Auditory and Other Non-verbal Expressions of Affect for Robots

Cindy L. Bethel and **Robin R. Murphy**
Institute for Safety, Search, & Rescue Technology
University of South Florida
Tampa, FL
clb1526@yahoo.com, murphy@cse.usf.edu

Abstract

This work focuses on the use of non-verbal methods of communicating affect in non-anthropomorphic *appearance-constrained* robots. The concepts discussed can be applied to other types of robotic systems. The paper gives a survey of relevant literature from the psychology community. Non-verbal methods of affective expression cannot be the sole means of communicating affect in robotic systems; however it can be appropriate as part of a multi-modal affective system. Another important factor is the distance between the robot and human with which it is interacting. From the psychology literature, there are four main zones of social interaction: *intimate* (contact – 0.46 m), *personal* (0.46 – 1.22 m), *social* (1.22 – 3.66), and *public* (3.66 and beyond). Sounds, tones, and/or music are appropriate to use in the *intimate* and *personal* zones; however it is not likely to be appropriate in the *social* and *public* zones due to environmental background noise. Communications are often challenging in search and rescue applications. Injuries that can impair language processing and limited visibility due to environmental conditions necessitate the use of non-verbal methods of communicating affect to survivors located during search and rescue operations.

Introduction

The use of affective expression and social interaction is an emerging area of importance in robotics, with the focus historically on facial expressions and/or animal mimicry (Breazeal 2002), (Caamero & Fredslund 2000), (Fong, Nourbakhsh, & Dautenhahn 2003), (Mizoguchi *et al.* 1997), (Scheeff *et al.* 2000), (Bethel & Murphy 2006a). However, a large number of mobile robots currently in use for applications such as search and rescue, law enforcement, and military are not anthropomorphic, do not have any method of projecting facial expressions, and cannot be re-engineered explicitly to support affective expression. This poses significant challenges as to how these *appearance-constrained* robots will support naturalistic human-robot interaction (Bethel & Murphy 2006a).

While facial displays are the most common mechanism for expressing affect, there is a need for a multi-modal approach to affective expression. There are situations and social interaction distances for which different methods of non-facial/non-verbal affective expressions are more appropriate (see Figure 2 (Bethel & Murphy 2006b) expanded from (Bethel & Murphy 2006a)). We have identified five non-facial/non-verbal methods of affective expression: *body movement, posture, orientation, color,* and *sound*. For the purpose of this work, auditory presentation methods such as sounds, tones, and/or music are the focus; however this is only one component of a multi-modal non-facial/non-verbal affective expression system. Some roboticists have used non-verbal sounds, tones, and/or music (Breazeal 2002), (El-Nasr & Skubic 1998), (Bartneck 2000), (Scheeff *et al.* 2000) as a secondary method of expression to provide affective expression redundancy. These robotic implementations were developed mainly for educational and/or entertainment purposes (Bethel & Murphy 2006a); however we believe that the use of non-verbal sounds, tones, and/or music can provide an effective method of affective communication for *appearance-constrained* robots used in search and rescue, military, and law enforcement applications.

Social Distance and Non-Verbal Affective Expression

In the psychology community, spatial distances between individuals socially interacting (*proxemics*) is divided into four main categories (see Figure 1): *intimate* (0.15 – 0.46 m), *personal* (0.46 – 1.22 m), *social* (1.22 – 3.66 m), and *public* (3.66 and beyond). Argyle (Argyle 1975) and Fast (Fast 1970) describe *personal* space to be the area that individuals must have surrounding them to feel comfortable, safe, and protected (Bethel & Murphy 2006a). Based on experiments conducted by Fast and Argyle, if an individual or object intrudes into the *personal* space of another it will produce a discomfort response from the person whose *personal* space was invaded (Bethel & Murphy 2006a), (Argyle 1975), (Fast 1970). Roboticists firmly support the notion that proxemics has a significant impact on the quality and comfort level of interactions between humans and robots and should be considered in the implementation of affective expressions (Bethel & Murphy 2006a), (Bickmore & Picard 2004), (Fincannon *et al.* 2004), (Dautenhahn, Ogden, & Quick 2002), (Mizoguchi *et al.* 1997), (Walters *et al.* 2005).

