Enhancing Interaction Through Exaggerated Motion Synthesis

Michael J. Gielniak
Georgia Institute of Technology
801 Atlantic Dr. NW
Atlanta, Georiga, USA
mgielniak3@gatech.edu

Andrea L. Thomaz
Georgia Institute of Technology
801 Atlantic Dr. NW
Atlanta, Georiga, USA
athomaz@cc.gatech.edu

ABSTRACT

Other than eye gaze and referential gestures (e.g. pointing), the relationship between robot motion and observer attention is not well understood. We explore this relationship to achieve social goals, such as influencing human partner behavior or directing attention. We present an algorithm that creates exaggerated variants of a motion in real-time. Through two experiments we confirm that exaggerated motion is perceptibly different than the input motion, provided that the motion is sufficiently exaggerated. We found that different levels of exaggeration correlate to human expectations of robot-like, human-like, and cartoon-like motion. We present empirical evidence that use of exaggerated motion in experiments enhances the interaction through the benefits of increased engagement and perceived entertainment value. Finally, we provide statistical evidence that exaggerated motion causes a human partner to have better retention of interaction details and predictable gaze direction.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Robotics—*Kinematics & dynamics, propelling mechanisms*; H.1 [**Models and Principles**]: [User/Machine Systems, Miscellaneous]

General Terms

Algorithms, Experimentation, Performance

Keywords

Exaggeration, cartoon-like motion, user study

1. INTRODUCTION

In human communication, the body is a high-bandwidth channel used for communicating spatial reference, disambiguating speech, inquiring for feedback, influencing others' behavior, identifying social goals or intentions, and directing attention. For a social humanoid robot to take advantage of its body like humans do, algorithms must autonomously modify its motion to include these social task elements.

Other than eye gaze and referential gestures (e.g. pointing), the relationship between robot motion and observer attention is not well understood. In addition to these explicit ways of directing attention, social robots need to communicate saliency and direct attention in all motions. Our research aims to do this with animation-inspired algorithms.

Exaggeration, a principle of animation [7], is defined in abstract terms as developing the *essence of an idea*, where a moderate version of something is replaced by a more extreme version. In theory, if a social robot needs to attract attention to a certain body part or region, that part should be exaggerated relative to the rest of its body. Exaggeration is a form of trajectory modulation used for emphasis.

We define a motion version to be exaggerated when it contains more contrast between two subspaces of motion (1) primary motion (i.e. task or intent of the motion) and (2) secondary motion (i.e. physics-based response to primary motion) as compared to the original motion that was used to create the exaggerated version. The open research question is how a social robot can autonomously generate exaggerated motion, rather than simply using such pre-designed motion.

In this paper, we present an algorithm to produce cartoon-like, exaggerated motion and present quantitative evidence from experiments that exaggerated motion used in social contexts can be used to control observer eye gaze. We show the benefits of exaggerated motion include: increased memory retention of interaction details, and improved interaction performance by keeping the human partner more engaged and more entertained throughout the interaction. We show that varying levels of exaggeration correlate to human expectations of robot-like, human-like, and cartoon-like motion.

2. RELATED WORK

Our literature survey uncovered no previous work adapting the animation principle of exaggeration to robots. However, the relationship between motion and subsequent observer attention direction has been studied in different contexts. E.g., unexpected changes in motion direction will attract attention [6]. And when watching videos, movement features, such as magnitude and vector, impact observer eye movement [2]. Humans attract the gaze of other humans by entering the space nearby; humans classify attention by variables such as head and body orientation [8]. In our work, we want to harness the relationships between motion and observer attention for use in social robots. Thus, we developed a motion synthesis algorithm, the output of which produces reliable, consistent predictability in observer gaze location.

Recently there has been an interest in applying animation principles to improve HRI. For example, surrounding action was designed to help communicate robot intent earlier [11]. In their related work, the robot motions were carefully designed by a Pixar animator. However, we are interested in making animation principles part of a robot's controller, so character-like motion can be generated autonomously, in real-time, during interaction with humans. Our own prior work examined autonomously generating other animation principles, namely secondary action [4] and anticipation [5].

In computer animation, there are many ways to create exaggeration. Small motion segments can be multiplied by a constant to create exaggeration [9]. Exaggeration of a motion signal can be created by changing filter amplitude in convolution [13]. Given an exaggerated and a moderate motion version, varying levels of exaggeration can be created by interpolation [12]. Motion can be transformed to the frequency domain, and exaggeration can be created by scaling only certain frequency bands [1]. Frequency domain tuning is difficult, non-intuitive and increases in complexity as the number of frequency bands increase. Often these techniques do not create noticable contrast between primary and secondary motion or they distort motion when all DOFs in motion are not systematically modulated together. Exaggeration in computer animation is an unfair comparison because breaking virtual world rules in the real-world can damage hardware (e.g., deforming bodies to exaggerate).

