Building HAL: Computers that sense, recognize, and respond to human emotion

Rosalind W. Picard
MIT Media Lab; E15-392; 20 Ames St; Cambridge, MA 02139;
http://www.media.mit.edu/affect

ABSTRACT

The HAL 9000 computer, the inimitable star of the classic Kubrick and Clarke film "2001: A Space Odyssey," displayed image understanding capabilities vastly beyond today's computer systems. HAL could not only instantly recognize who he was interacting with, but also he could lip read, judge aesthetics of visual sketches, recognize emotions subtly expressed by scientists on board the ship, and respond to these emotions in an adaptive personalized way. Of course, HAL also had capabilities that we might not want to give to machines, like the ability to terminate life support or otherwise take lives of people. This presentation highlights recent research in giving machines certain affective abilities that aim to make them more intelligent, shows examples of some of these systems, and describes the role that affective abilities may play in future human-computer interaction.

Keywords: affective computing, emotion recognition, human-computer interaction, emotion communication

1. INTRODUCTION

HAL in *2001* was affective: he had specific abilities relating to, arising from, and deliberately influencing people's emotions such as: he could sense and recognize human emotion, respond rationally to it, express emotion, and even give the appearance of "having" emotion. In fact, HAL was the most emotional character in the film *2001*. As the millennium dawns, we find that most computers today do not have affective abilities, but there is active research beginning to succeed in giving them a subset of such abilities.

Is it beneficial to make computers affective? Alternatively, is this just a theatrical gimmick, something that makes film characters entertaining but we wouldn't really want in real life? The latest neuroscience evidence supports the former: emotions are essential not only to dealing effectively with social-emotional interactions, but also they perform important regulatory and helpful biasing functions within the body and brain, even aiding in rational decision making (Damasio, 1994) and perception (LeDoux, 1996). These and a variety of other important roles of emotion, together with findings from neuroscience, cognitive science, and social-psychological sciences, have been argued to be important reasons for giving machines emotional abilities if they are to be intelligent (Picard, 1997).

What is the state of the art regarding giving machines emotional abilities? Can machines really "have" emotion? HAL's emotional state is associated with detrimental consequences for human life; thus, wouldn't giving machines emotion-like mechanisms be potentially dangerous? These are just a few of the questions that arise regarding *affective computing:* computing that relates to, arises from, or deliberately influences emotion. Affective computing also involves giving machines skills of emotional intelligence: the ability to recognize and respond intelligently to emotion, the ability to appropriately express (or not express) emotion, and the ability to manage emotions. The latter ability involves handling both the emotions of others and the emotions within one self. Because it is a large discussion whether computers can "have" emotions (and even a "self" to experience emotion) and the topic is addressed in a recent book chapter (Picard, 2001), this topic will not be addressed here. A discussion of the ethical issues related to HAL and an in-depth treatment of the state of the art as of 1997 regarding affective computing and HAL's abilities appear as chapters in the book *HAL's Legacy* (Dennett, 1997; Picard, 1997). [1]

[1] The reader is also referred to *HAL's Legacy* for its chapters addressing the state of the art (in 1997) of computer science with respect to HAL's abilities such as computer vision, playing chess, lip reading, natural language processing, and so forth (Stork, 1997).

Human Vision and Electronic Imaging VI, Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Editors,
Proceedings of SPIE Vol. 4299 (2001) © 2001 SPIE · 0277-786X/01/$15.00

Today, more than ever, the role of computers in interacting with people is of importance. Most computer users are not engineers and do not have the time or desire to learn and stay up to date on special skills for making use of a computer's assistance. One can argue: if the role of technology is to serve people, then why must the non-technical user expend so much time and effort getting technology to do its job? In *2001*, HAL's emotional abilities were intended to help address the problem of interacting with HAL. When the BBC reporter asks about HAL's emotional abilities, crewman Dave Bowman responds,

> "Well, he [HAL] acts like he has genuine emotions. Of course he's programmed that way to make it easier for us to talk with him…."