Our research focuses on the use of affective expression in

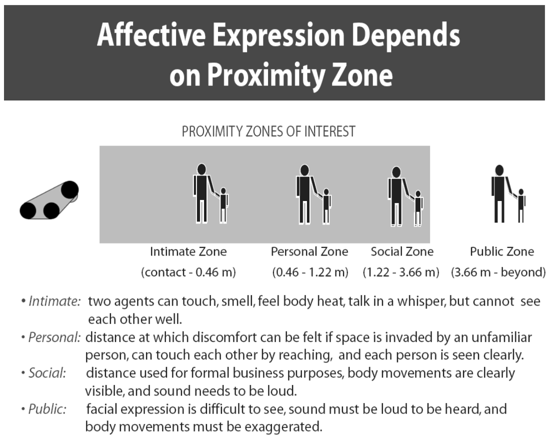

Figure 1: Proximity Zones

Appropriateness of Non-Facial and Non-Verbal Affective Expressions by Proximity Zones

Affective Expression Non-Facial and Non-Verbal	Proximity Zones		
	Intimate (contact - 0.46 m)	Personal (0.46 - 1.22 m)	Social (1.22 - 3.66 m)
Body Movement	Full body not visible at this distance	Small to medium movements	Large or exaggerated movements
Posture	Full body not visible at this distance	Postures visible at this distance	Exaggerated postures visible at this distance
Orientation	Orientation visible at this distance	Orientation visible at this distance	Orientation visible at this distance
Illuminated Color	Low intensity, dependent on small size to be visible	Medium to bright intensity	Bright intensity
Sound	Low to medium volume	Medium to loud volume, dependent on background environmental noise	Not audible, due to background environmental noise

Legend: ☐ Appropriate at this distance, ☐ May be appropriate at this distance, ■ Not appropriate at this distance

Figure 2: Appropriateness of non-facial/non-verbal affective expressions by proximity zone

non-anthropomorphic and *appearance-constrained* robots for human-robot interactions during urban search and rescue operations occurring within three meters (Bethel & Murphy 2006a). It is particularly motivated by a study conducted by Murphy, Riddle, and Rasmussen (Murphy, Riddle, & Rasmussen 2004) in using man-packable robots to act as a surrogate presence for doctors tending to victims. The study identifies how the robot will interact with the victim as one of the four major open research issues. They noted that the robots operating within three meters of the simulated victims were perceived as “creepy” and not reassuring (Bethel & Murphy 2006a). Therefore, three meters or less appears to be a reasonable distance for investigating the appropriateness of affective methods of communication such as sounds, tones, and/or music.

The use of non-verbal affective expression in non-anthropomorphic, *appearance-constrained* robots for human-robot interactions based on the appropriateness by proximity zone is a part of a multi-modal system of affective expression. As shown in Figure 2 from (Bethel & Murphy 2006b) expanded from (Bethel & Murphy 2006a), it is evident that different forms of non-facial/non-verbal methods of affective expression are more effective in different proximity zones. Body movement and posture are not effective in the intimate zone; however sounds, tones, and/or music will work well in this distance zone. Orientation is effective in any zone to indicate attentiveness and caring of those with which the robot is interacting. Illuminated colored lighting might be effective in all distance zones; however it has only been tested in the personal and social proximity zones. The use of sounds, tones, and/or music may not be effective in all proximity zones due to environmental conditions and therefore needs to be part of a multi-modal affective expression approach in robotic implementations.

The difference in the Sound entries in Figure 2 for the *personal* and *social* zones is in the likelihood of appropriateness of use. It is more likely that sounds, tones, and/or music would be audible in the *personal* zone versus the *social* zone. It is possible to use sounds in the *social* zone; however those sounds would need to be either loud to be heard or the environment would need to be extremely quiet which is not typical in robotic applications. Sound will be most appropriate in the *intimate* zone and likely appropriate for use in the *personal* zone depending on the environmental conditions.

Survey of Psychology Investigations into Auditory Affective Expression

The use of non-verbal sounds, tones, and/or music as a means for expressing affect has not been thoroughly investigated by the psychology community. Psychologists have focused in three primary directions: (1) sound related to animal communication (Argyle 1975), (Buck 1984); (2) vocal patterns and tone of voice in humans (Norman 2004), (Nass & Brave 2005), (Scherer 1979); and (3) expression of emotion through music and the associated physiological responses (Bartneck 2000), (Balkwill, Thompson, & Matsunaga 2004), (Schubert 2004), (Bartlett 1996), (Juslin 1997), (Khalfa *et al.* 2002), (Scherer & Oshinsky 1977), (Nass & Brave 2005).

