Subtle Expressivity of Characters and Robots

Proceedings of the CHI2003 Workshop
Monday, April 7th, Ft. Lauderdale, Florida, USA

Noriko Suzuki and Christoph Bartneck

ABSTRACT

Humans, both consciously and unconsciously, use subtle expressions to indirectly communicate their emotions and intentions through variations of the gaze direction, pitch of speech and gesture speed. Humans also perceive changes in the internal states of others from subtle changes in their expressivities while interacting with them. Subtle expressivity plays the supporting part to the leading role of explicit expressivity, such as contents of speech or category of facial expressions. However, subtle expressivity plays an important role to gently regulate the relationship among the participants of an continuous interaction.

The design and evaluation of subtle expressivity are challenges for designers and researchers of embodied characters. How do the subtle variations in expression-influence the interaction? What types of subtle expressions are most important for the design of interactive media? How can the effect of the expressions be reliably measured? To address these questions, we gathered related studies for this workshop. We hope that this workshop will serve as a forum for vivid discussions and hence point out future research directions.

ORGANIZERS

Noriko Suzuki and Christoph Bartneck

PROGRAM COMMITTEE

Paolo Petta, Yasunori Morishima, Toru Takahashi, Yasuyuki Sumi, B.J. Fogg, Katherine Isbister, Helmut Predinger, Hideyuki Nakansishi, Jan Allbeck , Marc Swerts, Atsuchi Shimojima, Shigeo Morishima, Kiyoshi Kogure, Jacques Terken, Joel Chenu, Kazuhiko Shinozawa, Futoshi Naya, Toru Nakata, Kenji Mase and Akira Utsumi.

Noriko Suzuki
Media Information Laboratory, ATR
2-2-2 Hikaridai Seika Soraku
Kyoto 619-0288
Japan
+81-774-95-1421 (phone)
+81-774-95-1408 (fax)
noriko@atr.co.jp

Christoph Bartneck
Technical University of Eindhoven
Den Dolech 2
5600 MB Eindhoven
The Netherlands
+31 (0)40 247 5175 (phone)
+31 (0)40 247 5376 (fax)
christoph@bartneck.de

OPENING 9.15 - 09.30

SYSTEMS/METHODOLOGY I 09.30 - 10.30

Karen Liu and Rosalind Picard: *Subtle expressivity in a robotic computer*

Toru Nakata: *Expression with informatical factor in human robot interaction*

COFFEE BREAK 10.30 - 11.00

SYSTEMS/METHODOLOGY II 11.00 - 12.00

Timothy Bickmore and Rosalind Picard: *Subtle expressivity by relational agents*

Toru Takahashi, Christoph Bartneck and Yasuhiro Katagiri: *Show me what you mean - expressive media for online communities*

LUNCH 12.00 - 13.30

EMPIRICAL 13.30 - 14.30

Yugo Takeuchi, Takuro Hada and Yasuhiro Katagiri: *Effects of CG synthesized facial expressions on social attitude* (not presenting)

Scott Brave: User responses to emotion in embodied agents

Aoju Chen and Carlos Gussenhoven: *Language-dependence in the signalling of attitude in speech*

THEORY/OPINION I 14.30 - 15.30

Nigel Ward: *On the expressive competencies needed for responsive systems*

Katherine Isbister: *Social signals: using principles and methods from social psychology to guide subtle expression design*

COFFEE BREAK 15.30 - 16.00

THEORY/OPINION II 16.00 - 17.30

Dirk Heylen: *Facial expressions for conversational agents*

John Cowell and Aladdin Ayesh: *Automatic analysis of facial expressions: problems and solutions to data collection*

Thomas Erickson: *Silence, murmurs and applause: reflections on expressions of collections*

DISCUSSIONS 17.30 - 18.00

CLOSING 18.00 - 18.15

Subtle Expressivity in a Robotic Computer

Karen Liu
MIT Media Laboratory
20 Ames St. E15-120g
Cambridge, MA 02139 USA
kkliu@media.mit.edu

Rosalind W. Picard
MIT Media Laboratory
20 Ames St. E15-020g
Cambridge, MA 02139 USA
picard@media.mit.edu

ABSTRACT
A Robotic Computer, which moves its monitor "head" and "neck" but has no explicit face, is being designed to interact with users in a natural way for applications such as learning, rapport-building, interactive teaching, and posture improvement. In all these applications, the robot will need to move in subtle ways that express its state, and that promote appropriate movements in the user, but that don't distract or annoy. Toward this goal, we are giving the system the ability to recognize subtle expressions as well as the ability to have them. This paper describes the design of this system, initial findings on the perceived qualities of its expressions, and planned future work aimed at measuring behavioral effects of its expressiveness.

Keywords
Subtle expressiveness, robot, affect, emotion

1. INTRODUCTION: A Robotic Computer

The Robotic Computer is a computer monitor with a mechanical neck that physically interacts with the user by recognizing and responding to social-emotional cues. The system takes in perceptual data – visual data through an IBM Blue Eyes camera system to track user facial expressions (Kapoor & Picard, 2002) and posture data through a posture recognition system, to track behaviors related to user interest level (Mota, 2002) – and is capable of responding to these cues through mechanical and auditory expressions. The mechanical expressions include postural shifts like moving closer to the user, and "looking around" in a curious sort of way, while the auditory expressions, designed to be similar in spirit to the fictional Star Wars robot R2D2, are non-linguistic but aim to complement the movements, e.g. electronic sounds of surprise. The long-term plan is for the robot to learn relationships between the expressive cues it makes and the ones its user(s) make, as part of an effort to learn what kinds of actions and behaviors are appropriate for facilitating user task goals while not annoying or distracting the user.

There are three application areas that have motivated the creation of a Robotic Computer. The first application is the construction of a computerized learning companion (Kort et. al., 2001), that would try to help a child persist and stay focused on a learning task, possibly also mirroring some of the child's affective states in a way to increase awareness of the role those states play in propelling the learning experience. For example, if the child's face and posture show signs of intense interest in what is on the screen, the robotic terminal would hold very still so as not to distract the child. If the child shifts her posture and glance in a way that shows she is taking a break, the computer might do the same, and might note that as a good time to interrupt the child and provide scaffolding (encouragement, tips, etc.) to help with the learning progress. In so doing, the system not only acknowledges the presence of the child, and shows respect for her level of attentiveness, but also shows subtle expressions that, in human-human interaction, are believed to help build rapport and liking (e.g., LaFrance, 1982). By increasing likeability, we aim to make the system more enjoyable to work with and potentially facilitate task outcome, such as how long the child perseveres with the learning task.

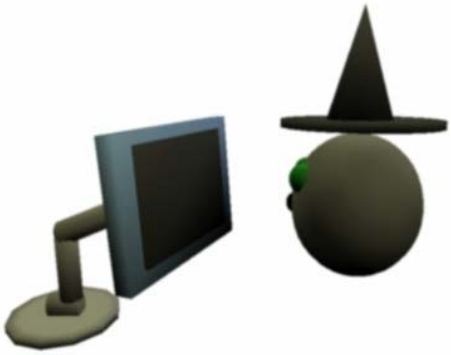


Figure 1: Virtual Robotic Computer Interacting with a Head Creature in a World Model

The second application area considers how a human can teach a robot new actions over time. In this application it is important that the robot's state be transparent to the user: for example, if the robot doesn't understand what action the person is requesting it to learn, it could look

confused. If the person hasn't shown it anything in a while, it could try to look curious. If its goal is to learn as much as possible from the user, without annoying the user, then its communication of its state via subtle expressions, in tandem with recognizing the user's expressions, will be important in achieving success. Additionally, in this application there is an issue of shared control (Breazeal, 2003): the user and character must negotiate, and subtle expressions will assist in this process.

The third application area is that of a health and posture improving system. As static sitting postures have become more prevalent in our workplaces (approximately 75 percent of the workforce has sedentary jobs) musculoskeletal problems -- in particular, low back pain and discomfort – have also increased (Faiks and Reinecke, 1998). Not only do these disorders contribute to higher medical expenses for corporations (Steelcase Inc., 2000), but they decrease worker productivity and increase fatigue. Experts agree that movement and alternating postures during sitting is beneficial. For instance, the research of Holm & Nachemson (1983) suggests that the flow of nutrient-rich fluids to and from the intervertebral discs increases with spinal movement. Adams (1983) found that alternating periods of activity and rest, and posture change, further boosts the fluid exchange, helping to nourish these spinal discs. Grandjean (1980) found that alternating unloading and loading of the spine (through movement) is ergonomically beneficial because this process pumps fluid in and out of the disc, thereby improving nutritional supply. To date, most efforts such as those from Steelcase Inc. (www.steelcase.com) and Herman Miller (www.hermanmiller.com) have focused on researching the ergonomic design of office furniture and environments that remain passive, such as chairs, tables, keyboards, office spaces etc. Successful designs, such as the LEAP chair by Steelcase, Inc. are able to show longitudinal benefits to health and productivity (Allie, P. and Palacios, N., 2002). In this case, we want to explore how similar benefits can be achieved through interactive technologies. The aim of the Robotic Computer would be to encourage movement and proper posture of the user, without distracting him from his primary task. Again, we think the combined perception and use of subtle mechanical and acoustic expressions will assist in achieving these multiple goals.

But why design a robotic computer? There has been increasing amounts of evidence that a physically present robot character, rather than an animated character on a screen, offers interesting advantages for applications such these. "Presence" is a term used to describe several dimensions of how similar a given interaction is to an actual social interaction between people (Lombard & Ditton, 2001). Endowing a technology with a strong social presence precedes the ability of the technology to develop a solid social rapport with the user (provided that the social presence is "strong" in a pleasing way and not in an annoying way). Further, we believe that there are interaction advantages to the physical presence of a machine that shares the same space as the user. For example, an important aspect of the mentor-student relationship is shared reference through cues such as directing attention, mutual gaze, etc. These cues are more easily communicated when both parties occupy the same physical space. Not only would a physical presence be advantageous in our application areas, but utilizing the computer terminal as a robotic interface will allow for natural communication of information to the user while not increasing the number of interfaces that a user would have to interact with.

New challenges arise with designing effective expressions within the physical limitations of a robot's motors and hardware. The rest of this paper describes the design of the current Robotic Computer (Sec 2), initial results on its expressive capabilities and how users perceive these (Sec 3), and future work (Sec 4.)

2. SYSTEM DESIGN AND USER PERCEPTIONS

2.1. System Design

The Robotic Computer system is built on the c4 behavior architecture system developed at the MIT Media Lab (Burke et. al., 2001). The IBM Blue Eyes system (Kapoor & Picard, 2002) in conjunction with the posture recognition system (Mota, 2002) sends the location coordinates of the user's pupil positions, whether the user is nodding or shaking his head, and the user's posture (leaning forward, leaning right, leaning left, leaning back, slumping, and sitting upright) to a virtual head creature in an internal world model which adopts the appropriate state based on sensor data as shown in Figure 1. The behavior engine contains a mental representation of the Robotic Computer that consists of a "perception and memory" section, an "action selection" section, and a "motor and graphic" section as shown in Figure 2. As data comes in from the world model to the cognitive architecture of the system, the "perception and memory" section affects the "action selection" section to determine which behaviors the terminal will take. The actions are passed to the "graphic and motor" section that controls the virtual computer and physical motors of the robot. The Robotic Computer's behavior patterns and actions adapt over time as the system builds a working memory with the user and the likelihood of taking certain actions changes based on past interactions and adaptations in the internal motivation system of the terminal.

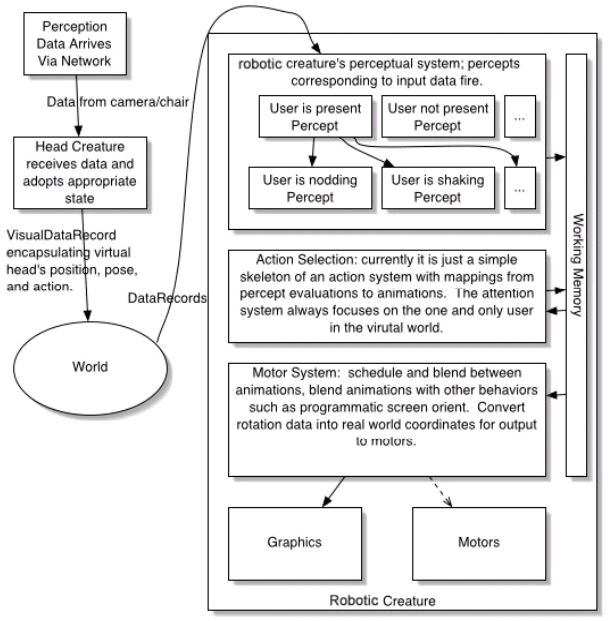


Figure 2: Robotic Computer System Architecture

2.2 Subtle Expressivity in the Robotic Computer

Unlike almost all other expressive robots built, this robot has no facial features that can be used to convey expression. Animation studios such as Disney have looked at how to convincingly portray life in inanimate objects; however, such essential animation tools such as "squash and stretch" (Thomas & Johnson, 1981) cannot be applied to a robotic terminal with limited degrees of freedom, as the motors and hardware are not malleable. What types of subtle behaviors can we use to convey attitude and emotional and cognitive state then? The Robotic Computer communicates all expression through the use of a mechanical neck with five degrees of freedom and the use of different mechanical sounds similar to those of R2D2. In order to evaluate if the robot has recognizable expressions, virtual simulations of the behaviors were created. These subtle expressions through use of audio and posture movements can be utilized to provide feedback and readable behavior to the user. Other questions would include how to effectively use these subtle expressions to manage expectations (help prevent annoyance, frustration, boredom in both user and computer)? The Robotic Computer could also manage the pace of the interaction through varying movement or audio speed as Kismet, the sociable robot, did (Breazeal, 2003).

2.3 User Perceptions

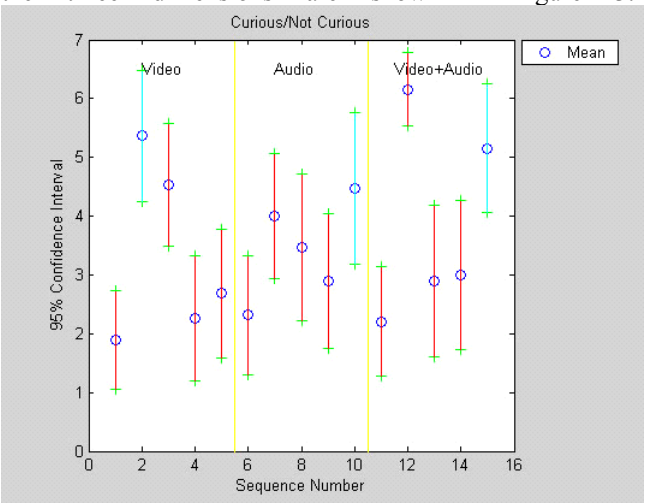
A preliminary study to measure if the first expressive behaviors of the Robotic Computer could accurately be recognized was conducted. While it is important to also measure if the expression actually made the user feel the intended way, this study only looked at user perceptions of the agent's behavior. The study consisted of an online survey where nineteen subjects viewed fifteen different video clips of an animated version of the Robotic Computer and/or heard sound sequences that the terminal would emit. Each of the five different video or audio sequences was designed to convey a certain expressive behavior. The fifteen sequences can be found at http://www.media.mit.edu/~kkliu/roco.html.

Preliminary Results

After each sequence, the user was asked to rate how expressive the computer appeared along five different dimensions. The dimensions used a seven-point scale and are listed below:

- Not Welcoming (1) – Welcoming (7)
- Sad (1) – Happy (7)
- Not Curious (1) – Curious (7)
- Not Confused (1) – Confused (7)
- Not Surprised (1) – Surprised (7)

For purposes of analyzing the data on the same scale, the happy/sad data was inverted to Not Sad (1) – Sad (7), as the sequence was designed with a sad expression in mind. For each dimension, the subjects were also given a "Not Applicable" choice. After performing a two-tailed t-test on this data, we found that there was significant recognition of the behavior sequences that were designed to express curiosity, sadness, and surprise. The results for the three dimensions are show in Figure 3.



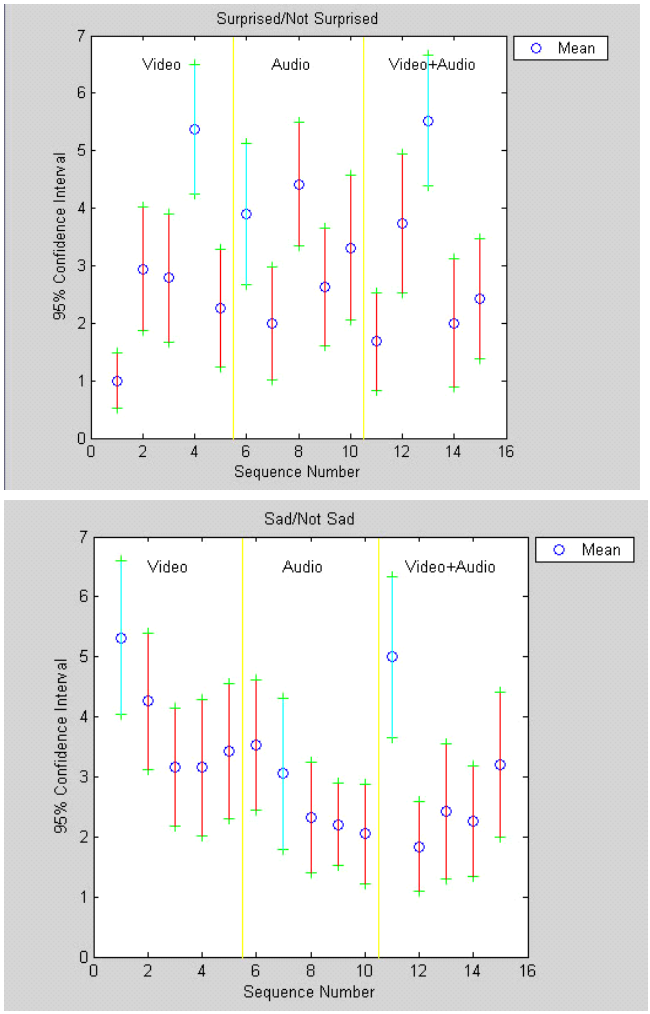


Figure 3: Mean and 95% Confidence Interval for Each Sequence on Curious, Surprised, and Sad Dimensions. Sequences designed for respective dimensions shown in blue.

The intended expression for each sequence is shown in Figure 4.

Sequence	Intent	8	Confused
1	Sad	9	Welcoming
2	Curious	10	Curious
3	Confused	11	Sad
4	Surprised	12	Confused
5	Welcoming	13	Surprised
6	Surprised	14	Welcoming
7	Sad	15	Curious

Figure 4: Intended Expression for Each Sequence

It was not surprising to find very little significant differences for the confused and welcoming expressive behaviors. As both confused and curious behavior could be seen as questioning behavior, there was expected overlap in the perception of how confused the computer appeared to be. As for a welcoming expression, this behavior could easily be interpreted as happy, which would explain some spread in the results. In the future, we are interested in measuring change in user behavior as well as perceived expression.

3. FUTURE WORK

Future directions include looking at how this Robotic Computer can create its own subtle expressions by learning over time which behaviors were effective in its different interactions as well as measuring the behavioral effects of its expressiveness. One major challenge will be to learn when an appropriate intervention time is in order to avoid annoying the user with a moving terminal. For example, when the posture recognition system (Mota, 2002) detects that the child is taking a break, this may signal a good time for the robotic computer to request permission for itself to take a break and then exercising its motors in a manner that resembles stretching. We plan on measuring effects related to the three application areas mentioned above in following questions:

- Does interacting with a robotic terminal that recognizes and responds to affective state cause a user to persist in the learning and teaching interaction?
- Does the system's movement encourage your own movement and alternating postures as you work at a terminal?
- Can the machine learn when its expression was ineffective or annoying? Can it be taught new behaviors? Can it learn when to hold still as people are working on a task?
- Do I like the computer more? Do I enjoy working with it more?

4. CONCLUSION

A Robotic Computer, having no explicit facial features, has been designed to recognize and exhibit a variety of subtle expressions. Additionally, its ability to communicate a set of expressions via motion and auditory cues has been evaluated with nineteen subjects. This paper has emphasized the importance of combining display of expressions with recognition of expressions, with the expectation that an appropriate interaction between these can lead to beneficial effects and behaviors in a variety of tasks.

ACKNOWLEDGMENTS

We thank Cynthia Breazeal, Jesse Gray, Ashish Kapoor, Cory Kidd, John McBean, and David Lafferty for their participation in building the Robotic Computer. This research was supported in part by NSF ROLE grant REC-0087768.

References

1. Adams, M. A. (1983). The Effect of Posture on the Fluid Content of Lumbar Intervertebral Discs. Spine, v8:n6.
2. Allie, P. & Palacios, N. (2002). Steelcase Leap Chair's Impact on Office Work Effectiveness, Productivity and Health. Steelcase Inc. Study Summary.
3. Breazeal, C. (2003). Social Interactions in HRI: The Robot View. To appear in IEEE Transactions in Systems, Man, and Cybernetics. MIT Media Lab. Cambridge, MA.
4. Burke, R., Isla, D., Downie, M., Ivanov, Y., Blumberg, B. (2001). CreatureSmarts: The Art and Architecture of a VirtualBrain. In Proceedings of the Game Developers Conference (pp.147-166). San Jose, CA.
5. Faiks, F. & Reinecke, S. (1998). Investigation of Spinal Curvature While Changing One's Posture during Sitting. Contemporary Ergonomics. M.A. Hanson, Taylor & Francis.
6. Grandjean, E. (1980). Fitting the Task to the Man. London: Taylor and Francis.
7. Holm S. & Nachemson A. (1983). Variations in Nutrition of the Canine Intervertebral Disc Induced by Motion. Spine, v.8, no.8.
8. Kapoor, A. & Picard, R. (2002). Real-Time, Fully Automatic Upper Facial Feature Tracking
   In the Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition Washington D.C.
9. Kort, B., Reilly, R., & Picard, R. (2001). An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy – Building a Learning Companion. In Proceedings of IEEE International Conference on Advanced Learning Technologies. Madision, USA.
10. LaFrance, M. (1982). Posture Mirroring and Rapport. In M. Davis (Ed.), Interaction Rhythms: Periodicity in Communicative Behavior (pp. 279-298). New York: Human Sciences Press, Inc.
11. Lombard, M., Ditton, T.B., Crane, D., Davis, B., Gil-Egul, G., Horvath, K. and Rossman, J. (2000). Measuring Presence: A Literature-Based Approach to the Development of a Standardized Paper-and-Pencil Instrument. In *Presence 2000: The Third International Workshop on Presence*. Delft, the Netherlands.
12. Mota, S. (2002). Automated Posture Analysis for Detecting Learner's Affective State. MS Thesis. MIT Media Lab. Cambridge, MA.
13. Reeves, B. & Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. New York: Cambridge University Press.
14. Reinecke, S., Bevins, T., Weisman, J., Krag, M.H. and Pope, M.H. (1985). The Relationship between Seating Postures and Low Back Pain. Rehabilitation Engineering Society of North America, 8th Annual Conference, Memphis, TN.
15. Steelcase Inc. (2000). Musculoskeletal Disorders (S10647). Grand Rapids, MI: Steelcase Inc.
16. Thomas, F., & Johnson, O. (1981). The Illusion of Life: Disney Animation. New York: Hyperion.

