Enabling a Robot to Understand a Human

Anja Austermann
The Graduate University for Advanced Studies (SOKENDAI), Japan

Seiji Yamada
National Institute of Informatics
The Graduate University for Advanced Studies (SOKENDAI), Japan

Goal: Adapting a robot to a user’s preferred way of giving commands and feedback

Method:
Biologically inspired learning method
Special computerized training tasks for adaptation:
Training tasks allow the robot to provoke commands and feedback from the user:
1) Robot and user know state of training task
2) Robot and user understand which actions are good/bad
3) Robot provokes and learns commands / feedback by making good or bad moves and prompting for instruction

Learning Method
Combining Hidden Markov Models and Classical Conditioning in a two-staged algorithm
1) Stimulus Encoding Stage: Training HMMs for speech and prosody and using duration model for touch
   Speech: Create utterance models from phoneme models to cluster similar utterances.
   Feedback/Object/Places: Represent as one HMM per symbol
   Command-Patterns: Determine parts of utterance that do not belong to expected object/place names. Train HMM(s) and model grammar with slots for objects/places
   Prosody: Based on features from pitch/energy contours and frequency spectrum
   Touch: Head and back touch sensors: < 0.5 sec | between 0.5 and 1 sec | > 1 sec
2) Associative Learning Stage: Associating HMMs with symbolic meanings
   Using the Rescorla-Wagner model of classical conditioning:
   Unconditioned Stimuli: Existing symbolic representations of commands, objects/places and positive/negative feedback
   Conditioned Stimuli: Encoded stimuli from user represented by HMMs

Training Tasks
Four different training tasks to provoke positive/negative feedback:
1) Finding a sample Image
2) Playing the game “Pairs”
3) Playing “Connect Four” against the computer
   * 4) Learning “dog commands”

Experiments
First application: Learning positive/negative user feedback
10 participants (ages 23-47)
5.5 hours of audio/video data

Experimental Setting

Results
95.97% recognition accuracy for distinguishing between positive/negative feedback based on speech, prosody and touch
Considerable improvement through multimodal integration. Single modality recognition accuracies:
- Speech: 83.53%
- Prosody: 84.27%
- Touch: 88.17%