HERB's Sure Thing: a rapid drama system for rehearsing and performing live robot theater

Garth Zeglin Aaron Walsman Laura Herlant Zhaodong Zheng Yuyang Guo Michael C. Koval
Kevin Lenzo Hui Jun Tay Prasanna Velagapudi Katie Correll Siddhartha S. Srinivasa

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA, USA

{garthz, awalsman, mkoval, lenzo, pkv, siddh}@cs.cmu.edu
{herlant,kal}@cmu.edu
{zhaodonz,yuyangg}@andrew.cmu.edu
{tay.hui.jun}@gmail.com

***Abstract*—**
In Spring 2014, the Personal Robotics Lab at CMU collaborated with the School of Drama to develop, produce and stage a live theatrical performance at the Purnell Center for the Arts in Pittsburgh. This paper describes some of our unique experiences collaborating with drama faculty, the director and the actor. We highlight the challenges arising from theatrical performance and specifically describe some of the technical tools we developed: a bidirectional Blender interface for robot animation, an interactive system for manipulating speech prosody, and a conductor's console for online improvisation and control during rehearsal and performance. It also explores some of the remaining challenges to our goal of developing algorithms and open-source tools that can enable any roboticist in the world to create their own dramatic performance.

I. Introduction

Theatrical drama involving robot and human actors provides an opportunity to explore techniques for seamless physical and verbal collaboration. As a first step, in the Spring of 2014 we collaborated with members of the Drama department to stage a live theatrical performance using a human actress and HERB, our butler robot (described in Section III-A). This process uncovered a rich set of questions as we extended our software infrastructure to support dramatic performance. We also gauged audience reactions to understand contemporary expectations of robot performance.

We chose to adopt as many conventions of the theater as possible to maximize our ability to collaborate with drama practitioners. Our goal was to replace one human actor with a robot within a conventional play and theater. A key challenge is supporting the dynamism of the rehearsal process in which the director and actors iteratively develop the interpretation of the text. This objective encouraged the development of flexible, improvisatory tools, also in keeping with our ultimate objectives of integrating expressive behavior into robot daily life.

The specific play we performed was "Sure Thing" by David Ives [1]. This play was selected because the comedic narrative structure involves a time-rewinding device which can be interpreted as a depth-first search through a dialogue. The play is very dialogue-intensive with minimal overall movement across the stage. This led to an emphasis on conversational gesturing in combination with prosodic speech synthesis.

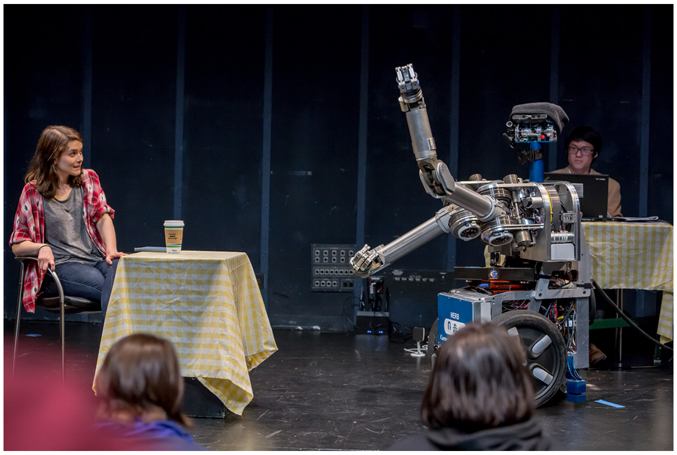


Fig. 1. HERB performing on stage with Olivia Brown. Don Zheng is the robot operator, posing as a cafe patron in the background. Photo credit: Michael Robinson.

The emphasis on rehearsal participation meant that extensive conversation was required with the director and actor while developing the performance, so we decided early on to include a human operator rather than focus on autonomy. The need for rapid rehearsal experimentation also prompted us to develop a motion graph approach which balances flexibility with the limitations of real-time operator input.

Our ultimate aim is not to replace actors but to understand how body movement and prosodic speech create an understanding of intention within a narrative. This can guide robot designers toward techniques for infusing the practical motions of everyday robot life with expressive gestures which convey the intention of the robot. With these, robots can not only perform daily tasks such as cleaning our houses, but move in a way which conveys the narrative of their purpose.

II. Related Work in Robot Drama

Machinery has a long history in theater, but recently a number of projects have used robots as machines in an acting rather than a scenic role. This project shares many individual traits with these performance projects. A motivation in common is the use of theater as a means of exploring human-robot social interaction [2] [3] [4] [5] [6]. We have chosen a canonical play [7] involving one robot and one human [8] and performed it on repurposed research hardware [7] [8] [9] [6] [10] combined with a human operator [7], emphasizing expression through physical animation [5] [6] but also incorporating dialogue and humor [10] [4], with a goal of developing the performance via rehearsal with humans [9]. We chose to substitute a robot for a human actor in a conventional setting rather than create a complete theater system [11] [12] or a non-narrative system oriented toward direct audience interaction [3].

Other work has focused more on specialized robots engineered for drama including marionettes [13], wheeled and legged humanoids [14], and highly articulated facial features [14]. We have not emphasized full autonomy [2] [3] [14] or automated interpretation of cues [15]. This sampling of projects should hint at the breadth of assumptions to be made in a scenario as complex as a dramatic production.

III. Objectives and Challenges

Our initial goal was defined as replacing a human actor within a conventional dramatic performance. We quickly discovered a number of objectives as this developed.