The emphasis when the film was released in 1968 was that HAL's emotional abilities would make things easier for people, leading naturally to a smoother interaction. Today, the tendency for people to interact socially with machines, even when the machine has no visible life-like face or audible voice, has been demonstrated via dozens of experiments (see, e.g., Reeves and Nass, 1996). The role of emotional skills is an essential part of social intelligence (Gardner, 1983) and the ability to perform a certain set of emotional skills has been argued to comprise a form of intelligence (Salovey and Mayer, 1990). An argument has even been made that such so-called emotional intelligence is more important for success in life than are the traditional mathematical and verbal capabilities that IQ tests attempt to measure (Goleman, 1995). A subset of these emotional skills: the ability to sense, recognize, and respond to human emotion, form the focus of the rest of this presentation: where does technology stand today with respect to giving machines these abilities?

2. EMOTION SENSING AND RECOGNITION

Emotions in people consist of a constellation of regulatory and biasing mechanisms, operating throughout the body and brain, modulating just about everything a person does. Emotion can affect the way you walk, talk, type, gesture, compose a sentence, or otherwise communicate. Thus, to infer a person's emotion, there are multiple signals you can sense and try to associate with an underlying affective state. Depending on which sensors are available (auditory, visual, textual, physiological, biochemical, etc.) one can look for different patterns of emotion's influence. The most active areas for machine motion recognition have been in automating facial expression recognition, vocal inflection recognition, and reasoning about emotion given text input about goals and actions.

HAL, with his superb visual abilities, could presumably perform facial expression recognition, although there are very few facial expressions made by any of the humans in the film *2001 (*most of the humans in the film were relatively impassive). Rosenfeld's chapter on computer vision in *HAL's Legacy* (Rosenfeld, 1997) includes an example where a computer algorithm recognizes one of Dave Bowman's rare facial expressions. There has been a fair bit of progress in this area since my 1997 treatment of the topic in *Affective Computing*. The research focus originally was on detecting six "basic" facial expressions (anger, sadness, happiness, disgust, surprise, fear) from still images, and then from video, with best results ranging for the latter from around 80-98% accuracy when the data was pre-segmented into one of the six categories and the lighting and position of the face were carefully controlled. More recent work (Bartlett et al, 1999; Cohn et al, 1999, and Donato et al, 1999) have focused less on a half dozen basic categories and more on recognizing dozens of facial actions, specific muscle movements that combine to form a much larger vocabulary of expressions. Some of these methods are now performing comparably to human ability to recognize facial actions (Donato et al, 1999). Problems remain, however, in tracking faces that move in front of the camera, handling changes in lighting, and recognizing facial expressions while a person is speaking. In short, current technology is still far behind what people can recognize from one another's faces.

Most pattern recognition researchers are familiar with a variety of tools for representation of patterns –including discrete categories, fuzzy or probabilistic categories, and dimensioned spaces, to name a few that are particularly relevant to emotion representations. Emotion theorists do not agree upon a definition of emotion, but most of them fall into one of two camps in how they describe emotion – either as basic discrete categories, e.g., fear, sadness, joy, etc., or as locations within a dimensioned space, the two foremost dimensions of which are usually termed "arousal" and "valence." The arousal dimension tends to refer to the overall excitement or activation of the emotion, while the valence dimension tends to refer to how pleasing (positive) or displeasing (negative) the emotion is. Peter Lang and his colleagues, for example, have measured how people respond to hundreds of images in terms of the dimensions of arousal, valence, and dominance, recording physiological patterns that exhibit significant differences especially with respect to the arousal and valence dimensions (Lang, 1995).

A given emotion can of course be represented in multiple ways. For example, anger can be represented as a discrete category, defined by some collection of attributes, such as by facial actions that typify its expression, or by some bodily parameters that lie within a negative valence, high arousal portion of a dimensioned space. In general, facial expressions are good at communicating valence (positive, negative) while vocal inflection (especially pitch and loudness) is good at communicating arousal. Combinations of facial and vocal analysis tend to strengthen the inference of the underlying emotion.

We know that HAL could detect emotions such as displeasure or distress from his lines such as,

> "Look, Dave, I can see you're really upset about this. I honestly think you ought to sit down calmly, take a stress pill, and think things over."