Expression with Informatical Factor in Human Robot Interaction

Toru NAKATA
Digital Human Laboratory
National Institute for Advanced Science and Technology (AIST).
also CREST Researcher, Japan Science and Technology Corporation.
2-41-6 Aomi,
Koto-ku, Tokyo, 145-0064, JAPAN
toru-nakata@aist.go.jp
http://staff.aist.go.jp/toru-nakata/

ABSTRACT
Besides meanings of verbal and nonverbal expressions, information transfer in communicational interactions effects impressions on opponents. Even on interactions between a human and a robot, a user may get some impression of temperament of the robot. This paper proposes a method of informatical analysis of human-robot-interaction. In addition, an experiment is conducted to show that informatical difference on robot's reaction selection effects impression that users have about the robot's temperament.

Keywords
Informatical analysis of interaction, tactile interaction, human-robot-interaction.

INTRODUCTION

There should be 2 ways to analyze communicational interaction: one is semantic analysis and the other is objective and non-semantic analysis.

Ethologists prefer to employ objective methods to analyze animal communications, because those methods do not require any assumptions on meanings of communicational signals. In non-semantic researches, animal interactions are measured regarding explicit values such as time-length of touch or frequency of utterances, and those data are analyzed with quantitative, statistical and Shannon informatical methods.

There are some interesting experiments that were analyzed with Shannon informatics method. Dingle[1] analyzed aggressive interactions between 2 mantis shrimps to find the existence of change on information flow during develop of dominance order. Likewise, Steinberg and Conant[2] analyzed the interaction of grasshoppers, and Hazlett[3] researched information flow among hermit crabs.

The common conclusion on informatic characteristic of animal interactions can be summarized as the followings:

1) Very few information flows on encounters of animals that are ignorant each other.
2) Very high information transfer efficiency is achieved on encounters of animals with social relationship or dominance order among them.
3) Intermediate information transfer efficiency is achieved on interactions when the animals are developing the social relationship or struggling for dominance.

The purpose of this paper is to examine that those facts is valid even on human-robot-interactions.

Meaning-neutral Informatical Analysis on Communicational Interaction

Precedent ethological researches[1,2,3] provide informatical analysis method of interaction. In Shannon informatics, transferring information is defined apart from the meaning of signals.

Consider a simple dyadic interaction between a human and a robot like shown in Fig. 1. The human make action toward the robot, and the robot make a reaction. Assume the frequencies of each pair of an action and a reaction are observed as shown in Table 1.

Informatical theory provides analyzing method on information flow in the dyad. Informatical uncertainty of the dyad (H_{dyad}) is calculated as formula (1):

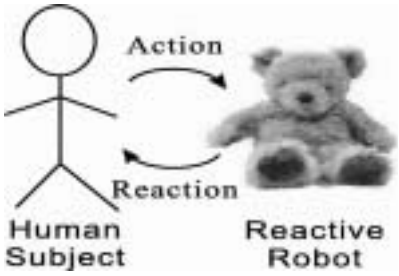


Figure 1: Dyad model of human-robot-interaction

Table 1: A fictional example of dyadic sociometric matrix

Frequency		Reaction		Sum
		Approch	Retreat	
Action	Call	6	2	8
	Touch	1	8	9
Sum		7	10	17

$$H_{dyad} = -\frac{6}{17}\log_2\frac{6}{17} - \frac{2}{17}\log_2\frac{2}{17} - \frac{1}{17}\log_2\frac{1}{17} - \frac{8}{17}\log_2\frac{8}{17} \quad \cdots(1)$$
$$= 1.65[\text{bit}]$$

Other informatical values are also calculated as the following.

$$H_{Action} = -\frac{8}{17}\log_2\frac{8}{17} - \frac{9}{17}\log_2\frac{9}{17} \quad \cdots(2)$$
$$= 1.00[\text{bit}]$$

$$H_{\text{Reaction}} = -\frac{7}{17}\log_2\frac{7}{17} - \frac{10}{17}\log_2\frac{10}{17} \quad \cdots(3)$$
$$= 0.98[\text{bit}]$$

$$I = H_{Action} + H_{\text{Reaction}} - H_{Dyad} = 0.33[\text{bit}] \quad \cdots(4)$$

$$D = I / H_{\text{Reaction}} = 0.34 \quad \cdots(5)$$

$$C = I / H_{Action} = 0.34 \quad \cdots(6)$$

H_{Action} is the uncertainty on distribution of the human's action. $H_{reaction}$ is the uncertainty of the robot's reaction. I is the amount of transferred information. D and C are information transfer efficiencies: D means the degree of how the robot selects reactions depending on the human reactions. In this example, D was 0.34. That means 34% of action selections was distinguished by the robot. C is controllability of the robot meaning the degree of how much the human can decide the robot reaction.

In general, high information transfer efficiencies mean that each action causes a stable reaction. In contrast, low efficiencies mean absences of communication order.

Hypotheses on Relationship between Impressions and Informatical Transfers in Human Robot Interaction

The following hypotheses on human-robot-interaction can be considered:

1) A robot that can react against a human user with high information transfer efficiency provides impression that the robot is clever so that the reaction is selected with some reason.
2) Information transfer efficiency effects on preference of a human user on reactive robots. For example, humans might favorite robots that behave with high-efficiency stable reaction rule.

An experiment is carried out to examine the hypotheses.

Experimental Equipment

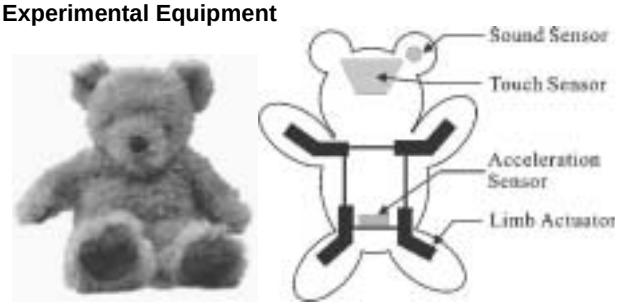


Fig. 2: Experimental Robot.

Fig 3: Scene of experiment.

A teddy bear robot shown in Fig. 2 is employed as an opponent of interaction with humans. The robot has tactile, acoustic and acceleration sensors to perceive human actions.

Each human subject holds the robot and make actions like shown in Fig 3.

The robot can make 4 types of reactions shown in Fig. 4 to 7.

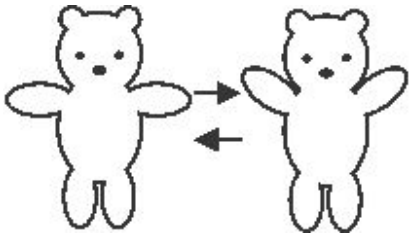


Fig. 4: Reaction 1; Raising arms.

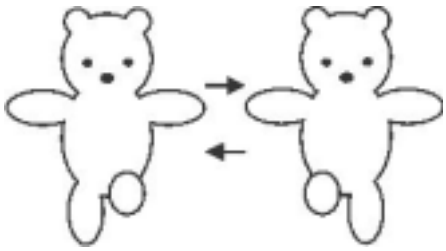


Fig. 5: Reaction 2; Moving legs.

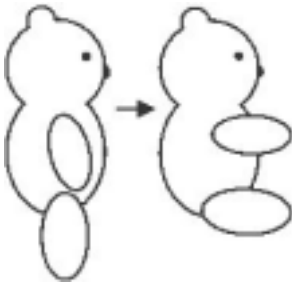


Fig. 6: Reaction 3; Curing.

Fig. 7: Reaction 4; Release and Rest.

In order to compare impression over different information transfer efficiencies, 3 rules on selection of robot's reaction are prepared.

The high information transfer efficiency reaction rule decides reaction without randomness. Table 2 shows the probabilities that each reaction is selected responding to each action. The estimated value of efficiency is 1 in this case.

Under the intermediate information transfer efficiency reaction rule, reactions are selected with some randomness described in Table 3. The efficiency is estimated as 0.5.

The low information transfer efficiency reaction rule selects reactions independently against human actions. The efficiency is estimated as 0.

Table 2: Selection Probabilities of the High Information Transfer Efficiency Reaction Rule.

Action of human	Reaction of Robot			
	Raise Arms	Move Legs	Curl	Rest
Call	1	0	0	0
Touch Forehead	0	1	0	0
Shake Body	0	0	1	0
Keep Silence	0	0	0	1

Table 3: Selection Probabilities of the Intermediate Information Transfer Efficiency Reaction Rule.

Action of human	Reaction of Robot			
	Raise Arms	Move Legs	Curl	Rest
Call	0.5	0.5	0	0
Touch Forehead	0	0.5	0.5	0
Shake Body	0	0	0.5	0
Keep Silence	0.5	0	0	0.5

Table 4: Selection Probabilities of the Low Information Transfer Efficiency Reaction Rule.

Action of human	Reaction of Robot			
	Raise Arms	Move Legs	Curl	Rest
Call	0.25	0.25	0.25	0.25
Touch Forehead	0.25	0.25	0.25	0.25
Shake Body	0.25	0.25	0.25	0.25
Keep Silence	0.25	0.25	0.25	0.25

Experimental Procedure

Each subject is explained about the robot's abilities of sensing human's action and making reaction. Then the subject is asked to hold the robot like Fig. 3, and the robot starts moving. The subject experiences 3 interaction sessions. Each session lasts 90 seconds, and intermissions last 30 seconds. The reaction rule is different in each session, so that the subject experiences all of the 3 different information transfer efficiencies, namely 1, 0.5, and 0. The order of presenting the reaction rule is randomized.

After finishing the 3 sessions, the following 2 impression is questioned to the subject; 1) impression of existence of the robot intelligence, and 2) loveliness of the robot. Answers are collected in relative-ordered evaluation in comparison with the 3 sessions, i.e. each subject answers in terms of “most impressive among the 3 sessions”, “secondly impressive” and “least impressive.”

Experimental Result and Discussion

The number of the subjects was 20 consisted of 7 female subjects and 13 male subjects.

Fig. 8 describes the result of impression intensity of existence of the robot intelligence. The robot reactions under high information transfer efficiency (D=1) obtained

11 most-impressive answers. Significant difference is found on impression intensity between D=0.5 type and D=1 type by using Mann-Whitney U-test with 10% significance level.

This result agrees with the hypothesis I described above: the stable and ruled selection of reaction produced the impression of the robot's intelligence. However, this result is not strong, while it is statistically significant. More investigations are required especially on the fact that strength of the impression did not increase in monotone against the information transfer efficiency.

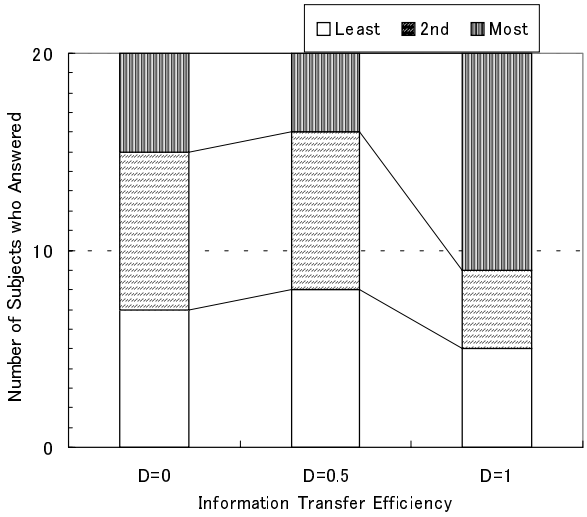


Fig. 8: Answer distribution on impression intensities of existence of the robot's intelligence.

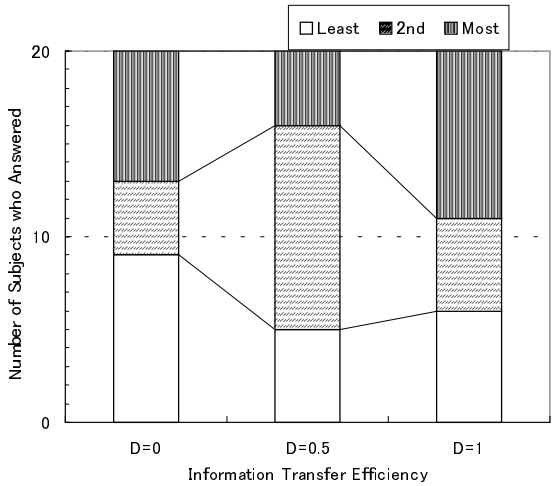


Fig. 9: Answer distribution on impression intensities of loveliness of the robot.

Fig. 9 shows the result of intensity of loveliness. According to binominal test, the probability of that 11 or more votes out of 20 votes concentrate in 1 category is 3.8%. So it is significant that the reaction rule with intermediate information transfer efficiency (D=0.5) obtained relatively less most-lovely and least-lovely answers.

Nine subjects felt most loveliness from D=1 type, while 6 other subjects answered D=1 is least lovely. Likewise D=0 type obtained relatively large number of best and worst answers.

This means that the subjects were relatively indifferent to D=0.5 type. However, more data are needed to detect some significant tendency.

In addition, some diversity is found on human's preference: some subjects like stable interactions, while the other people dislike it.

The hypothesis II states that information transfer efficiency effect preference of a human toward a robot. The author thinks this hypothesis will be supported when additional experiments carried out, since the data gathered so far do not disagree with it.

Conclusion

Informatical analysis on interactions between a human and a robot showed some relationship of the impression on the robot and informational transfer efficiency of the interaction.

This fact indicates the possibility of controlling impression of robots, since adequate information transfer efficiency produced intelligence-existence and loveliness impressions.

ACKNOWLEDGMENTS

This research was funded by Seiko Epson Corp.

REFERENCES

1. Dingle, H. A statistical and information analysis of aggressive communication in the mantis shrimp gonodactylus bredini manning. *Animal Behaviour*, 17 (1969), 561-575.
2. Steinberg, J.B. & Conant, R.C. An informational analysis of the inter-male behaviour pf the grasshopper chortophaga viridifasciata. *Animal Behaviour*, 22 (1974), 617-627.
3. Hazlett, B. Patterns of information flow in the hermit crab calcinus tibicen. *Animal Behaviour*, 28 (1980), 1024-1032.

Subtle Expressivity by Relational Agents

Timothy W. Bickmore
MIT Media Laboratory
20 Ames St. E15-120G
Cambridge, MA 02139 USA
+1 617 253 6341
bickmore@media.mit.edu

Rosalind W. Picard
MIT Media Laboratory
20 Ames St. E15-020G
Cambridge, MA 02139 USA
+1 617 253 0611
picard@media.mit.edu

ABSTRACT
Relational agents are computational artifacts designed to build long-term, social-emotional relationships with their users. In this paper we argue that subtle expressivity is especially crucial in human-computer interactions with relational agents in which social tasks such as relationship building or negotiation are being performed. We discuss these issues in the context of a relational agent designed to interact repeatedly with users during a one-month exercise adoption program.

Keywords
Subtle expressiveness, affect, emotion, non-verbal behavior, long-term interaction, embodied conversational agent

1. RELATIONAL AGENTS

Over the last three years we have begun investigating the development and use of Relational Agents; computational artifacts designed to build and maintain long-term, social-emotional relationships with their users (Bickmore, 2003). These can be purely software humanoid animated agents, but they can also be non-humanoid or embodied in various physical forms, from robots, to pets, to jewelry, clothing, hand-held, and other interactive devices. Central to the notion of relationship is that it is a persistent construct, spanning multiple interactions, thus Relational Agents are explicitly designed to remember past history and manage future expectations in their interactions with users. Finally, relationships are fundamentally social and emotional, and detailed knowledge of human social psychology--with a particular emphasis on the role of affect--must be incorporated into these agents if they are to effectively leverage the mechanisms of human social cognition in order to build relationships in the most natural manner possible.

Human relationships are primarily established in the context of face-to-face conversation. In addition to primacy of this interactional modality, face-to-face conversation affords many channels of subtle expressivity that are especially crucial in relational conversation. Thus, we have focused on developing relational agents that have anthropomorphic forms, implemented as embodied conversational agents (Cassell, Sullivan, Prevost, & Churchill, 2000)or sociable robots (Breazeal, 2002). The social and subtly expressive communicative cues used by these agents are gleaned from studies of human-human face-to-face conversation.

2. SUBTLE EXPRESSIVITY IN HUMAN RELATIONAL INTERACTIONS

Several studies have demonstrated what most business people already know: when the social aspects of an interaction are especially important--such as when you are getting to know a new client or negotiating an important deal--nothing beats face-to-face interaction. In a review of studies comparing video and audio-mediated communication, Whittaker and O'Conaill concluded that video was superior to audio only for social tasks while there was little difference in subjective ratings or task outcomes in tasks in which the social aspects were less important (Whittaker & O'Conaill, 1997). They found that for social tasks, interactions were more personalized, less argumentative and more polite when conducted via video-mediated communication, that participants believed video-mediated (and face-to-face) communication was superior, and that groups conversing using video-mediated communication tended to like each other more, compared to audio-only interactions. Obviously, some nonverbal communication must be responsible for these differences.

We define "subtle expressivity" to be those communicative behaviors used to convey any kind of meaning *except* for the primary propositional meaning of a communicative act.

The set of general functions of subtly expressive communicative behaviors studied in the literature on human-human communication is expansive, but includes:

- Interactional functions, such as: turn-taking (Duncan, 1974); engagement, disengagement, greeting and farewell (Kendon, 1990); and grounding (Clark, 1992).
- Framing (i.e., the use of "contextualization cues" to mark the type of interactional segment being initiated) (Tannen, 1993).
- Social deixis (i.e., marking relational stance) (Levinson, 1983).
- Conveying attitude (e.g., interpersonal attitude) (Argyle, 1988).

- Emphasis.
- Conveying emotional state (Argyle, 1988).

There is a correspondingly large array of communicative behaviors that have been found to be used to perform these functions, and there is a many-to-many mapping between them (a given behavior can be used to perform multiple functions and a given function can be performed by multiple behaviors). For example, emphasis can be marked using intonation, eyebrow raise, hand gesture or facial expression, whereas facial expression can be used not only for emphasis but for conveying attitude and emotional state. Further, most of these behaviors can also be used to convey non-subtle, propositional content (e.g., an isolated smile to indicate agreement).

Relatively little work has been done on studies of these behaviors in long-term interactions. It is known that entrainment (lexical, syntactic, prosodic, and postural)(Clark, 1992; LaFrance, 1982) occurs within a single interaction and likely continues to increase as a given dyad interacts over time. Gain-loss theory is a related phenomenon that posits that people who start out different but change to become more like each other over time along some trait or state dimension of personality will like each other more (Aronson & Linder, 1965). Forms of social deixis must necessarily change over time as the relationship between interlocutors evolves, and language use must take into account the increasing common ground between them as well as their shared (historical) discourse context. Relational partners also tend to develop idiomatic expressions (Bell & Healey, 1992), and it is likely that these idioms extend into the nonverbal domain of "subtle" behaviors.

3. AUTOMATICITY AND THE "SUSPENSION OF DISBELIEF"

Face-to-face conversation is hard work. Interlocutors must track task, conversational, and relational goals at varying levels of abstraction and respond to the dynamic moves of their partner by planning, re-planning and generating utterances to satisfy as many goals as possible, all within a few milliseconds (Berger, 1997; Waldron, Cegala, Sharkey, & Teboul, 1991). No wonder, then, that the production of most subtly expressive behaviors is completely automatic and unconscious (some researchers have even termed this level of interaction a "conversation between limbic systems" (Buck, 1993)).

Many researchers have argued that anthropomorphic agents must work actively to "suspend disbelief" in their users, in order for users to conduct natural, social interaction with them (Bates, 1994). We argue that this is exactly backwards. Studies by Reeves and Nass and others have demonstrated repeatedly that people respond to social cues from a computer in the same way that they respond to these cues from other people (Reeves & Nass, 1996). Further, people do this automatically and unconsciously; most people state emphatically that they would never behave according to social rules when interacting with a computer, immediately after completing an experiment in which they were observed to do just that.

Our experience has been that belief in a computer agent's acting like a person is automatic from the first moment of an interaction, and it is *this* belief which must be "suspended" by the user, when the agent fails to meet their expectations by behaving inappropriately. In a recent study of interactions with an animated real estate agent, we learned that her persona was inappropriate for the task (users rated her as unfriendly and cold) and that her nonverbal behavior was particularly inappropriate for social dialogue (users preferred conducting social dialogue with her over a telephone link, but preferred conducting real estate business "face to face") (Bickmore & Cassell, to appear). This experience taught us that, while it is easy to get users to readily engage an agent in social dialogue, it is an extremely challenging task to get the agent to maintain the illusion of human-like behavior over time; every aspect of the agent's appearance and verbal and non-verbal behavior must be correct or users will begin to discredit it.

4. LONG-TERM INTERACTIONS WITH RELATIONAL AGENTS

We have spent the last year developing and evaluating an exercise advisor system, in order to explore long-term relational interactions between people and relational agents. This system uses an embodied conversational agent who plays the role of an exercise advisor that users interact with on a daily basis during a one-month exercise adoption program. Exercise adoption was selected as a task domain because the current guidelines from the CDC and ACSM call for all Americans to accrue at least 30 minutes of moderate or better physical activity on most, if not all days of the week. This motivates a daily check in with an exercise advisor agent, thus giving the agent an opportunity to build a relationship with users over repeated interactions.