The non-verbal use of sound has been investigated by psychologists as it relates to animal communication (Argyle 1975), (Buck 1984) as a primary method of conveying affect between animals. For example, in primates, distress is indicated by a screeching sound, growling is used to convey anger, and soft grunts are used to keep touch with those in close proximity while traveling (Argyle 1975).

The research that has been conducted on non-verbal sounds related to affective expression in humans is primarily focused on vocal patterns and tone of voice (Argyle 1975), (Buck 1984), (Norman 2004). Norman (Norman 2004) indicates that vocal patterns and tone can be used to express affect and indicates that these patterns and tones can be interpreted even if the literal meaning is not understood. He also observes that pets can interpret a person’s affective state based on their vocal patterns and tone. Scherer (Scherer 1979) indicates that recognition of a person's emotion in content and context-free judgment situations, vocal cues have been as indicative of some emotions as facial cues.

The last primary area of psychology research related to non-verbal expression of affect is the use of music to express affect. Balkwill *et al.* (Balkwill, Thompson, & Matsunaga 2004) discusses that there is a strong association between music and emotions. Their quantitative results indicate that it is possible to detect emotions such as joy, anger, and sadness presented through the perception of acoustic cues in familiar and unfamiliar music transcending cultural boundaries (Balkwill, Thompson, & Matsunaga 2004). Juslin (Juslin 1997) discusses that music is a powerful method of emotional and non-verbal communication because it both induces and expresses emotion. Khalfa *et al.* (Khalfa *et al.* 2002) and Bartlett (Bartlett 1996) stress that music is a powerful elicitor of emotion and can induce responses from the autonomic nervous systems such as: heart and pulse rates, respiration rates, skin temperature, and blood oxygen levels. Some of the acoustic cues that are used to determine emotion in music ranked by associative strength are: amplitude variation, pitch variation, pitch contour, pitch level, tempo, envelope, and filtration cutoff (Scherer & Oshinsky 1977).

Robotic Implementations Using Non-Verbal Affective Expressions

The robotics community has focused on non-verbal affective methods of communication in a limited number of implementations (Breazeal 2002), (Bartneck 2000), (El-Nasr & Skubic 1998), (Scheeff *et al.* 2000). Sounds, tones, and/or music were utilized as secondary methods of affective expression for redundancy purposes. Kismet (Breazeal 2002), the robot developed by Cynthia Breazeal, uses expressive utterances that are child-like in nature to reinforce its personality and the emotions of: anger, fear, disgust, happiness, surprise, and sorrow. Scheeff *et al.* (Scheeff *et al.* 2000) utilizes a sound similar to muffled speech combined with a French horn to reinforce the emotional state of their robot Sparky. The sounds that Sparky makes fall somewhere between the beeps and chirps of Star Wars' robot R2-D2 and a human voice; however this was found to be confusing and unappealing to those interacting with Sparky. Bartneck (Bartneck 2000) used auditory stimuli on a computer avatar to express emotion; however his results concluded that auditory stimuli alone was not as convincing of affective expression as using visual (facial expression) or auditory/visual stimuli (sound/music and facial expression). El-Nasr and Skubic (El-Nasr & Skubic 1998) used a mobile robot that would make a crying sound to express pain, a shattering voice to express fear, and unspecified noise to express anger; however these sounds were used as reinforcement to the robot movements for affective expression.

Ben Burtt (Burtt 2001), the sound designer for the Star Wars films, discusses the difficulty in expressing emotions using only non-verbal sounds. His biggest challenge was in creating emotional warmth in the beloved R2-D2 robot. In the "Behind the Sounds" section of (Burtt 2001), Burtt discusses that "All sounds bring with them an association with something emotional. Music is the perfect example." He began his development of R2-D2's vocalizations with electronic beeps and tonalities; however he found that these sounds were mechanical and lacked warmth and aliveness. In order to produce the emotional warmth and the qualities that many love about R2-D2 he ended up "intercutting" an imitation of baby babbling, and the whistling sound of blowing through a small flexible plumbing tube with the electronic synthesized beeps and chirps. This combination of sounds and tones gave R2-D2 the emotional warmth and aliveness that he wanted to produce. Although that is a film application, the concepts need to be considered when implementing sounds, tones, and/or music to express affect in robotic systems.