From the Clarke and Kubrick novel (written after the 1965 screenplay), we learn that HAL detected stress by listening to voice patterns. Vocal analysis of emotion continues to be an area of active inquiry, although progress since my 1997 treatment of this topic in *Affective Computing* has not been dramatic. Machine and human recognition of affect in speech still remains around the same as I described then, typically well below 100% recognition accuracy, usually hitting around 60-75% when given data from one of six to eight categories. Polzin's thesis is perhaps the most recent work trying to recognize several categories of affect (Polzin, 2000) with our work providing one of the more recent efforts on stress recognition (Fernandez and Picard, 2000). In our work we built models of driver's speech under mild to moderate stress conditions, comparing methods such as factorial hidden Markov models (HMM's), hidden Markov decision trees, auto-regressive HMM's, a mixture of HMM's, Support Vector Machines, and a neural network. The mixture of HMM's gave the best performance to date, although the results are very person-dependent and the best results are still well below 100%. (See the references in these works for pointers to many more recent articles addressing vocal affect and stress recognition.)

Presumably, HAL could reason about emotion – knowing, for example, from inference about human value for the lives of ones colleagues, that Dave should be upset and stressed about the death of his crewmates. Although such reasoning does not always imply how somebody actually does feel, when combined with observations of behavior, such as Dave's seriousness and increasing tension, HAL's inference of Dave's state should be strengthened. Machine reasoning about affect is one of the areas of artificial intelligence that has been explored the longest, and my review in *Affective Computing* of this area is still fairly up to date. In my opinion, the real breakthroughs that are needed to improve emotion recognition are not so much in reasoning about emotion, but are in perceiving accurate information with which to reason, and in detecting the affective tone of the context. The latter relates more to problems in common sense learning (and generalization of what one has learned) than to reasoning per se. Thus, context sensing, perception of what the situation is, perception of the emotional nuances present in a situation, and perception of how the people are responding are critical inputs to combine with an affective reasoning system.

Although HAL apparently observed people through visual and vocal cues, most of today's computers still rely upon physical keyboard/mouse input, where sensors might attend to not only what is typed or clicked, but how it is typed or clicked (speed, pressure, and other skin-surface cues.) Recent efforts toward building wearable computers also open up a lot of new affect sensing possibilities – especially through skin-surface sensors that detect muscle tension, skin conductivity, heart activity, temperature, and respiration. Progress in these areas includes new sensors such as IBM's "emotion mouse" (Ark et al, 1999) and a variety of tangible and wearable interfaces designed and built by the Affective Computing Group at MIT (Picard, 2000). Using pattern recognition of physiology, we have achieved recognition rates of 81% accuracy for a set of eight emotions in a person-dependent forced-choice pattern recognition scenario (Vyzas and Picard, 1999) and rates of up to 96% accuracy in assessing level of stress in a subject-independent study of twelve Boston drivers (Healey, 2000).

3. RESPONDING TO EMOTION

Computers are in their infancy with respect to recognizing emotion; however, suppose that, like HAL, they could recognize some of our expressions of emotion; how then should they respond? Although one might argue that the answer to this is more of a social science or psychology issue than an engineering one, it is still an important question for us to consider with respect to thinking through the potential capabilities of affective systems and how they might be constructed and used, for better or worse. In *2001,* we saw that HAL's response to recognizing stress was to suggest that the stressful person (Dave) sit down calmly and take a stress pill. How would you feel if your computer, after being the source of your irritation, told you to sit down and take a stress pill? Chances are there would be a wide range of reactions, some of which might include increasingly negative feelings.

One of the advantages of a system that can recognize affective expressions, especially those of pleasure or displeasure, is that it can try out different responses on a user, to see which are most pleasing. Indeed, a core property of most learning systems is the ability to sense positive or negative feedback – affective feedback – and incorporate this into the learning routine. Most dogs are better than computers when it comes to sensing this feedback.

