

# CS CAPSTONE REQUIREMENTS DOCUMENT

NOVEMBER 10, 2017

## HOW TO MAKE AN EFFECTIVE ROBOT COMEDIAN

PREPARED FOR

OREGON STATE UNIVERSITY

HEATHER KNIGHT

\_\_\_\_\_  
*Signature*

\_\_\_\_\_  
*Date*

PREPARED BY

GROUP 13

AKAROBOTICS

ANISH ASRANI

\_\_\_\_\_  
*Signature*

\_\_\_\_\_  
*Date*

KEVIN TALIK

\_\_\_\_\_  
*Signature*

\_\_\_\_\_  
*Date*

ARTHUR SHING

\_\_\_\_\_  
*Signature*

\_\_\_\_\_  
*Date*

### Abstract

A comedian can observe an audience and improvise a delivery of a joke to connect the audience to the content. This makes the experience more authentic and genuine for the observer. The purpose of this project is to discover what makes an entertaining interaction by studying a robot that performs comedy. We propose that a performance is enhanced when (1) the comedian interacts spontaneously with the audience, (2) the comedian has and conveys a coherent, well-developed character, and (3) the comedian adapts its act to cater to an audience based on their reaction. This document covers the technical requirements for our project, as well as a description of software, hardware, and outside limitations.

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Previous Research</b>	<b>2</b>
<b>3</b>	<b>Hypothesis</b>	<b>3</b>
3.1	Crowdwork . . . . .	3
3.2	Character . . . . .	3
3.3	Adaptiveness . . . . .	4
<b>4</b>	<b>Research Approach</b>	<b>4</b>
4.1	Learning/Exploration Phase . . . . .	4
4.2	Prototyping/Testing Phase . . . . .	4
4.3	Evaluation Phase . . . . .	4
<b>5</b>	<b>Methods</b>	<b>5</b>
5.1	Tests . . . . .	5
5.2	Metrics . . . . .	5
<b>6</b>	<b>Conclusion</b>	<b>6</b>
<b>7</b>	<b>Glossary</b>	<b>8</b>
	<b>References</b>	<b>9</b>

# LIST OF TABLES

1	A gantt chart showing the projected timeline of the project. . . . .	7
---	--	---

## 1 INTRODUCTION

The field of human-robot interaction can learn a lot from stand-up comedy. A stand-up performance has a basis of scripted content, from which the comedian delivers jokes to engage the audience. Good Comedians can read the audience and sometimes adapt their delivery based on the mood of the room [1]. In social robotics, when a robot shares a space with a human, an interaction can influence the people's opinions of the robot. Additionally, evident character traits presented (through dialogue and non-verbal motion) by the machine can anthropomorphize itself, making it easier and more enjoyable to connect with for the human [2]. The purpose of this work is to explicitly evaluate what aspects of a robot comedian's performance are most salient to human audience.

## 2 PREVIOUS RESEARCH

Research on the improvement of HRI is indispensable for our project. In Heather Knight's *Eight Lessons*, gestures, liveliness, and joke timing are all aspects that can be incorporated into the robot [2]. Relatable and appropriate gestures significantly helps improve communication between the robot and the audience. If the actions are predictable, humans can relate to the robot. When watching someone perform an action, the human brain maps the actions onto itself and simulates the action in the best way possible. This is a physiological experience that should be replicated by the robot in order to enhance relatability. Simplicity is important as well. [2]

In addition, Knight observed that having the robot portrayed as a living character rather than just an object that is kept up on stage improved the overall experience for the observers. Having believable interactions can enhance the feeling of a living character. The goal of the audience tracking using sensors is to maximize enjoyment. The enjoyment levels were be read by the robot and used to modify upcoming jokes [2]. Pausing and letting the audience laugh is vital as well. Starting the next joke too early can break the rhythm and leave the audience baffled. Looking around and body poses should be used to fill the pause [2].

In a previous study of robot comedy [3], Katevas found that when a robot engaged the audience through eye contact, the audience was more receptive to the performance. Eye contact from the robot is important, as it is a non-verbal cue for direct interaction. The audience members can identify that the machine is making an attempt to engage with specific members of the audience. We will investigate this further by having the robot perform various non-verbal and verbal interactions using sensors. The idea is to have the audience be a part of the performance even if they are not the ones performing. This can be accomplished if the robot is socially intelligent.

