A Savvy Robot Standup Comic: Online Learning through Audience Tracking

Heather Knight
Carnegie Mellon University
Robotics Institute
5000 Forbes Ave
Pittsburgh, PA 15213 USA
heatherbot@cmu.edu

Scott Satkin
Carnegie Mellon University
Robotics Institute
5000 Forbes Ave
Pittsburgh, PA 15213 USA
ssatkin@andrew.cmu.edu

Varun Ramakrishna
Carnegie Mellon University
Robotics Institute
5000 Forbes Ave
Pittsburgh, PA 15213 USA
varunr@cmu.edu

Santosh Divvala
Carnegie Mellon University
Robotics Institute
5000 Forbes Ave
Pittsburgh, PA 15213 USA
santosh@ri.cmu.edu

TEI 2011, January 22-23, 2011, Funchal, Portugal.
ACM 978-1-60558-930-5/10/04.

Abstract
In this paper, we propose Robot Theater as novel framework to develop and evaluate the interaction capabilities of embodied machines. By using online-learning algorithms to match the machine"s actions to dynamic target environments, we hope to develop an extensible system of social intelligence. Specifically, we describe an early performance robot that caters its joke selection, animation level, and interactivity to a particular audience based on real-time audio-visual tracking. Learning from human signals, the robot will generate performance sequences on the fly.

Keywords
Audience Tracking, Human Robot Interaction, Online Learning, Entertainment Robots

ACM Classification Keywords
D.2.11. Domain-specific software architectures, J.5 Computer applications in the Fine and Performing Arts.

General Terms
Work-in-progress paper

Introduction
Robot theater is a rich new arena for developing and evaluating interaction capabilities between machines

and humans ([1][2]). It provides a constrained environment rich in data and attractive to general population feedback. By creating audience sensing technologies that can help a robot parse crowd response and developing more effective emotive, communicatory and animation capabilities for the robot itself, we hope to accelerate innovation in social robotics on and off the stage as well as to creating new forms of expression and collaboration for human performers.

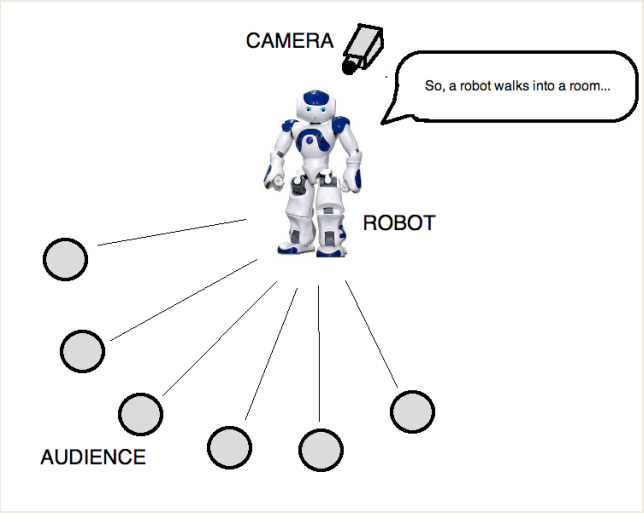

figure 1. Setup: Robot on stage with camera, facing audience

To this end, we have created a first-pass performing robot (Fig 1., based on the Nao platform, that caters its jokes, animation level and interactivity (e.g. soliciting information from the audience) to individual audiences using online learning techniques, as in [9].

Background

Storytelling is a central faucet of being human. It is common to spend much of our free time watching movies, sharing gossip and reading the newspaper. If robots are to integrate into everyday life, charisma will play a key role in their acceptance. Their embodied presence and ability to touch and move around in the physical world always us to communicate with them in novel ways which are more naturally human [4]. We believe that Robot Theater can provide a valuable stepping-stone to creating impactful robotic characters.