3. ALGORITHM

Exaggerated motion techniques developed for cartoon or virtual characters cannot be immediately applied to robots because fundamental differences exist in the real-world. The extent of exaggeration that can be produced safely on a robot is less than on cartoon or virtual character due to constraints such as torque or velocity limits of actual hardware. Our algorithm was designed to transform motion through exaggeration, while respecting real-world boundaries.

Exaggeration in joint-space can be produced by identifying the appropriate coordinates to modify for a given motion and then diminishing or amplifying these torque trajectories. Additional steps are necessary if the exaggerated motion must maintain certain features of the input motion (e.g. still be perceived as a representative exemplar of the input motion type). Exaggeration in other fields such as animation provide insight into how to appropriately adjust a trajectory for exaggerated motion effect. For example, important body parts must be amplified and less important body parts should be diminished so that gaze is directed away from regions of the motion which are less important.

We leverage actuation and timing information given in an input motion to appropriately exaggerate it, since the torques are arranged within this motion in magnitudes and directions relative to each other to produce a representative examplar. Logically then, relative DOF importance for exaggeration is measured by the amount of actuation in the degrees-of-freedom, which can be measured by torque.

Inspired by other research that examines motion decomposition by magnitude of actuation, we exploit the variance in actuation among all DOFs for a given motion to parameterize that motion according to a spectrum of actuation [14], [4], [3]. In doing so, relative importance of all DOFs remains constant in our algorithm, which ensures that as long as physical limitations (e.g. joint limits, torque limits) are not exceeded, the exaggerated version of the motion will also be an exemplar of the same type as the input motion.

Let q_j be the torque trajectory for DOF j of the original input motion with T equidistant time samples. For a robot with M degrees-of-freedom, the torque trajectories from the input motion are organized into an M x T column-stacked matrix $\tau = [q_0, q_1, ..., q_M]$. A singular value decomposition is performed on the covariance matrix in Equation 1 to obtain an M x M matrix of eigenvectors, denoted U, and an M x M matrix with eigenvalues along the diagonal, denoted Λ. The magnitude of the eigenvalues corresponds to a measure of the torque variance in the motion.

$$SVD((\tau - \mu)^T(\tau - \mu)) \quad (1)$$

where,
μ_i = mean torque of DOF i for the entire trajectory
μ = M x T stacked matrix, each column is $[\mu_1 ... \mu_M]^T$

The largest gap in the distribution of the eigenvalues defines a threshold, λ_{th}, which separates the corresponding eigenvectors into mostly actuated and near-unactuated eigenvectors. For reference, typically less than ten eigenvectors exist in the mostly actuated set. We define α_{algo} to quantify the amount of exaggeration for all DOFs in torque space. Larger values of α_{algo} equate to more exaggeration. Since exaggeration has a "polarizing" effect upon motion, pushing motion toward extremes (i.e. highly actuated increase in actuation, near-unactuated are diminished), all eigenvalues along the diagonal of Λ are modified according to the following four rules for DOF j, since $\alpha_{algo} \geq 1$:

1. If $\lambda_j < \lambda_{th}$ and $\lambda_j < 1$, then $\lambda_{j_{new}} = \lambda_j^{\alpha_{algo}}$.

2. If $\lambda_j < \lambda_{th}$ and $\lambda_j > 1$, then $\lambda_{j_{new}} = \lambda_j^{-\alpha_{algo}}$.

3. If $\lambda_j > \lambda_{th}$ and $\lambda_j < 1$, then $\lambda_{j_{new}} = \lambda_j^{-\alpha_{algo}}$.

4. If $\lambda_j > \lambda_{th}$ and $\lambda_j > 1$, then $\lambda_{j_{new}} = \lambda_j^{\alpha_{algo}}$.

These four rules are systematically designed to divide the two subspaces of a trajectory: primary and secondary motion. Rules one and two diminish the minimally existent actuation in near-unactuated torques in the original motion (i.e. reduce secondary motion); rules three and four exaggerate the torques in highly-actuated coordinates (i.e. amplifying primary motion). The composite effect of all four rules create the contrast necessary for exaggeration.

A new M x M diagonal matrix, denoted Λ_{new}, is formed from the new eigenvalues, and the original eigenvectors are used to determine the new torque matrix of exaggerated and diminished torques as shown in Equation 2. τ_{new} is the torque trajectory that is commanded to robot actuators to produce exaggerated motion.

$$\tau_{new} = U^T \Lambda_{new} U + \mu \quad (2)$$

The upper bound on α_{algo} is motion dependent and is a function of the maximum torque limits of the robot actuators. For safety, we place the upper bound on α_{algo} to be the minimum value that would cause any motor to exceed its torque limit.

Figure 1: SIMON: The hardware platform.

4. HARDWARE PLATFORM

The platform for this research is an upper-torso humanoid robot we call Simon (Figure 1). It has 16 controllable DOFs on the body and 4 per hand. Each arm has 7 DOFs (3 at the shoulder, 1 at the elbow, and 3 at the wrist) and the torso has 3 DOFs. Simon has 3 DOFs for the eyes, 2 per ear, and 4 for the neck.

