equation as a single step (i.e., forced entry). These variables accounted for a significant amount of the variance, $\underline{F}$ (4, 76) = 18.88, $\underline{p} < .01$, $\underline{R}^2 = .50$. However, mouth curve was the only significant predictor ($\underline{R}^2 = .47$). The beta-weights, $\underline{t}$-scores, and $\underline{R}^2$ for each of the variables appear in Table 2. The $\underline{R}^2$ values were change scores, starting with the most robust predictor and then adding on the statistical contributions of each incremental predictor. We also assessed the possibility of interactions among the four variables. Given that there were four predictor variables, there were six 2-way interaction terms, four 3-way interactions, and one 4-way interaction. However, the results revealed that none of the interactions were significant.

Mean valence ratings were calculated for the three levels of mouth curve. The mean rating when the mouth curve was down was 2.07 (SD = 0.92). When the mouth curve was partially up, the mean was 3.82 (SD = 0.99), and when the mouth curve was up, the mean was 5.14 (SD = 0.89).

Once again, the results can be interpreted simply. When the facial expression was the only form of feedback, the participants relied almost exclusively on the shape of the mouth in making their ratings. Participants interpreted the feedback as negative when the agent's mouth was curved downward, neutral when it was curved partially up (i.e., its default position), and positive when it was curved up.

Experiment 3

In this experiment, we included both verbal and facial cues in the same study to determine the features of these cues that are important for interpreting the valence of the feedback. For example, it might be the case that facial cues are the most important even when the participants simultaneously receive verbal cues. Alternatively, the verbal cues may overshadow the effect of the agent's facial expressions. Finally, it may be the case that these cues are additive, and that participants rely on both types of information in order to interpret the feedback. Experiment 3 was designed to explore these possibilities.

Method

Participants. Participants were 30 undergraduate introductory psychology students from The University of Memphis, participating for course credit. Participants had not participated in Experiments 1 or 2.

Materials and procedure. The student contributions were the same as in Experiments 1 and 2. AutoTutor's feedback was generated using Microsoft Agent, and included manipulation of both facial features and intonation parameters. Specifically, four feedback parameters were manipulated: linguistic expression (8 from Experiment 1; "well" was not used), rate (75 wpm or 150 wpm), pitch (50 Hz or 150 Hz), and mouth curve (down, partially up, or up), yielding 96 possible combinations.[1] All other parameters (i.e., pause, emphasis, eye size, eyebrow height, and eyebrow curve) varied randomly across items.

Experimental sessions were run in large groups of approximately 15 individuals. AutoTutor was projected from a laptop computer onto a 69 in. wide x 76 in. high screen via an LCD projector. A trial consisted of the presentation of a randomly chosen student contribution played from the laptop computer via external speakers, followed by a randomly chosen feedback item. Each of the 96 student contribution-feedback pairs was presented in a random order.

Because of limitations in Microsoft Agent, AutoTutor's speech and facial expressions could not occur simultaneously. Therefore, after the student contribution, AutoTutor's face first changed from its neutral expression, provided the verbal feedback, and then returned to its neutral expression.

Participants were given packets on which feedback scales were printed for each item. Participants rated the feedback holistically (i.e., considering both what was said and how it was said) on the six-point scale used in Experiment 2.

Results and Discussion

The mean rating across all feedback items was 3.81 (SD = 1.37). A multiple regression analysis that included linguistic expression category, mouth curve, rate, and pitch as the predictor variables, and subjects' ratings as the criterion variable, was conducted. The four predictor variables were entered into the regression equation at one step (i.e., forced entry). These variables accounted for a significant amount of the variance, F (4, 91) = 130.71, $p < .01$, $R^2 = .85$. Mouth curve ($R^2 = .54$) and feedback category ($R^2 = .31$) were the only significant predictors. None of the interactions were significant when these were assessed in follow-up multiple regression analyses. The beta-weights, t-scores, and R^2 for each of the four variables appear in Table 3. The R^2 values were change scores, starting with the most robust predictor and then adding on the statistical contributions of each incremental predictor.