Various forays have been made into the topic of robotic performance, including ([1][3][5][7]). The key addition in this project is the integration of intelligent audience sensing, which allows conscious and sunconscious human behaviors to motivate live performance generation. In this process, we transform the theater into a valuable arena for interaction research.

Because of this unique approach, our work extends and begins answer Breazeal's call to action in [1]:

"The script places constraints on dialog and interaction, and it defines concise test scenarios. The stage constrains the environment, especially if it is equipped with special sensing, communication or computational infrastructure. More importantly, the intelligent stage, with its embedded computing and sensing systems, is a resource that autonomous robotic performers could use to bolster their own ability to perceive and interact with people within the environment."

Informal post-performance surveys from Knight's previous work with the robot have helped inform the

behavioral design of the current joke choreographies. In August 2010, Knight implemented the predecessor to the standup comic presented here on the same physical robot [8]. The series was called 'Postcards from New York' and displayed publicly to strangers in Washington Square Park, coinciding with professional lunch breaks. Though individual comedy sketches were preset, visitors (including students, professionals and tourists, both adult and children) were invited to choose the topic by selecting a Postcard, then showing it to the robot. Upon recognizing the Postcard, the robot would go on to perform a two-minute long sketch relating to its 'personal experiences' in that neighborhood. Viewers enjoyed its humor and physical form, but were very sensitive to sound quality. Their almost universal favorite feature was seeing the robot move, so we animate all the jokes in this series and will also amplify its sound.

System Overview

In the software we are developing now (see Fig. 2), the robot caters its jokes and joke sequence to its viewers using online learning techniques. Individual jokes have labeled attribute sets including: topic, length, interactivity, movement-level, appropriateness and hilarity. As the robot is telling a joke, we aggregate the sensor data, categorizing the total enjoyment at the end of the joke as positive or negative on a -1 to 1 scale. Using that number, the Audience Update reweights its model of what the audience likes and dislikes based on the attributes present in the last joke, increasing those weights if the response was good and visa-versa. The Joke Selector then finds a best match joke given the latest audience model, also accounting for the current story phase desired. The process iterates until the show is done.

We have a scheduled performance coming up for TEDWomen in Washington DC on December 8, 2010. As of this paper's submission date, we will direct an on-stage microphone and high-resolution camera toward the audience to collect the audio and video data. To aid the latter, we will also distribute red-green indicator paddles among attendees, which gives them an explicit modality for communication in expressing approval and

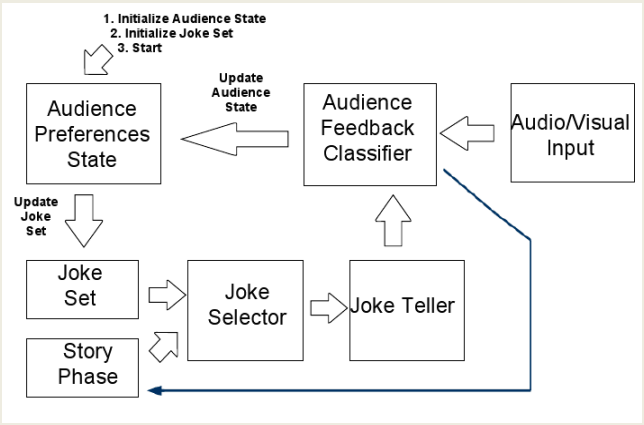

figure 2. System for Online Joke Sequencer

disapproval, or if the robot asks them a question. Previous to that, we are testing our system locally at Carnegie Mellon lecture halls, gathering audio data unobtrusively during live presentations and performances, and testing vision capabilities using ourselves and our friends.

Software Architecture

Because our application scenario is in a large lecture hall, we decided to use an external HD camera and microphone to improve the resolution of our data collection. As displayed in Fig. 3, the robot stores the

full set of jokes and corresponding animations in his head, awaiting computer command via wifi. Off-board, communication between modules is moderated by a mother python script and shared data files.