5. HYPOTHESES

We have several hypotheses about the algorithm that we developed, the exaggerated motion it produces, and the benefits that it brings to HRI. These hypotheses are as follows:

- H1: Humans can perceive the difference between exaggerated and unexaggerated motion.
- H2: Exaggerated motion will appear to be more cartoon-like than unexaggerated motion.
- H3: Humans will feel more engaged and prefer interacting with a robot that displays exaggerated motion.
- H4: Exaggerated motion will enable human partners to remember more of the interaction details.
- H5: Exaggerated motion changes which body parts are salient for a given motion and directs attention to body parts that have higher kinetic energy.

6. EXPERIMENTAL DESIGN

To test these five hypotheses, we conducted two experiments using a storytelling task. Storytelling is suitable for our experimental goals since it minimizes the intellectual load of the human in the interaction, allowing humans to devote more attention to the robot. It engages the human enough to prevent overanalysis of motion. Furthermore, storytelling provides a context for the interaction and gives us a mechanism to test the effects of exaggerated motion on memory, to show instrumental HRI benefits.

Unlike other work that has focused on the benefits of *robot* gaze for storytelling [10], we manipulate robot motion to test the effects of *human* gaze without being responsive to the human. Robot gaze is irrelevant to our work because the robot is an actor in the story, not a storyteller or narrator.

In both experiments we use one of Aesop's fables adapted to include the robot as a main character. There are 13 motions that were designed to match the computer-synthesized story spoken in the background, so that the robot acted out what the narrator said. Context was completely omitted from the story. No objects were used, and no other characters in the story physically existed while the robot acted out the story. So as not to elicit attention, the robot made no intentional eye contact with participants.

Both experiments include two versions of 13 motions: original unexaggerated motions (UN), and exaggerated versions of each (EX), created using our algorithm (Section 3).

The 13 motions were shown in the same order to all participants because it was a story. The 13 motions (in story order) were: walk, look up, reach, look down, beckon, scan, shucks, grrr..., phooey, leave, point, hands open, and wave.

Four different experimental conditions were created for the story. They had identical story text, but differed in the motion types that accompanied the speech.

- AE: All 13 exaggerated motions in the story.
- EU: 7 EX followed by 6 UN.
- UE: 7 UN followed by 6 EX.
- AU: All 13 unexaggerated motions in the story.

6.1 Experiment 1: Testing Memory

Experiment 1 was designed to test whether exaggerated motion improves people's recollection of their interaction with the robot. 54 participants (36 male, 18 female; age 19-30) were recruited. Participants had no prior story knowledge. After participants watched one of the four story conditions, they rated the story motions according to 16 variables, each on a 7-level Likert scale. The 16 variables were: subtle, entertaining, realistic, exaggerated, expressive, stiff, accentuated, cartoon-like, life-like, emphatic, emphasized, natural, noticeable, engaging, smooth, and human-like.

After the Likert questionnaire, participants answered six fill-in-the-blank (FIB) questions and three short answer questions. The six FIB questions were verbatim from the narrator-spoken story, so that the objective was to fill in the exact word spoken from the story. Three were from each half of the story so that in conditions EU and UE, participants answered three FIB questions each from the UN and EX portions of the story. The answers were paired in both story halves so that these three answers were a location, an object, and an emotion. The correct answer pairs were synonyms in the story context. I.e., location: {high-above, vine}; object: {grapes, fruit}; emotion: {discouraged, unhappy}.

These pairings test the effect of exaggeration on memory substitution. In conditions where the same participant saw motions with and without exaggeration, we wanted to see if people more frequently substitute one answer for the other in the pair for EX or UN. A full analysis of this effect requires a high failure rate on correct responses to the FIB questions or a large number of participants so that enough data is collected to use appropriate statistical techniques.

Three short answer questions were designed to test H4:

- SA1: What was the title of the story that the robot just told you?
- SA2: What was the moral of the story?
- SA3: What was the color of the object that the robot was trying to reach?

Following the short answer questions, participants were asked to tell their favorite part of the story and their favorite motion from the story, and they were prompted for reasons why they selected these as their favorites.

6.2 Experiment 2: Testing Gaze Direction

The same exact story, motions, experimental layout, and four experimental conditions were repeated in Experiment 2. However, the faceLAB[1] system was used to track participant eye gaze direction. Left and right eye trajectories were captured for the duration of the entire story, from which we calculate the exact location on the robot's body that each eye of the participant was looking at during the story.

For Experiment 2, 68 participants (44 male, 24 female; age 19-30) were recruited. After the story finished, participants were asked to rate the robot according to each criteria using the same 16 variables on the 7-level Likert. Participants were then prompted for their favorite part of the story, their favorite motion from the story, and their reasons for both.