4.1 Subtle Expressivity in the Exercise Advisor

The subtle behaviors used by this agent include nonverbal markers of relational stance and framing. One of the most consistent findings on relational stance is that the use of "immediacy" behaviors--including close conversational distance, direct body and facial orientation, forward lean, increased and direct gaze, smiling, pleasant facial expressions and facial animation in general, nodding, frequent gesturing and postural openness--projects liking for the other and engagement in the interaction, and is correlated with increased solidarity (Argyle, 1988; Richmond & McCroskey, 1995). The specific relational cues implemented in the exercise advisor agent include: increased proximity, more frequent communicative headnods, eyebrow raises, and hand gestures, and less frequent gaze aways.

Based on a series of pilot studies of human fitness trainers and their clients, four conversational frames were

developed for the agent: a task frame, for information delivery; a social frame, for greetings, farewells, and social dialogue; an empathetic frame, for empathy exchanges (following Klein (Klein, Moon, & Picard, 2002)); and an encouragement frame for coaching and motivating users. Contextualization cues were primarily encoded in proximity, facial expression and prosody.

These nonverbal behaviors were implemented as extensions to BEAT, an extensible text-to-embodied-speech translator (Cassell, Vilhjálmsson, & Bickmore, 2001). The extensions were implemented in a "Stance Manager" module that takes relational stance and conversational frame as inputs, and outputs modifications to be applied to the agent's default nonverbal behavior. Figure 1 shows examples of the exercise advisor agent in various relational stances and conversational frames.

4.2 Long-Term Changes in Behavior

The exercise advisor agent changes its behavior over time as a function of the number of interactions with a subject and increasing common ground. The daily interactions with the agent are scripted using Augmented Transition Networks (Woods, 1986) and are designed to increase relational closeness over time, for example by increasing the amount and intimacy of social dialogue used. In addition, the agent learns facts about the subject (stored in a database between interactions) and modifies its future dialogues accordingly. While neither of these long-term adaptations directly impact the subtle behaviors described in the previous section, they do change the frequency with which different conversational frames are used. Further, social dialogue itself may be viewed as a type of subtle behavior ("phatic communion" being the best exemplar (Malinowski, 1923)) in that little propositional meaning is typically conveyed in this frame.

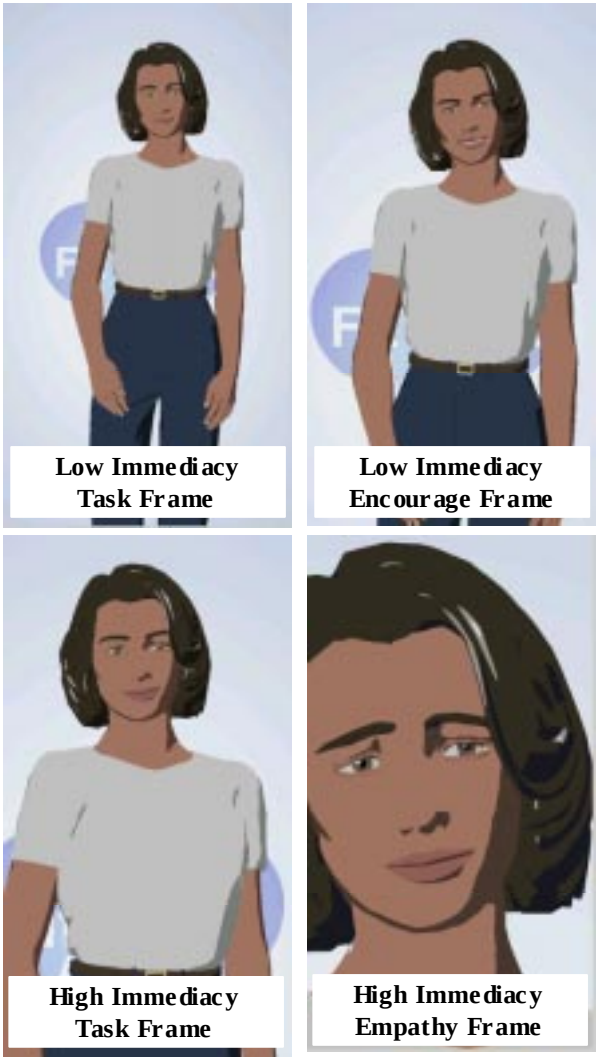


Figure 1. Effects of Relational Stance and Frame on Proximity and Facial Expression

5. EVALUATION

Evaluation of subtle expressivity can take place on multiple levels. First, evaluations can be performed that determine whether users can correctly perceive the expressive behaviors or not. Along these lines, a series of surveys was conducted on the Exercise Advisor agent to determine which nonverbal behaviors (postures and facial expressions) conveyed the intended relational stance and emotional displays.

Second, the impact of subtly expressive behaviors on user's attitudes towards the agent and the interaction can be assessed using self-report instruments and behavioral measures. A large-scale evaluation of the exercise advisor agent was recently completed, in which 100 users interacted with it on a daily basis for a month. The study used a between-subjects experimental design, with differences between two of the conditions intended to demonstrate the efficacy of long-term relationship-building strategies used by the agent. Results indicate that when the agent used these relational strategies, users reported liking, trusting and respecting the agent more, feeling that it liked, trusted, respected and cared about them more, and an increased desire to continue working with it, compared with users in the non-relational condition. The primary instrument used to assess these relational effects was the Working Alliance Inventory (Horvath & Greenberg, 1989), and we found that relationship building strategies resulted in significantly greater ratings on the bond dimension of this scale on day 7 ($t(69)=2.10$, $p<.05$) and on day 27 ($t(60)=2.54$, $p=.007$) of the intervention.

Finally, and most importantly, the impact of subtly expressive behaviors on task outcome should be measured. In the Exercise Advisor study we measured levels of physical activity through both self-report questionnaire and pedometer readings, for results refer to (Bickmore, 2003).

6. CONCLUSION

This paper has argued that Relational Agents, which build long-term social-emotional relationships with their users, need to *appropriately* employ subtle expressive capabilities. A relational agent with several subtle expressive capabilities has been designed, built and tested with over a hundred users, and shown to increase their liking of, trusting in and respecting of the agent, their feeling that it liked, trusted, respected and cared about them more, and an increased desire to continue working with it, relative to users who interacted with a non-relational agent.

ACKNOWLEDGMENTS

Thanks to Justine Cassell, Amanda Gruber, Candy Sidner, and the many folks who contributed to the development and evaluation of the Exercise Advisor system.

References

1. Argyle, M. (1988). Bodily Communication. New York: Methuen & Co. Ltd.
2. Aronson, E., & Linder, D. (1965). Gain and loss of esteem as determinants of interpersonal attractiveness. Journal of Experimental and Social Psychology, 1, 156-171.
3. Bates, J. (1994). The role of emotion in believable agents. Communications of the ACM, 37(7), 122-125.
4. Bell, R., & Healey, J. (1992). Idiomatic communication and interpersonal solidarity in friends' relational cultures. Human-Communication-Research, 18(3), 307-335.
5. Berger, C. (1997). Planning Strategic Interaction. Mahwah, NJ: Lawrence Erlbaum Associates.
6. Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships. MIT, Cambridge, MA.
7. Bickmore, T., & Cassell, J. (to appear). Social Dialogue with Embodied Conversational Agents. In N. Bernsen (Ed.), Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems. New York: Kluwer Academic.
8. Breazeal, C. (2002). Designing Sociable Robots. Cambridge, MA: MIT Press.
9. Buck, R. (1993). The spontaneous communication of interpersonal expectations. In P. D. Blanck (Ed.), Interpersonal expectations: Theory, research, and applications (pp. 227-241). New York: Cambridge University Press.
10. Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). Embodied Conversational Agents. Cambridge: MIT Press.
11. Cassell, J., Vilhjálmsson, H., & Bickmore, T. (2001). BEAT: The Behavior Expression Animation Toolkit. Paper presented at the SIGGRAPH '01, Los Angeles, CA.
12. Clark, H. H. (1992). Arenas of Language Use. Chicago, IL: University of Chicago Press.
13. Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. Language in Society, 3, 161-180.
14. Horvath, A., & Greenberg, L. (1989). Development and Validation of the Working Alliance Inventory. Journal of Counseling Psychology, 36(2), 223-233.
15. Kendon, A. (1990). A Description of Some Human Greetings, Conducting interaction: Patterns of behavior in focused encounters (pp. 153-207). Cambridge: Cambridge University Press.
16. Klein, J., Moon, Y., & Picard, R. (2002). This Computer Responds to User Frustration: Theory, Design, Results, and Implications. Interacting with Computers, 14, 119-140.
17. LaFrance, M. (1982). Posture Mirroring and Rapport. In M. Davis (Ed.), Interaction Rhythms: Periodicity in Communicative Behavior (pp. 279-298). New York: Human Sciences Press, Inc.
18. Levinson, S. C. (1983). Pragmatics. Cambridge: Cambridge University Press.
19. Malinowski, B. (1923). The problem of meaning in primitive languages. In C. K. Ogden & I. A. Richards (Eds.), The Meaning of Meaning: Routledge & Kegan Paul.
20. Reeves, B., & Nass, C. (1996). The Media Equation. Cambridge: Cambridge University Press.
21. Richmond, V., & McCroskey, J. (1995). Immediacy, Nonverbal Behavior in Interpersonal Relations (pp. 195-217). Boston: Allyn & Bacon.
22. Tannen, D. (1993). Introduction (Framing in Discourse). In D. Tannen (Ed.), Framing in Discourse (pp. 3-13). New York: Oxford University Press.
23. Waldron, V. R., Cegala, D. J., Sharkey, W. F., & Teboul, B. (1991). Cognitive and tactical dimensions of conversational goal management. In K. Tracy & N. Coupland (Eds.), Multiple goals in discourse (pp. 101-119). Clevedon: Multilingual Matters.
24. Whittaker, S., & O'Conaill, B. (1997). The Role of Vision in Face-to-Face and Mediated Communication. In K. Finn & A. Sellen & S. Wilbur (Eds.), Video-Mediated Communication (pp. 23-49): Lawrence Erlbaum Associates, Inc.
25. Woods, W. A. (1986). Transition Network Grammars for Natural Language Analysis. In B. J. Grosz & K. S. Jones & B. L. Webber (Eds.), Readings in Natural Language Processing (pp. 71-88). Los Altos, CA: Morgan Kaufmann Publishers, Inc.

Show Me What You Mean – Expressive Media for Online Communities

Toru Takahashi **Yasuhiro Katagiri**
ATR Media Information Science Research Labs.
2-2-2 Hikaridai, Keihanna Science City,
Kyoto, 619-0288 JAPAN
{toru, katagiri}@atr.co.jp

Christoph Bartneck
Technical University of Eindhoven
Faculty of Industrial Design
Den Dolech 2, 5600 MB Eindhoven
The Netherlands
christoph@bartneck.de

ABSTRACT
In this paper, we describe the TelMeA2002 asynchronous online community, which uses embodied characters as expressive media to communicate messages. The functionality of the system and the challenges faced in designing it are discussed. Furthermore, we present the results of its first user evaluation.

Keywords
Online communities, Personified Media, Embodied Characters

INTRODUCTION
In face-to-face conversations, people use all of their natural modalities, such as speech, body language (gesture, pose, etc.) and facial expressions (gaze, emotion, nodding, etc.) to communicate with each other. Every conversation takes place in a shared context that may include the presence of other people and objects. The conversation is supported by the embodiment of all of its participants. This embodiment is still directly supported in videoconferences, but in Internet chat systems, only indirect representations of each participant, so called avatars [2], are available. These avatars help the participants to build a shared conversation context in a virtual chat environment. However, when people have a conversation with others through online communities, such as newsgroup and bulletin board systems, people are restricted to using textual information. Misinterpretations of these textual messages are common. The widespread application of emoticons [:-)] demonstrates that pure textual information lacks human embodiment and their communication modalities. Furthermore, it misses the conversational context, which might be compensated by including multimedia content in the messages. Although people are currently able to implicitly share context information by including links to web pages in their messages, they cannot include the web content explicitly inside the message itself.

One can distinguish two types of communications in online communities: asynchronous communication and synchronous communication. In the latter, the participants are present throughout the communication and react in real time to messages. Videoconference systems and online chats are good examples of synchronous communication. In asynchronous communication, the participants are not present during the communication and several days may pass before a reaction to a message is posted. Newsgroups and bulletin board systems are instances of asynchronous communication. This study focuses on asynchronous communication because we regard it is still having considerable untapped potential for human-based new information society designs and the analysis of social conversation.

Asynchronous communication poses stronger restrictions on communication modalities and awareness than synchronous communication. One way to overcome the restrictions is to employ avatar like embodied characters and let the characters express all non-textual information. The Media Equation [7] suggests that user will treat such characters as social actors and hence communicate with them as they would with other humans [12, 9]. The anthropomorphic appearance of the characters also helps the users to identify other participants and hence makes it easier to follow a discussion. Furthermore, the characters can help the users to understand the context of the conversation, including the involved personalities and their social relationships toward each other [10].

Based on this theoretical framework, we employed anthropomorphic characters in a prototypical asynchronous online community system called TelMeA2002. These characters play the role of personal representations (avatars) and present the conversational contribution of its representative user. We call such characters *personified media*. The TelMeA2002 system enables us to investigate the effect that expressive personified media have on the user's conversational behaviors. By analyzing logs of long-term online community activities, we hope to be able to find rules of social conversation from the viewpoint of usage and effects of expression with personified media. Such rules would be particularly helpful for autonomous character agents and enable them to act naturally and hence support them fulfilling their purpose, such as stimulating discussion between unacquainted users [4].

In this paper, we describe TelMeA2002's functionality and its design challenges as well as report results of a usability test that preceded the upcoming long-term case study.

FUNCTIONALITIES OF TelMeA2002

The basic functionality of TelMeA2002 is based on a bulletin board system. Users can post their messages in an online community, and these messages become available to all other users.

The main improvement is the use of personified media to present messages. Figure 1 shows the conversation process in TelMeA2002. In the beginning, the users can choose to create a new topic or they can reply to an existing message.

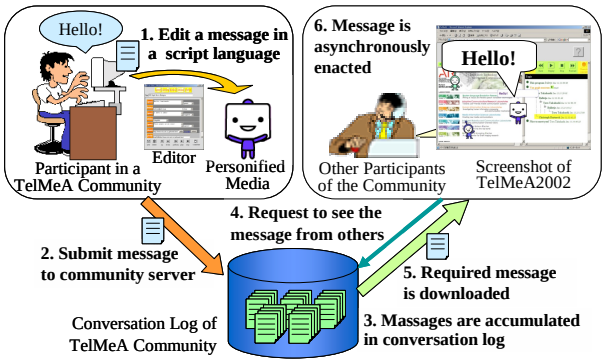


Figure 1. Conversation process in TelMeA2002.

In both cases, the TelMeA2002 Editor opens to allow the users to design their messages (Figure 2). The Editor provides five types of behaviors for the user's personified medium; speech, affective expression, interpersonal attitude, document reference, and comments on document. The behavior components can be used multiple times and arranged in any sequence. This enables the users to create even complex messages

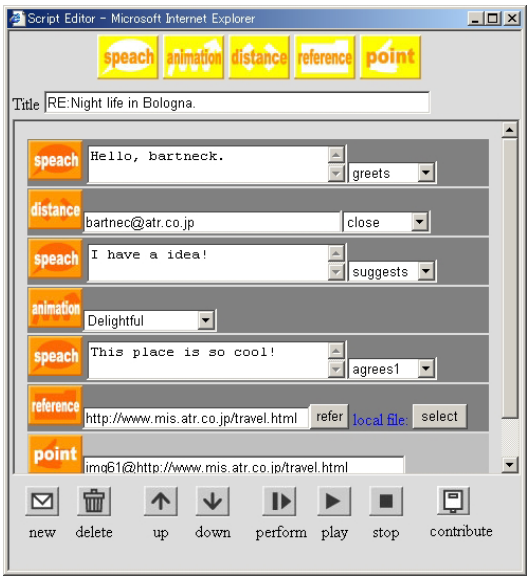


Figure 2. TelMeA2002 Editor

The user may, for example, type in the text to be spoken ("Hello, bartneck") and choose a performative verb that describes the intention of the utterance ("greets"). To take advantage of the full potential of non-verbal cues, we defined a fine-grained set of 35 performative verbs. The verb "agree," for example, is further sub-instantiated to represent the entire range of agreeing from smiling to nodding and thumbing up. Next, the user may want to direct the attention of the audience to a certain website. Therefore, the user selects the web component and enters a URL. This web page will become the new background of the TelMeA2002 stage. Afterwards, the user enters text again ("This is my favorite website") and chooses an affective expression for the personified medium ("happiness"). The user could have chosen from a range of 48 other affective expressions or 13 types of interpersonal attitudes. Finally, the user submits the message. The user's message is transformed into a script language format [3] and sent to the community server, where it becomes available to all other users. Another user may select this message, for which only this script is transferred and executed on the client's computer. The posted messages are archived in a conversation log that presents the basis for further analyses of the communication, such as summarization of social conversations.

CHALLENGES

While building the TelMeA2002 prototype, we encountered several problems that challenged the expressiveness and believability of the online community and its personified media.

Personalization

Each participant in the TelMeA2002 community is represented by a character, i.e., his/her personified medium. To quickly identify the various participants in a conversation, a unique embodiment for each personified medium is necessary. Therefore, a variety of eight personified media is available in the TelMeA2002 system, where these media vary in shape and color. We would like to expand this set to allow further personalization, but the high number of expressions required by each personified medium puts a heavy load on the development resources. Consequently, we first focus on an analysis of the usage of the various personified media before extending the grade of personalization.

Communication features

The expressions of the personified media should cover all four features of human communication: facts, relationship, appeal and self-revelation [8]. The facts feature contains the content of the message. and the relationship feature contains the sender's opinion of the receiver and the sender-receiver relationship. The appeal feature contains the information on what the sender wants the receiver to do (intention), and the self-revelation feature contains information on the state of the sender, in particular his or her emotional state. The

Figure 3. Examples of the expressivity of the TelMeA2002 characters. From left to right: greeting, happiness and complaint.

relationship, appeal and self-revelation features are not communicated through *what* is said but through *how* it is said.

The sender encodes all four features into his or her message and the receiver interprets the four features of the perceived message. Successful communication requires that the sent features of a message be similar to the interpreted features. A mismatch between the sent and interpreted features of the message can explain many failures of communication.

The TelMeA2002 system enables the user to communicate facts through the spoken content of the messages. The relationship of the users toward each other can be expressed through the relative spatial distance and position of their personified media. The user might, for example, stand right next to a befriended user. The spatial distance might be a good indication of social distance. The appeal feature might be expressed through the various performative verbs, such as *asks*, *agrees* and *declares*. The self-revelation feature is communicated through the emotional expressions of the personified media, such as happiness, sadness and anger.

Expressive Repertoire

Humans have a wide repertoire of conversational and emotional expression, ranging from subtle frowns to ecstatic dances of joy. Personified media need to cover the entire scale of expression to become believable entities. Unfortunately, many of the current implementations of personified media exaggerate their emotional expressions or do not have enough variations in their expressions and are therefore perceived as comic characters. The TelMeA2002 system employs 35 performative verbs (explains, agrees, complains, etc.), 48 affective expressions (likes, sadly, worries, etc.), and 13 interpersonal attitudes (yes, I know, forgotten, etc.). This variety should enable to the users to find a suitable expression for almost any situation. Figure 3 shows some examples.

In addition, the TelMeA2002 system has certain conversational expressions to direct attention, such as pointing to objects and the distances of personified media from each other and objects. The relatively higher importance of the conversational expressions, such as glance and nods, over emotional expressions [1] might not be the same in the TelMeA2002 system since turn taking is regulated automatically.

Communication modalities

Personified media should express their emotions consistently through all modalities available to them to ensure high believability [6]. It would not be convincing if the personified medium showed a sad face but talked with a neutral voice. Systematical manipulation of emotions in speech remains difficult, and unfortunately the speech synthesizer used for TelMeA2002 is not able to perform this task. Therefore, we are planning to make some rough manual adjustments in the pitch and speed of the synthesized speech to acquire a minimum level of consistency.

Logging and analysis

The TelMeA2002 community enables its participants to use rich personified media for their messages. All messages are encoded in an XML-based script language [11] and stored in log files on the TelMea2002 server. The highly structured nature of this scripting language is optimized comprehensive analyses of the messages, including their content, performative verbs, affective words and animations. Ideally, such analysis will enable us to gain a better understanding of social communication.

EVALUATION

A qualitative usability test of the TelMeA2002 system was performed at the Technical University of Eindhoven, Holland, in December 2002. The goal of the test was to identify major usability problems and suggest design solutions. Five participants were given the representative task to show their favorite website to another user. The other user was the second experimenter, located in Kyoto, Japan. The experimenter played the conversation partner of the participant. He took a passive role and thus only reacted to the messages of the participant. A videoconference link connected the experiment room in Eindhoven with the experimenter in Kyoto and enabled him to observe the progress from a distance and gain insight into the activities of the user. Only when he observed that the participant appeared to be stuck would he take the initiative and send a new message. Two cameras filmed the participants and their screen activities. The participants used the "Thinking-out-loud" method [3] to allow the experimenter to gain insight into their goals and activities. The experimenter in Eindhoven also observed the participants and made notes during the experiment. Afterwards he reviewed the videotapes to cross check the initial notes. Several usability problems could be identified and classified into general graphical user interface (GUI) problems, technical problems, and communication problems.

The GUI problems included problems with missing or unclear labels, wrong visualization of buttons, and redundant interface elements. Most of these problems were easily resolved by a redesign of the respective elements. The technical problems of the system consisted of excessively long response times, instability of the servers,

and scripting problems in the client software. If the user, for example, wanted to compose a new message he or she would click on the compose button, which would bring up a composer window. The loading of all elements of this window took several seconds, and the window was only operational if the loading was complete. Many participants clicked on elements before the completion of the loading process and hence caused a scripting error that in some cases disabled the entire interface. As the result, the participant would have to go back to the login screen and start over. An ongoing effort is being put into the technical improvement of the system, and we hope to have solved most problems before the upcoming case study.

The most interesting but also most difficult to solve problems are the communication problems. Several participants had problems understanding that TelMeA2002 is an asynchronous communication system (bulletin board) and not a synchronous system (chat). They tried to use the system as they would use chat systems, which resulted in several process problems. The replies to their messages, for example, appeared too late. We believe that the participants might have been misled by the constant presence of the personified media. Since the personified medium of the other user was visible all the time, the participants assumed that the other user himself or herself was online all the time and hence that they could chat with the other user. The constant presence also had the effect of making the participant believed that they could literary show a certain webpage to the other user by showing it to his or her personified medium. They put the other user's personified medium on top of a page and scrolled it up and down to show it to the other user. The participants assumed that the interface would be a shared space and that the other user could see exactly what they themselves saw on the screen. Another problem was the expectations of the participants toward the conversational abilities of the personified media. Due to the anthropomorphic gestalt of the personified media and their ability to synthesize speech, they expected the personified media to also be able to recognize speech. The participants started to talk back to the personified media after they finished their utterances.