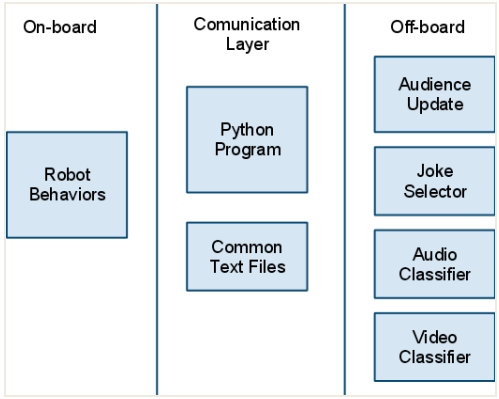

figure 3. Software Architecture

On-board Robot Behaviors

A library of possible jokes and animations are pre-loaded onto the robot. An individual joke will trigger when the robot receives its respective 'joke id' from the communication layer. If communication breaks down between jokes, there is also a timeout behavior in which it will automatically forward to a randomly picked next joke. If time permits, we could also add stalling capabilities, where it pretends to drop something or smoke a cigarette while waiting for communication to resume.

Sample Joke:
"Waiter! Waiter! What's this robot doing in my soup?" "It looks like he's performing human tasks twice as well, because he knows no fear or pain."

Audio-Visual Audience Feedback Classifier

The goal of the audience feedback classifier is to determine how much the audience is enjoying the robot comedian's performance at any given time using a weighted combination of the sensor data. This is the central module of the project, as it highlights the new possibilities for feedback and interaction in the space of robot-audience interaction and tracking. We are beginning with a small subset of all possible audience sensors, evaluating the relevant weights of each by training Support Vector Machines [10] on hand-labeled data, a technique that is extensible to various future sensors. The output of this module will be a real-valued number on the interval [-1, 1]. An output of 1 represents a "maximum enjoyment" classification. An output of -1 indicates a "minimum enjoyment" prediction.

Real-time Vision Module
The current system allows the audience to give direct yes/no feedback by holding up a colored paddle (red on one size and green on the other). During high interactivity settings, the robot might prompt the audience to answer a question that requests information (yes/no) or valence (like/dislike).

Real-time Audio Module
In contrast to the explicit response targeted above, the audio captures meaning about the audience's ambient response in the form of laughter, applause or chatter and is the highest weight feedback mode for the upcoming performance. A simple baseline metric assumes that all audio feedback from the audience is positive (i.e., only laughter and applause, no booing or heckling) and measures the duration and amplitude (volume) of the audio track to predict a level of

enjoyment. We are also computing amplitude, time and frequency spectrum statistics on test auditorium recordings to further refine our estimates.

Audience Model Update

Our current model assumes that the joke attributes are known. Thus, the purpose of this module is to use the audience's enjoyment-level of the previous joke to update the robot"s estimates of what attributes the audience likes and dislikes. In mathematical terms, we use a technique called online convex programming [10]. We use the previous audience estimate, as summed up by the weight vector *w(t),* and increase or decrease each attributes" value by multiplying the valence of the response, y, with the characteristics of the previous joke *J(t)* and a learning-rate constant *a*. Thus audience model is updated to the next timestep, *w(t+1),* using the equation below:

$$w(t+1) = w(t) + ayJ(t)$$

Joke Selector

The simplest way to select the next joke is to use the latest audience model to find joke that maximizes the "enjoyment score" based on individual joke attributes, by taking the dot product of the two. This technique aggregates all the possible aspects of each joke that the audience may like or dislike and chooses the one with the best score.

Exploration versus Exploitation

One danger of the above technique is that if the robot finds one successful set of attributes, it will not continue to try out other subjects or performance modes that the audience might also enjoy. We suspect that the audience will appreciate diversity of performance, thus, we are also incorporating a bandit algorithm strategy in which we will blindly choose a new joke, every so often, particularly toward the beginning of the set. This is called the 'epsilon-decreasing strategy' because the probability of choosing the best scored joke is 1-ε, where ε decreases over time [10].