Then, participants were seated at a virtual model of the robot and given two controls: + (plus) and - (minus). These buttons indirectly controlled the value of α_{algo} for the exaggeration using a mapping function. Participants were asked to use the buttons to select the value of α_{exp} that produces the motions that they considered to be:

- HL: Most Human-like
- VP: Most Visually Pleasing
- CL: Most Cartoon-like
- RL: Most Robot-like
- LB: They Liked Best

Order of all the five values was randomized for each participant, but each participant saw the same order of the five values for all motions. Motion order was randomized for each participant. In order to intentionally bias our experiment away from the results we expected to achieve, the value of α_{exp} was always initially set at the value that corresponded to unexaggerated motion. The range on α_{exp} was limited between positive and negative one, which prevents exceeding the torque limits of any motor on the robot. This part of Experiment 2 was not done on hardware for safety.

7. RESULTS

The results from Experiments 1 and 2 are discussed together in support of the five hypotheses presented in Section 5 because data from both experiments is used to support the hypotheses. Discussion of the statistically significant variables from the Likert scales is distributed so that results are presented in support the appropriate hypotheses. The mean values from the Likert variables are shown in Table 1.

7.1 EX and UN are Perceptibly Different

First we test our algorithm to ensure it accomplished our goal of creating motion that is exaggerated. Although many of our Likert scales achieved statistical significance between participants in the AE and AU groups, three of the sixteen scales add evidence to support that our algorithm creates exaggerated motion: subtle, exaggerated, and accentuated.

First, ANOVAs were conducted on all the data from both experiments assuming that all four conditions (AE, EU, UE, AU) belong to the same distribution. F_{crit} for each variable in the Likert is 2.68. Respectively, the F-values achieved from these ANOVAs are 23.0 (subtle), 3.18 (exaggerated), and 4.85 (accentuated). Thus, for each measure at least two of the conditions is statistically different.

[1]faceLAB is a trademark of Seeing Machines.

Table 1: Average subjective responses to 16 Likert scale variables. 1 (not). 7 (variable).

Variable	AE	EU	UE	AU
Subtle	2.63	3.48	3.40	5.35
Entertaining	5.20	4.02	3.92	3.67
Realistic	2.87	3.84	3.59	4.21
Exaggerated	4.86	3.82	3.86	3.13
Expressive	4.59	4.08	4.22	3.39
Stiff	3.18	3.26	3.31	3.49
Accentuated	4.62	2.88	3.55	2.90
Cartoon-like	3.34	2.40	2.72	2.50
Life-like	3.24	3.85	3.70	3.62
Emphatic	4.36	3.86	3.77	3.05
Emphasized	4.47	3.76	3.57	2.87
Natural	3.90	3.52	3.50	3.70
Noticable	4.91	3.92	4.00	3.98
Engaging	5.24	3.84	3.67	3.53
Smooth	5.49	5.43	5.32	5.26
Human-like	4.12	3.56	3.72	3.67

Post-hoc pairwise t-tests were performed. For the subtle variable, three of the six pairings exhibit statistically significant results ($p<0.01$): (AE,AU), (EU,AU), and (UE,AU). And for both the accentuated and exaggerated Likert variable, three of the six pairings exhibit statistically significant results ($p<0.01$): (AE,EU), (AE,UE), and (AE,AU).

Participants who saw at least half of the motions modified by our algorithm indicated that the motions are less subtle than participants who saw only the original motions, and participants who saw only motions modified by our algorithm indicated that the motions are more accentuated and more exaggerated than participants who saw at least half of the original motions. Thus, we conclude that motions produced by our algorithm are less subtle, more accentuated, and more exaggerated than the input motions.

7.2 EX Appears More Cartoon-like Than UN

A good measure of success is whether the output of our algorithm maintains the same qualities and characteristics of its inspiration. H2 is a logical hypothesis, since the inspiration for exaggerated motion comes from animated and virtual characters. And by testing if EX is more cartoon-like than UN, we also are evaluating whether our algorithm accomplished one of its most fundamental goals. Two of the Likert variables provide evidence in support of H2, the hypothesis that exaggerated motions produced by our algorithm are more cartoon-like: realistic and cartoon-like.

ANOVAs conducted on all the data from both experiments assuming that all four conditions (AE, EU, UE, AU) belong to the same distribution yielded F-values of 3.84 and 4.26 for realistic and cartoon-like respectively. Both of these values exceed the F_{crit} of 2.68, indicating that for both measures at least one of the conditions is statistically different.

For the realistic Likert variable, three of the six pairings exhibit statistically significant results ($p<0.05$): (AE,EU), (AE,UE), and (AE,AU). These three results are the three pairings that compare participants who saw only motions produced by our algorithm compared with participants who saw at least half of the original motions. And for the cartoon-like Likert variable, one of the six pairings exhibits statistically significant results ($p<0.02$): (AE,AU). These results are the pair that compares participants who saw only mo-

Table 2: Average α_{exp} Participant Responses for Most Human-like (HL), Most Cartoon-like (CL), Most Robot-like (RL), Most Visually Pleasing (VP), and Liked Best (LB).