A. The HERB Robot

Our actor was the personal robot HERB [16], visible in Fig. 1. The acronymic name stands for Home Exploring Robotic Butler and reflects its usual purpose: research into mobile manipulation for assistive household tasks. HERB is built on a two-wheeled Segway RMP base and incorporates a pair of seven-axis Barrett Technology WAM arms with three-fingered Barrett BH280 hands. Above the body is a sensor mast with a high-speed pan-tilt head supporting monocular and RGBD cameras and a DLP projector. Within the body is a stereo speaker system for voice output. HERB contains three onboard computers running Robot Operating System (ROS) [17] and an extensive set of motion planners, controllers, and vision modules comprising the lab research infrastructure.

From a dramatic standpoint, the primary anthropomorphic elements of HERB are the pair of arms, the presence of a head, and basic mobility. However, HERB is hardly an android: HERB lacks a recognizable face, exhibits an arm motion range that is substantially non-anthropomorphic, and locomotes slowly with non-holonomic diff-drive constraints.

Our general approach was to emphasize coordinated arm and head movement over driving around the stage. This approach assumes that affective body language can provide effective dramatic expression despite the lack of an articulated face, an idea well-supported in puppetry. [13]

B. Dialogue

The play we selected, "Sure Thing", depends on expressive dialogue delivery with comedic timing for an effective performance. The usual HERB speech synthesizer is a general-purpose text-to-speech product from Cepstral [18]. The plain text of a play provides enough semantic information for the synthesizer to produce utterances rendering informational content, but the renderings lack the prosodic nuances required for effective drama.

To generate more nuanced speech, we instead capture a set of prosodic utterance structures from a human actor to drive a custom synthesizer based on the open-source Festival system. [19] [20] This is essentially the speech equivalent of motion capture: the prosodic structure of an utterance can be rerendered with the opportunity for manipulating parameters. This approach provides a rich performance as the starting point for dialogue manipulation in rehearsal.

C. Motion

The motion controller for HERB had previously only been used to execute automatically planned paths for functional manipulation tasks without tight latency requirements. However, the play has no actual manipulation and only minimal driving motion, so the primary purpose of movement is to disambiguate and accentuate dialogue delivery. This requires highly responsive and expressive movement not easily captured in planning constraints.

Our general approach was to create a set of expressive motions in advance using a combination of direct pose capture on the robot and animation trajectory editing. The arms and head of the robot are easily positioned by hand when running a low-impedance gravity-compensation controller, allowing an animator to apply their physical intuition and real-world observation to choose effective poses using the robot itself as a master control.

D. Operator Interface

Theatrical rehearsal and live performance both emphasize immediacy and reactivity either to experiment with variation in performance or to accommodate mistakes and variations. The goal for the operator interface was to reduce the required input to a single cue action per line of dialogue while still allowing for flexible control in the event of a major mistake.

Our general approach was to constrain the overall movements to animated transitions between prespecified poses in a motion graph. This keeps the amount of operator input to a feasible level and eliminates the need for real-time motion planning. On a dramatic level, it emphasizes the elements of timing and tactics over detailed motion variation. The graphical operator interface includes both high-level sequencing controls for performance as well as separate panels to trigger individual motions and script lines. In normal operation, the operator cues the start of each short sequence, but can take over and puppet the robot in more detail if necessary.