The ability of dogs to not only sense positive and negative affect, but also to respond to it, was part of what motivated our work (Klein et al, 2001) where we showed that a system that responds to user frustration with (the appearance of) active listening, and appropriately gauged expressions of empathy and sympathy, has a significant effect on user behavior with respect to decreasing user frustration. The basic idea is that a system reflect that it has somehow understood the user's emotion, even in a limited way – much like a dog might put its ears back and tails down if it sees its master is upset. Such a display of apparent empathy, even by a dog, can have a powerful impact toward alleviating the strong negative feeling of a person, in this case the dog's master. The ability of a computer to not only detect its user's emotion, but to influence it by choosing a careful response, is an important one, which raises many ethical and social issues. We address many of these in a forthcoming article (Picard and Klein, 2001), but it is important to raise this issue here as well, so that designers of these systems can be aware of at least one potentially powerful way such technology may be used.

Above, I mentioned success we have had in detecting stress in drivers. The automobile environment is another place where the response of the system needs careful consideration. For example, if the system threatened the user's privacy by reporting driving behavior to the insurance company, this would not be acceptable to most users. However, if the system processed data in real time, saved no identifying or potentially incriminating information, and used the affective cues only to determine its own behavior – like routing an incoming cell phone call to voicemail during stressful attention-demanding driving situations, or adjusting its presentation of neighborhood or navigation information – then it might be considered of benefit to the consumer's safety and peace of mind. Such considerations influence how we design recognition algorithms – for example, aiming for real time analysis with minimal storage of state information.

Respect for the user's privacy and sense of control has influenced many of our design decisions. For example, we have built a pair of glasses that senses changes in the brow muscles, to detect furrowing of the brow. This wearable sensor, while initially seeming more awkward, can also be perceived as less intrusive than video. A camera pointed at one's face is nice in that you don't have to do anything but be visible; however, it may also record and extract information that you may not want to share – such as who you are and what you look like. In contrast, a small wearable sensor that just registers muscle tension presents the computer only with the furrowing information, while having the advantage that the user is in complete control of whether or not the sensor is allowed to operate. (It is easy to remove the glasses, or to detach the sensor from them without awareness of such to the system.) Items that are worn, that exist in the user's personal space, tend to give a greater sense of empowerment to the user. In contrast, when the sensing is "in the walls" like HAL's red eyeball, then the user may have information sensed without their awareness, as when HAL read the lips of Dave and Frank, a capability they did not know that he had.

4. CLOSING REMARKS

This presentation has highlighted some of the affective abilities that the HAL 9000 computer had, with emphasis on the sensing and recognition of emotion, and on responses to human emotion. HAL also had many other affective abilities, such as the ability to express emotion (through the emotive human voice of actor Douglas Rain), an ability that speech synthesizers still cannot emulate. Additionally, HAL acted as if he "had" emotions, especially fear and paranoia, as expressed not only through his behavior but also his famous words to Dave, "I'm afraid, Dave, …, I'm afraid, …" as he was being disconnected.

Although work in affective computing includes all of these aspects of emotion, our work at the MIT Media Lab has focused on giving machines a subset of affective abilities especially related to improving interaction with people, improving the machine's skills related to social-emotional intelligence. With this focus, we have tried to steer away from some of the hard AI problems of "understanding and experiencing" emotion, reposing the issues as problems in sensing, signals analysis, and pattern recognition. However, this is only part of the frontier where work needs to be done – researchers are needed from many areas, including engineering, social sciences, psychology, and cognitive science, to work in collaboration to build systems that are truly (following the words of Bowman) easier for us to interact with. In particular, careful regard must be made in the design of these systems so that they do not further irritate, annoy, or bring about unwanted stress to their users; such consequences would be antithetical to the goals of affective computing, which involve honoring the emotions of people above any such abilities that might be given to machines.

ACKNOWLEDGMENTS

I would like to thank the graduate students in Affective Computing, who have contributed significantly to this research, and the Media Lab sponsors in the Things That Think and Digital Life consortia, who have generously supported our investigations and our design and construction of prototypes.

REFERENCES

1. W. Ark, D. C. Dryer, and D. J. Lu (1999), “The Emotion Mouse,” *Proc. of HCI International ’99, Munich Germany, August 1999.*

2. M. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, ``Measuring facial expressions by computer image analysis,’’ *Psychophysiology*, vol. 36, pp. 253--263, 1999.