To make users tacitly understand the asynchronous nature of TelMeA2002, we are planning to change the interface concept to the metaphor of a theatre. The various personified media would only be visible on the stage and perform their acts on it. This metaphor appears to better suit the concept of TelMeA2002. A second usability test will be necessary to confirm this assumption.

CONCLUSIONS

We presented a first prototype of an asynchronous online community that enables its participants to communicate by using personified media in the form of embodied screen characters. Participants can enrich their messages with a wide range of animations and expressions. A first usability test was performed and resulted in several redesign suggestions. Currently, we are working on the implementation of these suggestions to prepare the system for a long-term case study. Some challenges remain, such as the consistent role of the characters. Characters are commonly used for assistants, such as the Microsoft Office Assistants [5], and as avatars to represent the user in a virtual environment. Personified media in TelMeA2002 are avatars of their respective human users. In addition, the personified medium of each individual user also fulfill the role of an assistant. The media help the user with problems and gives suggestions on how to overcome them. This dual role caused some confusion in the usability test, and we intend to make the special role of the users' own personified media clearer through their continuous presence. All other characters should only be present on the TelMeA2002 stage. This is a major conceptual change in the system, and a second usability test will be necessary to assess its success.

ACKNOWLEDGMENTS

We thank to Tsutomu Kanegae and Hideaki Fujii for their technical support and Keiko Nakao for her design of TelMeA2002. This research was partly supported by the Telecommunications Advancement Organization of Japan.

REFERENCES

1. Cassell, J. & Thorisson, K. R., The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. Journal of Applied Artificial Intelligence, 13(3), 519-538, 1999.
2. Damer, B., Avatars: Exploring and Building Virtual Worlds on the Internet. Berkeley: Peachpit Press, 1997.
3. Dumas, J. S. & Redish, J. C., A Practical Guide to Usability Testing. Portland: Intellect Books, 1993.
4. Isbister, K., Nakanishi, H., Ishida, T., & Nass, C., Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space. Paper presented at CHI2000, The Hague, 2000
5. Microsoft Corporation, *Microsoft Agent Documentation*, 1998. Available at http://msdn.microsoft.com/workshop/imedia/agent/alldocs.zip
6. Nass, C. & Gong, L., Is Maximization or Consistency the More Social? The Case of Synthesized Voices and Faces. Paper presented at the CHI 2000, The Hague, 2000.
7. Nass, C. & Reeves, B., The Media Equation. Cambridge: SLI Publications, Cambridge University Press, 1996.
8. Schulz, F. v. T., Miteinander Reden - Stoerungen und Klaerungen. Reinbeck bei Hamburg: Rowolth Taschenbuch Verlag GmbH, 1981.
9. Takahashi, T., Takeuchi, Y., & Katagiri, Y. Change in Human Behaviors Based on Affiliation Needs – Toward the Design of Social Guide Agent System –, Proc. of KES2000, Vol. 1, pp. 64-67, 2000.
10. Takahashi, T. and Takeda, H., TelMeA: An Asynchronous Community System with Avatar-like Agents, Proc. INTERACT2001, pp. 480-487, 2001.
11. Takahashi, T. and Takeda, H., Proposal of a Script Language for Embodied Conversational Agents as Asynchronous Conversational Media, Proc. AAMAS2002, pp. 1387-1388, 2002.
12. Takeuchi, Y. & Katagiri, Y., Social Character Design for Animated Agents, In Proc. RO-MAN'99, pp. 53-58, 1999.

Effect of CG Synthesized Facial Expressions on Social Attitude

Yugo Takeuchi
Faculty of Information
Shizuoka University
3-5-1 Johoku, Hamamatsu,
Shizuoka 4328011 JAPAN
+81 53 478 1455
takeuchi@cs.inf.shizuoka.ac.jp

Takuro Hada
Faculty of Information
Shizuoka University
3-5-1 Johoku, Hamamatsu,
Shizuoka 4328011 JAPAN
+81 53 478 1451
cs8070@cs.inf.shizuoka.ac.jp

Yasuhiro Katagiri
ATR Media Information
Science Laboratories
2-2-2 Hikaridai, Soraku,
Kyoto 6190288 JAPAN
+81 774 95 1480
katagiri@atr.co.jp

ABSTRACT
In this study, we examine the effect of synthesized emotional facial expressions on human social attitudes. Our experimental result indicate that when the synthesized face shows a positive expression immediately after a response from a subject, the subject tends to give an answer that would lead to a positive expression by the synthesized face in the subsequent dialogue. On the other hand, if the synthesized face reacts negatively to the subject's response, the subject unconsciously tries to avoid further negative responses by the synthesized face. These observations suggest that subjects infer the mental state of the synthesized face from its expressions and try to maintain a friendly relationship with it.

Keywords
CG synthesized face, Facial expression, Emotion, Social interaction

INTRODUCTION
A human's voice and facial expression naturally express his or her real intention. Therefore, people often deliberately control their voice and facial expression to conceal their real intention [1,2]. This intentional behavior can sometimes be useful in avoiding social conflict with others. For example, common social etiquette requires that people express a sorrowful attitude toward the bereaved. People are expected to talk mournfully with others in an undertone and make their facial expression sorrowfully with downcast eyes. In this way, people who consistently display voice and facial expression grounded on the same emotional attitude can make other persons believe that they genuinely feel sorrow even if they do not. On the other hand, it would cause serious social discord if people talked cheerfully or expressed a delighted facial expression in the situation of someone's death. In general, when a person expresses his or her voice or facial expression in a way that deviates from accepted social manners, people assume that the deviated expression reveals his or her real intention in any situation.

In this study, we examined the effect of simultaneously presented synthesized emotional voice and facial expressions as reflected in human social attitudes. Previous studies have examined whether people perceive an impression from either synthesized emotional voice or facial expressions when these are tested separately [3]. No study, however, has simultaneously examined these two synthesized emotional expressions in a meaningful attempt to investigate how people respond to them. Furthermore, our experiment was set up to observe the process of attitude transition while a subjects socially interact with synthesized entity.

ATTITUDE TOWARD SOCIAL MORALS
Questions concerning social morals were given to subjects as listed in **Table 1**. In the case of an official employment interview, people would generally answer "yes" in accordance with social desirability. This is probably because interviewee expects that being accepted by the interviewer is an expedient way to expand one's opportunities.

Table 1 Sample questions on social morals.

Social Situation	Question
When you buy something at a shop, the salesperson gives you too much change by mistake.	Would you repay the extra change to the salesperson?
When you walk on an empty street, you happen to meet with an aged person carrying heavy baggage with an unsteady walk.	Would you help him or her carry the baggage?

People, however, do not always decide their answer based on the content of question but often attempt to reach a consensus on the question with the interviewer. Therefore, people sometimes give a different answer from his or her actual attitude toward the question concerning social morals.

EXPERIMENT

Materials

In order to synthesize CG facial expression, we used the CSLU Toolkit [4], which incorporates emotional facial animation technology and speech synthesis technology called CHATR [5]. We applied these voice and facial expression synthesis systems to infuse our CG human-like entity (**Figure 1**) with an identity so that it could emotionally speak with people through its voice and facial expression.

Figure 1 Synthesized CG facial expressions (Sad, Neutral, and Happy expressions from left to right).

Procedure

In our experiment, subjects were instructed that they would have to take an aptitude test to be employed as subjects of this experiment. Then they were interviewed for a job with our CG human-like entity as the interviewer. Eight social situations were described to them, and they were asked a question on each that inquired about their sense of social morals, as seen in the examples of **Table 1**.

Table 2 Facial expression feedback in each experimental condition.

Experimental Condition	Facial Expression Feedback	
	"YES"	"NO"
NN	Neutral	Neutral
HH	Happy	Happy
HS	Happy	Sad
SH	Sad	Happy
SS	Sad	Sad

Experimental Conditions

The interviewer expressed three types of facial expressions when the subject answered "YES" or "NO" to a question (**Table 2**). In the NN experimental condition, the interviewer expressed a neutral expression for either answer. In the HH (SS) experimental condition, the interviewer expressed a positive (negative) expression with a happy (sad) face for either answer. In the HS experimental condition, the interviewer expressed a positive expression with a happy face when subjects answered "YES" and a negative expression with a sad face when they answered "NO"; conversely, in SH, the interviewer expressed a negative expression with a happy face to "NO". After the subject answered the question, the interviewer said "I see" and expressed either a positive face with a high-pitched voice or a negative face with a low-pitched voice.

Results

The results of the experiment are shown in **Figure 2**, which indicates the ratios of four transition types of the subjects' answer in each experimental condition. For example, when a subject answers "YES" at the N^{th} (N>1) question and answers "NO" at the $(N+1)^{th}$ question, this is illustrated as <YES to NO>.

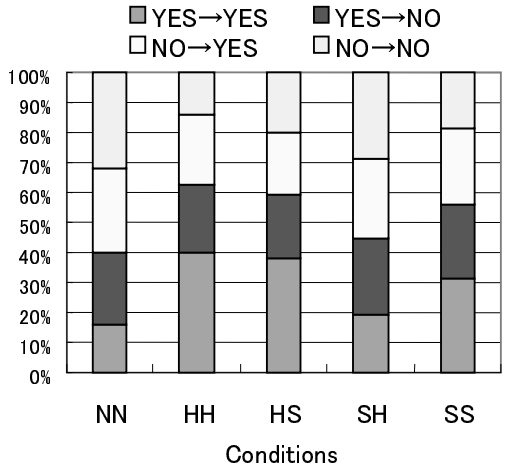


Figure 2 Ratios of four transition types for subjects' answers in each experimental condition.

CONCLUSIONS

The large ratios of <YES to YES> transition types of subject answers in the HH and HS experimental conditions indicate that subjects tend to respond positively when an interviewer's CG synthesized face expresses a positive expression when they expect a positive feedback. Moreover, the large ratio of <NO to NO> transition types of subjects' answer in the SH experimental condition indicate that subjects tend to respond negatively when an interviewer's CG synthesized face expresses a negative expression when they expect a negative feedback.

These experimental results suggest that people tend to prefer receiving a positive expression in any case when they interact with a synthesized entity. This is valuable evidence that people perceive human-like intentions from the emotional expressions of the synthesized entity.

REFERENCES

1. Russell, J. A. and Fernandez-Dols, J. M. *The psychology of facial expression*, Cambridge Univ. Press, 1997.
2. Ekman, P. *Telling lies*, Yale University Press, 1985.
3. Nagao, K. and Takeuchi, A. Speech dialogue with facial displays: Multimodal human-computer conversation, *Proceedings of ACL-94*, 102-109.
4. The CSLU Toolkit: A Platform for Research and Development of Spoken-Language Systems. Available at http://cslu.cse.ogi.edu/toolkit.
5. Black, A. and Taylor, P. CHATR: A Generic Speech Synthesis System, *Proceedings of COLING-94*, 983-986

User Responses to Emotion in Embodied Agents

Scott Brave
Dept. of Communication
Stanford University
Stanford, CA 94309 USA
+1 650 723 5499
brave@stanford.edu

ABSTRACT
In this paper, we investigate the effect of agent emotion on users' perceptions of an embodied agent. A between-subjects experiment (N = 88) evaluates user responses to four types of embodied agent: an agent exhibiting no emotion, an agent exhibiting self-concerned emotion only, an agent exhibiting user-concerned emotion only, and an agent exhibiting both self-concerned and user-concerned emotion. The results show users' opinions of the agents are significantly affected by the type of emotion exhibited. Implications for the design of interface agents are discussed.

Keywords
Emotion, agents, characters, social interfaces

INTRODUCTION
The topic of emotion has gained increasing attention in the HCI community. Pioneers such as Bates [1] and Picard [12] have repeatedly emphasized the importance of emotion for believability, social intelligence, effective communication, and natural interaction. Agents that exhibit human-like emotion have now become commonplace, both on and off screen. Virtual characters found in the Oz Project [2], Comic Chat [11], the Finali netSage™ [8], and the Virtual Theatre [9], for example, as well as the humanoid robot Kismet [4], all include emotion as a fundamental component of their interaction with users.

Although much progress has been made in creating agents that display believable emotions, little is known about how emotion in agents affects users. This paper presents an experimental study aimed at understanding the effect that emotion has on a user's perceptions of and opinions about an embodied agent.

Orientation of Emotion
The psychological literature generally views emotion as a reaction to events deemed relevant to the needs or goals of an individual [3]. Happiness, for example, results when an individual's goals are met and sadness when they are not met. In an HCI context, there are two broad categories of goals which an agent might be oriented toward: 1) The agent's own goals and 2) the user's goals. Emotion in agents can thus be considered along two dimensions: *agent-concern* and *user-concern*.

Agent-concern is present when an agent responds emotionally to an event deemed relevant to its own internal needs and goals. Endowing agents with such self-concern lends an air of lifelikeness and believability to their behavior. *User-concern* is present when an agent responds emotionally to the needs and goals of the *user*: Becoming happy, for example, when the *user's* goals are met and sad when they're not. Such sympathetic emotional responses require the agent to possess either a model of the user's needs and goals or a method of detecting the user's emotional state (often considered essential for effective social interaction[5][7][10]).

Although normal humans have both self-concern and concern for others, in the case of agents, the two dimensions are orthogonal, allowing us to construct four categories of agent (see Table 1).

METHOD
In order to investigate the psychological effects of emotion in agents upon users, we conducted an experiment for which each participant observed an embodied agent exhibiting one of the four emotional dispositions described in Table 1. We then asked the users' opinions of the agent and observed the level of altruism users displayed toward the agent.

Participants
Participants consisted of 88 adults, randomly assigned to condition, with an equal number of men and women in each condition.

Table 1. Agent Emotional Dispositions

		User Concern	
		Absent	Present
Agent Concern	Absent	*Neither-emotional*	*User-only-emotional*
	Present	*Agent-only-emotional*	*Both-emotional*

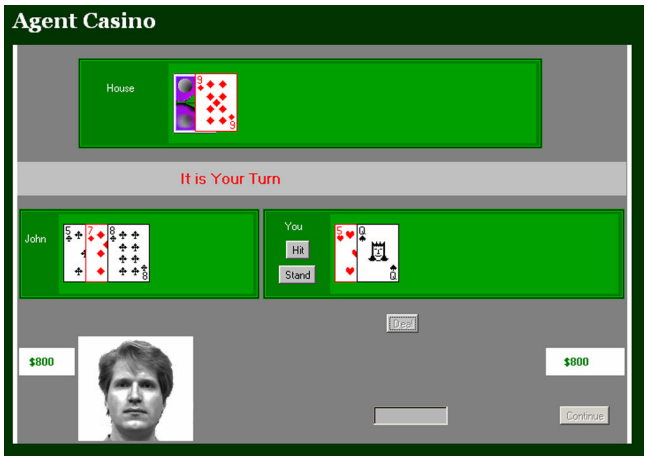


Figure 1. Blackjack Screenshot.

Procedure

The experiment was a 2 (self-concern: present vs. absent) by 2 (user-concern: present vs. absent) by 2 (gender: male vs. female), between-participants design. Participants were directed to a website that gave instructions and ran the experiment. Participants were then paired up with an agent (always the same face), exhibiting one of the four emotional dispositions described in Table 1 (selected at random), for ten hands of blackjack, lasting approximately five minutes.

The participant and agent "sat" next to each other at the blackjack table, both playing against a disembodied dealer (see Figure 1). Hands were dealt at random (i.e., not fixed) and the agent's "hit or stand" decisions were made using a simple rule-based algorithm, with some randomness for believability. The agent played each round first, followed by the participant, followed by the dealer. The announcement of winning or losing hands was given through text messages at the center of the game playing area.

After each round was complete, the agent always reacted with an observation about his performance, followed by an observation of the participant's performance. In conditions where *agent-concern* was present, the agent's observation of his own performance included an emotional reaction (both in text and in facial expression; see Figure 2a). The agent would express positive emotion if he won and negative emotion if he lost. In conditions where *agent-concern* was absent, the agent simply made an unemotional observation of his performance (Figure 2b).

Likewise, in conditions where *user-concern* was present, the agent's observation of the *participant's* performance included an emotional reaction; otherwise, it did not.

At the conclusion of the blackjack game, participants were asked to complete an online questionnaire (measuring their opinions of the agent). Participants were then told that the experiment was officially complete, but were given the opportunity to help the agent by training him to play blackjack better. Random blackjack hands were presented one at a time and the participant was asked to advise the agent what he should do (hit or stand). The number of hands completed served as a measure of altruistic behavior.

RESULTS

Likeability

Likeability (Fig. 3) was comprised of three 10-point semantic differential items: likable-unlikable, pleasant-unpleasant, and comforting-disturbing (Cronbach's alpha = .74).

There was a significant 3-way interaction (user-concern x agent-concern x gender) with respect to likeability of the agent, $F(1, 88)=12.0$, $p<.001$. This effect was due primarily to women's strong preference for an emotionally symmetric agent (either both-emotional or neither-emotional), $F(1,44)=21.1$, $p< .001$. Men showed no such preference for symmetry and, in fact, seemed to like the asymmetric, user-only-emotional agent the most.

Competence

Competence (Fig. 4) was comprised of four items: competent-incompetent, intelligent-unintelligent, reliable-unreliable, and careful-careless (alpha = .46). Although the alpha was low, factor analysis maintained that the items comprised a single index (all eigenvalues > 1.0).

There was a significant 2-way interaction (user-concern x gender) with respect to competence, $F(1, 88)=7.0$, $p<.01$. For women, the presence of user-concern significantly decreased the agent's perceived competence, $F(1, 44)=7.2$, $p <.01$. Men, on the other hand, showed a non-significant tendency in the opposite direction, such that user-concern increased perceived competence.

Trustworthiness

Trustworthiness (Fig. 5) was comprised of three items: trustworthy-untrustworthy, honest-dishonest, and sincere-insincere (alpha = .71).

There was a significant cross-over interaction (user-concern x agent-concern) with respect to trustworthiness, $F(1, 88)=5.5$, $p<.02$. Post hoc tests (LSD) showed that this effect was due to the general opinion that the neither-emotional agent was the most trustworthy of the agents.

a)

b)

Figure 2. a) Examples of present *agent-concern* b) Examples of absent *agent-concern*.

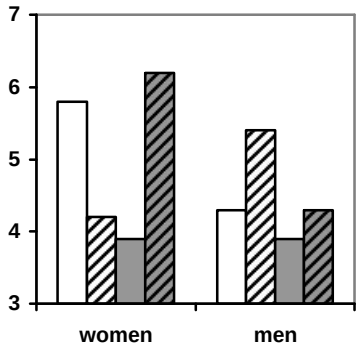


Figure 3. Likeability.

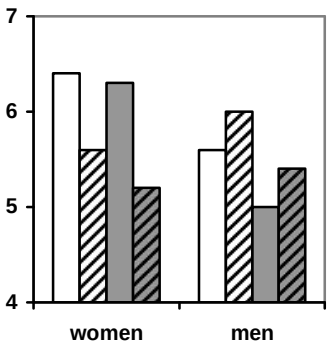


Figure 4. Competence.

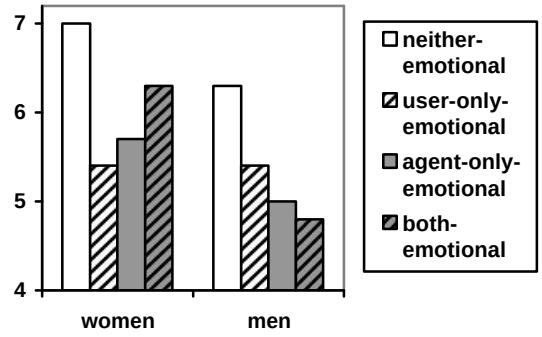


Figure 5. Trustworthiness.

(Note: presence of user-concern indicated by stripes; presence of agent-concern indicated by darker color)

Likelihood of Use

We also asked participants a series of questions rating task appropriateness; we asked how likely they would be to use the agent for each of nine tasks (e.g., web searching, online banking). Responses were given on a 10-point scale anchored by "very unlikely" at one end and "very likely" at the other. Based on theory and confirmed by factor analysis, the nine tasks were divided into three groups based on perceived risk to the user. These indices were reliable.

Low Risk was comprised of three items: game playing, web searching, and searching for MP3s (alpha = .74).

Medium Risk was comprised of three items: personal assistant (for scheduling appointments, etc.), online auctions (like eBay), and shopping for CDs online (alpha=.78).

High Risk was comprised of three items: online banking, travel reservations, and trading stocks online (alpha=.78)

Likelihood of using the agent for *high ris*k applications (Fig 6) showed effects similar to those seen for trustworthiness. There was a significant 2-way interaction (user-concern x agent-concern), $F(1, 88)=6.7$, $p<.01$. Post hoc tests again showed that the effect was due to an overall higher rating for the neither-emotional agent.

With respect to likelihood of using the agent for *medium risk* applications (Fig. 7), there was a significant 2-way interaction (user-concern x gender), $F(1, 88)=6.7$, $p<.01$. In this way the effect was similar to that found for competence. Here, however, the effect was due mainly to the men. For men, the presence of user-concern increased the reported likelihood of using an agent for medium risk applications, $F(1,44)=6.9$, $p<.01$. Women showed a non-significant tendency in the opposite direction, with the presence of user-concern decreasing the likelihood of use.

There was also a main effect for agent-concern, such that the presence of agent-concern decreased likelihood of agent use for medium-risk applications, $F(1, 88)=8.6$, $p<.005$.

There were no significant differences among agents with respect to likelihood of use in *low risk* applications.

Altruistic Behavior

Altruistic behavior (Fig. 8) was measured as the total number of training hands played by the participant, at the close of the experiment.

The behavioral altruism measure showed a significant 3-way interaction, $F(1, 88)=4.2$, $p<.05$. This result was due primarily to a significant 2-way interaction (user-concern x agent-concern) for men, $F(1, 44)=4.7$, $p<.05$. Men helped agents exhibiting asymmetric emotion more than agents exhibiting symmetric emotion.

