Robot Nonverbal Communication as an AI Problem (And Solution)

Henny Admoni and Brian Scassellati
Department of Computer Science
Yale University
New Haven, CT 06520
{henny,scaz}@cs.yale.edu

Abstract

In typical human interactions, nonverbal behaviors such as eye gazes and gestures serve to augment and reinforce spoken communication. To use similar nonverbal behaviors in human-robot interactions, researchers can apply artificial intelligence techniques such as machine learning, cognitive modeling, and computer vision. But knowledge of nonverbal behavior can also benefit artificial intelligence: because nonverbal communication can reveal human mental states, these behaviors provide additional input to artificial intelligence problems such as learning from demonstration, natural language processing, and motion planning. This article describes how nonverbal communication in HRI can benefit from AI techniques as well as how AI problems can use nonverbal communication in their solutions.

Introduction

In human-human interactions, nonverbal behaviors (NVBs) like eye gazes (Argyle 1972) and gestures (McNeill 1992) play an important role in communication. NVBs reveal mental states, reinforce spoken communication, and provide additional information beyond what is being said (Goldin-Meadow 1999).

In human-robot interactions, NVBs play a similarly important role in supporting and augmenting communication. Robot NVBs can improve the fluency of human-robot conversations (Andrist et al. 2014; Kirchner, Alempijevic, and Dissanayake 2011; Mutlu et al. 2012), increase the teaching effectiveness of robot tutors (Huang and Mutlu 2013; Saerbeck et al. 2010; Szafir and Mutlu 2012), and improve the efficiency of human-robot physical collaborations (Admoni et al. 2014; Breazeal et al. 2005; Huang and Thomaz 2011; Huang and Mutlu 2013; Lockerd and Breazeal 2004; Zheng et al. 2015).

Generating effective robot NVB is challenging, however. The success of specific NVBs depends greatly on context, and the wrong kind of NVB can actually serve to hinder interactions (Choi, Kim, and Kwak 2013; Wang and Gratch 2010).

To address the challenge of generating effective NVBs, researchers use artificial intelligence techniques such as machine learning and cognitive modeling. The following section of this paper describes some of these AI techniques and their benefit to HRI.

Because NVBs are so integral to human interaction, they also provide a channel of information that can help address AI problems that involve direct interactions with people. The final section of this paper describes a few of these AI problems that can benefit from the incorporation of NVB input.

AI Techniques for Robot Nonverbal Behavior

AI has developed tools to build intelligent machine behavior. These tools can be successfully applied to generating and recognizing NVBs in human-robot interactions.

Learning Many researchers have approached the problem of generating NVBs for robots by basing them on observed human behavior. Data from annotated human-human interactions are inputs to machine learning algorithms such as SVMs, decision trees, and Bayesian networks. Such data-driven models have been shown to successfully enable human-robot conversation (Andrist et al. 2014; Liu et al. 2012), support robot narration (Huang and Mutlu 2014), and inform object handovers (Strabala et al. 2013).

Similar learning techniques can be used to interpret human behavior in human-robot interactions. By modeling NVBs and their communicative intent in human-human interactions, systems can predict the context of a human partner's communicative actions (Admoni et al. 2014) and identify when a robot should offer assistance (Sakita et al. 2004).

Machine learning is also useful when taking a developmental approach to robotics. For example, a learning model that is given demonstrations of attention to salient objects can develop the ability to perform joint attention (Nagai et al. 2003), a key nonverbal behavior.

Cognitive Modeling People's NVBs reveal their knowledge, goals, and intentions—their *mental state*. In AI, cognitive modeling builds systems that mimic human mental states. By linking NVBs to the underlying cognitive models, robots can generate NVBs that communicate their own mental states to a human partner.

For example, using the ACT-R/E cognitive architecture, a robot can engage in a multi-party conversation by switching its visual attention to the current speaker in a natural way (Trafton and Bugajska 2008). By tightly integrating gaze behaviors with an underlying cognitive model that controls reasoning, dialogue management, and goals, the Rickel Gaze Model generates real-time gaze shifts that provide insights into an agent's internal processes (Lee et al. 2007). Understanding a robot's mental states leads to more efficient human-robot interactions (Breazeal et al. 2005).

Perception Robots can use visual perception to guide their generation of NVBs in human-robot interactions. Human attention is dependent on low-level visual features of the environment (like color and intensity) together with high-level contextual information about the task at hand (Itti and Koch 2001). Neurobiological computer vision models that mimic human visual attention yield human-like gaze behaviors when looking at a variety of visual scenes (Itti, Dhavale, and Pighin 2004). Combining low-level visual feature maps with high-level motivational information creates a behavior model that realistically directs a robot's attention based on its current task (Breazeal and Scassellati 1999). Robots can use biological vision models to perform shared attention and gaze imitation, enabling cognitive learning from a human teacher (Hoffman et al. 2006).