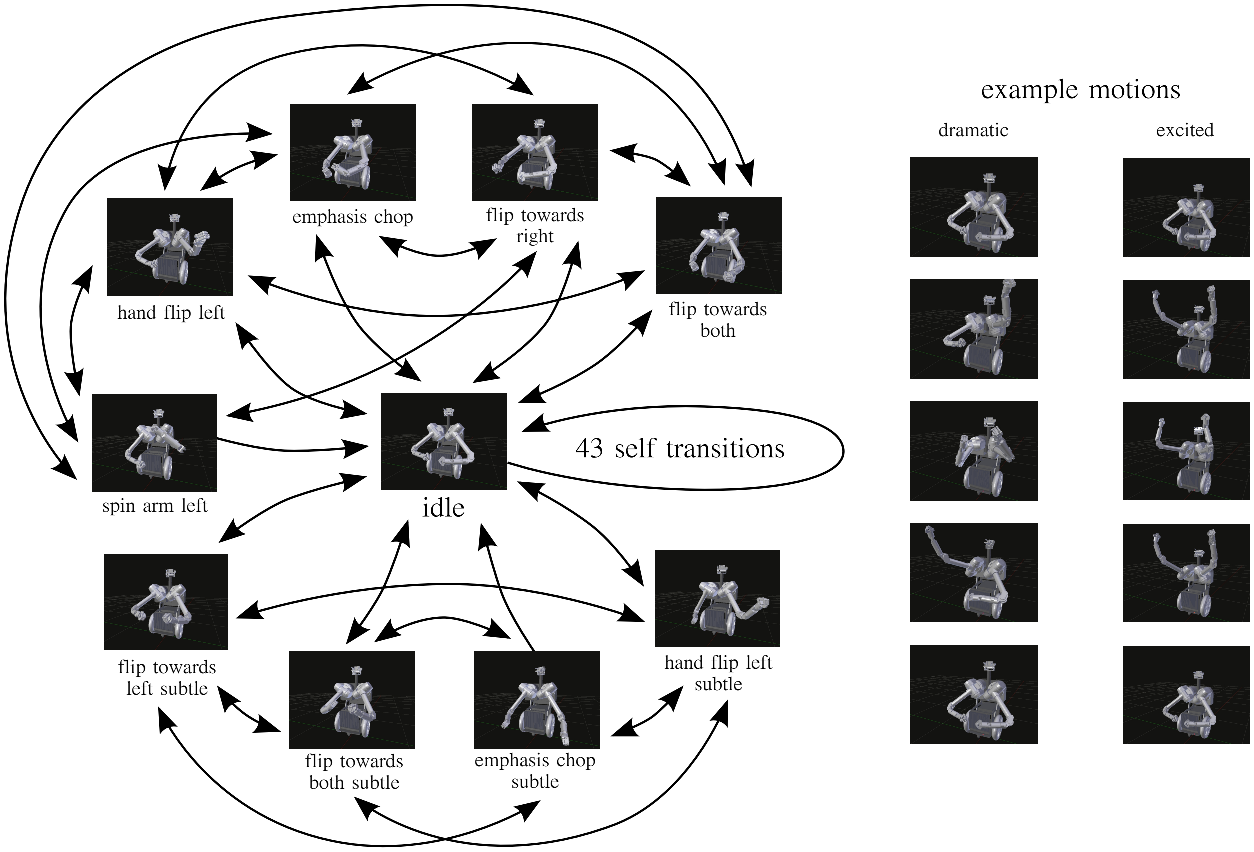


Fig. 2. Animation motion graph. Each of the 91 transitions is a hand-animated trajectory beginning and ending at a reference pose. There are 43 self-transitions from the idle state representing trajectories which begin and end at a common pose.

IV. Motion

The overall design of the motion system was motivated by the requirement for flexible gestural improvisation during rehearsal. This prompted us to create a library of hand-tuned general-purpose gestures organized into a graph of trajectories transitioning between a set of specific poses, as shown in Fig. 2. This design is a balance between real-time puppeteering and scripted motion. The overall dramatic outcome emphasizes the choice and timing of gestures as the basis for expression.

A. The Motion Graph

We entered the motion construction process with the notion that most of our gestures, in and of themselves, should not suggest any particular emotion. In different contexts, the same gestures should represent different intents. A sweeping arm gesture, such as motion `hand_flipL` from "idle" to "hand flip left", could be a display of incredulity or simply a reference towards an object. A pointing gesture, such `point_sbtl` self-transitioning from "idle" to "idle", can be an accusation or an expression of recognition.

The motion graph consists of static poses linked by actions that transition between them (Fig. 2). This is a common technique in crowd and video game animation [21]. Our graph includes ninety-one specific motions connecting ten rest poses, comprising every gesture the robot was required to make during the performance. We gained substantial flexibility by including 43 motions that started and stopped at the same idle position, as these can be used in any order.

The animations fall logically into several pragmatic categories which we informally termed as "keep-alive", "referencing", and "specific meaning." These are loosely related to work by Cassell [22] and Nehaniv [23] categorizing gestures used in combination with speech.

The first and most heavily used category was keep-alive motion [24] designed to add subtle speech cues and prevent the robot from appearing too static. This category is similar to Cassell's beat gestures. Referencing gestures, a combination of Cassell's deictics and Nehaniv's pointing gestures, were used to specifically draw the audience attention to a particular item or physical location. The "specific meaning" gestures are idiomatic motions with commonly-understood meaning, such as a head scratching motion to indicate confusion.

The referencing gestures and keep-alive motions proved to be the most often-used gestures in the performance. We found that the more elaborate idiomatic gestures would frequently take too much time to execute to fit within the short time frame allowed by fast dialogue.

Throughout the rehearsal process, the motion graph approach allowed the operator to rapidly assemble and perform different gesture combinations with little difficulty. This ease of creation also meant that less-successful motions could be discarded without too much lost time. Ultimately, out of the 91 animations produced, we used 36 motions in 185 instances. This ability to experiment with expression is a very human characteristic that directors expect of actors.

B. F-Curves

Prior to this project most of the motion executed on the HERB robot was purely functional and produced by automatic planners [25] [26]. These planners can take several seconds to generate new motion, and the resulting trajectories often do not convey intent or exhibit predictable timing. While much research has gone into making these planners more legible [27], they have not been designed to produce affective gestures making them ill-suited to dramatic performance.

In order to achieve the desired level of control and dramatic potential, we turned to the tools and methodology of animation [28]. We used the open-source computer graphics software Blender [29] as a virtual environment for generating motion. A common representation for joint motion in computer graphics is a cubic spline with two dimensional control points specified in time and position known as an F-Curve [30]. The spline is defined by keyframes which specify the pose of a joint at specific points in time. Between these key poses, the joint interpolates smoothly according to a cubic polynomial function [31]. This gives an animator a great deal of control as the position of each joint may be specified exactly at these keyframes. Keyframes may also be created at arbitrary intervals, allowing tight spacing of keys for high frequency motion and broader spacing for smooth motion. This technique has been successfully used both for dramatic [5] and functional [32] robotic motion, and worked well for our purposes. In order to execute this animation on the robot, we built a new trajectory type for the OpenRAVE [33] environment that samples position, velocity and acceleration of these curves at 500 Hz to interface with HERB's existing motion controllers.

C. Animation Interface

Generating F-Curve animation can be difficult and time-consuming for non-experts. These techniques allow a great deal of control, but consequently require substantial effort in specifying many joint positions over the duration of an action.

To streamline the keyframing process we developed a plugin for Blender enabling the creation of animation poses via manipulation of the physical robot in a low-impedance gravity-compensated mode. This provides direct manual control of the joints which is much easier than controlling a virtual character, and unlike the Blender animation interface, does not require any special skills or training. Working with the actual robot provides immediate visual feedback on pose and scale and reduces the iterations required to discover a satisfactory result.