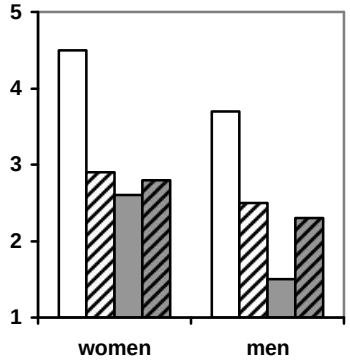


Figure 6. High Risk.

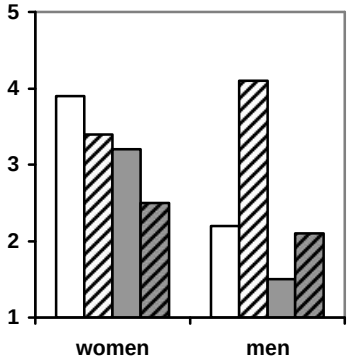


Figure 7. Medium Risk.

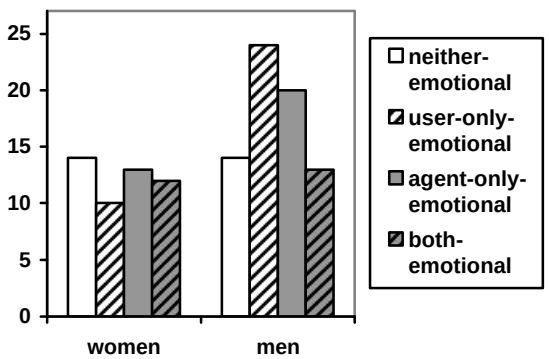


Figure 8. Altruistic Behavior.

DISCUSSION

The results clearly indicate that endowing an agent with emotion has psychological effects upon the user. More importantly, *who* the agent has emotion about (itself vs. the user) makes a significant difference. Finally, a user's gender plays an important role in mediating psychological reactions to embodied agent emotion.

Women liked agents that exhibited emotional symmetry. Women rated the symmetric neither-emotional and both-emotional agents as more likable than the asymmetric agent-only-emotional and user-only-emotional agents (Fig. 3). Men, in contrast, showed no such preference and if anything seemed to like the user-only-emotional agent best. This result is consistent with popular notions of gender differences in relationships: Women are often viewed as preferring more reciprocal relationships while men seen as preferring hierarchical relationships (particularly if they're in charge) [6].

A second pattern shows itself in the competence and medium risk measures (Fig 4 and Fig. 7). In both cases, the presence of user-concern lead to a more positive rating of the agent with men, but a less positive rating with women. Although further studies are needed to better understand this result, the effect could be due to differing standards held by men and women regarding empathetic behavior. While men may have seen the mere ability to empathize as an indication of competence, women may have considered the agent's empathetic behavior to have been overly simplistic and even juvenile, lowering perceived competence. Competence may then have been the main deciding factor in a user's willingness to use the agent for medium risk applications, explaining the similar effect in that measure.

Perhaps the most interesting result is that, regardless of other opinions, both men and women *trusted* the unemotional agent most. This effect showed itself in both the trustworthiness (Fig. 5) and high-risk (Fig. 6) measures, with the neither-emotional agent standing apart with significantly higher ratings in both cases. One possible explanation for this is that the presence of emotion suggests a more complex agent "psychology" and therefore less predictability, bringing issues of trust to the forefront.

CONCLUSION

Endowing agents with emotion has complex psychological implications. As the results in this paper have shown, user responses to an agent can be strongly influenced both by the presence or absence of emotion, as well as the orientation (agent vs. user) of that emotion. In addition, gender stands out as a strong mediator of these effects. More work is clearly necessary to fully understand user's psychological responses to emotion in agents and ultimately devise a set of design principles. However, at this stage, it seems clear that creating believable *empathetic* (i.e., user-concerned) emotion should occupy at least as high of a research priority as creating believable agent-concerned emotion. Further, designers of virtual characters are advised to carefully consider application characteristics (e.g. high vs. low risk) before deciding the emotional behavior of their agent. In this way, the interface designer occupies a crucial place in creating an emotionally intelligent agent, for intelligent *design* must dictate the emotional disposition appropriate to context.

REFERENCES

1. Bates, J. The role of emotions in believable agents. *Communications of the ACM, 37* (7). 122-125.
2. Bates, J., Loyall, A.B. and Reilly, W.S. An architecture for action, emotion, and social behavior. *in Artificial social systems: Fourth European workshop on modeling autonomous agents in a multi-agent world*, Springer-Verlag, Berlin, 1994.
3. Brave, S. and Nass, C. Emotion in human-computer interaction. in Jacko, J.A. and Sears, A. eds. *Handbook of human-computer interaction*, LEA Press, New York, 2002.
4. Breazeal, C. *Designing Sociable Robots*. The MIT Press, Cambridge, MA, 2002.
5. Canamero, L., Building Emotional Artifacts in Social Worlds: Challenges and Perspectives. *in Proceedings of Emotional and Intelligent II: The Tangled Knot of Social Cognition*, (North Falmouth, MA, 2001), AAAI Press, 22-30.
6. Coats, E.K. and Feldman, R.S. Gender differences in nonverbal correlates of social status. *Personality and Social Psychology Bulletin, 22* (10). 1014-1022.
7. Elliott, C. and Ortony, A., Point of view: Modeling the Emotions of Others. *in Proceedings of the 14th Annual Conference of the Cognitive Science Society*, (1992), Lawrence Erlbaum, 809-814.
8. Finali. netSage™, www.finali.com.
9. Hayes-Roth, B. and van Gent, R., Story-Making with Improvisational Puppets. *in Proc. 1st Int. Conf. on Autonomous Agents, (Marina del Rey, CA, 1997)*, 1-7.
10. Klein, J., Moon, Y. and Picard, R.W., This computer responds to user frustration. *CHI'99 extended abstracts*, (New York, 1999), ACM Press, 242-243.
11. Kurlander, D., Skelly, T. and Salesin, D. Comic Chat. *in Proceedings of SIGGRAPH'96*, ACM Press, New York, 1996, 225-236.
12. Picard, R.W. *Affective Computing*. The MIT Press, Cambridge, MA, 1997.
13. Yale Face Database, http://cvc.yale.edu/projects/yalefaces/yalefaces.htm

Language-dependence in the signalling of attitude in speech

Aoju Chen **Carlos Gussenhoven**
Centre for Language Studies, University of Nijmegen
6500 HD Nijmegen, the Netherlands
aoju.chen@let.kun.nl c.gussenhoven@let.kun.nl

ABSTRACT
From cross-cultural studies on emotion and intonational meaning, there emerges universality in the signalling of emotion and attitude. However, findings from our cross-linguistic studies on the universality of intonational meaning suggest that listeners with different language backgrounds differ in the extent and manner in which they exhibit these universal tendencies. Accordingly, we suggest that there is a language-dependent component in the universal signalling of attitudes. To illustrate this point of view, we describe two perceptual studies in which we compared the perception by Dutch and British English listeners of the attitudes *friendly*, *confident*, *surprised* and *emphatic*, and we suggest explanations for the differences we found.

Keywords
Attitude, emotion, speech, F0, universality, language-dependence, perception, pitch span, pitch register

INTRODUCTION
Cross-linguistic studies on emotion recognition suggest that although strong universal tendencies in the signalling of emotion exist, there are nevertheless cross-linguistic differences [e.g. 4]. A recent investigation of emotion in speech across languages on a large scale is [10]. In order to establish whether the vocal changes produced by emotional and attitudinal factors are universal or language and culture specific, Scherer and his colleagues examined the recognition of five emotions, including anger, fear, joy, sadness and neutrality, in nine countries on three continents. The stimuli were generated from two nonsense utterances composed of phonological units from different Indo-European languages and read by professional German actors. Judges from the nine countries recognised the emotions with much better than chance accuracy, providing support for the claim of universality in the signalling of emotion in speech. However, there were also clear differences in the performance of the judges from different countries. For example, judges with a Germanic language background judged the emotions with greater accuracy than judges with a Romance language background, and in particular judges with a non-Indo-European language background. This suggests that although the emotions examined can be universally recognised, there are cross-linguistic differences in the way they are expressed. Because the stimuli were composed of phonological units from various Indo-European languages and presented out of context, Scherer suggests that the cross-linguistic differences may be either located at the segmental level (e.g. formant structure) or at the suprasegemental level (e.g. F0). On the basis of these findings, Scherer forcefully argues for the necessity of incorporating rules capturing cross-linguistic differences into multilingual speech modelling.

The two important and still unexplored issues here are (1) how exactly languages differ in the signalling of emotion; and (2) what the linguistic factors are that lead to these differences. Regarding the first issue, a question arises as to whether languages exhibit the universal tendencies observed across languages to the same degree. We therefore narrow down the first issue to (1)': how languages differ in the universal signalling of emotion. Because F0 has been shown to be a consistent factor in the signalling of emotion [7], this paper attempts to address issues (1)' and (2) in the light of findings from two cross-linguistic studies on the perception of speaker attitudes in speech stimuli varying in F0.

Our research concerns perceived attributes of the speaker, like 'surprised', 'confident', 'sad', which we distinguish from attributes of the linguistic message, like 'request for information', 'significance', 'turn-finality' [5]. We believe that this distinction between 'affective' and 'informational' meanings provides a workable framework for the study of intonational meaning, but are at the same time aware that our speaker attributes may not be considered either to be consistently 'emotions' or consistently 'attitudes' by other researchers. However, we believe we are justified in blurring this latter distinction. For one thing, the two notions are closely related. For another, research on emotion in speech regularly includes speaker states that others might consider attitudes, like 'dominance', 'shyness', 'surprise', as shown by the survey in [3]. In the remainder of this paper, the terms 'speaker attitude' and 'affective state of speaker' will be used interchangeably.

UNIVERSALITY IN THE SIGNALLING OF ATTITUDE

Universals in the signalling of attitude can be derived from a body of universal intonational meanings proposed by Gussenhoven (e.g. [5]). This theory of paralinguistic meaning in intonation holds that the universal meanings are based on biological conditions that are responsible for F0 variation. Three metaphors have been identified that describe these biological conditions and the corresponding interpretations in human vocal communications: the Frequency Code [8], the Effort Code and the Production Code. As the Production Code does not concern itself with speaker attributes, we will only consider the Frequency Code and the Effort Code in some detail here.

The Frequency Code (which might also have been termed the 'Size Code') is based on the fact that smaller larynxes produce higher notes than larger ones. In human vocal communication, the Frequency Code therefore associates high pitch with the primary meaning of 'small vocalizer' and secondary meanings like appealing, submissive, friendly, etc. and associates low pitch with the primary meaning of 'large vocalizer' and secondary meanings like aggressive, assertive, confident, etc. The relevant pitch variation can be implemented through varying prosodic parameters, such as peak height, end pitch and overall pitch level.

The Effort Code is based on the fact that greater articulatory effort tends to create more explicit phonetic realisations. In the context of F0, greater explicitness leads to wider pitch movements, less explicitness to narrow-range movements. As a result, a wider pitch span is associated with meanings that can be derived from speakers' motivations for the expenditure of articulatory effort. An affective interpretation is 'surprised', derived from the perception that the speaker shows agitation, thereby spending more effort on speech production. Although in Gussenhoven's original account, the interpretation 'emphatic' is regarded as an attribute of the linguistic message, we think that 'emphatic' can be interpreted as a speaker attribute too.

LANGUAGE-DEPENDENCE IN THE SIGNALLING OF ATTITUDE

In an attempt to address the question how a difference in the mean pitch range between two languages may influence the use of these universal codes, we conducted two cross-linguistic perception studies in which we examined whether Dutch and British English (BrE) differ in the use of the Frequency Code [1] and the Effort Code [2]. According to [9], the mean pitch range of Dutch is approximately 70Hz and that of BrE is approximately 100Hz. We use the term 'standard pitch range' to refer to the mean pitch range of a language. The general finding was that although both Dutch and BrE employ these codes, they differ systematically in the extent and manner in which they are used, suggesting that there is a language-specific component in implementing the universal meanings of intonation. We would therefore like to put forward the view that there is a language-specific component in the universal signalling of attitude. In the next two sections, we will consider findings from these studies that demonstrate how languages can differ in the universal signalling of attitude. Moreover, we will try to account for these differences, and suggest that speech technology needs to incorporate such findings in multilingual applications. Methodology will be dealt with only briefly; for more detail see [1] and [2].

Signalling of 'friendly' and 'confident'

[1] examined how Dutch and BrE listeners differed in the perception of 'friendly' and 'confident', two attitudinal meanings deriving from the Frequency Code. Equivalent perception experiments were carried out in which native speakers of Dutch and BrE judged stimuli in their native language, on the scales CONFIDENT vs. NOT CONFIDENT and FRIENDLY vs. NOT FRIENDLY. The magnitude estimation method was used to obtain the perceptual judgements [11].

Stimuli

As the perception of attitude may be a function of pitch contour as well as of the speech act, we designed 12 source utterances exemplifying three speech acts that are likely to be interpretable in terms of the attributes 'confident' and 'friendly': INFORMATION, REQUEST and INSTRUCTION. The speech acts were implemented by means of wh-interrogatives, yes-no interrogatives and declaratives. The stimuli, which were lexically equivalent across the two languages, were varied in pitch contour (H*L L% vs. L*H H%) and pitch range. Pitch range was varied in both pitch span and pitch register. Pitch span variation refers to variation between the lowest and the highest F0 in a stretch of speech; pitch register variation refers to variation in F0 averaged over the utterance (e.g. [6]). This gave us two subsets of the stimuli: the Pitch Span set and the Pitch Register set.

Predictions

We hypothesised that the most likely difference between Dutch and BrE would be due to a projection of different F0 scales onto the same semantic scale, a hypothesis we refer to as the Relative Scale Hypothesis. Under this view, listeners were expected to project their habitual pitch range onto the Frequency Code scale, such that a given F0 value would get a higher score on the Frequency Code scale in Dutch than in BrE. The higher score in the case of the Dutch listeners would be due to the fact that they project a smaller F0 scale onto the same semantic range. A higher score on the Frequency Code scale means a higher degree of 'small' meanings such as 'friendly', and a lower degree of 'big' meanings such as 'confident'.

Statistical Analyses[1]

Two sets of data were obtained for each semantic scale. Two ANOVA's were performed on the Pitch Span data and another two ANOVA's on the Pitch Register data, for each of the two dependent variables, the Friendliness score and the Confidence score. The analyses comprised one between-subject factor: Language (2 levels) and four within-subject factors: Pitch Contour (2 levels), Speech Act (3 levels) and Pitch Span (5 levels) or Pitch Register (5 levels). We adopted a significance level of .05.

[1] We thank Toni Rietveld for his contribution during the design stage of these experiments and for dealing with the statistical processing of the results.

Results

As for the attribute 'confident', the three-way interaction of Language × Speech Act × Pitch Register was found to be significant ($F_{8,496}$=3.70, p<0.05). By pooling the results of the three-way interaction over the speech acts, we found that, in both Dutch and BrE, the degree of perceived confidence decreased when the Pitch Register was increased. This is a clear manifestation of the Frequency Code. However, at identical pitch registers, Dutch listeners perceived a lower degree of confidence than BrE listeners, as shown in Figure 1. This difference was anticipated and can be accounted for as due to the smaller standard pitch range of Dutch. We referred to this type of range-induced language-dependence as a Type1 difference.

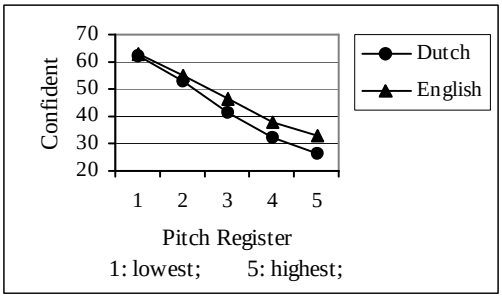


Figure 1. The interaction of Language × Speech Act × Pitch Register pooled over speech acts.

As for the attribute 'friendly', a significant interaction was found between Language and Pitch Register ($F_{4,248}$=3.90, p<0.05). In both languages, the higher the pitch register, the higher the degree of perceived friendliness, again a straightforward manifestation of the Frequency Code. Interestingly enough, at identical pitch registers, Dutch listeners perceived a lower degree of friendliness than BrE listeners, as shown in Figure 2. Although this might at first sight be regarded as a Type 1 difference, this finding cannot be accounted for by the smaller standard pitch range of Dutch, since a given F0 pattern would have been perceived as signalling a higher degree of friendliness in Dutch than in BrE. At this point, we may tentatively conjecture that pitch register variation is employed to express some other attitudinal meaning than friendliness in Dutch, and that the perception of this meaning undermines the perception of friendliness.

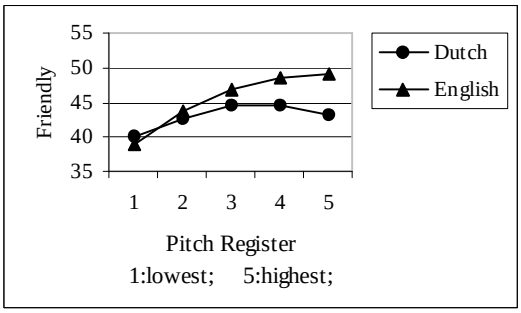


Figure 2. The interaction of Language × Pitch Register.

Signalling of 'surprised' and 'emphatic'

In [2], we examined the way Dutch and BrE differed in the signalling of 'surprised' and 'emphatic', two interpretations of the Effort Code. Because the Effort Code relies on the expression of pitch range differences, we expected Dutch and BrE to behave differently for these perceived attributes, due to Type 1 differences.

As in [1], native speakers from Dutch and BrE were asked to listen to a number of sentences in their native language and to judge these sentences on the semantic scales SUPRISED vs. NOT SURPISED and EMPHATIC vs. NOT EMPHATIC. The magnitude estimation method was used to obtain the perceptual judgements.

Stimuli

The stimuli were generated from source utterances representing two sentence modes, interrogatives and declaratives, which were expected to be readily interpretable in terms of the attributes 'surprised' and 'emphatic'. The stimuli, which were lexically equivalent across the two languages, were varied in peak height, peak alignment, end pitch and pitch register. Each stimulus was assigned the contour H*L L%. Depending on which variables were varied and which were controlled for, the stimuli could be divided into four subsets.

Statistical analyses

A set of data was obtained from each of the four sets of stimuli, consisting of perceived surprise scores and perceived emphasis scores. ANOVA's were performed on each set of data for two dependent variables Surprise and Emphasis. These eight ANOVA's comprised one between-subject factor: Language Background (2 levels) and within-subject factors: sentence mode (2 levels) plus the prosodic variables varied in each data set. We adopted a significance level of .05.

Results

As for the attribute 'surprised', we found significant interactions between Peak Height and Language ($F_{4,192}$=7.22, p<0.05) as well as between Pitch Register and Language ($F_{4,192}$=4.37, p<0.05). The two-way interaction of Peak Height × Language and the two-way interaction of Pitch Register × Language give us similar pictures. By and large, in both languages, the higher the peak height, the higher the degree of perceived surprise, while the same was true for Pitch Register. These are evident manifestations of the Effort Code. However, at identical peak heights and pitch registers, Dutch listeners perceived a higher degree of surprise in Dutch than BrE listeners did in BrE. This is a Type 1 difference, triggered by a difference in the standard pitch range.

As for the attribute 'emphatic', we found significant two-way interactions between Peak Height and Language ($F_{4,192}$=47.41, p<0.05) and between Pitch Register and Language ($F_{4,192}$=15.16, p<0.05). The two-way interaction of Peak Height by Language is similar to that found for 'surprised'. In both languages, there is a positive correlation between peak height and perceived emphasis. However, at identical peak heights, Dutch listeners perceived a higher degree of emphasis in Dutch than BrE

listeners did in BrE. Again, this is a Type 1 difference, due to a difference in the standard pitch range.

However, the two-way interaction of Pitch Register by Language presents a rather different picture than the three two-way interactions considered above. More specifically, we observed reversed interpretations of pitch register variation. In Dutch, a higher pitch register led to higher degrees of emphasis, while in BrE, a higher pitch register led to lower degrees of emphasis, as illustrated in Figure 3. This is a new type of difference, which we refer to as a Type2 difference.

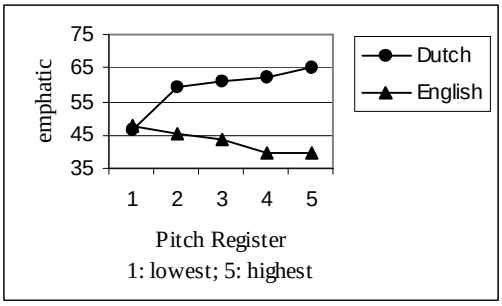


Figure 3. The interaction of Pitch Register × Language

A type 2 difference cannot be accounted for as due to different standard pitch ranges. High pitch register can in principle signal 'friendly' by the Frequency Code as well as 'emphatic' by the Effort Code. In view of our research results, it is likely that BrE listeners may have perceived pitch register variation as expressing friendliness, but that Dutch listeners may have perceived the same variation in terms of the Effort Code, i.e. as variation in emphasis. The meaning 'friendly' undermines the meaning 'emphatic', and as a result, when the pitch register, and thus perceived friendliness, increases, the degree of emphasis would decrease in BrE. Accordingly, we would expect a higher degree of friendliness to be perceived in BrE than in Dutch at identical pitch registers. As it turned out, this is indeed what we found in [1]. Therefore, a Type2 difference was triggered by a difference in the primary meaning signalled by register variation, and for this reason BrE listeners perceived a higher degree of friendliness than Dutch listeners for a given pitch register. By contrast, Dutch listeners not only perceived a higher degree of emphasis than BrE listeners, but also showed a positive correlation between register and emphasis, as opposed to a negative correlation for the BrE listeners. This implies that rival interpretations based on different biological codes may be treated differently by different speech communities.

CONCLUSIONS

In the light of findings from two cross-linguistic investigations, we have demonstrated how languages can differ in the implementation of universal ways of signalling of attitude. Two types of language dependence have been identified: Type 1 differences amount to different 'strengths' of perceived universal attitude; Type 2 differences amount to reversed form-meaning relations between a given prosodic variable and its corresponding universal attitude. Two factors have been proposed to account for the language-dependence in the universal signalling of attitude: (1) a difference in the standard pitch range; (2) a difference in the kind of meaning a language chooses to express by means of a given prosodic variable. The latter situation may arise when two rival meanings are potentially present, and the interpretation of one of these precludes the interpretation of the other. To conclude, although the signalling of attitude across languages exhibits universal tendencies, there is a distinctive language-specific component to be dealt with. Knowledge of how and why languages can differ in the universal signalling of attitude will no doubt facilitate language-specific adjustments in multilingual speech modelling.