Motion	HL	CL	RL	VP	LB
Walk	0.131	0.910	0.056	0.877	0.844
Look Up	0.230	0.967	0.074	0.812	0.754
Reach	0.164	0.787	0.041	0.385	0.501
Look Down	0.607	0.959	0.434	0.8361	0.869
Beckon	0.230	0.956	0.066	0.771	0.574
Scan	0.517	0.869	0.246	0.689	0.877
Shucks	0.680	0.851	0.098	0.762	0.844
Grrr...	0.443	0.967	0.197	0.911	0.899
Phooey	0.639	0.926	0.080	0.836	0.756
Leave	0.541	0.754	0.221	0.803	0.639
Point	0.508	0.853	0.148	0.836	0.910
Hands Open	0.623	0.951	0.180	0.885	0.541
Wave	0.320	0.861	0.107	0.615	0.525
Average	0.433	0.893	0.150	0.771	0.733

tions produced by our algorithm compared with participants who only original motions. From these results, we conclude that motions exaggerated by our algorithm are less realistic but more cartoon-like than the input motions.

There is further quantitative evidence relevant to H2 from the data of experiment 2 where participants selected α_{exp} values that pertain to their subjective ratings of most human-like, most cartoon-like, most robot-like, most visually pleasing, and liked best (HL, CL, RL, VP, and LB, respectively).

The analysis was initially performed assuming that the qualities of HL, CL, RL, VP, and LB for α_{exp} are motion dependent. Thus 13 separate ANOVAs were performed, 1 per story motion, assuming all 5 groups belonged to the same distribution. 13 of 13 ANOVAs yielded F-values greater than F_{crit}, which means that for each motion, subjective settings for α_{exp} according to each of the 5 groups held at least one statistically independent pair of groups. 10 post-hoc pairings were performed for each motion, using all possible pairs of the five subjective qualities, yielding a total of 130 post-hoc pairings. 122 of the 130 pairings were statistically significant ($p<0.05$). The 8 pairings that failed statistical difference tests were all between VP and LB for the motions of reach, beckon, scan, shucks, phooey, leave, hands open, and wave. We conclude that by modulating α_{exp} (i.e. exaggeration) we can consistently create motions which are more robot-like, more human-like, and more cartoon-like, and the values of α_{exp} (and α_{algo}) are largely motion independent.

Average values across all motions for the α_{exp} setting from Experiment 2 are shown in the bottom row in Table 2. For consistency, prior to Experiment 2, all unexaggerated input motions to be used in the story were selected to have α_{exp} values between -0.2 and 0.0, and no α_{exp} for any exaggerated motion used in the story was less than 0.7. For reference, exaggerated motion values with an α_{exp} value of 1.0 correspond to exaggerated motion so that any more exaggeration would cause at least one motor to exceed its torque limit.

The average value of 0.89 for cartoon-like is consistent with the Likert results, which also found that the exaggerated motions were cartoon-like; it indicates that exaggerating motions by using our algorithm adds a cartoon-like quality to the motions. Robot-like motion, as defined by participants' expectations, tends to lack exaggeration, which may help explain the results regarding entertainment and engagement of exaggerated motion that we discuss in Section 7.4. Human-like motion tends to exhibit moderate levels of exaggeration, not near the torque limits for any motors, but also far from the values of robot-like exaggeration.

Table 3: P-values from post-hoc pairwise t-tests of the data in Table 2 for all 5 groups. Tests with p>0.05 are in gray. x = elsewhere in the table.

	HL	RL	VP	LB
CL	0.008	0.004	0.024	0.016
RL	0.023	x	0.009	0.012
VP	0.021	x	x	0.072
LB	0.027	x	x	x

The results from the motion-by-motion analysis for α_{exp} lead us to suspect that human-like, cartoon-like, and robot-like are motion independent. To show that these measures of HL, CL, and RL for exaggerated motion are truly independent of motion when equalized on the scale of α_{exp}, a singular ANOVA was performed on all the data in Table 2. The F-value of 59.71 from this analysis is greater than F_{crit} of 2.53, which means there is statistical difference between at least two of these measures independent of motion. 10 pairwise t-tests grouping data averages across all motions were performed to compare all possible pairings of HL, CL, RL, VP, and LB (p-values shown in Table 3).

Table 3 shows that across all motions, the qualities of human-like, cartoon-like, and robot-like are statistically different from all other variables in the study. Only statistical difference tests fail between visually pleasing and liked best (shaded gray in Table 3), which indicates that these two subjective measures for exaggerated motion may not come from independent distributions.

The final analysis that we performed in support of H2, was to evaluate 130 additional pairwise t-tests (10 per motion) for the α_{exp} values that correspond to the story motions and the distributions provided in Experiment 2 based upon HL, CL, RL, VP, and LB, to find out if the UN or the EX motions used in the story are statistically different from the human subjective measures. The ten pairings are all possible combinations of one member from each of the sets: {UN, EX} and {HL, CL, RL, VP, LB}. 65 of 65 pairwise t-tests for the UN motion pairs have $p<0.05$, which indicates that the choice of unexaggerated motion in the story did not coincide with our participants distributions of any of the five measures. The 26 motion pairs for EX with HL and RL have $p<0.05$, which indicates that our exaggerated motions used in the story were not robot-like or human-like, using participants' subjective responses as the measures of these two variables. However, for 13 of 13 (EX,CL) pairs there is no statistical difference ($p<0.05$) between exaggerated story motion and participants' expectations of cartoon-like motion.