Current Work

Our approach is to use a combination of sounds, tones, and/or music to express affect in our *appearance-constrained* search and rescue robots. These non-verbal methods of communication are only part of a multi-modal affective expression system since sounds, tones, and/or music are only appropriate for use in the *intimate* and possibly *personal* proximity zones in the search and rescue environment. Sounds, tones, and/or music developed for the search and rescue application would be applicable to any robotic system to express affect.

It appears that if sounds, tones, and/or music are correctly used and implemented on robots, they could be valuable tools for affective expression. Due to the fact that search and rescue robotic applications may occur internationally, the use of sounds, tones, and/or music can be used and understood across cultures when verbal communication may not be viable. Additionally, these robots may interact with victims who may have brain injuries in which sounds, tones, and/or music may be interpreted, but language processing may not be possible.

This work focuses on the use of synthesized sounds, tones, and/or music because we are not trying to make these functional search and rescue robots anthropomorphic. The use of synthesized sounds similar to R2-D2's beeps, whistles, and sounds would be recognizable and allow the victim to realize that assistance is near, but is in the form of a mechanical-type device. These sounds and tones need to be a blend of sounds that express warmth and emotionality. We are currently going through a process similar to that experienced by Burtt (Burtt 2001) in the creation of the emotional sounds for R2-D2. This requires trial and error to develop the right blend of sounds and tones to express the desired emotion. It is a matter of finding the right musical expressions and sounds to convey an emotional quality to the robot system.

A limitation that we have experienced in this process is the sound quality of the robot's speaker. Sounds developed are distinctly different when played through the speaker. In the case of these functional search and rescue robots, it is not possible to add external, better quality speakers to the system. This requires determining the best sound and tone sequence for the particular hardware involved, which is a trial and error process of development.

Future Work

Sounds and music can elicit physiological responses in the body that can be measured with the proper equipment. This can make determining the appropriateness of an emotional expression using sounds, tones, and/or music more easily verifiable. We believe that by using self-assessments and physiological measurements, it is possible to determine a viable mapping between sounds, tones, and/or music to specific emotions.

Human testing will be required to determine the appropriateness of using non-verbal methods of affective expression by proximity zone. This testing will be conducted using only non-verbal sounds, tones, and/or music in addition to testing the appropriateness of this method as part of a multimodal system of affective expression. These tests will include videotaping interactions with the robot and an experimental participant as well as having the participant complete a self-assessment questionnaire. Additionally, physiological responses of the participant will be monitored to determine if the video and self assessments correlate with the physiological measurements obtained through human studies.

Conclusion

Although non-verbal methods for affective expression cannot be the sole means of communicating affect, they can be appropriate when implemented as part of a multi-modal affective system. Based on (Bethel & Murphy 2006a), sounds, tones, and/or music would not be appropriate to use in the *social* zone; however would likely be appropriate for use in the *personal* and *intimate* zones. The robot systems we are designing sounds, tones, and/or music for affective expression are *appearance-constrained*, but the results will be applicable to other robot implementations. In search and rescue applications, communications are sometimes difficult and the use of verbal communication is not always possible; therefore the development of non-verbal methods of communication is essential.

Additionally, search and rescue robots must be socially interactive and communicate emotion without depending on a human-like form. The goal is not to anthropomorphize the robots used in this application, but rather to make a non-anthropomorphic robot sociable, to interact with victims, to keep them calm, and to serve as a companion until help can arrive. These robots are “creepy” in appearance and our goal is to improve the comfort level of those interacting with them. We will determine which sounds, tones, and/or music can be beneficial to victims especially in the *intimate* and *personal* proximity zones.

Acknowledgment

This work is supported in part under a National Science Foundation Graduate Research Fellowship – Award Number DGE – 0135733.

References

Argyle, M. 1975. *Bodily communication*. London: Methuen.

Balkwill, L.-L.; Thompson, W. F.; and Matsunaga, R. I. E. 2004. Recognition of emotion in japanese, western, and hindustani music by japanese listeners. *Japanese Psychological Research* 46(4):337–349.

Bartlett, D. L. 1996. Physiological responses to music and sound stimuli. In *Handbook of Music Psychology*, volume 1996. San Antonio: IMR Press. 343 – 385.