REFERENCES

1. Chen, A.J., Rietveld, A., and Gussenhoven, C. Language-specific Effects of Pitch Range on the Perception of Universal Intonational Meaning, in *Proceedings of Eurospeech'01* (Aalborg, September 2001) II: 1403-1406.
2. Chen, A.J., Gussenhoven, C., and Rietveld, A. Language-specific Uses of the Effort Code, in *Proceedings of the Speech Prosody '02* (Aix-en-Provence, April 2002), 215-218.
3. Douglas-Cowie, E., Campbell, N., Cowie, R. and Roach, P. Emotional speech: Towards a new generation of database. *Speech Communication* (to appear).
4. Graham, C.R., Hamblin, A.W., and Feldstein, S. Recognition of emotion in English voices by speakers of Japanese, Spanish and English. *International Review of Applied Linguistics in Language Teaching* 39 (2001), 19-37.
5. Gussenhoven, C. Intonation and interpretation: phonetics and phonology, in *Proceedings of the Speech Prosody 2002* (Aix-en-Provence, April 2002), 45-55.
6. Ladd, D.R. *Intonational Phonology*. Cambridge: Cambridge University Press, 1986.
7. Mozziconacci, S.J.L Modeling Emotion and Attitude in Speech by Means of Perceptually Based Parameter Values. *User Modeling and User-Adapted Interaction* 11 (2001): 297-326.
8. Ohala, J.J. Cross-language use of pitch: an ethological view. *Phonetica* 40 (1983): 1-18.
9. de Pijper, J. *Modelling British English Intonation*. Dordrecht: Foris Publications, 1983.
10. Scherer, K.R. A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology, in *Proceedings of ICSLP '00* (Beijing, October 2000), II: 379-382.
11. Zraick, R., and Liss, J.M. A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *Journal of Speech, Language, and Hearing Research* 43, 979-988.

On the Expressive Competencies Needed for Responsive Systems

Nigel Ward
Computer Science, University of Texas at El Paso
El Paso TX 79968-0518 USA. nigelward@acm.org

ABSTRACT
Subtle emotions and their expression often arise in the context of managing involvement levels and turn-taking in task-oriented interactions. This paper presents some thoughts regarding their importance for effective and efficient interaction, their essentially real-time nature, and their relation to social conventions.

Keywords
reflex, attitudes, feelings, social conventions, real-time, communication

RESPONSIVENESS
In human-human interaction, people sometimes are able to pick up and respond sensitively to the other's internal state as it shifts moment by moment over the course of an exchange. Table 1, taken from [6], suggests what some of these feelings might be. A literature survey and systematic inventory appears in Cowie et al. (2001).

People who can do this are generally known as good communicators and sensitive listeners. We would like computer systems ultimately to be able to do the same.

One exploration of this was a semi-automated tutoring-type spoken dialog system [6]. The system inferred information about the user's 'ephemeral emotions', such as confidence, confusion, pleasure, and dependency, from the prosody of his utterances and the context. Then, for each user utterance, the system adopted an appropriate "emotional" posture, such as being business-like, patiently supportive, encouragingly supportive, sharing in the user's triumph, being reassuring, and so on. This was conveyed by selecting an appropriate acknowledgement form, such as *yes*, *yeah*, *mm-hm*, *right*, *okay*, and *that's right!*. Although the differences in meaning between these expressions are quite subtle and hard to identify, even after careful analysis, users prefered the system with this ability to use these expressions appropriately.

In building this system the initial aim was only to mimic human behavior, on the belief that the pleasantness of dialog was due, in part, to successful exchanges of information about the participants' states, in real time as they change moment-by-moment during the dialog. However it became clear that doing so was in fact also functional: endowing the system with this sort of subtle emotional expressivity can not only make interactions more pleasant, it can make them more effective and more

I want to express my thoughts (by taking a turn soon)
I'm uncomfortable (with this topic)
I'm amused (by your story)
I'm frustrated (that I've not been able to convince you)
I'm pleased (that you appreciate the irony in my words)
I'm missing something (so you need to be more explicit)
I need a moment (to digest that statement)
I know what I'm talking about (so just listen a minute)
I'm not committed to any opinion (so you're welcome to keep talking)
I'm bored (so let's talk about something else)
I'm concerned (that I'm not expressing myself well enough)
I'm really interested (in your opinion on this)
I'm aware of that already (so we can go on to talk about something else)
I'm getting restless (so let's close out this conversation)
I'm feeling a twinge of irritation (at the tone of your last remark)

Table 1: Examples of Feelings that Occur as 'Ephemeral Emotions' in Dialog, as suggested by studies of prosody, back-channel lexical items, disfluency markers, and gestures, as they occur in tutorial-like dialogs, casual conversations and narrations (Bavelas *et al.* 1995, Ward and Kuroda 1999, Ward 2000)

CHI 2003 Workshop on Subtle Exppressivity for Characters and Robots.

efficient.

In particular, it seems that these sorts of expressions often convey attitudes regarding the flow of conversation and the general cast of the conversation, with implications for conversation control functions, such as determining who will speak how much and how slowly and at what level of detail. To summarize these meanings in terms of a communications engineering metaphor, they are out-of-band, and like out-of-band signals in communications systems, they are generally priority messages, status indicators, and control signals relating to the transmission of the main message [9].

DISCUSSION

While "subtle expressivity" is necessarily an imprecise term, it is worth attempting to roughly characterize what is involved.

It is often **task-related**: in comparison to expressions of classic emotions such as anger, fear, and joy, it can be closely related to task achievement.

It is often purely **communicative**, rooted in guiding and responding to the user, rather than in manifesting some deeply felt internal state. Producing subtle expressions usefully, or even just avoiding inappropriateness, may require a system to monitor and direct the dialog at a very fine grain, and involves dimensions of interaction different that those usually handled by user models or by dialog managers.

It is often a reflection of correctly following **social conventions**, rather than being doing anything clever, creative, or distinctively original. This may need to be programmed at a near-reflex level, where system expressions are directly determined by prosodic, gestural, and contextual properties of the user's actions. In a sense, it may be part of a low-level reactive sub-system, in the spirit of models where appropriate social behavior is explained and implemented without use of inference about the other's internal state, and without implementing any internal state for the agent [2, 4, 7]. On the other hand, even if subtle expressivity is reflex-level, when building a system it is often useful and appropriate to relate it to the expression of feeling or emotion. Certainly, when trying discuss peoples' perceptions of system behavior it is hard to avoid explaining it in terms of intentions and emotions.

It is often highly **real-time** -constrained. At least in some applications, if subtle expressions appear within the window of acceptability they are convincing and effective, but if they come even a fraction of a second too late, users may fail to relate them to the proper context, and their meaning can be weakened or changed.

It is of course **subtle**, by definition. This has several implications, including the difficulty of measuring their value. One technique that is sometimes useful is to have users evaluate them off-line, in a second evaluation phase. That is, after interacting with a system, if the user can then observe a video or audio recording of his own interaction, while following along on a automatically generated transcript, he may be able to more accurately judge the quality of of the system's contributions. This technique can be an effective way to amplify weakly-detected user preferences [5].

References

1. Janet Beavin Bavelas, Nichile Chovil, Linda Coates, and Lori Roe. Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, 21:394–405, 1995.

2. Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

3. Roddy Cowie, Ellen Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollais, W. Fellenz, and J. G. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18:32–80, 2001.

4. Alan J. Fridlund. The new ethology of human facial expressions. In J. A. Russell and J. Fernandez Dols, editors, *The Psychology of Facial Expression*, pages 103–129. Cambridge, 1997.

5. Wataru Tsukahara and Nigel Ward. Evaluating responsiveness in spoken dialog systems. In *International Conference on Spoken Language Processing*, pages III: 1097–1100, 2000.

6. Wataru Tsukahara and Nigel Ward. Responding to subtle, fleeting changes in the user's internal state. In *CHI '01*, pages 77–84. ACM, 2001.

7. Nigel Ward. Responsiveness in dialog and priorities for language research. *Systems and Cybernetics*, 28(6):521–533, 1997.

8. Nigel Ward. The challenge of non-lexical speech sounds. In *International Conference on Spoken Language Processing*, pages II: 571–574, 2000.

9. Nigel Ward. A model of conversational grunts in American English. submitted to *Cognitive Linguistics*, 2002.

10. Nigel Ward and Takeshi Kuroda. Requirements for a socially aware free-standing agent. In *Proceedings of the Second International Symposium on Humanoid Robots*, pages 108–114, 1999.

Social Signals:
Using Principles and Methods from Social Psychology to Guide Subtle Expression Design

Katherine Isbister, Ph.D.
katherineinterface
1904 23rd Street
San Francisco, CA 94107 USA
+1 415 722 1945
ki@katherineinterface.com

ABSTRACT
This paper describes some principles and evaluation techniques from social psychology that may be of use in guiding subtle expression design. Suggestions include grounding expression design in social goals and roles, and considering situational variables and cultural specificity of expression. A mix of subjective and direct behavioral measures to evaluate the efficacy of subtle expressions is proposed.

Keywords
Nonverbal communication, social roles, social goals, situational cues, social rhythms, cultural differences, evaluation techniques.

INTRODUCTION
This workshop highlights the importance of reliably developing subtle expressivity for robots and other characters that will engage in communication and tasks with humans. In this brief paper, I suggest that our research community can benefit from both the theories and evaluation methods of social psychologists. Nonverbal communication is grounded in the functional contexts of communication—social situations. Understanding the dynamics of social situations and making use of existing concepts for bounding nonverbal communication in these contexts can help us to direct research and development efforts.

FUNCTIONS OF NONVERBAL COMMUNICATION
Nonverbal expression plays a crucial role in everyday human communication. Beyond providing informational content [11], nonverbal communication helps to build trust and relationship. As Ting-Toomey points out: "nonverbal messages are often the primary means of signaling our emotions, attitudes, and the nature of our relationships with others" [15, p. 115]

Subtle changes in gaze [3], pitch of voice, speed and range of gesture [1], help to establish the rhythms of the individual, so that his/her communication partner can learn to align and harmonize [5]. In addition, the performance of culturally situated nonverbal behaviors, such as greeting and farewell rituals, and the maintenance of appropriate interpersonal distance [8], help to maintain smooth and satisfactory interaction rhythms, minimizing embarrassment and potential loss of face [7].

GROUNDING NONVERBAL COMMUNICATION IN ROLES AND GOALS
Just as there are many layers of meaning, communication, and potentials for error in language [2], there are many complex layers of meaning, communication, and potentials for error in nonverbal communication. As researchers, we must select from among these many layers, areas for development that have strong promise and potential. I would propose that we use social goals and roles as a focal point for making selections of expressive behaviors and strategies to develop.

This is in contrast to an approach that would emphasize crafting some sort of reading of the internal state of the robot or character, in a parallel to the biologically expressive function of expressions like emotions [4]. It is also in contrast to an approach which seeks to simply maximize informational communication through the use of paralinguistic cues.

Social Goals in Social Encounters
At the risk of stating the obvious, social encounters are not purely about attaining functional ends, such as getting assistance with a task or obtaining some information. Human beings enter social encounters with ongoing relational goals, and it is these that I feel we should look toward in exploring the potential of subtle expressivity in adding value to interactions with robots and characters.

Receiving reassurance that what I am doing is correct, absorbing and reflecting enthusiasm, getting a sense that I will be protected, or that I am approved of—these are all parallel goals and cues operating in social encounters. A robot or character that does not perform this sort of parallel communicative work risks at worst alienating the interaction partner, at best seeming cold and unengaging. In addition, a robot or character that does not take part in

this parallel work is not likely to motivate continued interaction, once an initial informational goal is met.

Social Roles and Person Perception as a Road Map for Communication Strategies

If we believe that these sorts of relational goals are worth exploring and engineering expressions to facilitate, then where shall we turn to learn more about what to develop? There is a large body of findings in social psychology about what people notice and evaluate when they meet one another, and how it impacts their social judgment [e.g. 6]. Some key variables include how dominant the other person may be, whether they are friendly or not, how intelligent they might be, and also, the assessment of what social role the other is currently performing. Social roles, such as teacher, or employee, are sets of common expectations for behavior and potential relational support. I believe a combination of social roles and key person perception variables can go a long way toward guiding wise development of expressivity in robots and characters.

IMPORTANCE OF SITUATION

Another crucial truth from social psychology is the tremendous impact of situations upon behavior [12]. Human beings tend to have a bias toward seeing their social partners as possessing very stable traits that explain their behavior and choices (such as personality), when in fact, much of behavioral variance can be explained by situational factors. This is a reason to devote significant attention in the development of expressivity in robots and characters to awareness of context, and performance of situationally appropriate behaviors.

Some situational variables that we could consider include: significant events in the setting itself; duration and frequency of interactions with any given human partner; presence of other humans or robots; events within the institutional setting that may affect everyone's mood; and micro-cultural norms of behavior within the range of venues the robot or character will need to inhabit.

IMPORTANCE OF CULTURE

The robot and character design community is an international one. It is important to acknowledge and factor into our efforts, the extreme cultural specificity of many expressive behaviors, as well as our social roles and expectations [8, 11, 15]. Interestingly, to the extent that our creations display disfluencies and awkwardness, they may be read as 'from another culture' by their human interaction partners. Thus, applying what is known about inter-cultural communication and its issues can be of use both in working together to set international agendas for expression development, and also, in interpreting how end users will react.

EVALUATION STRATEGIES

I have suggested a variety of factors from social psychology to consider incorporating into expressive robot and character research. Just as important as understanding these concepts, is understanding how to evaluate when they have been effectively applied. Studying nonverbal behavior is notoriously difficult [13]. One cannot simply rely on self-report of effects of nonverbal behavior on impressions—in empirical studies, subjects may deny observing nonverbal cues, and yet display statistically significant differences in impressions [9]. It is important to employ a careful combination of objective and subjective measures to confirm effects. Direct observational measures that may be helpful include:

- Videotaping/observation and coding of nonverbal behaviors and activities [13].
- Use of sensors to track and correlate nonverbal behaviors of humans and robots [10].
- Tracking of gaze and eyebrow movement of humans.
- Observation/recording of post-interaction behavior.

Subjective measures can be obtained by having participants fill out questionnaires after their interactions with robots or characters. Typically, social scientists limit the variables under study in any given trial, and use subjective measures to compare reactions to target behaviors versus some control condition [14].

Given these evaluative tools, how shall we choose which to use in any given study? Each has strengths for certain types of research questions. A detailed overview is beyond the scope of this paper, but here are some preliminary connections between evaluative methods and questions:

Nonverbal behavioral tracking and coding:

Can be used to observe interpersonal distance, orientation, touching, mimicry of robot postures, leaning forward or away, dominant or submissive body language, facial expressions, reactions to awkward behaviors. These results can be used to examine whether there is increasing trust, warmth, and engagement with the robot, and when it has done something socially awkward. These measures can be used in tandem with subjective measures (questionnaires, see below) to get a clearer picture of participants' conscious and unconscious social reactions to the robot behaviors.

Eyebrow tracking can be used to monitor surprise, and thus can be used to detect awkward or unexpected behavior by the robot. Gaze tracking can be used to monitor

ongoing engagement and dominant or submissive reactions to the robot.

Activity tracking and coding:
Can be used to observe whether participants change their behavior in different experimental conditions—are they more helpful toward the robot, do they follow its suggestions on tasks, do they perform target tasks better and/or longer. These results can be used to establish the practical value of a robot's social expressivity and appropriateness in actual behavioral outcome.

Questionnaires:
Can be used to ask participants about the reactions and attitudes toward the robot or character, after an interaction. Measures can include: trust, liking, perceived competence, perceived dominance and friendliness, cultural attribution, social role appropriateness and competence, and other variables. Participants can also be asked about their general affective state, and attitude toward the task as a whole. These measures can be used to assess whether the robot's behaviors 'read' as hoped for, and whether the presence of such subtle expressivity has a positive impact on the perception of functional encounters. As mentioned above, these measures will be strongest when combined with some direct observational measures that break out behavior and reaction during actual encounters with some granularity. So for example, if we seek to create reassurance in a participant with the robot's behavior, we would both look for direct nonverbal cues of reassurance and trust (relaxed body postures in subject, mimicry of robot body language and greater proximity) as well as attitudinal measures (perceived trust, greater reassurance about a task in a questionnaire given after the task than in one given before). In this way we can both confirm the effect we seek, and understand more specifically how to refine the expressive behaviors of the robot to maximize social effectiveness.

CONCLUSION

This paper has highlighted social psychological theory as well as evaluative strategy, to consider in directing the expressive robot and character community's development efforts. I hope these ideas will serve as a springboard for discussion and collaboration.

REFERENCES

1. Allbeck, J., and and Badler, N. *Toward Representing Agent Behaviors Modified by Personality and Emotion*. Workshop on Embodied Conversational Agents—Let's Specify and Evaluate Them. AAMAS, Bologna, Italy, 2002.
2. Clark, H. *Using Language*. Cambridge University Press, Cambridge UK, 1996.
3. Colburn, A.R., Cohen, M.F., and Drucker, S.M. *The Role of Eye Gaze in Avatar Mediated Conversational Interfaces*. Technical Report MSR-TR-2000-81, Microsoft Research, 2000.
4. Darwin, C. *The Expression of the Emotions in Man and Animals*. The University of Chicago Press, Chicago IL, 1965.
5. Davis, M. (ed.). *Interaction Rhythms: Periodicity in Communicative Behavior*. Human Sciences Press, Inc., New York NY, 1982.
6. Ekman, P., Friesen, W., O'Sullivan, M., and Scherer, K. Relative Importance of Face, Body, and Speech in Judgments of Personality and Affect. *Journal of Personality and Social Psychology 38, 2*: 270-277, 1980.
7. Goffman, E. *The Presentation of Self in Everyday Life*. Anchor, New York NY, 1959.
8. Hall, E.T. *The Hidden Dimension*. Anchor Books, Doubleday, New York NY, 1966.
9. Isbister, K., and Nass, C. Consistency of Personality in Interactive Characters: Verbal Cues, Nonverbal Cues, and User Characteristics. *International Journal of Human-Computer Studies 53, 2,* 251-267.
10. Kanda, T. *A Constructive Approach for Communication Robots*. Dissertation, Kyoto University Department of Social Informatics, 2003.
11. Leathers, D.G. *Successful Nonverbal Communication: Principles and Applications*. Macmillan Publishing Company, New York NY, 1986.
12. Ross, L., and Nisbett, R.E. *The Person and the Situation: Perspectives of Social Psychology*. McGraw-Hill, Inc., New York NY, 1991.
13. Scherer, K.R., and Ekman, P. (eds.) *Handbook of Methods in Nonverbal Behavior Research*. Cambridge University Press, Cambridge UK, 1982.
14. Stempel, G.H., and Westley, B.H. *Research Methods in Mass Communication*. Prentice Hall, Englewood Cliffs, NJ, 1989.
15. Ting-Toomey, S. *Communicating Across Cultures*. The Guilford Press, New York NY, 1999.
16. Webb, E.J. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Rand McNally & Co., New York NY, 1966.

Facial Expressions for Conversational Agents

Dirk Heylen
University of Twente
PO Box 217
7500 AE, Enschede, NL
+31 53 4893745
heylen@cs.utwente.nl

ABSTRACT
In this paper, we argue that the choice of facial expressions of conversational agents should be guided mainly by parameters in the dialogue context: the ideational, interpersonal, and meta-communicative functions, and not so much by the emotional state of the agent.

Keywords
Facial expressions, conversational agents.

INTRODUCTION
It is perhaps questionable whether the facial expressions of synthetic characters should be modeled to any realistic detail on the example of that of their human creators. For instance, Blumberg's work at MIT on Synthetic Characters takes inspiration from the Disney guidelines for animating characters which involves taking the real humans as a basis but modifying the behavior by exaggeration and simplification (Kline and Blumberg, 1999). However, even in this case the human face serves partly as a model for the synthetic behavior. For other purposes the human model may even be more relevant. Suppose then, that we want to make synthetic characters that resemble real humans as much as possible. What behavior should they display? Or in other words, how do humans act with their face? In this paper, we take another look at the literature on human facial expressions to come up with a list of guidelines for implementing synthetic faces. We will focus on the use of synthetic faces in a conversational setting.

There is a core repertoire of facial displays that appears in most work on synthetic faces. Often, this core constitutes almost the whole of the repertoire of facial displays. It is familiar to anyone who has been concerned with this topic. The displays are based on the work of Ekman on the universal expressions of what he claims to be the six basic emotions *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. Sometimes the repertoire is extended with variations on the emotions by allowing differences in intensity or blends of displays. Symptomatic of this dominance of attention to basic expressions are the predefined displays that appear in facial animation software or toolkits (see, for example: cslu.cse.ogi.edu/toolkit, interface. digital. com, www. miralab. unige. ch, mrl. nyu. edu / ~perlin/ facedemo). Often one can only work with the predefined displays.

A number of non-emotional expressions, mostly conversational signals have been worked on by researchers on conversational agents. For instance, the topic of gaze and eye-movements has been investigated by quite a few authors examining the function in discourse, the affective meanings, etcetera (Cassell et al. (1994), Cassell et al. (1999), Chopra-Khullar, N.I. Badler (1999), Colburn et al. (2000), Fukayama et al. (2002), Garau et al. (2001), Heylen et al. (2002), Novick et al. (1996), Poggi et al. (2000), Thórisson and Cassell (1996), Vertegaal et al., 2001).

If one is interested in an ecologically valid model of facial expressions, it is important to look beyond the core set of emotional expressions. Besides the affective displays and the conversational signals there are also other functions that should be taken into account. By considering what one would miss out on when focussing on this core set only, we can arrive at a set of guidelines for building synthetic models of facial expressions. In this paper we will try to do this by putting the theory of Ekman on universal expressions of basic emotions in its proper perspective. It is easy to do this just, by looking at a few pages of Ekman's "Tellling Lies". We want to claim that precisely the part of his work on the universal expression of basic emotions is perhaps the least useful for modeling embodied conversational agents because its focus is too limited for this particular context. The ecological settings we want to consider for the case of embodied conversational agents are precisely those for which other rules hold. The refinements involve two aspects. On the one hand, the repertoire of facial displays to be considered needs to be enlarged. Secondly, the motives that trigger the appearance of facial displays should not be limited to the involuntary basic emotion system.

UNIVERSAL EXPRESSIONS OF BASIC EMOTIONS

What are the claims made by Ekman in his writings on which the core set of expressions has been based? We consider a summary taken from Ekman (2001, pages 124-125).