Other studies by Katevas et al. [4] that involved evaluating the social dynamics of a live performance by a robot have used SHORE<sup>TM</sup> vision framework software to analyze and detect faces in the audience. SHORE<sup>TM</sup> allows for facial expression recognition, estimated age, gender, and eye or mouth openings [5], giving the study a heterogeneous audience model. These allowed for the robot to interact directly with specific audience members. However, usage of SHORE<sup>TM</sup> involves expenses and funds that are unavailable to us, so we will encounter behavioral limitations dealing with a homogeneous audience model.

A study by Guy Hoffman [6] noted the importance of anticipation in human-robot interaction (HRI). The timing and meshing of anticipatory action and perception are a useful framework for HRI. The greatest challenge when designing a robot that will perform on stage is to enable the robot to be both - expressive and responsive. Robot models in the past have ended up on either extreme; they are either real-time and do not allow for continuous expression, or they are very animated but do not allow well times reactive behavior.

Researchers have also proposed multiple design patterns to promote sociability in Human-Robot Interaction (HRI). Some of these include having an initial introduction, some sort of didactic communication, including personal interests or history, and recovering from mistakes [7]. These patterns in design are proposed to allow for more effective and meaningful social interactions. While there is yet to be much data or research on the validity of these claims, they may still prove to be useful in guiding the designs of our project.

Robots utilizing non-verbal communication and statically written audience engagement have been attempted in robot comedy. In particular, Knight has observed the importance of character and spontaneous interactions in creating effective comedy [2]. However, there is little research on the actual effectiveness of character and spontaneous interactions [4]. This project will aim to examine the effectiveness of character and spontaneous interactions in robot comedy.

### 3 HYPOTHESIS

Our research will be guided by the following questions:

- 1) How can the robot make the audience feel like a part of the performance?
- 2) How can the robot convey a coherent and well-developed character?
- 3) How can the robot adapt and influence to the audience?

We hypothesize that comedy scripts with greater degrees of (1) crowdwork, (2) character, and (3) adaptiveness, will create more effective comedy, and have a more positive response from the audience.

#### 3.1 Crowdwork

Keeping the audience engaged is vital in creating an entertaining performance. Katevas et al. had some success leading to a better audience response when using gestures and acknowledging the audience's presence. A successful performance manages the dynamics of various aspects of interaction to the benefit of both the performer and the audience [3]. Interaction with the audience will make the comedian robot feel more authentic, like an entity, and less like an object. Knight's research found that the robot's connection with the audience is stronger if the robot is able to convey social intelligence. This helps the robot display that it is not just an inanimate object [2].

#### 3.2 Character

Expressing character will be understood from the robot under the "theory of the mind" [8]. Understanding the agents meta-representation of a behavior is helpful for the audience in relating to the robot's intent, desires and knowledge. A robot attempts this to understand and relate its desires and intent better to the audience [9].

The appearance of character in a robot may create more effective comedy. Research has shown that expressive behaviors in a robot may cause interacting humans to favor the robot [10]. One could argue that expressive behaviors are behavioral actions formed by an inner character. While robots currently are not capable of having intrinsic character qualities, we postulate that a robot which behaves as if it has character could be more effective at engaging an audience. The task of conveying a sense of character may also benefit from social behaviors developed in researching the effectiveness of crowdwork, as social interactions may increase the sense of agency [?] and thus aid the audience in grouping behaviors as acts of character.

### 3.3 Adaptiveness

When the robot tells a joke, it needs to make adaptive transitions that correlates with the response. For example, if a joke does well and is received with laughter, the robot needs to time the next joke so that it is delivering it when the audience is ready, and can hear. However, if a joke is not well received by the audience, the time a robot needs to wait for the next joke will be different than if there is laughter. This is important for the effectiveness of a performance, as the connotation of the next joke is determined by the result of the previous joke; an audience that is told a bad joke will be hesitant to enjoy a joke if the previous jokes were bad. A robot comedian needs to be able to stay in a joke if the audience likes it, or address/recover the bad joke before starting a new sequence. Timing, or anticipation for a new joke, when coordinated correctly, positively influences the fluidity of the task (the performance) [6].