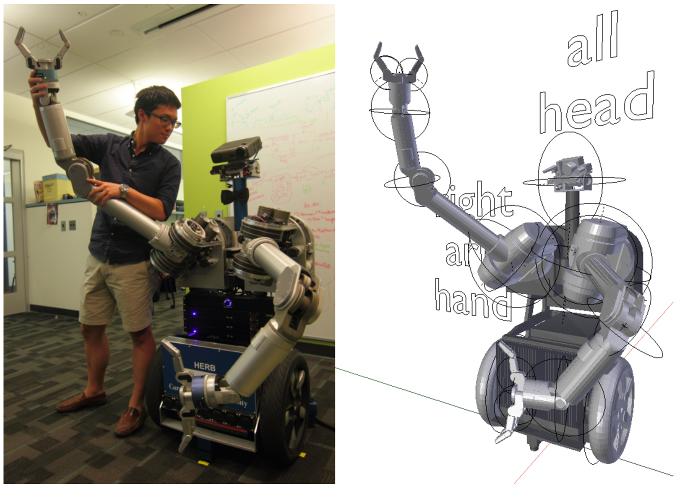


Fig. 3. Manipulating the HERB robot to create an animated motion, and the corresponding pose in Blender.

Once a pose sequence is captured, it can be modified and retimed into a trajectory using existing keyframing tools in Blender. In this process the speed and duration of each motion is set, and the individual keyframes can be fine-tuned for appearance and the elimination of self-collisions. Often this process would be iteratively refined as we tested on the robot and with the actor. Each individual movement is developed separately in the Blender system and then exported into the motion graph performance system.

The interface and keyframe capture system were designed to be quick and easy to use. Our primary animator, who had no prior animation experience, was able to create most of the motions in the play over the course of 3 months, and was able to make edits in rehearsal. We found that using the actual robot as a master to specify animation poses was far less time-consuming and produced superior results compared to using the animation interface alone.

D. Animation Process

The primary inspiration for creating a varied gesture library was human motion. We relied on videos of individuals in conversation as reference for creating convincing human-like gestures, and although HERB is not fully humanoid, we found that most human motions translated effectively onto the robot.

While refining the animations for performance, we found most adjustments were made not in the poses themselves, but instead in correcting the timing or the speed of the motions. During work sessions with the director, most adjustments made were in retiming gestures to achieve more pronounced inflection points to fit with the dialogue.

The correct timing, however, was difficult to gauge even in Blender. The simulation presented an idealized robot: it would not shake when it made sudden moves, it was always calibrated correctly, and it never failed to execute any animation. We discovered how many motions that seem benign on humans or in simulation are perceived as threatening when played on the actual robot with its large arms. Seeing an graphical animation wave an arm is different from seeing a heavy aluminum tube sweep through the air. In general, during testing on HERB we found most of the preliminary animations too dramatic and scaled them gradually into less pronounced movements, as well as creating additional subtle versions of many poses.

V. Dialogue

Our primary objective for the spoken dialogue system was to produce clearly understandable speech conveying the humor and character of the written text, with a secondary goal of supporting dynamic adjustment of the line readings in rehearsal. Our experience with the existing Cepstral system was that although understandable, the relatively flat delivery would not be satisfying for performance. We set out to satisfy both of these goals by building a new voice model specifically for the play from recordings of an actor providing a performative and inflected reading of the text. We used the open-source festvox system [19], which enabled us to expose the captured utterance data in an interface allowing dynamic adjustment of the pitch and duration contours while retaining the essential prosody of the actor.

A. Unit-Selection Speech Synthesis

Our system uses a concatenative speech synthesis method in which the speech of a person is recorded and stored. During synthesis, the target audio is produced from a concatenation of segments of the recording via a specific scheme. The technique that we used is *unit selection*, which was proposed by Hunt and Black [34] and implemented in the Festival software package [20]. An alternative approach is to use Linear Predictive Coding (usually through HMMs), which has the advantage of speed and a smaller footprint, but the disadvantage of more muffled, buzzy output speech.

In unit selection, a database is formed from a corpus of known utterances and audio files. To generate the audio for a new text utterance, the utterance is broken down into phonemes, which are distinct units of sound, and an utterance structure, which describes the ordering of the phonemes. These two parts are used to find the database entries that most closely match the text to be generated. The audio files in these database entries are then concatenated together and signal processing is performed to reduce artifacts.

B. The Challenge of Prosody

The way a particular phoneme is pronounced depends on the *prosody*, which is the contextual patterns of stress and intonation in language. The prosody changes with meaning, context, and emotion. Context from word placement within a sentence can be modeled with the utterance structure. Emotion and meaning are much more difficult to capture, as it would require having both a corpus and input text labeled with either emphases or emotions. For the corpus, the audio is accessible, for which there are techniques of automatic labeling for emphasis [35] [36] and more recently for emotions [37]. However, when generating speech from text, the text itself is the only source for finding the emphasis and emotional cues. Lacking the semantic understanding, it is currently not feasible to automatically infer these labels on the raw text.

C. Copy-Prosody aka Voice Performance Capture

Not having emphasis or emotional labels on the target text, we used the technique of copy-prosody, in which the prosody is simply copied to the generated target audio from the nearest units in the database. In a general-purpose corpus this results in a monotone voice since the average emphasis will be used. Instead, we recorded a corpus which was not a neutral reading of all common phonemes but instead a dramatic reading of the lines of the play. When generating speech the prosody from the original recordings is copied through and roughly captures the actual desired prosody. Note that this works well in the limited context of a play, but if the same corpus is used to generate speech not within the script the synthesized voice is highly distorted because of the dramatic nature of the reading.

The model-building process began with a clean recording of the voice actor. The tracks were split into phrases and matched with the corresponding text segment from the script. The text was broken into phonemes using a pronunciation dictionary. An utterance structure was built for each text phrase, and the corresponding durations and fundamental frequencies were recorded from the audio file for each of the phonemes.