7.3 EX Improves Interaction Performance

Thus far, the results have not established the benefits of exaggerated motion. In support of H3, HRI performance will be measured by how well participants can remember the story over a short period of time. Thus, evaluation of performance is based upon SA and FIB question answers.

Correct answers were verbatim from narrator speech to the 6 FIB questions (Table 4). Each of the 6 FIB answers were associated to one sentence and one motion in the story (either UN or EX). Thus, question answers can be grouped

Table 4: Percent of participants in each of four conditions that correctly answered the fill-in-the-blank question. EX and UN columns are percent of correct FIB question responses grouped according to motion associations cumulatively across all four conditions.

Answer	AE	EU	UE	AU	EX	UN
high above	71.4	84.6	30.8	42.9	77.8	37.0
grapes	92.9	76.9	38.5	64.3	85.2	51.9
discouraged	85.7	69.2	69.2	64.2	77.8	66.7
fruit	75.9	30.8	69.2	42.9	74.1	37.0
unhappy	85.7	61.5	84.6	71.4	85.2	66.7
vine	85.7	53.8	92.3	64.3	88.9	59.3

Table 5: Percent of incorrect answers for FIB question answer pairs substituted across contextual synonyms (1) location {high above, vine}, (2) object {grapes, fruit}, (3) emotion {discouraged, unhappy} in each of four conditions. EX and UN columns are percent of incorrect FIB question responses substituted across contextual synonyms, grouped according to the motion type executing when the substituted word was heard. The left column still defines the question by the intended correct response.

Answer	AE	EU	UE	AU	EX	UN
high above	75.0	50.0	66.7	12.5	71.8	31.3
grapes	0.0	66.7	75.0	20.0	37.5	43.3
discouraged	100.0	50.0	75.0	20.0	87.5	35.0
fruit	66.7	77.8	75.0	25.0	72.2	37.5
unhappy	0.0	80.0	50.0	0.0	40.0	25.0
vine	100.0	66.7	100.0	20.0	83.3	60.0

based upon associated motion type. The percent of correct FIB question responses grouped according to motion associations are shown cumulatively across all four conditions in the two columns furthest to the right in Table 4.

Comparing columns of (AE, EU, and UE) to (AU), more participants remember the story better when they watch some exaggerated motions. In the two columns on the right in Table 4 where the participant responses are grouped according to the associated motion type (EX or UN), the trend is stronger, which demonstrates the benefit that exaggerated motion helps people to remember the story better.

In Table 5, EX and UN are accumulated over the motion type that was executing when the substituted word was heard; this is different than in Table 4 because in Table 4, the UN and EX columns are tallied over the motion type associated with the text in the question.

As evident in Table 5, when participants saw an exaggerated motion, they remembered the words associated with the motion better than when they saw unexaggerated motion. When answers were incorrect, contextual synonyms heard during exaggerated motions more often get substituted for incorrect responses than from unexaggerated motions. This suggests that one benefit of exaggerated motion is to help retain interaction details in memory.

We see similar trends in the short answer questions. The moral of the story was explicitly mentioned during the 'hands open' gesture, which was the second to last motion in the story. The moral of the story was "Do not speak disparagingly of things that you cannot attain." Table 6 shows the percent of correct answers as a function of all four conditions and the two most common substitutions grouped according to the motion type seen when the participant-given answer was spoken by the narrator in the story.

Table 6: Percent of participants vs. SA2 (moral) answers. EX and UN columns are responses tallied across all four conditions, grouped by the motion type that was executing when the participant-provided answer was spoken by the narrator. #1 = "reach" substituted for "attain." #2 = "be discouraged by" substituted for "speak disparagingly of."

Answer	AE	EU	UE	AU	EX	UN
Verbatim Correct	35.7	7.7	30.8	0.0	33.2	3.8
Substitution 1	35.7	15.3	7.7	0.0	25.5	3.8
Substitution 2	7.1	30.7	23.1	7.1	19.0	15.1

Table 7: Percent of participants vs. SA3 (object color) answers. Not in Story = participant explicity wrote that the narrator did not say.

Answer	AE	EU	UE	AU
Not in Story	21.4	0.0	7.7	0.0
Blank; Knew	64.3	15.3	7.7	0.0
Blank; Didn't know	0.0	7.7	0.0	42.9
Purple	14.3	61.5	38.9	7.1
Green	0.0	15.3	38.4	0.0
Other	0.0	0.0	0.0	42.9

Four instances of the word "reach" are heard in the story, and all occur during the first half of the story. One instance of "discouraged" occurs in the story and it occurs in the first half of the story. The two substitutions in Table 6 are "attain" replaced by "reach" and "speak disparagingly of" replaced by "be discouraged by." The short answer question about the story moral shows the same results as for the FIB questions. Exaggerated motions are associated to both (1) more correct responses and (2) a higher percentage of the answer substitions for incorrect responses made with contextual-synonyms heard during exaggerated motions.