## 4 RESEARCH APPROACH

This project will be carried out in three phases; A learning and exploration phase in Fall term, a prototyping and testing phase in the Winter, and our evaluation phase Spring term.

### 4.1 Learning/Exploration Phase

This phase of our development will focus on understanding Social Robotics and the technology of the robot. The three of us will become familiar with stand-up comedy and the dynamics of an audience-comedian interaction. The NAO robot behaviors are programmed in the software Choregraphe, which has an API for python. We will test primitive scripts of decision making and non-verbal behavior. This is to learn how the coding environment works and to familiarize ourselves with hardware limitations. Additionally, we will learn to work with the sensors on the robot, and how they function (microphone, camera, etc). To become familiar with the format of a stand-up performance, we intend to study jokes and comedy devices.

### 4.2 Prototyping/Testing Phase

In the prototyping and testing phase, we will develop early sets for the robot. These implementations need to reflect and support our research questions. Crowd-work will involve audience sensing, as well as jokes that incorporate a measurement of response from the audience. Character implementation will involve testing the differences in effectiveness of robot vs human joke delivery, and the effectiveness of robo-centric jokes. As a stretch goal, we also hope to prototype and test the effectiveness of adapting a set to the audience, using intelligent calibration of the sensors.

These prototypes will be in the form of 3-6 minute set scripts with several variants in Choregraphe. The variants will help us evaluate the research questions and can be tested in front of a small sample of humans, or in the form of a video recording. For example, the robot may perform in front of a handful of friends, or recordings may be to show online live viewers. Non-mechanical feedback from our testing will influence the direction of our prototyping, meaning that the implementation of our research questions will adapt according to the audience response. By the time this phase is completed, we will have working sets of robot stand-up.

### 4.3 Evaluation Phase

While doing the research, we will perform 6 shows with audiences ranging from 10-30 people. Tests in this phase will be at a greater scale and with a more realistic environment. Each stand-up performance, or set, will contain bits, or

sections of content that will be categorized as crowd work, characterization dialogue, and jokes. We will be testing on a live human audience to learn the effectiveness of each bit in a set. Based on the effectiveness of each set, we will modify the set and behavior of the robot. By the end of this phase, we hope to have a working, effective robot comedian.

## 5 METHODS

### 5.1 Tests

Our research questions will be how we evaluate the effectiveness of a robot comedian. We will create performances with varying degrees of (1) crowdwork, (2) character, and (3) adaptiveness. We will vary the presence of each of the factors of the performance to determine which factors are most influential to the comedian. To test crowdwork, we may create one script that may include no references to the audience (low degree of crowdwork), while another may include many instances of interactions with the audience (high degree of crowdwork). To test the effectiveness of adding character, we may create a script with random jokes (low degree of character), and one with a coherent character throughout the set (high degree of character).

Likewise, testing the effectiveness of adaptivity will include scripts with no adaptivity and scripts that adjust to audience response. These scripts will be run on the NAO robot, and performed in front of a small audience in the testing phase, and later a larger live audience in the evaluation phase.

The audio sensors on the NAO bot will be used to evaluate audience reception to a joke, which the robot can then adapt to. The microphone on the NAO does not perform well when receiving input from a large audience, as it is designed to handle smaller scale interactions. NAO interprets speech with the `ALSpeechRecognitionProxy` and `ALTextToSpeechProxy` [11].

### 5.2 Metrics

To evaluate the effectiveness of our robot comedian, we will measure the response of the performance with the audio captured during the set, and surveys for the audience afterwards. Measuring the audio will help us understand a broad audience response, as well as data for adaptive functionality. Surveys will give a more in depth information on factors of the show that are not covered by audio sensing, such as opinions of the perceived character and quality of the jokes. For example, to see if the robot's character was conveyed coherently, the audience will fill out a questionnaire prompting them to describe its character, as well as some humanizing questions, e.g. "Would you invite this robot to dinner?" These responses will be used to study if the robot matched the expected persona and gauge how comfortable the people are with the robot.