D. Dialogue Operator Interface

The voice model needs to be dynamic and easily adaptable to achieve the same dynamics as a human actor. To that end, we created a plugin for the open-source WaveSurfer [38] sound manipulation tool in which the pitch of each phoneme is plotted versus time for any line of the play. The audio operator can drag the pitch mark vertically to increase or decrease the pitch of the individual phoneme. Dragging the pitch mark horizontally increases or decreases the duration of the phoneme. Emphasis is created by combining duration, pitch, and amplitude changes. We chose to control just two dimensions (duration and pitch) for each phoneme to keep the interface simple and fast.

E. Performance Results

The copy-prosody technique generated a baseline voice which performed to the director's satisfaction in many cases. When a change was requested, the director's feedback was typically very abstract, such as "can you say that more hesitatingly?" While these complex instructions are easily understood by a human, it is a difficult task to deconstruct the abstract concept of hesitancy down to the level of which syllable to elongate or diminish. As such, many iterations of small changes followed by listening and evaluation of the resulting waveform were needed to reach a satisfactory output. Unfortunately this slowed down the speech editing process to a point where it could not effectively be done live during rehearsal. To manage this gap in the level of abstraction versus level of control, either the audio operator needs to learn what low level changes map to high level effects, or the controls need to be raised to a more abstract level.

VI. Operator Interface

HERB operating as an actor in a conventional drama requires a hybrid human-robot system in which an operator is responsible for human-level perception tasks, translating those observations into cues for a semi-autonomous robot. For this particular play we used no robot-level perception, so the operator was solely responsible for translating directorial input, actor responses, and audience reactions into robot cues.

The overall objective of the operator system is to enable a sliding autonomy between direct puppeting and high-level cues. In practice, this was implemented as a combination of graphical interfaces for cueing individual motions and cueing sequences, shown in Fig. 4.

The graphical interface was built using the open-source Pure Data (Pd) [39] performance system. This selection was motivated by the real-time design of Pd and the highly interactive development process of the graph-based language. The connection between the Pd event system and the robot was made by creating a simple Python plugin for Pd to ease attachment to the existing Python robot API. We created a back end database using Python for storing all performance data, which includes poses, animation trajectories, the motion graph, dialogue text, and cue sequences.

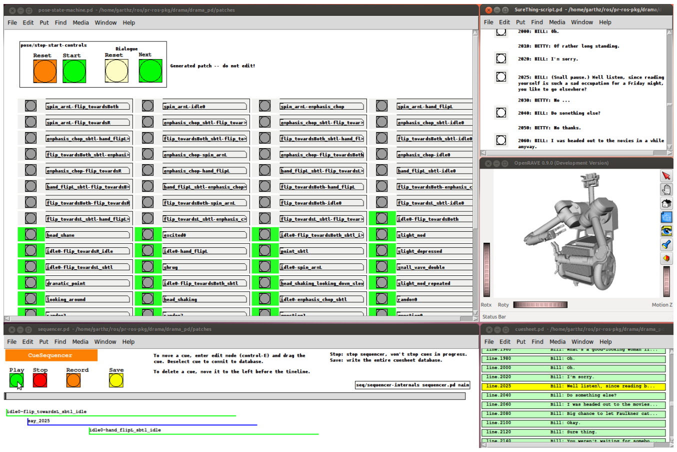
The sequencer interface records interface events into cue sequences to build each dramatic beat using the basic puppeting interface. It has controls for graphical editing of cue timing, switching between sequences, and loading and saving the database. During performance, all events are logged to allow post-hoc reconstruction of the robot actions.

Pd was an effective choice for rapidly prototyping the control interfaces but poorly suited for building the sequencer editor due to the limited support for manipulating structured data. The sequencer editor works but lacks many direct manipulation features common to non-linear editing systems due to the difficulty of implementation in Pd. The graph-oriented nature of Pd is better suited to semi-autonomous performance and improvisatory programming in keeping with its typical usage in computer music. All in all, the system was very quick to develop and proved to be reliable in performance, a key goal for a live show.

The real-time performance of Pd was more than adequate for our event-driven robot API. However, the latency for initiating trajectories is noticeable due to limitations of the robot control pipeline. In practice, the timing of the events in the sequencer was adjusted to compensate, but the operator still needed to practice and memorize the timing for initiating each sequence.

VII. Rehearsal

The rehearsal process in general consists of four stages: (1) table work (talking through the script and making basic choices), (2) exploring and rehearsing different possible deliveries, (3) pruning the options, and (4) perfecting the final choice into a consistent version. We built our system to allow the director to follow this process as closely as possible.