Nonverbal Behavior for AI Problems

Social robots become increasingly useful as they move out of the lab and into natural human environments like homes and schools. But interacting socially in dynamic, unpredictable real-world environments requires real time intelligence, including the ability to perform natural language processing, to learn from demonstrations, and to plan motions.

Nonverbal communication is a subtle, multimodal channel that can be used to augment and support intelligent behavior during a human-robot interaction. As an additional input or as an additionally expressive motor output, NVBs like eye gaze and gestures can simplify other AI problems.

Natural Language Processing Researchers have developed a model for understanding natural language commands to robots performing navigation and manipulation (Tellex et al. 2011). The model first grounds the components of the natural language command to specific objects, locations, or actions in the environment. This grounding operates exclusively on verbal inputs. Incorporating referential eye gaze into the grounding model would potentially increase the confidence of symbol groundings by providing additional, multi-modal command input.

Eye gaze could disambiguate between two similar objects. For example, if there are two available groundings for the word "truck," incorporating eye gaze into the model for the command "put the pallet on the truck" clarifies the reference without needing additional spatial speech such as "the one on the left." This increases efficiency by requiring fewer verbal commands from the user and less language processing from the system.

Eye gaze knowledge could also increase the speed (and thereby the efficiency) of the interaction. Because people naturally fixate on objects about one second before they verbally reference them (Griffin and Bock 2000), gaze could be used for pre-processing, allowing the system to eliminate some potential groundings before the whole command is even received.

Learning from Demonstration Learning from demonstration (LfD) is an approach to machine learning in which a robot develops a policy for how to complete a task by watching demonstrations of that task being performed (Atkeson and Schaal 1997). LfD has been used widely to train robots in numerous domains (Argall et al. 2009).

Expressive robot NVBs have already been found to improve LfD. Robot NVBs provide feedback to human teachers, revealing the robot's knowledge and focus of attention (Lockerd and Breazeal 2004). People are sensitive to the robot's mental states when they are teaching it, and will adjust their behavior (in terms of pauses, speed, and magnitude of motions) to account for the robot's visual attention (Pitsch, Vollmer, and Mühlig 2013). This subtle but natural feedback mechanism leads to teaching that has fewer errors, faster recovery from errors, and less repetition of material (Huang and Thomaz 2011).

Future LfD systems will benefit from continuing to incorporate NVBs. Robot systems that do so will closely tie gaze and gesture behaviors to knowledge, intentions, and goals. For example, integrating a NVB controller into a logical planner will allow a robot to fixate its gaze on an object it is currently reasoning about, transparently reveal its intentions to a human partner.

Legibility and Predictability of Motion When collaborating with a robot, it is important that a robot's motion clearly reflect its intentions and future action. Legibility and predictability of a robot's motion trajectories can be mathematically defined (Dragan, Lee, and Srinivasa 2013). The equations for legibility and predictability model the user's inferences between motion trajectories and goal locations.

People use gaze behavior to perform similar inferences about where a collaborator will reach. In collaborations, people can recognize and respond to eye gaze that indicates spatial references, successfully predicting the target of their partner's reference (Admoni et al. 2014; Boucher et al. 2012). Expressive nonverbal behavior that reveals mental states makes cooperative task performance faster, with errors detected more quickly and handled more effectively than purely task-based nonverbal communication (Breazeal et al. 2005).

Incorporating eye gaze into equations for predictability and legibility would allow robots to take advantage of this natural, subtle, communicative behavior. Combining eye gaze with motion trajectories would generate multimodal behavior that is even more communicative than motion trajectories alone.

Acknowledgments

This work was supported by NSF grant #113907 and ONR grant #N00014-12-1-0822.

References

Admoni, H.; Dragan, A.; Srinivasa, S.; and Scassellati, B. 2014. Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In *HRI 2014*. ACM.

Andrist, S.; Tan, X. Z.; Gleicher, M.; and Mutlu, B. 2014. Conversational gaze aversion for humanlike robots. In *HRI 2014*. ACM.

Argall, B. D.; Chernova, S.; Veloso, M.; and Browning, B. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57(5):469–483.

Argyle, M. 1972. Non-verbal communication in human social interaction. In Hinde, R. A., ed., *Non-verbal communication*. Oxford, England: Cambirdge University Press.

Atkeson, C. G., and Schaal, S. 1997. Robot learning from demonstration. In *ICML 1997*.

Boucher, J.-D.; Pattacini, U.; Lelong, A.; Bailly, G.; Elisei, F.; Fagel, S.; Dominey, P. F.; and Ventre-Dominey, J. 2012. I Reach Faster When I See You Look: Gaze Effects in Human-Human and Human-Robot Face-to-Face Cooperation. *Frontiers in neurorobotics* 6(May):1–11.