The results support the claim that verbal and nonverbal cues are additive. Specifically, participants relied on both the linguistic expression and the mouth curve Each variable explained a significant amount of variance in Experiment 3, with degrees of impact that are comparable to those found in Experiments 1 and 2. Finally, each feedback cue explained a unique portion of the variance. Therefore, more of the variance in the ratings was explained in this experiment, when the cues were presented together, than when they were presented alone. A graphical representation of the participants' ratings as a function of mouth curve and feedback category is shown in Figure 2. It is interesting to note how these features affect each other. For example, a neutral feedback term combined with an upturned mouth results in a positive rating. In addition, when the modalities conflict (e.g., a positive term combined with a downturned mouth), participants tended to provide neutral feedback ratings.

General Discussion

Although the results of these experiments seem straightforward, it should be kept in mind that we chose to examine a fairly small number of features of each modality. For example, we did not manipulate forehead wrinkling, mouth openness, or nose wrinkling, even though these have been shown to be important for the expression of emotion (Ekman & Friesen, 1984). However, many of the emotions studied by these researchers (e.g., disgust or fear) are not relevant to tutorial feedback (Fox, 1993; Graesser et al., 1995; Johnson, Rickel, & Lester, 2000). It was possible to factorially manipulate these features by using simple faces in order to assess their relative importance.

In addition, there are also relevant vocal features that we did not manipulate, including pitch change (e.g., upward or downward) and articulation (e.g., tense or slurring), as described by Murray and Arnott (1993). It is plausible that changes in these parameters are more salient features than absolute levels because human information processing systems are more attentive to changes than constancies. Unfortunately, these parameters cannot be manipulated in the Microsoft speech engine. However, some of the parameters that have been discussed in the literature that investigates intonation in conversational dialog are absolute parameter levels rather than contours that dynamically change over time. For example, in the context of tutorial dialog, Fox (1993) reported that expert tutors have a delay in the onset time of short feedback after student errors whereas Graesser et al. (1995) reported a short onset time in novice tutors. Nevertheless, the present study has at least established that absolute intonational parameters have no impact on the perception of short feedback in pedagogical conversational agents. Instead, it is the linguistic expression that reigns supreme.

It is also the case that we chose relatively arbitrary values for each feature (e.g., low brow curve and pitch of 50 Hz). These values were chosen to be psychologically distinguishable, but a larger range of values could have been used. Once again, however, the goal was to factorially manipulate all of the values, and a large number would have been unwieldy.

With these caveats in mind, the results of these experiments may be informative for developers of pedagogical agents. It has been claimed that facial gestures have context-independent meanings (Wierzbicka, 2000), so these results may have utility in other domains as well. We found that mouth curve and feedback category alone explain a large portion of the variance in participants' evaluations of feedback. This suggests that even a simple, two-dimensional head can capture the expressive range required for varying types of feedback.

References

André, E., Rist, T., & Müller, J. (1998). Integrating reactive and scripted behaviors in a life-like presentation agent. In Proceedings of the Second International Conference on Autonomous Agents (pp. 261-268). New York: ACM.

Brown, P., & Levinson, S. (1987). Politeness: Some universals in language usage. Cambridge: Cambridge University Press.

Cassell, J., Pelechaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., & Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture, & spoken intonation for multiple conversational agents. In Proceedings of the 21st Annual International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '94) (pp. 413-420). New York: ACM.

Cassell, J., Sullivan, J. Prevost, S., and Churchill, E. (Eds.). (2000). Embodied conversational agents. Cambridge, MA: MIT Press.

Cassell, J., & Thórisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. Applied Artificial Intelligence, 13:519-538.

Crystal, D. (1975). The English tone of voice. Bristol, UK: Edward Arnold.

Davitz, J. R. (1964). Personality, perceptual, and cognitive correlates of emotional sensitivity. In J. R. Davitz (Ed.), The communication of emotional meaning (pp. 57-68). New York: McGraw-Hill.

Ekman, P. (1993). Facial expression and emotion. American Psychologist, 48:384-392.

Ekman, P., & Friesen, W. F. (1984). Unmasking the face: A guide to recognizing emotions from facial expressions. Englewood Cliffs, NJ: Prentice-Hall.

Ellison, J. W., & Massaro, D. W. (1997). Featural evaluation, integration, and judgment of facial affect. Journal of Experimental Psychology: Human Perception and Performance, 23:213-226.

Fox, B. (1993). The human tutorial dialog project. Hillsdale, NJ: Erlbaum.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. Psychological Bulletin, 97:412-429.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. Applied Cognitive Psychology, 9:1-28.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group (1999). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1:35-51.

Graesser, A. C., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Person, N., & the Tutoring Research Group (2000). Using latent semantic analysis to evaluate the contributions of students in AutoTutor. Interactive Learning Environments, 8:129-147.