pose-state-machine.pd - /home/garthz/ros/pr-ros-pkg/drama/drama_pd/patches
File Edit Put Find Media Window Help
pose/stop-start-controls
Reset
Start
Dialogue
Reset
Next
spin_armL-flip_towardsBoth
spin_armL-idle0
spin_armL-emphasis_chop
spin_armL-hand_flipL
spin_armL-flip_towardsR
emphasis_chop_sbtl-flip_towar>
emphasis_chop_sbtl-flip_towar>
emphasis_chop_sbtl-idle0
emphasis_chop_sbtl-hand_flipL>
flip_towardsBoth_sbtl-flip_to>
flip_towardsBoth_sbtl-hand_fl>
flip_towardsBoth_sbtl-idle0
flip_towardsBoth_sbtl-emphasi>
emphasis_chop-spin_armL
emphasis_chop-flip_towardsBoth
emphasis_chop-idle0
emphasis_chop-flip_towardsR
emphasis_chop-hand_flipL
hand_flipL_sbtl-flip_towardsL>
hand_flipL_sbtl-idle0
hand_flipL_sbtl-flip_towardsB>
hand_flipL_sbtl-emphasis_chop>
flip_towardsBoth-hand_flipL
flip_towardsBoth-emphasis_c
flip_towardsBoth-flip_towardsR
flip_towardsBoth-spin_armL
flip_towardsBoth-idle0
flip_towardsL_sbtl-idle0
flip_towardsL_sbtl-hand_flipL>
flip_towardsL_sbtl-emphasis_c>
flip_towardsL_sbtl-flip_towar>
idle0-flip_towardsBoth
head_shame
excited0
idle0-flip_towardsBoth_sbtl_i>
slight_nod
idle0-flip_towardsR_idle
idle0-hand_flipL
point_sbtl
slight_depressed
idle0-flip_towardsL_sbtl
shrug
idle0-spin_armL
small_wave_double
dramatic_point
idle0-flip_towardsBoth_sbtl
head_shaking_looking_down_slow
slight_nod_repeated
looking_around
head_shaking
idle0-emphasis_chop_sbtl
random0
SureThing-script.pd - /home/garthz/ros/pr-ros-pkg/drama/
OpenRAVE 0.9.0 (Development Version)
Rotx
Roty
Motion Z
Status Bar
sequencer.pd - /home/garthz/ros/pr-ros-pkg/drama/drama_pd/patches
CueSequencer
Play
Stop
Record
Save
seq/sequencer-internals sequencer.pd main
idle0-flip_towardsL_sbtl_idle
say_2025
idle0-hand_flipL_sbtl_idle
cuesheet.pd - /home/garthz/ros/pr-ros-pkg/drama/drama_pd

Generated patch -- do not edit!

2000: BILL: Oh.

2010: BETTY: Of rather long standing.

2020: BILL: I'm sorry.

2025: BILL: (Small pause.) Well listen, since reading yourself is such a sad occupation for a Friday night, you like to go elsewhere?

2030: BETTY: No ...

2040: BILL: Do something else?

2050: BETTY: No thanks.

2060: BILL: I was headed out to the movies in a while anyway.

To move a cue, enter edit mode (control-E) and drag the cue. Deselect cue to commit to database.

To delete a cue, move it to the left before the timeline.

Stop: stop sequencer, won't stop cues in progress.
Save: write the entire cuesheet database.

line.1960	Bill: What's a good-looking woman li...
line.1980	Bill: Oh.
line.2000	Bill: Oh.
line.2020	Bill: I'm sorry.
line.2025	Bill: Well listen\, since reading b...
line.2040	Bill: Do something else?
line.2060	Bill: I was headed out to the movies...
line.2080	Bill: Big chance to let Faulkner cat...
line.2100	Bill: Okay.
line.2120	Bill: Sure thing.
line.2140	Bill: You weren't waiting for somebo...

Fig. 4. Primary graphical interface for drama operator. Panels clockwise from upper left: pose graph transition controls, script dialogue cues, robot state viewer, sequencer cuesheet, and sequence editor. Not shown are panels for hand controls, low-level robot debugging, or auxiliary cuesheet controls. The interface is designed to require minimal operator cue input for normal performance but allow interactive puppeting to improvise sequences and quickly adjust timing during rehearsal.

While the exploration phase for human actors consists of coming up with new performance variations spontaneously, changing HERB's motions and voice takes longer, thereby reducing the number of different directions that could be explored. To allow the maximum amount of freedom to the director, we prepared many gestures and variations ahead of time to enable quickly swapping sequences options during rehearsal. Our system can also be used to generate new motions in a few minutes during the rehearsal if required.

Although robots are less spontaneous than human actors, one advantage is *repeatability*. The consistency helped the human actress rehearse since she knew what the robot was going to do and could rely on the same reaction every time.

VIII. Audience Response to Performance

We evaluated our system through its the final product: a *performance*. We framed the performance into two parts. First, HERB played a lead role in the one-act play "Sure Thing" by David Ives. Immediately afterward, we held an "open rehearsal" on the stage, where we discussed the software tools and acted out a mock rehearsal.

Audience members completed a two-part paper survey. Before the play we asked them if they expected the robot to perform as well as a human actor and after the play we asked them if the robot had performed as well as a human actor, to measure the perceived performance capability. We also asked if they related to the human actor and if they related to the robot actor. All questions were on a 7 point Likert scale. After the play, we invited the audience to respond freely to the prompt "Comments on performance? Did you enjoy it? How did the robot contribute? How did robot compare to the actress? Issues?".

Ninety audience members responded with the following primary affiliations: computer science (17), robotics (24), drama (18), art-related (3), science-related (4), and other (24).

Ideally we would have liked to see the audience equating the robot to the human actor. On average they neither agree nor disagree that the robot performed as well as a human actor, but their ratings did increase after seeing the show $t(88) = 6.37, p < .001$, as shown in Fig. 5. It is important to note that the audience's expectations were very low in the first place, which could be explained by low exposure to robots in similar social settings. Even though the robot did not reach a performance level equaling that of a human, this does not mean that the robot did not deliver an effective performance. Of the 28 people who commented on the performance itself, the top words used were enjoyed (8), great (4), and exceeded expectations (3), with 26 out of the 28 comments being positive. The audience most frequently wrote that there was a problem with the robot's timing. There was some confusion as to whether the occasional pauses before the robot delivered his lines were intentional for dramatic effect or mistakes. One audience member even went so far as to say that HERB "almost 'forgot' one dialog."

The affiliation of audience members did have an impact on how much they related to the actors as shown in Figure Fig. 5. Viewers with a background in drama or the arts related significantly more to the human actor, $t(56) = 4.51, p < 0.001$, than viewers with a scientific or technical background. This bias does not appear as strongly with the robot actor, $t(43) = 0.17, p < .87$.