> The involuntary facial expressions of emotion are the product of evolution. Many human expressions are the same as those seen on the faces of other primates. Some of the facial expressions of emotion – at least those indicating happiness, fear, anger, disgust, sadness, and distress, and perhaps other emotions – are universal, the same for all people regardless of age, sex, race, or culture. These facial expressions are the richest source of information about emotions, revealing subtle nuances in momentary feelings. […] The face can show:
>
> – which emotion is felt – anger, fear, sadness, disgust, distress, happiness, contentment, excitement, surprise, and contempt can all be conveyed by distinctive expressions;
> – whether two emotions are blended together – often two emotions are felt and the face registers elements of each;
> – the strength of the felt emotion – each emotion can vary in intensity, from annoyance to rage, apprehension to terror, etc.

In this quotation we find the notions of universal expressions of basic emotions (although not referred to these as such) along with important modulations pointing to a number of claims that are sometimes inferred from the work of Ekman but which are not properly claims made by him. So what are some of the claims *not* made by Ekman in his work?

1) The face is only an involuntary emotional signal system.
2) All facial expressions of emotion are universal and each emotion is always expressed in the same way.
3) Only basic emotions are displayed on the face.
4) Only emotions are displayed on the face.

We will now look at each of these points in turn and see what Ekman and others have to say about them.

1) The face is only an involuntary emotional signal system

Ekman does not claim that the face is only an involuntary emotional signal system. The claim would not be right because the face also signals emotions voluntarily and also because it signals other things besides emotions. Here we deal with the first reason. The second reason is discussed below when refuting (non-)claim 4.

Voluntary control of emotional expressions is accounted for in Ekman's model by introducing so-called display rules.

> But as I said, the face is not just an involuntary emotional signal system. Within the first years of life children learn to control some of these facial expressions, concealing true feelings and falsifying expressions of emotions not felt. Parents teach their children to control their expressions […]. As they grow up people learn *display rules* so well that they become deeply ingrained habits. After a time many display rules for the management of emotional expression come to operate automatically, modulating expression without choice or even awareness. […] I believe that those habits involving the management of emotion – display rules – may be the most difficult of all to break. (Ekman, 2001, page 125).

Typically, the display rules are said to be learned ways of hiding the real emotions or faking emotions in a social setting. They are also used to account for cultural differences. The well-known experiment with American and Japanese students that are shown emotion-arousing films in the presence and absence of others is used to collaborate this claim. The Japanese would mask expressions of negative emotion with a polite smile when a person in authority would be present much more than the Americans (Ekman and Friesen, 1971).

In the search for universal expressions of basic emotions, Ekman takes care to consider precisely those circumstances in which truly felt emotions are expressed without consideration of display rules. This means that in many experiments a situation is created in which the subjects are alone and care is taken that it is not obvious to them that they are observed. But this precisely entails that for an embodied conversational agent it does not make sense to be equipped only with a system that shows how it feels, because conversations are social encounters in which agents, like people, should not be expected to show how they really feel without further ado. In Ekman's terms, the research on facial expressions for embodied conversational agents should not be concerned with the voluntary, universal expression of basic emotions but concentrate on the identification of the kinds of display rules that are appropriate for the nature of the agent, its character, culture and the kind of encounter: the nature of the exchange and the nature of the interlocutor. In other words, for an engineer building synthetic faces to be used in conversational settings the research program as formulated by Bavelas and Chovil (1997) as below would be more to the point.

> The research […] represents the first stage in a program of study to investigate facial displays as discourse-oriented actions. In this research, facial displays are regarded as linguistic elements of a message rather than outputs or "spillover" of emotion processes.

Discourse-oriented actions should not be construed too narrowly as being related merely to the linguistic exchange; which would mean looking specifically at conversational signals. Rather, one should look at conversation as a form of social interaction and the role of facial expressions in terms of such things as impression management (Goffman, 1959; for instance. In this respect it is also important to mention Fridlund's critique on Ekman's model and his views on facial expressions and social interaction. See, for instance, Fridlund (1994).

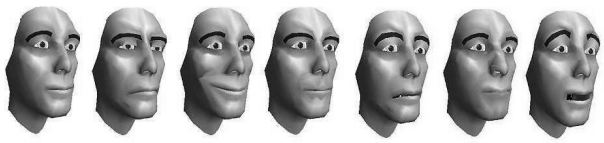
This way of approaching the subject of facial expressions would pay more attention to the social setting in which the conversation proceeds. The importance of the social situation and the function of display rules is also pointed out by the oft-cited observations by Kraut and Johnston (1979). They observed that people that were bowling would

smile when they hit a strike only if there was someone present that would take note of their happiness. Also related to this is the phenomenon of *mimicry* in which persons "enact" (out of sympathy or whatever) the feeelings of the person who is talking to them or the feelings of the character talked about (Bavelas et al. 1986). Scherer (2001, p. 180) summarises the issue as follows.

> Rather, emotional expression is multifaceted – expression is determined both by a person's reaction to an event and by the attempt to manipulate this expression for strategic reasons in social interaction.

2) All facial expressions of emotions are universal

It is already obvious from the first quote from Ekman that he does not consider all expressions to be universal. For one he only considers the involuntary expressions and says that "*some*" of the facial expressions of at least some (so perhaps not all) emotions are universal. The system of display rules is another way to account for non-universal expressions. There are differences between cultures but also between individuals. It would therefore not be wise for an embodied conversational agent to have his repertoire of facial expressions limited to the set below: the generic Everyman's face. For most purposes that use synthetic faces, a more individual, personalised repertoire of facial expressions would be more suited.



Facial expressions for emotions are certainly not always identical. For instance, the face displays differences in the intensity of how the emotion is felt and the expression of an emotion may be blended with the expression of other emotions. Furthermore, Ekman (2001, pages 127/128) writes the following.

> There is not one expression for each emotion, but dozens and, for some emotions, hundreds of expressions. Every emotion has a family of expressions, each visibly different one from another. This shouldn't be surprising. There isn't one feeling or experience for each emotion, but a family of experiences. […] Already we have evidence that there are more different facial expressions than there are different single words for any emotion.

Of course, one could claim that it is not always necessary for an embodied conversational agent to be true to life and to be able to express a multitude of nuances. In certain cases, it is perhaps easier to express emotions unambiguously, always in the same manner. However, one can also imagine situations in which variation in expression means that the agent becomes more engaging, entertaining, and true to life.

3) Only basic emotions are displayed on the face

This claim does not hold for a number of reasons. For one, it would not be true if other emotions besides basic ones are displayed on the face. Secondly, it would not be true if the face shows other things besides emotions. This will be discussed further below in 4).

It would also not be true if there were no such things as basic emotions. The issue whether there exist basic emotions and if so how many and which has been hotly debated in the psychological literature (Ortony & Turner, 1990).

Following Tomkins, Ekman considers a limited set of about six emotions to be basic (happiness, fear, anger, disgust, sadness, and distress) and associated with universal expressions. However, it is clear from the above already that he considers variations in terms of the existence of blends.

Others (Frijda & Tcherkassof; Smith & Scott and Russell in the same book, 1997, for instance) have argued that one should not associate facial expressions directly with these basic emotions but that the emotions can be decomposed into several dimensions and facial displays should be similarly decomposed. This leaves open the way to combine the components in such a way as to obtain the expressions associated with the emotions as Ekman proposes (See Breazeal (2000), for instance, for a discussion.).

Note that in much of the work on emotion models for agents, the work by Ortony Clore and Collins (1988) is used, in which about 20 emotion types are distinguished.

4) Only emotions are displayed on the face

Perhaps the face is the most important channel to signal aspects of the emotional state of a person. However, that does not mean that other channels do not signal affect, nor that the face only signals affect. In fact, Chovil (1991) remarks:

> Although facial displays are undoubtedly used at times to convey information about how a person is feeling or reacting, emotion displays do not account for the majority of displays that occur. Ekman and Friesen […] found that in nearly 6,000 facial displays of psychiatric patients, less than one-third were classifiable as facial expressions of emotions. This suggests that, although some facial displays may convey information about emotion, there are a substantial number of displays that we know very little about.

From her own research Chovil (1991) concludes that hardly 20% of the displays in face-to-face conversations are affective.

Ekman has written quite extensively about non-emotional displays. In several writings he has presented his classification of facial expressions. Again we take the summary from Ekman (2001, page 127).

> There are thousands of facial expressions, each different from one another. Many of them have nothing to do with emotions. Many expressions are what we call conversational signals, which, like body-movement illustrators, emphasize

speech or provide syntax (such as facial question marks or exclamation points). There are also a number of facial emblems: the one-eye closure wink, the raised eye-brows-droopy upper eyelid-horseshoe mouth shrug, the one-eyebrow-raised skepticism, to mention a few. There are facial manipulators, such a lip-biting, lip sucking, lip wiping, and cheek puffing. And then there are the emotional expressions, the true ones and the false.

This shows that what appears on the face is not just related to the emotional state of a person but to many other aspects. There are quite a few expressions that provide cues about the body state of a person. Other displays relate to information and symbol-processing: the conversational signals, often with a meta-communicative function, but also illustrators and emblems. Together with the fact that emotional displays are also used voluntarily to communicate emotions rather than to express them, this motivated Bavelas and Chovil to consider facial expressions as discourse-actions (see above). Related to the emotional state are signals that indicate the mental state of speakers and listeners. The work by Baron-Cohen (1995) on the language of the eyes, shows the importance of interpreting expressions in terms of a person's mental state (such as attention) and more specifically in terms of "reading" the state of mind of the other. Similarly, Interpersonal functions of facial displays (Argyle, 1993), showing how one feels and thinks, about the other and the relation to self are also very important in conversations.

SUMMARY

From the discussion above, we can deduce some guiding principles to use when building embodied conversational agents, be they animated talking heads or robot faces.

1) Because embodied conversational agents are involved in social interaction, researchers should pay special attention to voluntary control of emotional expressions. In Ekman's terminology the "display rules" should take first place.
2) To make embodied conversational agents behave more like individuals, variation in expression is preferred above universal expression.
3) The emotional life of humans involves more than six basic emotions. To show what goes on in the mind of the conversational agent, the signs should not be limited to "basic" emotions, only. Signals of affective states of agents should also be richer. Also, depending on the situation, the mood, or the choice of style, a different kind of expression may be chosen for the same or a similar affective state.
4) Ecologically valid simulations should pay attention to those expressions that occur most in conversations. Meta-communicative, conversational and interpersonal functions of facial displays should be brought to the fore.

Above, we have emphasised the importance of the situation in which the expressions on an agent are to be displayed. Focussing on embodied conversational agents, this involves a social setting in which the facial expressions are used in communicative situations. The face is not simply used to display the emotional turmoil, but rather as an actively controlled communication channel. Great attention has to be paid therefore to conversational signals and display rules. However, it also implies a number of other things. As conversation is a form of social interaction one should pay special attention to the way people present themselves to others; i.e. impression management. What kinds of expressions would the other think are appropriate? How exaggerated should they be to be understood. This means that the agents should be able to hide their feelings or be able to lie.

What does this mean for the way we build our agents and robots? To enable rich, human-like expressivity it is important not just that the characters can display these expressions, but also that their internal states are appropriate to what is expressed. The displays should be correct expressions of the states the different components of the agent are in: its body and mental state, its knowledge about conversation, and most importantly perhaps the models of self-awareness and awareness of the other.

As far as affect influences the choice of expression, it should be realized that in a conversation the affective state is co-determined mainly by the events that happen during the conversation (the goals of the conversation and how they are realized or get disturbed in an argument, how well one likes the interlocutor, how one wants to present oneself in the conversation and how well one succeeds, etcetera). This, again points out that in conversational agents, the interactional factors are of major importance.

CONCLUSION

In this paper we have tried to summarize important observations and theories that should be taken into account when implementing synthetic faces for embodied conversational agents. We have tried to argue for a more ecologically valid model that takes into account also the individual characteristics of agents on the one hand and the fact that conversations are social encounters. We have taken the claims of the universal expression of basic emotions as proposed by Ekman as a starting point and then we critically examined what the shortcomings would be if we only considered this part of his writings as a sufficient model for the facial displays of an embodied conversational agent. We have shown how this part of Ekman's writings should be complemented with many other observations he has made and we have argued why precisely these are so important for our attempts to build synthetic talking faces. On one hand this means that an extended repertoire of facial expressions should be considered and on the other hand that the main factors that motivate facial displays in conversational agents relate to the conversational and interpersonal, social setting.

REFERENCES

1. M. Argyle (1993) *Bodily Communication*. Routledge, second edition.
2. S. Baron-Cohen (1995). Mindblindness. An Essay on Autism and Theory of Mind. MIT Press.
3. J.B. Bavelas, A. Black, C. R. Lemery and J. Mullett. "I show how you feel": Motor mimicry as a communicative act. In: *Journal of personality and social psychology*, 50(2):322-329, 1986.
4. Bavelas, J. B., & Chovil, N. (1997). Faces in dialogue. In J. Russell & J.-M. Fernandez-Dols (Eds.), *The psychology of facial expression (pp. 334-346).* Cambridge, U.K.: Cambridge University Press.
5. Breazeal, C. (2000), "Sociable Machines: Expressive Social Exchange Between Humans and Robots", Doctoral Dissertation. Department of Electrical Engineering and Computer Science. MIT. (Book version forthcoming from MIT Press).
6. J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, M. Stone (1994). Animated Conversation. Rule Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In: *Computer Graphics* (p. 413-420).
7. J. Cassell, O. Torres, S. Prevost (1999). Turn Taking vs. Discourse Structure, in *Machine Conversations* (p. 143-154).
8. N. Chovil (1991). Discourse-Oriented Facial Displays in Conversation. In: Research on Language and Social Interaction, Vol. 25, 163-194.
9. S. Chopra-Khullar, N.I. Badler (1999). Where to look? Automating attending behaviors of virtual human characters. In: *Proceedings of Autonomous Agents*. Seattle (p. 9-23).
10. R.A. Colburn, M.F. Cohen, S.M. Drucker (2000) Avatar Mediated conversational interfaces. Microsoft Technical Report. MSR-TR-2000-81. July 2000.
11. P. Ekman. Telling Lies: Clues to Deceit in the Marketplace, Marriage, and Politics. New York: WW Norton, 2001.
12. P.Ekman and W. V. Friesen (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17, 124-129.
13. A.J. Fridlund, Alan (1994) Human Facial Expression. An Evolutionary View. San Diego, Academic Press.
14. Frijda, N. H., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. A. Russell & J. M. Fernandez-Dots (Eds.), The psychology of facial expression (pp. 78-102). London: CUP.
15. A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, N. Hagita (2002) Messages Embedeed in Gaze of Interface Agents. Impression management with agent's gaze. ACM, CHI 2002, 41-48.
16. M. Garau, M. Slater, S. Bee, M.A. Sasse (2001) The impact of eye gaze on communication using humanoid avatars. In: *CHI* 2001 (p. 309-316).
17. E. Goffman (1959). The Presentation of Self in Everyday Life. Garden City, NY Doubleday Anchor.
18. D. Heylen, I. Van Es, B. van Dijk, A. Nijholt (2002) Gaze Behavior of Talking Faces Makes aDifference. Proc. *CHI 2002: Changing the World, Changing Ourselves*, L. Terveen & D. Wixon (eds.), Minneapolis, April 2002, 734-735.
19. C. Kline and B. Blumberg. (1999) The Art and Science of Synthetic Character Design Proceedings of the AISB Symposium on AI and Creativity in Entertainment and Visual Art, Edinburgh, Scotland.
20. Kraut, R. E. & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology, 37*, 1539-1553
21. D.G. Novick, B. Hansen, K. Ward (1996) Coordinating Turn-Taking with Gaze. In: *Proceedings ICSLP* (p. 1888-1891).
22. A. Ortony, G.L. Clore and A. Collins (1988). The Cognitive Structure of Emotions. CUP.
23. A. Ortony and T.J. Turner (1990). What's basic about basic emotions? Psychological Review, 97, 315-331.
24. I. Poggi, C. Pelachaud, F. de Rosis (2000) Eye Communication in a Conversational 3D synthetic Agent. Special Issue on Behavior Planning for Life-Like Characters and Avatars. AI Communications Vol 13(3) (p.169-181).
25. K.R. Thórisson, J. Cassell (1996) Why Put an Agent in a Body: the importance of communicative feedback in human-humanoid dialogue. Presented at Lifelike Computer Characters, Utah, October 1996.
26. K. R. Scherer (2001). Emotion. In: M. Hewstone and W. Stroebe (eds.) Introduction to Social Psychology, Malden Massachusetts, Blackwells p. 151-195.
27. Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expressions. In J. A. Russell & J. M. Fernandez-Dols (Eds.), The psychology of facial Expression (pp. 229-254). New York, Cambridge University Press.
28. R. Vertegaal, R. Slagter, G. van der Veer, A. Nijholt (2001) Eye Gaze Patterns in Conversation. There is more to conversational agents than meets the eyes. In: *Proceedings of CHI 2001 Anyone. Anywhere*. ACM

Automatic Analysis of Facial Expressions: Problems and Solutions to Data Collection

Dr John Cowell
Dept. of Computer Science,
De Montfort University,
The Gateway, Leicester, LE1 9BH
England.
jcowell@dmu.ac.uk

Dr Aladdin Ayesh
Dept. of Computer Science,
De Montfort University,
The Gateway, Leicester, LE1 9BH
England.
aayesh@dmu.ac.uk

ABSTRACT
Humans convey information on their emotional state by their facial expression. A system capable of determining the emotions of a person based on their expression would have many applications including security applications, assessment of satisfaction of commercial products and improvement in the way humans and robot systems interact.
In this paper, we provide an analysis of the current state of emotion classifiers identifying the extent of the two main problems: data collection and emotion classification. In response, we propose an experimental procedure which will allow rapid collection of high quality statistically viable data. This data will be analyzed for the development of an automated connectionist based classifier using models of fuzzy cognitive maps.

KEY WORDS
facial expression, facial action encoding, facial expression emotional classification

INTRODUCTION
Humans communicate through their facial expressions, speech, movement and hand gestures, but of these only one is always active: the expression on the face. Even a neutral display apparently showing no emotion conveys information to the viewer. Extensive research on the relative importance of these modes of communication indicate the importance of facial expression. Mehrabian [1] experiments indicated that the contribution of the spoken part of a message was only 7%, while the intonation account for 38% and the facial expression 55%. The ability to detect and classify the emotional state of humans from their facial expressions is an attractive one which has many application of a social, economic and political nature. While automatic classification systems are unlikely to perform better than a trained human observer, automatic systems are cheaper, do not grow tired, or lose interest and therefore can be used in a more extensive way. Identifiable applications include:

- Assessment of commercial products, particularly where the user interacts with the product over a long time period, such as a robot system, a vehicle or a film.
- Interactive systems capable of identifying the users' degree of satisfaction could adapt to provide greater satisfaction.
- Security applications such as at airports where individuals exhibiting particular facial expressions identified with stress could be identified.

Recent work on the automatic analysis of facial expressions focuses on three areas as identified by Pantic[2],

- Detection of the face in an image.
- Extracting data on the expression of the face.
- Classification of the extracted data.

Pantic[2] gives an gives an extensive review of the state of the art in automatic analysis. The first two areas are within the field of image analysis and pattern recognition and satisfactory systems can be developed if a constrained environment is used, for example a static frontal view of the face with a single colour background, however the classification and interpretation of the extracted data is the most complex area, with an extensive body of literature beginning with Darwin[3] in 1872.

CATEGORIZATION OF EMOTION
There are some basic questions regarding the relationship between facial expressions and emotions which must be addressed before it is possible to consider the automation of this process. Fundamental questions include:

- What are the basic categories of emotion?
- What is the effect of culture on the association between emotions and facial expressions. Is there a universality of expression?
- Are the same emotions stimulated by different events in different cultures?
- Does age and gender affect how emotions are expressed?

These questions have been answered with varying degrees of success by empirical work of varying quality over the

past fifty years. By the 1980's this field was dominated by the Facial Expression Program. This consists of a set of assumptions, theories and methods. Russell[4] captures the essence of the Program with a list of 14 items which contain essential information required for the development of an automated system. The first item in the list states that there are a small number of emotions (seven plus or minus two). These emotions are happiness, surprise, fear, anger, contempt, disgust and sadness. There is still some controversy over the inclusion of contempt and over the distinction between surprise and fear. Shame is also a possible contender for addition to the list. This group of emotions has been used by most automatic emotion detection systems. There are, however, alternatives: some researchers of automated systems have used. Picard[10] uses the system proposed by Clynes[11] which describes eight emotions: no emotion, anger, hate, grief, platonic love, romantic love, joy and reverence.

The Facial Expression Program has broad but by no means universal acceptance, however it is clear that it is the most widely presented set of corollaries and form an integral part of current theories of emotion[8][9]. In addition it is presented as the dominant paradigm in text books [5][6] and advisory documents for practicing psychologists[7]. Another of the most significant items of the Facial Expression Program are the three related propositions as described by Russell[4]:

1. "The same pattern of facial movement occurs in all human groups.
2. Observers in different societies attribute the same specific emotions to those universal facial patterns
3. Those same facial patterns are, indeed, manifestations of those very emotions in all human societies"

While there is extensive research to indicate that for western literate people there are a common set of facial expressions which can be correctly interpreted with varying degrees of success (although better than chance), some expressions are interpreted more correctly than others. Notably happiness is the easiest to identify and disgust the hardest. There are strong suggestions from researchers such as Russell[12], Fernández-Dols[13] and Fridlund[14] that the earlier work of researchers such as Ekman[15][18] which pointed to a universality of expression is flawed. A useful discussion of this is given Russell and Fernández-Dols[2]. Further empirical research is required with non-western people to indicate whether the three propositions listed above from the Facial Expression Program are in fact true or not.

Less controversial is the view that the same stimulus may evoke different emotional responses and facial expressions in different cultures.

Kaiser, Wehrle and Schmidt [20] discuss the two contrasting theoretical approaches to the interpretation of emotional expressions. The first is the work of discrete emotion theorists such as Ekman and Izard who postulate a set of basic emotions. Emotional expressions which do not match one of these basic emotions are the result of blending two or more together. The second approach is suggested by Kaiser and Scherer who state that the expressions described by Ekman and Friesen in the Emotion Prediction tables in the FACS manuals do not often occur and argue that facial expressions play an important role in social communications and do not reflect solely an internal emotional state[21].