Generating a Joke Sequence

A natural extension to the single next-joke selector is to generate a coherent joke sequence. Proposed solutions to this challenge thus far include: creating coherent joke groups and/or modeling a performance histogram. In the former, the audience model would be used to select and sequence several jokes at a time, though still exiting early in the case of dramatic failure (very low audience enjoyment score). In the latter, instead of just looking for a single best-fit joke, we look for the best-fit current story phase joke, e.g. Phase I: Grab Attention, Phase II: Interstitial and Phase III: Climax, (using a simple Markov model to decide when to transition between them).

Metrics for Success

This project presents an early realization of a robot performance with an emphasis on real-time audience tracking, personalization, and live performance generation. We will judge the system successful if it realizes the following:

- Real time integration of sensing, processing and actions.
- Domain knowledge of audience tracking feature set: exploring and discarding sensing modalities based on the added value to the system's effectiveness.
- Improved enjoyment levels for the proposed system versus a randomly-sequenced control

Conclusion and Future Work

In our case, the metaphor of robotic theater has already served to develop algorithm concepts that apply directly to everyday robotics research.

For example, we plan to adapt this software to a tour-guide robot that will individualize the sequence, style and content of its tours based on self-supervised audience tracking as part of an autonomous robots initiative here at CMU. It will incorporate online sequencing of content and some forms of spontaneous interaction, which is novel for robotic guides, although [5] does propound the importance of non-verbal interaction. Such experience validates the claim Demers makes in his survey paper motivating robot performance [2]:

"By combining AI with Theatre, new questions will be raised about how a presentation of an experiment from an AI lab can differ from a theatre presentation of the same machine. Researchers from these disciplines operate from different perspectives; art can become the "new" experimental environment for science because it the world does not only consists of physical attributes but also of intangible realities."

In follow-up projects, we hope to partner with and learn from those in the arts community. By adding a model of performance attributes to the overall system diagram, we could begin to evaluate the success of the robot's delivery. Parameters might include: tone of voice, accent, costuming, props, gestures, timing, LED illumination and pose. In this process, we are beginning to translate human behavior into rules a machine can understand. Ultimately, Knight hopes to extend these learning to full theatrical performances, deepening the understanding of characters, and even, relationships with other actors on state. The work has only just begun.

Acknowledgements

Thanks go to our Statistical Methods for Robotics Professor Drew Bagnell and TA Felix Duvallet.

References and Citations

[1] Breazeal, C., et al. Interactive Robot Theatre. *Comm. of ACM,* (2003), 76-85.

[2] Demers, L. Machine Performers: Neither Agentic nor Automatic. *HRI Workshop on Collaborations with Arts,* (2010).

[3] Hoffman, G., Kubat, R. A Hybrid Control System for Puppeteering a Live Robotic Stage Actor. In *Proc. RoMan 2008,* (2008), 1-6.

[4] Knight., H. et al. Real-time social touch gesture recognition for sensate robots. In *Proc. IROS 2009,* (2009), 3715-3720.

[5] Kobayashi, Y., et al. Museum guide robot with three communication modes. IROS 2008. (2008) 3224-3229.

[6] Lin, Chyi-Yeu, et al. The realization of robot theater: Humanoid robots and theatric performance. In *Proc. ICAR 2009*, Advanced Robotics (2009), 1-6.

[7] Murphy, R., Hooper, A., Zourntos, T. A Midsummer's Night Dream (with Flying Robots). *HRI Workshop on Collaborations with Arts,* (2010).

[8] Postcards from New York http://www.marilynmonrobot.com /?page_id=197.

[9] Sofman, B. et al. Improving robot navigation through self-supervised online learning. *Journal of Field Robotics,* (2006), Vol 23, 1059-1075.

[10] Thrun, S., Burgard, W, and Fox, D. *Probabilistic Robotics.* The MIT Press, (2005).