3. J. F. Cohn, A. J. Zlochower, J. Lien, and T. Kanade, ``Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding,’’ *Psychophysiology*, vol. 36, pp. 35-43, 1999.

4. A. R. Damasio, *Descartes’ Error: Emotion, Reason, and the Human Brain.* Gosset/Putnam Press, New York, NY 1994.

5. D. Dennett, “Did HAL commit murder? ” In D. G. Stork, editor, HAL’s Legacy: 2001’s Computer as Dream and Reality, The MIT Press, Cambridge, MA 1997.

6. G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, ``Classifying facial actions,’’ IEEE *T. Patt. Analy. and Mach. Intell.*, vol. 21, pp. 974--989, October 1999.

7. R. Fernandez and R. W. Picard, “Modeling Driver’s Speech under Stress, *Proc. ISCA Workshop on Speech and Emotions, Belfast, 2000.*

8. D. Goleman, *Emotional Intelligence*, Bantam Books, New York, 1995.

9. J. Healey, *Wearable and automotive systems for the recognition of affect from physiology*, Ph.D. thesis, Electrical Engineering and Computer Science, MIT, 2000.

10. J. Klein, Y. Moon, and R. W. Picard, “This Computer Responds to User Frustration,” *Interacting with Computers,* To appear, 2001.

11. P. J. Lang “Cognition in emotion: Concept and action. In C. E. Izard, J. Kagan, and R. B. Zajonc, editors, *Emotions, Cognition, and Behavior*, pages 192-226, Cambridge, 1984. Cambridge University Press.

12. J. E. LeDoux, *The Emotional Brain*, Simon & Schuster, New York, 1996.

13. R. W. Picard, “Does HAL cry digital tears? Emotion and computers.” In D. G. Stork, editor, *HAL’s Legacy: 2001’s Computer as Dream and Reality*, The MIT Press, Cambridge, MA 1997.

14. R. W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, 1997.

15. R. W. Picard, “Toward computers that recognize and respond to user emotion,” *IBM Systems Journal*, Vol. 39, Nos. 3&4, 2000.

16. R. W. Picard, “What does it mean for a computer to “have” emotions?” Chapter in *Emotions in Humans and Artifacts*, Eds. R. Trappl and P. Petta, MIT Press, 2001, to appear.

17. T. Polzin, *Detecting Verbal and Non-verbal Cues in the Communication of Emotions*. Ph.D. thesis, Carnegie Mellon, School of Computer Science, June 2000.

18. B. Reeves and C. Nass, *The Media Equation*, Cambridge University Press; Center for the Study of Language and Information, 1996.

19. A. Rosenfeld, “Eyes for Computers: How HAL Could `See’” In D. G. Stork, editor, *HAL’s Legacy: 2001’s Computer as Dream and Reality*, The MIT Press, Cambridge, MA.

20. P. Salovey and J. D. Mayer, “Emotional Intelligence,” *Imagination, Cognition, and Personality*, Vol. 9, no. 3, pp. 185-211, 1990.

21. Scheirer, J., Fernandez, R. and Picard, R.W. (1999). “Expression Glasses: A Wearable Device for Facial Expression Recognition,” *CHI ’99 Short Papers*, Pittsburgh, PA.

22. D. G. Stork, editor, *HAL’s Legacy: 2001’s Computer as Dream and Reality*, The MIT Press, Cambridge, MA 1997. Several chapters can be accessed at http://mitpress.mit.edu/e-books/Hal/.

23. J. D. Velasquez, “A Computational Framework for Emotion-Based Control,” In *Workshop on Grounding Emotions in Adaptive Systems, part of Fifth Int. Conf. on Simulation of Adaptive Behavior*, Zurich, Aug. 1998.

24. E. Vyzas and R. W. Picard, “Online and Offline Recognition of Emotion Expression from Physiological Data,” *Workshop on Emotion-Based Agent Architectures at the Int. Conf. on Autonomous Agents*, Seattle, WA, 1999.