As each set will derive content from the three research questions, we need to measure the effectiveness of each portion compared across the 6 performances. The survey that the audience members take after the show will gauge the response to sections. Additionally, there will need to be questions to establish a pretense of how an audience member felt before the show, and how the performance has influenced their opinions of robot comedy. We want to see if someone who has seen a robot comedian would recommend the show to others.

The microphone on the NAO robot is designed for small environment settings. It will be difficult to distinguish speech from separate sources in a noisy environment [12]. Audio levels will be important data to collect from a crowd, where a louder crowd response could correlate to a level of enjoyment. However, a problem could be that a crowd could be booing very loud, and if we do not distinguish between different sounds that the audience can make, we

could accidentally associate a negative response with a good response. The background noise could be very different depending on the room size. The density of people in a room making noise may return different audio level [13]. It will be important to test the effectiveness of a microphone to receive input.

## **6 CONCLUSION**

A stage presence for a comedian is important because it connects the audience to the content, making it more effective than soulless delivery. A robot has a disadvantage in this; being soulless is the essence of being a robot. For a robot to establish a stage presence, the machine needs to make efforts to connect with the crowd, present a cohesive character, and dynamically adapt to a response. We think that a round character personifies a relatable agent for an audience. If a robot can give insight to its desires, behavior, and preferences during a performance, the robot will humanize itself, connecting itself to the crowd.

