Understanding the emotion in the context of a conversation is an essential skill for robots to develop. This project aims to develop a service that helps robots recognize acoustic patterns in tone of voice by classifying approving, neutral and prohibitive affect. Prosodic Therapy (patterns of rhythm and speech) has been proven to aid in social inclusion for high functioning children with autism. Infant directed speech is a good source of exaggerated affective prosody for encouraging and prohibiting voice tones

Affective Prosody Recognition for Human-Robot Interaction

http://cs.yale.edu/~eliskim

Yale University

Brian Scassellati

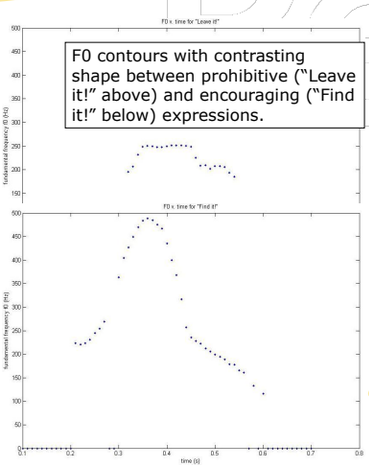
When I yell "Leave it!" my dog Jay Jay instantly drops whatever is in his mouth. If I encourage him to "Find it!" his attention perks up, and he searches for a snack or toy. Domesticated pets are responsive to people's *affective* vocalizations. We readily use this intuitive communication mechanism with our pets, with each other, and especially with our children.

Robots are emerging that interact with everyday people in play, work, caregiving and hazardous situations. Pleo plays tug of war with its caretakers' fingers, Roomba vacuums homes, disaster-relief robots will provide companionship to trapped victims, and intelligent wheelchairs promise to endow their users with greater independence from human caregivers. However, these robots understand neither what is said nor the important affective content of a shouted "no" or a whispered "quiet."

In this project, we aim to make robots sensitive and responsive to untrained, affective vocalizations, by recognizing acoustic patterns in our *speech prosody*, or tone of voice.

Prosodic therapy for children with autism

One of the primary application areas driving our current research is the use of social robots to aid in the training of appropriate prosodic understanding and production in children with autism spectrum disorder. Individuals with autism often express affect with atypical prosodic intonation and have difficulty in understanding the important affective content of stress, prohibition, and praise in spoken utterances. These stereotypical vocal patterns are often used during instruction, and lack of prosodic comprehension is a barrier to social inclusion.[1] Robots have been particularly promising tools for therapeutic intervention as they often engender a high degree of motivation, prolonged interactions, and at times provoke novel social responses from children with autism. Our current pilot studies have used a commercial robot (Pleo, shown above) in interactions with high-functioning children with autism spectrum disorder aged 9-13. The children must use an encouraging tone of voice to have the robot move across areas of "water" which the robot fears. Children in these trials have shown increased appropriate prosodic production with the robot trials as opposed to equal sessions with a human instructor.



F0 contours with contrasting shape between prohibitive ("Leave it!" above) and encouraging ("Find it!" below) expressions.

Affect classification from speech prosody

In infant-directed speech, which features exaggerated affective prosody, characteristic pitch contours have been observed for approving and prohibitive expression.[2] When speaking to robots, speakers have been observed to use prosody exaggerated in affective expression similar to that used in infant-directed speech.[3] Statistical work has achieved up to 97% accuracy in recognizing prohibitive prosody, and anywhere from 57% to 92% accuracy over five affect classes (including soothing and attention-seeking), depending on the genders and number of speakers analyzed, with higher accuracy for female speakers and for fewer speakers.[3456] Machine learning algorithms have been demonstrated to learn from prosodic affect decisions as input.[6]

Affective prosody recognition for everyday or therapeutic use will require the development of more portable and robust classifiers, as well as a finer understanding of differences in prosodic expression among application domains.

Corpora for prosodic instruction

The goal of this project is to implement a service for Microsoft Robotics Studio which classifies approving, neutral, and prohibitive affect in robot-directed speech. Our software will be efficient enough to run on platforms with low computational power, including handheld computers. To further that goal, we have collected a large corpora of affective prosody spoken to Pleo robots by healthy adults engaged in a training exercise. This work, to be presented at HRI 2009, identified variations in the natural utterances that humans use when training robots with varied histories of success at a given task.[7] The number of words spoken to express positive affect and the intensity of the prosodic feedback varies with a robotic learner's performance, in contrast to expectations commonly made by machine learning techniques. This corpus will be an essential tool in building multi-platform recognition systems compatible with Microsoft Robotics Studio.

Yale

LUX ET VERITAS

References

1. Paul, R., Shriberg, L., McSweeney, J., Cicchetti, D., Klin, A., & Volkmar, R. (2005). Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 35, 861 – 869.
2. A. Fernald, "Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages" *Child Development*, 64, 657-67, 1993.
3. C. Breazeal and L. Aryananda. "Recognizing affective intent in robot directed speech," *Autonomous Robots*, vol. 12, no. 1, pp. 83-104, 2002.
4. A. Robinson-Mosher and B. Scassellati, "Prosody recognition in male infant-directed speech," in *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sendai, Japan, 2004.
5. M. Slaney and G. McRoberts, "Baby ears: a recognition system for affective vocalizations," *Speech Communication*, vol. 39, no. 3-4, pp. 367-384, Feb. 2003.
6. E. S. Kim and B. Scassellati, "Learning to refine behavior using prosodic feedback," in *Proceedings of the 2007 IEEE International Conference on Development and Learning (ICDL)*, London, 2007.
7. E. S.. Kim, D. Leyzberg, K. Tsui, and B. Scassellati, "How People Talk When Teaching a Robot," in *Proceedings of the ACM SIGCHI/SIGART Human-Robot Interaction Conference (HRI)*, San Diego, 2009.