In the open-ended comments, several audience members noted aspects of the play that were absent. Finger movements, which were intentionally not used, were noted as being absent by audience members who noticed that HERB has fingers. Similarly, audience members desired facial expressions from the area that moved like a head. They also pointed out a lack of movement around the stage by drawing attention to the fact that HERB has two wheels and yet remained stationary for most of the play.

With these comments, we see a repeated theme of assumed capability. Because the audience could see what appeared to be functional mechanical or body parts, they expected that all of the parts would be involved in the performance.

IX. Future Work

We would like to apply the drama techniques we are learning to daily life interactions with the robot so that HERB can be an effective collaborator, responding to the intentions of a human and clearly exhibiting intention and other internal state. With clear communication of intent, humans could easily compensate for the physical limitations of the robot.

To this end, drama can serve as a model for creating embodied communicative gesture from functional motions. Many of the tactics of physical acting involve movements related to functional action taken in order to reveal a character's intentions. The robot's goals are considerably less abstract that a human character, but the same thinking can be used to derive gestures which reveal the computational state of the robot from motions which are native to the normal robot tasks.

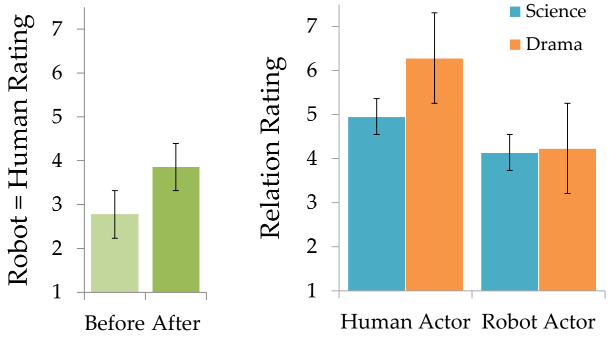


Fig. 5. The rating of whether the robot performs as well as a human is shown on the left, both before and after the play is performed. A rating of how much the viewer related to the human and robot actors, broken down by the viewer's primary affiliation is shown on the right.

Similar thinking may apply to estimate human intention. Rather than estimating abstract emotional state from human pose or facial expression, we would like to infer internal state from the functional properties of human pose. E.g., if a person's hand is moving toward for an object, it is likely they will attempt to grasp it. This movement implies desire and purpose in the implicit narrative of a practical collaborative task.

We would also like to continue working on dramatic performances. On the artistic side, working with a playwright to create a new text could draw out the essential character of the robot. This might involve more pantomime and functional motion as part of the story, since concrete actions such as moving and touching objects can tell more of a story with less emphasis on dialogue and verbal gestures. Navigation around the stage would become more important.

On the technical side, many members of the audience expressed an interest in seeing more autonomy. The operator could be eliminated if robot reactions were cued via direct sensing, especially in a narrative based more on physical drama. We would like to explore simple pattern-matching techniques for tracking scripted cues using vision and audio perception to provide dynamic reactivity on stage using simple robust tools.

For both drama and daily life, the motion graph vocabulary could be expanded by introducing parametric variation, even as simple as varying the animation rate to control overall duration. Simple random gestures (like fidgeting) might also allow the robot to appeare more active in listening and make for a more natural performance.

X. Conclusions

Robotics researchers working in live drama need to carefully balance artistic and technical goals. Producing an effective performance requires prioritizing artistic needs, and the demands of a live show place an emphasis on simplicity and reliability. Events on stage happen quickly and the operator workload must remain manageable.

Our emphasis on rehearsal kept the project focused. Ceding artistic authority to the director and frequently running performance tests provided early feedback on features and allowed dramatic goals to take priority over technical aspirations. The director provided an outside viewpoint on the interpretation of specific animations and substantially guided the development of the motion vocabulary.

Our goal of substituting a robot for an actor in rehearsal and performance proved to be a substantial engineering challenge, but one which we were able to solve by repurposing and extending existing tools. The simplicity of the approach highlights how robotic storytelling can succeed with a limited set of motion primitives and careful timing, at least when placed under the guidance of experienced dramatists.

Acknowledgments

This work has been developed in collaboration with Sam French, Olivia Brown and Kyle Wilson from the School of Drama; and the 'bowtie' team from the Entertainment Technology Center: Rachina Ahuja, Kwong Fai Lew, Ruobing Li, Michael Lotze, Juan Ramirez, Jiahao Xia, and Xiaoyi Zhang. Special thanks to Jiahao 'Simon' Xia for his extra effort on performance animation.

References

[1] D. Ives, *All in the Timing: Fourteen Plays*. Vintage Books, 1994.

[2] A. Bruce, J. Knight, S. Listopad, B. Magerko, and I. Nourbakhsh, "Robot improv: Using drama to create believable agents," in *IEEE International Conference on Robotics and Automation*, vol. 4. IEEE, 2000, pp. 4002–4008.

[3] C. Breazeal, A. Brooks, J. Gray, M. Hancher, J. McBean, D. Stiehl, and J. Strickon, "Interactive robot theatre," *Communications of the ACM*, vol. 46, no. 7, pp. 76–85, 2003.

[4] H. Knight, "Eight lessons learned about non-verbal interactions through robot theater," in *Social Robotics*. Springer, 2011, pp. 42–51.

[5] G. Hoffman, R. Kubat, and C. Breazeal, "A hybrid control system for puppeteering a live robotic stage actor," in *IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2008, pp. 354–359.

[6] L. Takayama, D. Dooley, and W. Ju, "Expressing thought: Improving robot readability with animation principles," in *ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '11. New York, NY, USA: ACM, 2011, pp. 69–76.