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. Archives of General Psychiatry, 43:280-283.

Johnson, W. L., Rickel, J. W., Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education, 11:47-78.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis of acquisition, induction, and representation of knowledge. Psychological Review, 104:211-240.

Lester, J., Towns, S., & FitzGerald, P. (1999). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. The International Journal of Artificial Intelligence in Education, 10:278-291.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. Psychonomic Bulletin & Review, 3:215-221.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. Journal of the Acoustical Society of America, 93:1097-1108.

Ochsman, R. B., & Chapanis, A. (1974). The effects of 10 communication modes on the behavior of teams during co-operative problem solving. International Journal of Man-Machine Studies, 6:579-619.

Paiva, A., & Machado, I. (1998). Vincent, an autonomous pedagogical agent for on-the-job training. In Proceedings of the Fourth International Conference on Intelligent Tutoring Systems (pp. 584-593). Berlin: Springer-Verlag.

Person, N. K., Bautista, L., Kreuz, R. J., Graesser, A. C. & the Tutoring Research Group (in press). The dialog advancer network: A conversation manager for AutoTutor. To appear in the ITS 2000 Proceedings of the Workshop on Modeling Human Teaching Tactics and Strategies. Montreal.

Person, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. Cognition and Instruction,13:161-188.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. Psychological Bulletin, 99:143-165.

Scherer, K. R. & Scherer, U. (1981). Speech behaviour and personality. In J. K. Darby (Ed.), Speech evaluation in psychiatry (pp. 115-135), New York: Grune and Stratton.

Trower, P., Bryant, B., & Argyle, M., (1978). Social skills and mental health. London: Methuen.

TruVoice 1.5 [Computer software]. (1997). Burlington, MA: Lernout & Hauspie Speech Products.

Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (1999). Improving an intelligent tutor's comprehension of students with Latent Semantic Analysis. In Artificial Intelligence in Education (pp. 535-542). Amsterdam: IOS Press.

Wierzbicka, A. (2000). The semantics of facial expressions. Pragmatics & Cognition, 8:147-184.

Author Notes

This research was funded by the National Science Foundation (SBR 9720314) in a three-year grant awarded to The Tutoring Research Group at The University of Memphis. Portions of this research were presented at the 1999 International Workshop on Affect in Interactions, Siena, Italy. The authors would like to thank Lee McCauley and Derek Harter for their assistance in programming the experiments.
Correspondence concerning this article should be addressed to Kristen E. Link, Department of Psychology, The University of Memphis, Memphis, Tennessee 38152-6400 (klink@memphis.edu).

Footnote

[1]Pitch and rate were selected as representative intonation variables to be included in this experiment. Feedback category and mouth curve were included because they were significant predictors in the previous experiments.

Table 1

Results from the Regression Analysis from Experiment 1

Variable	β	t	R^2
Feedback Category	-0.56	8.09	.318*
Rate	0.07	1.07	.006
Pitch	0.07	0.98	.005
Pause	-0.03	0.45	.001
Emphasis	0.01	0.18	.000

* $p < .01$

Table 2

Results from the Regression Analysis from Experiment 2

Variable	β	t	R^2
Mouth curve	0.69	8.48	.474*
Brow height	0.15	1.82	.022
Brow curve	-0.02	0.29	.001
Eye size	-0.04	0.52	.002