The two short answer questions that asked participants to tell the story title and color of the object that the robot was trying to reach were not explicitly given in the story. Across all four conditions, the participants who saw only EX had the highest correct response rates in both Tables 7 and 8. Furthermore, feasible answers for color of the grapes such as green or purple for grapes are concentrated in the conditions that had at least half of the motions exaggerated. Blank answers were sorted during during the interview: *knew* (participant knew that the narrator did not say the answer) and *didn't know* (ambiguous or uncertain answers).

Based on the results presented, interaction task performance is improved with exaggerated motion because exaggerated motion leads to higher correct response rates for questions about the interactive experience with the robot. There is a distinct benefit of using exaggerated motion, when remembering details about the interaction is important.

7.4 Humans Prefer EX Over UN

In support of H4, two Likert variables provide evidence that humans prefer exaggerated motion: entertaining and engaging. ANOVAs conducted on all the data from both experiments assuming that all four conditions (AE, EU, UE, AU) belong to the same distribution yielded F-values of 3.56 and 3.17 for entertaining and engaging respectively. Both of these values exceed the F_{crit} of 2.68, which indicates that at

Table 8: Percent of participants vs. SA1 (story title) answers. Not in Story = participant explicitly wrote that the narrator did not say.

Answer	AE	EU	UE	AU
Not in Story	21.4	0.0	0.0	0.0
Blank; Knew	64.3	46.1	46.1	7.1
Blank; Didn't know	7.1	7.7	0.0	28.6
Adaptation of an Aesop Fable	7.1	46.1	46.1	14.3
The Robot & the [Sour] Grapes	0.0	0.0	7.7	42.8
Other Incorrect	0.0	0.0	0.0	7.1

least one of the four conditions is statistically different.

For the both the entertaining and engaging Likert variables, three of the six post-hoc t-test pairings exhibit statistically significant results ($p<0.05$): (AE,EU), (AE,UE), and (AE,AU). Participants who saw at least half of the story motions exaggerated through modification by our algorithm indicated that the motions are more entertaining and more engaging than participants who saw only original motions. Thus, motions produced by our algorithm are more entertaining and more engaging than the input motions.

Returning to the subjective α_{exp} data discussed in Section 7.2, in which participants were asked to find values of α_{exp} that are most human-like, most cartoon-like, most robot-like, most visually pleasing, and liked best, only 3 of 13 (EX,VP) and 5 of 13 (EX,LB) pairs showed statistical significance $p<0.05$; thus, the majority of motions showed no statistical difference between exaggeration and visually pleasing or exaggeration and liked best.

In regard to H4, participants also expressed a preference for exaggerated motion in favorite motion and story part. If all four conditions are included (i.e. AE, EU, UE, AU) from both experiments, then 62.2% of participants chose a favorite part of the story associated with an exaggerated motion and 64.8% of participants selected an exaggerated motion from one of the thirteen as their favorite motion in the story. However, in only two conditions did participants actually have a choice (i.e. for EU and UE). Excluding the other two conditions, these percentages increase to 75% and 80% respectively. Participants selected exaggerated motion or favorite story parts because these were described as "animated," "expressive," or "emotional."

7.5 EX Can Be Used To Direct Attention

The final set of results addresses whether exaggerated motion can be used to direct attention to salient body parts (H5). To draw meaningful conclusions from the eye gaze data collected with the faceLAB system, we need a measure of exaggeration that evaluates body part exaggeration, since the eye gaze data is a function of body part. α_{algo} is not sufficient because it corresponds to joint-space exaggeration.

Exaggeration in joint-space corresponds to an increase in energy for the motor moving the exaggerated DOF. Since the exaggeration for a particular body part is a function of all of the parent joints on the hierarchy relative to the body part, cumulative energy ratio (EX over UN) is used to represent exaggeration in Cartesian space.

The robot was discretized into 12 segments (i.e. non-overlapping body parts). Using the data from Experiment 2, the intersection of each of the two eye gaze trajectories with the 12 body part segments was determined and accumulated over all participants. Using these trajectories, percent of total trajectory time that each participant spent watching each body part could be determined for each motion in the story. Assuming both groups of motions (UN and EX), all body parts, and all motions belong to the same distribution, the ANOVA yielded an F-value of 67.4, which was greater than $F_{crit} = 3.9$; therefore post-hoc analysis is required.