Bartneck, C. 2000. *Affective Expressions of Machines*. Masters thesis, Eindhoven University of Technology.

Bethel, C. L., and Murphy, R. R. 2006a. Affective expression in appearance-constrained robots. In *1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI2006)*, 327–328. Salt Lake City, UT: ACM Press.

Bethel, C. L., and Murphy, R. R. 2006b. Non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man, Cybernetics, Part C* Under Revision.

Bickmore, T. W., and Picard, R. W. 2004. Towards caring machines. In Dykstra-Erickson, E., and Tscheligi, M., eds., *Conference on Human Factors in Computing Systems CHI '04 extended abstracts on Human factors in computing systems*, 1489–1492. Vienna, Austria: ACM Press, New York, NY, USA.

Breazeal, C. L. 2002. *Designing sociable robots*. Intelligent robots and autonomous agents. Cambridge, Mass.: MIT Press.

Buck, R. 1984. *The communication of emotion*. New York: Guilford Press.

Burtt, B. 2001. *Star Wars, Galactic Phrase Book & Travel Guide*. New York: Ballantine Publishing Group.

Caamero, L. D., and Fredslund, J. 2000. How does it feel? emotional interaction with a humanoid lego robot. In Dautenhahn, K., ed., *AAAI 2000 Fall Symposium - Socially Intelligent Agents: The Human in the Loop*, volume Technical Report FS-00-04, 23–28. AAAI, Menlo Park, CA.

Dautenhahn, K.; Ogden, B.; and Quick, T. 2002. From embodied to socially embedded agents - implications for interaction-aware robots. *Cognitive Systems Research* 3(3):397.

El-Nasr, M. S., and Skubic, M. 1998. A fuzzy emotional agent for decision-making in a mobile robot. In *The 1998 IEEE International Conference on Fuzzy Systems Proceedings, IEEE World Congress on Computational Intelligence*, volume 1, 135 – 140.

Fast, J. 1970. *Body language*. New York: M. Evans; distributed in association with Lippincott.

Fincannon, T.; Barnes, L. E.; Murphy, R. R.; and Riddle, D. L. 2004. Evidence of the need for social intelligence in rescue robots. In *IEEE/RSJ International Conference on International Conference on Intelligent Robots and Systems (IROS 2004)*, volume 2, 1089 – 1095.

Fong, T.; Nourbakhsh, I.; and Dautenhahn, K. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42(3-4):143.

Juslin, P. N. 1997. Emotional communication in music performance: A functionalist perspective and some data. *Music Perception* 14(4 [Summer 1997]):383 – 418.

Khalfa, S.; Isabelle, P.; Jean-Pierre, B.; and Manon, R. 2002. Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters* 328(2):145 – 149.

Mizoguchi, H.; Sato, T.; Takagi, K.; Nakao, M.; and Hatamura, Y. 1997. Realization of expressive mobile robot. In *1997 IEEE International Conference on Robotics and Automation*, volume 1, 581 – 586.

Murphy, R. R.; Riddle, D.; and Rasmussen, E. 2004. Robot-assisted medical reachback: a survey of how medical personnel expect to interact with rescue robots. In *13th IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2004.*, 301 – 306.

Nass, C. I., and Brave, S. 2005. *Wired for Speech: how voice activates and advances the human-computer relationship*. Cambridge, Mass.: MIT Press.

Norman, D. A. 2004. *Emotional design: why we love (or hate) everyday things*. New York: Basic Books.

Scheeff, M.; Pinto, J.; Rahardja, K.; Snibbe, S.; and Tow, R. 2000. Experiences with sparky, a social robot. In *Proceedings of the Workshop on Interactive Robot Entertainment*.

Scherer, K. R., and Oshinsky, J. S. 1977. Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion* 1(4):331 – 346.

Scherer, K. R. 1979. Nonlinguistic vocal indicators of emotion and psychopathology. In Izard, C. E., ed., *Emotions in Personality and Psychopathology*. New York: Plenum Press. 495 – 529.

Schubert, E. 2004. Modeling perceived emotion with continuous musical features. *Music Perception* 21(4):561 – 585.

Walters, M. L.; Dautenhahn, K.; te Boekhorst, R.; Kheng Lee, K.; Kaouri, C.; Woods, S.; Nehaniv, C.; Lee, D.; and Werry, I. 2005. The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In *IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005*, 347 – 352.