* $p < .01$

Table 3

Results from the Regression Analysis from Experiment 3

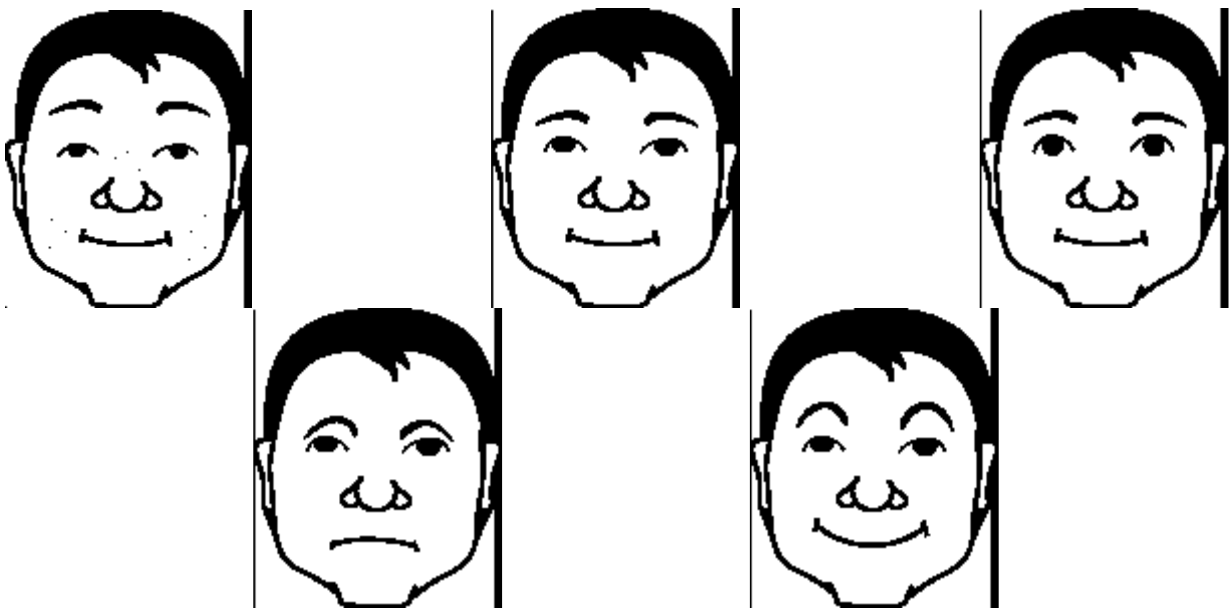
Variable	β	t	R^2
Mouth curve	0.73	18.12	.535*
Feedback category	0.56	13.88	.314*
Rate	0.05	1.10	.002
Pitch	-0.03	0.81	.001

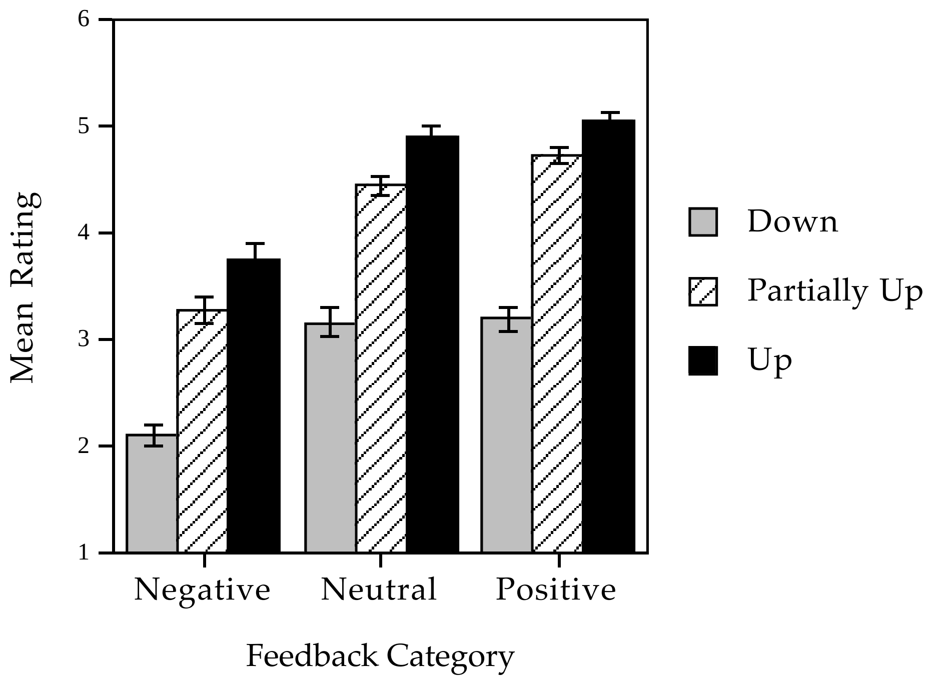
* $p < .01$

Figure Captions

Figure 1. Facial parameters manipulated in Experiment 1. The face in the center of the first row is AutoTutor's default facial expression. The faces in the top row show variations in eye size and brow height. The faces in the bottom row show variations in brow curve, brow height, and mouth curve.

Figure 2. Mean feedback rating for levels of mouth curve and feedback category from Experiment 3. Error bars depict 95% confidence intervals.




6
5
4
3
2
1
Mean Rating
Negative
Neutral
Positive
Feedback Category
Down
Partially Up
Up