[7] R. Murphy, D. Shell, A. Guerin, B. Duncan, B. Fine, K. Pratt, and T. Zourntos, "A midsummer nights dream (with flying robots)," *Autonomous Robots*, vol. 30, no. 2, pp. 143–156, 2011.

[8] S. Lemaignan, M. Gharbi, J. Mainprice, M. Herrb, and R. Alami, "Roboscopie: a theatre performance for a human and a robot," in *ACM/IEEE International Conference on Human-robot Interaction*. ACM, 2012, pp. 427–428.

[9] D. V. Lu, C. Wilson, A. Pileggi, and W. D. Smart, "A robot acting partner," in *ICRA Workshop on Robots and Art*, Shanghai, China, 2011.

[10] J. Zlotowski, T. Bleeker, C. Bartneck, and R. Reynolds, "I sing the body electric an experimental theatre play with robots [video abstract]," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2013, pp. 427–427.

[11] S. Cavaliere, L. Papadia, and P. Parascandolo, "From computer music to the theater: The realization of a theatrical automaton," *Computer Music Journal*, vol. 6, no. 4, pp. 22–35, Dec. 1982.

[12] N. Mavridis and D. Hanson, "The IbnSina center: An augmented reality theater with intelligent robotic and virtual characters," in *IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2009, pp. 681–686.

[13] E. Jochum, J. Schultz, E. Johnson, and T. D. Murphey, "Robotic puppets and the engineering of autonomous theater," in *Controls and Art*. Springer, 2014, pp. 107–128.

[14] C.-Y. Lin, L.-C. Cheng, C.-C. Huang, L.-W. Chuang, W.-C. Teng, C.-H. Kuo, H.-Y. Gu, K.-L. Chung, and C.-S. Fahn, "Versatile humanoid robots for theatrical performances," *Int J Adv Robotic Sy*, vol. 10, no. 7, 2013.

[15] M. I. Sunardi, "Expressive motion synthesis for robot actors in robot theatre," M.S. Thesis, Electrical and Computer Engineering, Portland State University, 2010.

[16] S. Srinivasa, D. Berenson, M. Cakmak, A. Collet Romea, M. Dogar, A. Dragan, R. A. Knepper, T. D. Niemueller, K. Strabala, J. M. Vandeweghe, and J. Ziegler, "Herb 2.0: Lessons learned from developing a mobile manipulator for the home," *Proceedings of the IEEE*, vol. 100, no. 8, pp. 1–19, July 2012.

[17] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, 2009, p. 5.

[18] "Cepstral." [Online]. Available: http://www.cepstral.com

[19] "festvox." [Online]. Available: http://www.festvox.org/festival/

[20] The Centre for Speech Technology Research, "The Festival speech synthesis system," The University of Edinburgh. [Online]. Available: http://www.cstr.ed.ac.uk/projects/festival/

[21] G. Davis and B. Far, "Massive: Multiple agent simulation system in a virtual environment." [Online]. Available: http://www.enel.ucalgary.ca/People/far/Lectures/SENG697/PDF/tutorials/2002/Multiple_Agent_Simulation_System_in_a_Virtual_Environment.pdf

[22] J. Cassell, "A framework for gesture generation and interpretation," *Computer vision in human-machine interaction*, pp. 191–215, 1998.

[23] C. L. Nehaniv, "Classifying types of gesture and inferring intent," in *Procs of the AISB 05 Symposium on Robot Companions*. AISB, 2005.

[24] R. Williams, *The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators*. Macmillan, 2009.

[25] D. Berenson, S. S. Srinivasa, D. Ferguson, and J. J. Kuffner, "Manipulation planning on constraint manifolds," in *IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 625–632.

[26] N. Ratliff, M. Zucker, J. A. Bagnell, and S. Srinivasa, "Chomp: Gradient optimization techniques for efficient motion planning," in *IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 489–494.

[27] A. Dragan and S. Srinivasa, "Generating legible motion," in *Proceedings of Robotics: Science and Systems*, June 2013.

[28] F. Thomas, O. Johnston, and F. Thomas, *The illusion of life: Disney animation*. Hyperion New York, 1995.

[29] "Blender." [Online]. Available: http://www.blender.org

[30] J. Lasseter, "Principles of traditional animation applied to 3d computer animation," in *ACM Siggraph Computer Graphics*, vol. 21, no. 4. ACM, 1987, pp. 35–44.

[31] N. Burtnyk and M. Wein, "Computer-generated key-frame animation," *Journal of the SMPTE*, vol. 80, no. 3, pp. 149–153, 1971.

[32] J. Angeles, A. Alivizatos, and P. J. Zsombor-murray, "The synthesis of smooth trajectories for pick-and-place operations," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 1, pp. 173–178, Jan. 1988.

[33] "OpenRAVE." [Online]. Available: http://www.openrave.org

[34] A. J. Hunt and A. W. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, vol. 1. IEEE, 1996, pp. 373–376.

[35] B. Arons, "Pitch-based emphasis detection for segmenting speech recordings," in *Recordings*, 1994, pp. 1931–1934.

[36] F. R. Chen and M. Withgott, "The use of emphasis to automatically summarize a spoken discourse," in *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 1. IEEE, 1992, pp. 229–232.

[37] J. Xu and L.-H. Cai, "Automatic emphasis labeling for emotional speech by measuring prosody generation error," in *International Conference on Intelligent Computing, 2009. ICIC, 2009 LNCS 5754*, 2009, pp. 117–186.

[38] KTH Royal Institute of Technology, "Wavesurfer," Sweden. [Online]. Available: http://www.speech.kth.se/wavesurfer/

[39] M. Puckette *et al.*, "Pure Data." [Online]. Available: http://puredata.info