DATA COLLECTION

There are also serious problems of how empirical data is collected for analysis. The inherent variation in human faces requires a large data set to extract meaningful results. Ideally, examples of individuals expressing all of the basic emotions in a pure form are required, however this is extremely difficult to obtain. A possible solution is to request individuals to simulate emotions. Many studies ask individual to pretend to experience certain emotions and to show this by their facial expression. There is strong evidence that the facial expressions which are produced in these circumstances are different from those spontaneously produced when those emotions are experienced. Any system which uses this type of input as the basis for interpreting emotions is flawed and cannot produce accurate results. It is possible to collect data by using video cameras, however this is fraught with difficulties:

- The environment may have a big impact on the expression produced. Laboratory or an unfamiliar environment may add an underlying tension or anxiety to the subject.
- The presence of the camera may affect the subject: hidden cameras may be used, however there are ethical considerations and most countries have legislation which either prohibits the use of hidden cameras or requires the subject to be informed after they have been filmed and given the opportunity to destroy the film produced.
- If the subject is aware of the purpose of the experiments he may behave in a different manner.
- The collection of data is time consuming and it may take weeks of observation to observe the required emotions. Every minute of video film must be scrutinized in detail. This is a serious problem since large numbers of subjects must be used to produce viable data.
- The choice of subjects may influence the results. Many studies use young literate educated students as the subjects rather than a carefully selected cross section of the entire population which reflects the age, gender and social groupings of that group.

In addition to these complexities there is the difficulty of creating situations in which the subject can experience a wide range of emotions. Techniques such a stress interviews or showing the subject a variety of images can produce emotions, however the interpretation of these is difficult. If a subject is filmed expressing an emotion how

do we know what emotion is being expressed? An obvious answer is to ask the subject, however this does not always lead to an true reflection of the individuals emotional state. Some people are lacking in insight to their own or others emotional state and also people may not tell the truth. Experiments designed to induce a feeling of disgust in subject include the dissection of rats, however in some people this may produce a feeling of anger, surprise or in some cases pleasure. In these situations subjects may report that they experienced an emotion which they feel is a socially acceptable rather that what they actually felt.

An alternative strategy to determine which facial expression relates to a particular emotion is to show the film to people and ask what emotion they think is being shown. Surprisingly there is often no agreement between what the individual reports he experienced and what observers report. This is an area where there are strong cultural considerations. Japanese and Chinese subjects find it harder to express anger and disgust compared to their Western counterparts: it is therefore essential that the subject and the observers who are asked to report on the emotional state they perceive are of the same cultural group.

A further difficult in obtaining data is that subjects rarely express pure emotions, subjects may report that they felt angry and surprised, and produce facial expressions which are a composite of the two.

There have been attempts to solve some of these problems and to produce an objective coding system which measures the position of facial components involved in emotional expression and relates combinations of these to emotions. Early examples include the Facial Affect Scoring System (FAST) developed by Ekman, Friesen and Thompson [17] which has 77 different descriptions of expression related to emotional states. Similar schemes have been produced by Izard who developed the Maximally Descriptive Facial Movement Coding System (MAX). These encoding schemes have their limitations. Neither scheme allows for the intensity of emotion to be measured and the MAX system was developed for infants and may omit expressions which children and adults produce. A refinement of these systems was produced by Ekman [16] the Facial Action Coding System FACS [16]. FACS breaks each facial movement into 44 action units (AU). Each AU is assigned a five point intensity value. FACS is the dominant encoding system in use, however one of the problems associated with it, is that it is very time consuming to apply FACS to faces. If this process could be automated it would be of significant benefit. Most of the automated systems being developed use FACS as their basis. A review of current systems and a discussion of their capabilities is given by Pantic [2].

The extraction of AUs from faces is a complex process especially if the data is collected in natural environments from video film; the background is likely to be complex and the face may have any size and position and may be partially obscured. This complex image processing problem coupled with the difficulty of interpreting the AUs and relating them to emotional state makes this a complex multi-disciplinary research problem which currently has no satisfactory solution.

PROPOSED SOLUTION

In order to develop a robust system capable of relating emotional state to facial expressions experiments must be designed which allow automated collection of data in natural environments. Simulated emotions collected from small numbers of individuals or data collected in unfamiliar stress-inducing environments cannot yield high quality data. The easier it is to collect the raw data, the more the inherent variations in facial expression can be represented and the easier the analysis into cultural, gender and age related groups. By contrast many experiments typically use few (less than 10) individuals with no consideration of their culture, age or gender. The invasive nature of the laboratory environment which affects outcomes must be reduced. The experiments must be repeatable by other researchers. We believe that this could be achieved by observing a group watching commercial films, many of which are designed to evoke strong emotions. Small cameras could be positioned so that each camera observed one person. This approach has several advantages. The environment is sufficiently constrained to make the automatic extraction of the face and hence the expression easier compared to completely unconstrained environments. The controlled nature of the environment also focuses the attention onto the film, (indeed cinemas are designed to do this) so that it is more likely that the emotion experienced is related to the film. The experience of watching a film is common to most people in most cultures and there the familiarity of the experience means that people will not be exhibiting the underlying stress emotions which occur in unfamiliar surroundings. The length of the film also means that after a few minutes the presence of small unobtrusive cameras will be forgotten. The emotional experience is shared and therefore tend to be stronger and easier to detect. The film may be shown many times to different audiences and the emotional experiences at the same points in the film recorded. It should be noted that most countries have controls on the collection of such data and people filmed must be given the opportunity to view the film and have it destroyed if they wish.

In the attempt of developing automated emotion classifiers, variations of rule-based expert systems have been used[2]. The problem of such solutions is that they assume a perfect world model scenario, ignoring the cultural and personal differences in expressing emotion. The availability of large amounts of high quality data will allow an alternative approach to tackling this problems based on [19] and uses fuzzy cognitive maps to relate perceptions and emotions. Once the nodes representing emotions of expressions are established, the fuzzy functions will fine-tone the

relationship between these nodes for a better emotional classification.

CONCLUSIONS

This paper provides a brief discussion of some of the problems facing the development of a system for determining emotions from facial expressions. The following are the key points of the paper:

- The facial expressions which correspond to emotional states are likely to be culturally dependent, although there is still debate in this area. It is clear however, that the interpretation of these states can be most reliably achieved by a member of the same cultural and social group.
- A set of seven basic emotions have been identified by psychology researchers, and are widely used by researchers, however there is inevitably some ambiguity in the definition of these states.
- The collection of data for use in a recognition system is fraught with difficulties: people cannot experience emotions and express them on demand; the environment and the choice of subjects has a bearing on the type and manner of emotional expression.
- There is little research on how combinations of the basic emotions are expressed facially. Some researchers suggest that facial expressions are a reflection of social communication in addition to internal emotion.
- One of the pre-requisites for an automated system capable of identifying emotional state from facial expressions is that data must be collected from a broad cross section of people. This needs extremely careful design of the environment used to collect valid data.
- The FACS system provides a method of measuring the AUs involved in emotional expression, however the interpretation of these is still the subject of research particularly for non western cultures.
- We propose an experimental approach which will allow the collection of a large amount of high quality data and allow this date to be related to emotional states.
- The analysis of this data can be achieved by fuzzy cognitive maps.

REFERENCES

1. Mehrabian A. *Communication without words*, Psychology Today, vol. 2 no. 4. pp53-56 1968.
2. Pantic Maja, Rothkrantz Leon J. M. *Automatic Analysis of facial expressions: The State of the Art.* IEE transactions on Pattern Analysis and Machine Intelligence, vol. 22 no. 12, pp 1424-1445 Dec. 2000
3. Darwin C. *The expressions of the emotions in man and animals.* Chicago: University of Chicago Press. 1965 (Original work published 1872)
4. Russell James A, Fernández-Dols José Miguel. *The psychology of facial expressions*. Cambridge University Press. 1997 ISBN 0-521-49667-5
5. Carlson J .G., Hatfield E. *Psychology of emotion*. New York: Holt, Rinehart & Winston.
6. Ingram J. *The burning house : Unlocking the mysteries of the brain*. London: Penguin.
7. Behavioral Science Task Force of the national Advisory Mental Health Council 1995. *Basic behavioral science research for mental health: A national investment: motivation and emotion*. American psychologist 50 pp838-845.
8. Damasio A.R. *Descartes' error*. 1994. New York: G.P. Putnam & Sons.
9. Oatley K. *Best laid schemes: the psychology of emotions.* Cambridge : Cambridge University press.
10. Picard Rosalind et al. *Towards machine emotional intelligence: analysis of affective physiological state.* IEEE Transactions on pattern Analysis and Machine Intelligence. vol. 22 no. 10 pp1175-1191. Oct. 2001.
11. Clynes. D. M. *Sentics: The touch of emotions.* Anchor Press/Doubleday 1977.
12. Russell James A. *Is there a universal recognition of emotion from facial expression*? Psychological Bulletin 115 pp102-141 1994,
13. Fernández-Dols J.M. Ruiz-Belda M. A. *Are smiles a sign of happiness? Gold Medal winners at the Olympic games*. Journal of personality and social psychology 69. pp1113-1119 1995.
14. Fridlund A. J. *Human facial expressions: An evolutionary view*. New York Academic Press 1995
15. Ekman P. *Universals and cultural differences in facial expressions of emotions*. Nebraska Symposium on Motivation vol. 19 pp207-283 1971. University of Nebraska Press.
16. Ekman P. *Emotions in the human face.* Cambridge University Press 1982.
17. Ekman P. Friesen W., Tomkins S.S. *Facial Affect Scoring Technique : a first validity study*. Semiotica, 3, 37-38. 1971.
18. Ekman P., Friesen W., Ellsworth P. *Emotions in the human face*. Pergamon Press 1972.
19. Ayesh A. A connectionist approach to perception and emotion based reasoning. IASTED Conference Artificial Intelligence applications (AIA 2002) Malaga Spain 2002.
20. Kaiser S. Wehrle T Schmidt S *Emotional episodes, facial expressions and reported feelings in human computer interactions.* In A.H. Fischer(Ed.), proceedings of the Xth Conference of the International Society for Research on Emotions.
21. Kaiser S and Scherer K R Models of 'normal' emotions applied to facial and emotional expressions in clinical disorders. In W.F. Flack Jr. and J D Laird (Eds) Emotions in Psychopathology (pp81-98) New York : Oxford University press 1998.

Silence, Murmurs and Applause: Reflections on Expressions of Collections

Thomas Erickson
IBM T.J. Watson Research Center
snowfall@acm.org

INTRODUCTION

To start out, I should say that my interest in this workshop comes from a very different perspective than, I suspect, those of the other applicants to the workshop. My hope is that there is enough in common that the difference in my goals and perspective will prove to be a source of fruitful discussion.

The area of commonality has to do with the issue of subtle expressivity. I am interested in means—particularly as manifested in visual and auditory portrayals—of conveying complex and subtle expressions. And much of my work draws upon studies—principally from sociology—of the ways in which people notice and interpret large arrays of subtle cues expressed by 'natural systems.' The area of difference is that I am *not* interested in human facial expression, *per se*, or in designing characters or robots to make them more expressive. Instead, I am interested in designing 'expressions' for complex social and computational systems, particularly multi-user systems within which large collections of people are active.

I define an "expression" as a large array of subtle cues that recipients are able to interpret holistically. A canonical example is, of course, human facial expression, produced by the reconfiguration of skin by large numbers of muscle groups, which serve (among other things) to portray emotional state (e.g. [1]). Although the facial muscles produce literally thousands of distinct configurations of the skin, people are quite good at mapping these complex configurations into a relatively small number (6 to 10) of basic emotional states. In addition, facial expressions may also represent blends of emotional states, as seen, for example, in the classic illustration of a dog's face exhibiting a mixture of the expressions of rage and fear. While researchers are divided as to whether emotions evolved for the *purpose* of communicating internal states, it is quite clear that people make use of expressions to infer internal emotional states and, as well, that people attempt to manage their own expressions to control the inferences of others (e.g. Goffman, 1959).

In my view, "expressions" are not limited to individual humans and animals,[1] but are—at least potentially—a feature of any complex social or technical system, particularly those that are used by, inhabited by, or otherwise include large numbers of people. That is:

> *Any complex social or technical system that provides an interpretable portrayal of its internal state by making the fine structure of its internal activities visible, may be said to have an expression.*

The sorts of complex systems whose "expressions" (and their interpretations by people) that I have tried to understand include groups of people in auditoriums, city streets and urban plazas. The sorts of systems for which I have attempted to design expressions (or, as I prefer it, expressive representations) include online classrooms, virtual auctions, and multi-room discussion environments.[2]

In the remainder of this position paper, I will do two things. I will elaborate on the concept of expressions as attributes of complex systems rather than of individuals, and then I will discuss my approach to designing expressions for such systems.

COLLECTIVE EXPRESSIONS

In his discussion of the ways in which people manage their expressions [4], Eving Goffman distinguishes between cues that are *given* (that is, that are used deliberately and solely to convey information), and cues that *given off* (that is, behavioral cues that are not produced primarily for communicative purposes, and may therefore be presumed to be inadvertently released). Thus, when I encounter a colleague at a conference, the words of my greeting (*given*) may express pleasure at the encounter, even as an involuntary grimace (*given off*) suggests that I feel otherwise. Of course, hopefully I am better at managing my expression than that, and so in spite of my feelings I will successfully enact a smile, thus *giving off* a feigned expression. In general, our expressiveness as individuals is an ongoing effort to control the 'face' we present to the world, the information we are *giving* undermined or reinforced by the expressions we *give off*.

[1] Those who agree with Minsky's Society of Minds theory [7] might argue that even individual expressions are not the product of a single entity, but rather produced by a large set of cooperating and competing mental agencies.

[2] I am not the first to suggest that the notion of 'expression' is applicable to systems. For instance, Don Norman, in *Turn Signals are the Facial Expressions of Automobiles* [8], draws a parallel between human facial expressions and the function of automobile signals as a means of conveying real or feigned intent.



Figures 1, 2 and 3: The expression of state in physical systems

Awkward Silences and Standing Ovations

It is interesting, and not, I think, entirely coincidental, that Goffman [4] often relies on the metaphor of theatrical drama to illustrate his ideas. Just as an individual makes a concerted effort to control his expression and ensure that the information *given off* reinforces that which is *given*, so does the company of actors putting on a play make an effort—behind the scenes—to coherently depict a scene or situation to the audience.

More interestingly, the audience is similarly engaged. When the play is ready to begin, the house lights are lowered, and the audience responds, their collective murmur subsiding into silence, punctuated by the occasional cough. Similarly, when the play ends, the audience makes an attempt—each individual intentionally acting on his or her own—to *give* signs of their enthusiasm. Typically the result of this is applause, an individual's hand-claps quickly taken up by others, swelling into a uniform texture of sound. Occasionally, if the play is well received, one or a few individuals may stand up, perhaps leading the rest of the audience to stand as well. On the other hand, if the play is not so well received, a very different situation can result: the attempt at a standing ovation may fail, with a few scatted people standing as the rest remain seated; or worse, even applause may fail to catch on, with distinct isolated claps echoing loudly in the largely silent theatre.

These situations are uncomfortable, indeed, being the failure of an audience to express a collective response in a coherent fashion. All of theses cases—in the ways in which their blends of *given* information (individual claps, standing up) and *given off* information (the unanimity and synchrony of the audience's collective action that may be presumed to be spontaneous and uncontrolled) reflect the audience's reception (real or feigned) of the play—bear a very strong resemblance to the workings of individual facial expression.

Streets and Markets

While the case of an audience applauding (or not) in a theatre seems particularly apropos, it is easy to identify situations in which the activities of a collection of individuals—often generated without awareness of their collective impact—produce a global or holistic impression.

We are adept, particularly in situations with which we are familiar, at judging the state and level of activity of the system at a glance (figures 1, 2 and 3, above). And it is not simply that these 'expressions' of collective activity support us in our instrumental activities, enabling us to decide whether we have arrived too late, have arrived at a good time to accomplish a task efficiently, or are in for a wait. Rather, the expressions of systems affect how we feel about them: we take in a streetscape, noticing that it is lively and interesting; we are attracted to markets filled with people and a sustained murmur of conversation (figure 4). As Jane Jacobs argues, in writing about the effect of activity (or its absence) on a street, the impressions conveyed by mundane activity can have more profound and longer-reaching effects:

> The sum of such casual, public contact at a local level—most of it fortuitous, most of it associated with errands, all of it metered by the person concerned and not thrust upon him by anyone—is a feeling for the public identity of the people [of the neighborhood], a web of public respect and trust, and a resource in time of personal or neighborhood need. ([5], p. 57)

Similarly, Kevin Lynch, an urban designer, wrote:

> …a distinctive and legible environment not only offers security but also heightens the potential depth and intensity of human experience. … Potentially, the city is in itself the powerful symbol of a complex society. If visually well set forth, it can also have strong expressive meaning. ([6], p. 5)

There is more to be said about the roles of collective expressions, but space requires us to move on.

Figure 4: A lively market.

DESIGNING COLLECTIVE EXPRESIONS

Recently, I've been involved in designing multi-user systems that provide public visualizations of the activities of participants. We call the sort of visualizations we design, social proxies, and suggest that—by revealing the fine structure of individuals' activities within the system in a form which can be readily taken in— social proxies can play a variety of roles in supporting public behavior in the system. In this paper, we discuss two examples[3] in which the aim is to convey a feeling for the overall state of the system.

Social proxies are minimalist graphical representations that typically consist of a geometric background figure that depicts a particular activity or situation, and small colored dots that represent participants. Movements of the dots relative to the background figure provide information about the individual activities of the participants, and express the overall state of the system. Let's look at two examples.

Auctions

In the physical world of face to face interaction, auctions are social events. A crowd gathers, inspects the items being offered, and participates in a public bidding process. Participants not only look at what is being auctioned—they also observe who is interested in what, and who bids for what; and they are conscious that their own actions and gazes are watched by others. That is, people not only bid *for* items, they also bid *against* other participants. All this contributes to making auctions intensely social and dramatic experiences, as well as enabling them to function as social mechanisms for computing the value of items, asserting the social or professional status of the bidders, and, of course, actually carrying out transactions.

However, when we look at online auctions, the social cues that make their face-to-face counterparts such rich and engaging experiences have vanished. The social proxy shown in figure 5 is an attempt to restore some of these cues. The large circle represents the auction 'room,' the center circle a clock, and each dot a participant. People who look at information about the to-be-auctoned itme are shown around the outside of the circle; when they place bids, their dots move into the circle. Thus, the auction proxy shows how many have shown interest, how many have bid, and how much time remains. Also, a dot is shown in color if the user has recently hit the web page: thus, the proxy also indicates how many people are 'present' and thus, perhaps, are candidates for entering the bidding at the last minute. This visualization expresses some of the drama that characterizes face-to-face auctions.

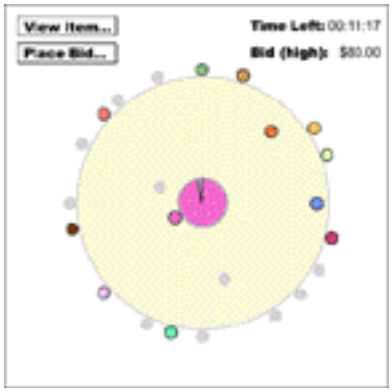


Figure 5. An auction proxy

Lectures

Imagine an on-line talk or lecture delivered as part of a conference call and accessed by people using screen phones. The Lecture proxy, shown in figure 6, assumes that we have some way of identifying who has spoken. The background figure represents the lecture 'room;' dots represent people; and the positions of the dots reflect how much they've spoken during the last five minutes. If the lecture is going as it 'ought'—with the lecturer speaking and the audience being quiet—the dots in the proxy assume a very regular pattern. However, if a person interrupts with a question or a comment, his or her dot will move a bit to the left, and if the interruptions continue, that person becomes, quite literally, 'out of line' (as shown in figure 5). If multiple people interrupt, their dots move forward as well, imparting a 'raggedness' or incoherence to the visual image that is not unlike that experienced when an audience fails to enthusiastically applaud for a play.


Figure 6: The lecture proxy

Because the proxy is seen by everyone, everyone knows (and knows that everyone knows) what is happening: it makes the state of the system public. How the group makes use of this information is up to it: making the fact that people are interrupting the lecture public may act to encourage the return to the norms of the lecture interaction, or it may encourage more people to interrupt. A social proxy is a means of expressing the system's state, not a means of control, and it dictates a response no more (and no less), than an expression of surprise on someone's face requires a particular sort of response.

Summary

I've argued that expressions are not simply phenomena produced by individual actors, but that they may also be seen as the products of the actions of a large collection of people. Both individual and collective expressions consist of a large number of subtle cues, both reflect internal characteristics of the actor(s) as they shift dynamically over time, and both can be used as a grounds for holistic interpretations about the state of the system. This approach to designing social proxies, or "expressions" for complex systems has been deliberately minimalist: we believe that our use of large number of simple shapes to represent the activities of a system's components aids the viewer in generalizing or holistically interpreting the state of the system. A more detailed discussion of this approach to designing visualizations may be found in [3].

[3] These examples, and some of the accompanying text, is taken from Erickson, et al. [3].

References

1. Darwin, C. *The Expression of the Emotions in Man and Animals.* London, John Murray, 1872. (See "The writings of Charles Darwin on the web" (ed. John van Wyhe), http://pages.britishlibrary.net/charles.darwin3/expression/expression_intro.htm.)
2. Erickson, T., Halverson, C., Kellogg, W. A., Laff, M. and Wolf, T. "Social Translucence: Designing Social Infrastructures that Make Collective Activity Visible." *Communications of the ACM* (Special issue on Community, ed. J. Preece), Vol. 45, No. 4, pp. 40-44, 2002.
3. Erickson, T. and Kellogg, W. A. "Social Translucence" Using Minimalist Visualization of Social Activity to Support Collective Interaction." *Designing Information Spaces: The Social Navigation Approach* (eds. K. Höök, D. Benyon and A. Munro). London: Springer, 2003, pp. 17-42.
4. Goffman, E. *The Presentation of Self in Everyday Life.* New York: Anchor Books, 1959.
5. Jacobs, J. *The Death and Life of Great American Cities.* New York, Vintage Books, 1961.
6. Lynch, K. *The Image of the City.* Cambridge, MA: MIT Press, 1966.
7. Minsky, M. *The Society of Mind*. New York: Simon and Schuster, 1985.
8. Norman, D. *Turn Signals are the Facial Expressions of Automobiles.* Reading, MA: Addison-Wesley, 1992.