Breazeal, C., and Scassellati, B. 1999. How to build robots that make friends and influence people. In *IROS 1999*, 858–863.

Breazeal, C.; Kidd, C.; Thomaz, A.; Hoffman, G.; and Berlin, M. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *IROS 2005*, 708–713. IEEE.

Choi, J. J.; Kim, Y.; and Kwak, S. S. 2013. Have you ever lied?: the impacts of gaze avoidance on people's perception of a robot. In *HRI 2013*, 105–106.

Dragan, A. D.; Lee, K. C.; and Srinivasa, S. S. 2013. Legibility and predictability of robot motion. In *HRI 2013*, 301–308.

Goldin-Meadow, S. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences* 3(11):419 – 429.

Griffin, Z. M., and Bock, K. 2000. What the Eyes Say About Speaking. *Psychological Science* 11(4):274–279.

Hoffman, M. W.; Grimes, D. B.; Shon, A. P.; and Rao, R. P. 2006. A probabilistic model of gaze imitation and shared attention. *Neural Networks* 19:299–310.

Huang, C.-M., and Mutlu, B. 2013. The Repertoire of Robot Behavior: Designing Social Behaviors to Support Human-Robot Joint Activity. *Journal of Human-Robot Interaction* 2(2):80–102.

Huang, C.-M., and Mutlu, B. 2014. Learning-based modeling of multimodal behaviors for humanlike robots. In *HRI 2014*, 57–64. ACM.

Huang, C.-M., and Thomaz, A. L. 2011. Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *RO-MAN 2011*, 65–71.

Itti, L., and Koch, C. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience* (2):194–203.

Itti, L.; Dhavale, N.; and Pighin, F. 2004. Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention. In Bosacchi, B.; Fogel, D. B.; and Bezdek, J. C., eds., *SPIE*, volume 5200, 64–78.

Kirchner, N.; Alempijevic, A.; and Dissanayake, G. 2011. Nonverbal robot-group interaction using an imitated gaze cue. *HRI 2011* 497.

Lee, J.; Marsella, S.; Traum, D.; Gratch, J.; and Lance, B. 2007. The Rickel Gaze Model: A Window on the Mind of a Virtual Human. In *IVA*, volume LNAI 4722, 296–303. Springer-Verlag.

Liu, C.; Ishi, C. T.; Ishiguro, H.; and Hagita, N. 2012. Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In *HRI 2012*, 285–292. ACM.

Lockerd, A., and Breazeal, C. 2004. Tutelage and socially guided robot learning. In *IROS 2004*, volume 4, 3475–3480.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: The University of Chicago Press.

Mutlu, B.; Kanda, T.; Forlizzi, J.; Hodgins, J.; and Ishiguro, H. 2012. Conversational gaze mechansims for humanlike robots. *ACM Transactions on Interactive Intelligent Systems* 1(2).

Nagai, Y.; Hosoda, K.; Morita, A.; and Asada, M. 2003. A constructive model for the development of joint attention. *Connection Science* 15(4):211–229.

Pitsch, K.; Vollmer, A.-L.; and Mühlig, M. 2013. Robot feedback shapes the tutor's presentation: How a robot's online gaze strategies lead to micro-adaptation of the human's conduct. *Interaction Studies* 14(2):268–296.

Saerbeck, M.; Schut, T.; Bartneck, C.; and Janse, M. D. 2010. Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *CHI 2010*, 1613–1622.

Sakita, K.; Ogawara, K.; Murakami, S.; Kawamura, K.; and Ikeuchi, K. 2004. Flexible cooperation between human and robot by interpreting human intention from gaze information. In *IROS 2004*.

Strabala, K.; Lee, M. K.; Dragan, A.; Forlizzi, J.; Srinivasa, S. S.; Cakmak, M.; and Micelli, V. 2013. Toward Seamless Human-Robot Handovers. *Journal of Human-Robot Interaction* 2(1):112–132.

Szafir, D., and Mutlu, B. 2012. Pay attention! Designing adaptive agents that monitor and improve user engagement. In *CHI 2012*, 11–20.

Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M. R.; Banerjee, A. G.; Teller, S.; and Roy, N. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. In *AAAI 2011*.

Trafton, J., and Bugajska, M. 2008. Integrating Vision and Audition within a Cognitive Architecture to Track Conversations. In *HRI 2008*, 201–208.

Wang, N., and Gratch, J. 2010. Don't just stare at me! In *CHI 2010*, 1241–1249. Atlanta, GA USA: ACM.

Zheng, M.; Moon, A.; Croft, E. A.; and Meng, M. Q.-H. 2015. Impacts of Robot Head Gaze on Robot-to-Human Handovers. *International Journal of Social Robotics*.