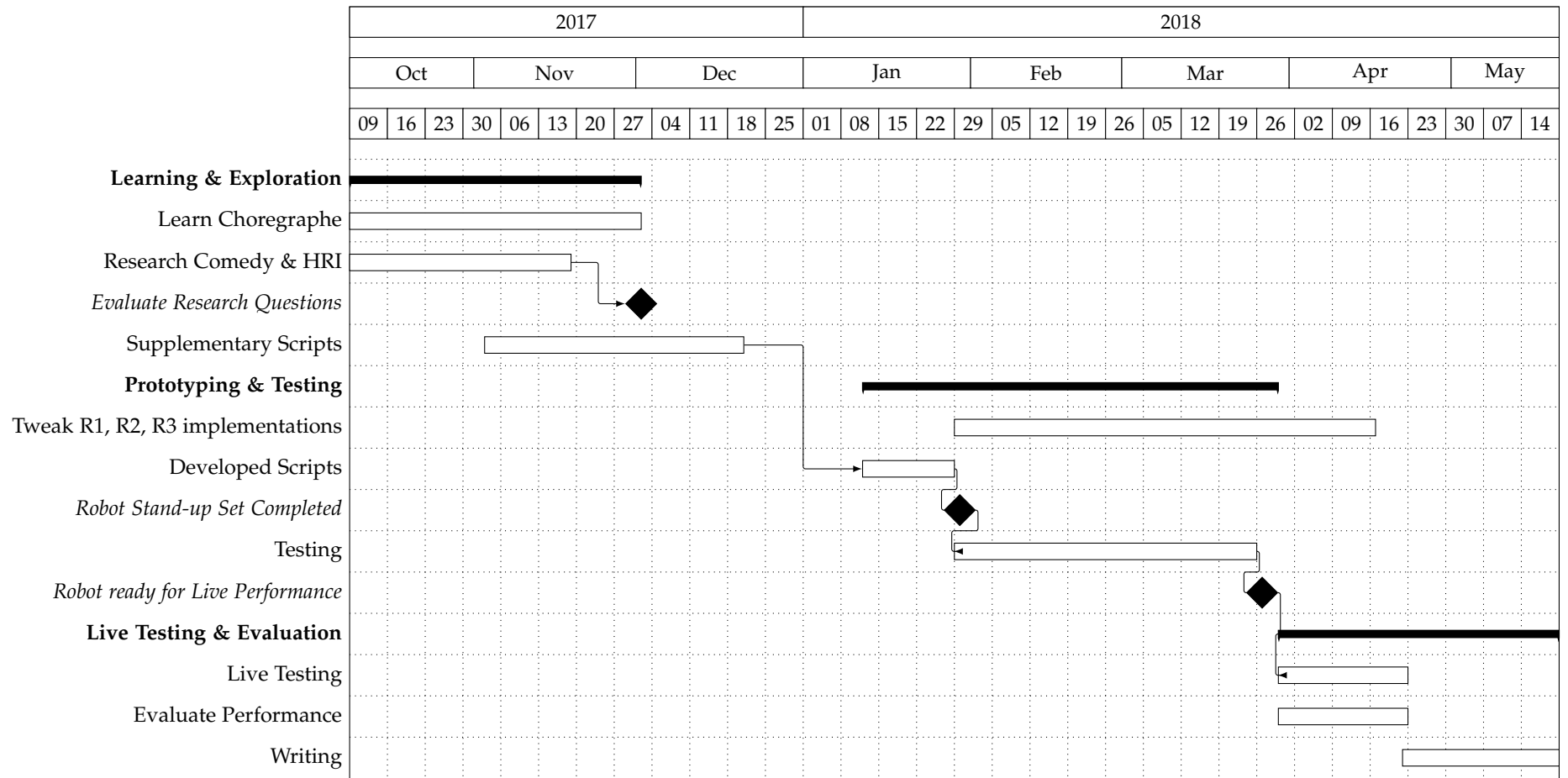


TABLE 1  
A gantt chart showing the projected timeline of the project.



## 7 GLOSSARY

ALSpeechRecognitionProxy

API for the NAO that converts received speech to text

ALTextToSpeechProxy

API for the NAO that converts text to speech

API

Application Programming Interface

Choregraphe

Software used to program behavior and performance sets, made by SoftBank Robotics

Crowdwork

Part of a Comedian's performance that involves content from the current audience

NAO

Model of Robot that will be used as the Comedian Agent, made by SoftBank Robotics

NAOqi

The API library for all functions used to program the NAO robot

Python

An Object-Orientated programming language

SoftBank Robotics

Manufacturer of the NAO robot, NAOqi API, and Choregraphe software

## REFERENCES

- [1] Chris Rock, Ricky Gervais, "Talking funny," Home Box Office, dec 2012.
- [2] Knight, "Eight lessons learned about non-verbal interactions through robot theater," in *International Conference on Social Robotics*. Springer, 2011, pp. 42–51.
- [3] Katevas, Healey, and Harris, "Robot comedy lab: experimenting with the social dynamics of live performance," *Frontiers in psychology*, vol. 6, 2015.
- [4] Katevas, Healey, and Harris, "Robot stand-up: engineering a comic performance," in *Proceedings of the Workshop on Humanoid Robots and Creativity at the IEEE-RAS International Conference on Humanoid Robots Humanoids (Madrid)*, 2014.
- [5] Ernst, Ruf, and Kueblbeck, "A modular framework to detect and analyze faces for audience measurement systems," in *2nd Workshop on Pervasive Advertising at Informatik*, 2009, pp. 75–87.
- [6] Hoffman, "Anticipation in human-robot interaction." in *AAAI Spring Symposium: It's All in the Timing*, 2010.
- [7] Kahn, Freier, Kanda, Ishiguro, Ruckert, Severson, and Kane, "Design patterns for sociality in human-robot interaction," in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '08. New York, NY, USA: ACM, 2008, pp. 97–104. [Online]. Available: <http://doi.acm.org/10.1145/1349822.1349836>
- [8] Leslie, "Pretense and representation: The origins of theory of mind," *Psychological Review*, vol. 94, no. 4, pp. 412–426, 1987.
- [9] Scassellati, "Theory of mind for a humanoid robot," vol. 12, pp. 13–24, 01 2002.
- [10] Anand, Luetke, Venkatesh, and Wong, "Designing expressive behaviors to improve human-robot relationships," *World Academy of Science, Engineering and Technology, International Journal of Mechanical and Mechatronics Engineering*, vol. 4, no. 8, 2017.
- [11] "Aldebaran documentation." [Online]. Available: <http://doc.aldebaran.com/2-1/naoqi/audio/index.html>
- [12] "Aldebaran documentation." [Online]. Available: <http://doc.aldebaran.com/2-1/naoqi/audio/alsounddetection-api.html>
- [13] "Aldebaran documentation." [Online]. Available: <http://doc.aldebaran.com/2-1/naoqi/audio/alsoundlocalization.html>