Assuming each of the distributions are different according to body part and motion from the story, (UN, EX) data pairs exist. Each of these 156 pairs (12 body parts x 13 motions) represent distributions for the amount of time participants spent watching a particular body part for a particular motion. 151 of 156 pairwise t-tests were statistically significant ($p<0.05$), which indicates that participants watch the same body part in the same motion for different lengths of time. The five tests that failed to yield statistical difference were (scan, left hand), (scan, torso top), (scan, torso mid), (scan, head), and (wave, left hand). Scan and wave are both perfomed with the left hand; during gestures, humans focus on the symbol formed by the salient hand. Additionally, the head is near the left hand for a significant portion of the scan gesture, which could help explain these results.

The 156 pairwise t-tests were repeated using only data from EU and UE to create the distributions to determine if the same participants will change their behavior when they see exaggerated and unexaggerated motion. This time, 153 of 156 pairwise t-tests were statistically significant ($p<0.05$). The three pairings that were not statistically different with respect to percent of gaze time for a specific motion and body part were (scan, left hand), (scan, head), and (wave, left hand). Thus, the *same* participants will change their watching behavior when watching UN or EX.

To test whether attention is directed toward exaggerated body parts, we employ our Cartesian measure for exaggeration and plot the average percent time watching both exaggerated and unexaggerated motions against cumulative energy ratio (CNR) for the point trajectory that represents the body part centroid as a function of time (Equation 3).

$$CNR = \frac{\sum_{t=0}^{t=T_f} (v_x(t)_e^2 + v_y(t)_e^2 + v_z(t)_e^2)}{\sum_{t=0}^{t=T_f} (v_x(t)_u^2 + v_y(t)_u^2 + v_z(t)_u^2)} \tag{3}$$

where,
$v_i(t)$ = velocity of body part centroid in i-direction at time t, where i={x, y, or z}
T_f = final time sample for discrete motion trajectory
u = denotes unexaggerated motion trajectory
e = denotes exaggerated motion trajectory

The two plots of average percent time spent watching versus CNR for UN and EX are shown in Figures 2 and 3, respectively. Since trajectories for body part motions with higher CNR are more exaggerated, the horizontal axis is a measure of exaggeration. Each point on the plots represents the data for one body part from one motion. In Figure 3, on average participants spent more time watching body parts with more exaggeration. The absence of a trend in Figure 2 provides evidence that exaggeration produces this effect upon participant behavior. I.e. exaggerated motion is used to direct attention to body parts that are exaggerated more, and attention is directed away from body parts with diminished motion. We conclude that hypothesis H5 holds true.

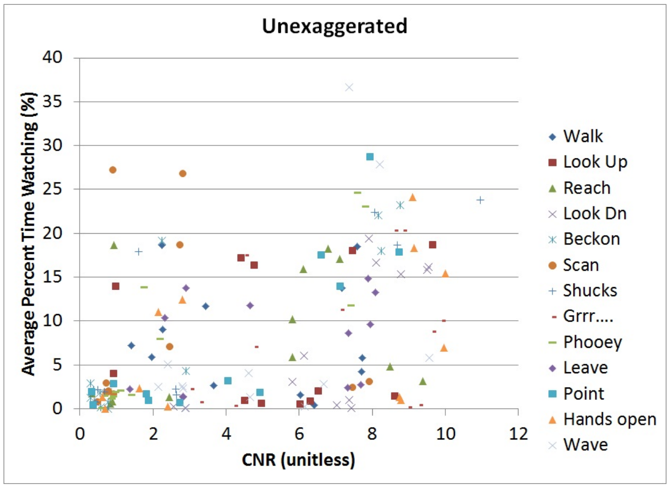

Figure 2: Average percent of time participants spent watching a specific body part in an unexaggerated trajectory vs. cumulative energy ratio for that body part centroid trajectory.

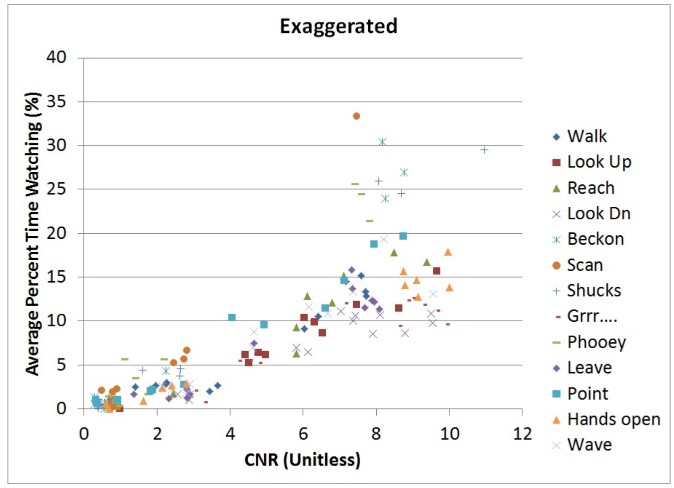

Figure 3: Average percent of time participants spent watching a specific body part in an exaggerated trajectory vs. cumulative energy ratio for that body part centroid trajectory.

8. DISCUSSION

Some might claim that the existence of a tunable parameter, α, is a disadvantage of our algorithm. Perhaps a systematic method for determing an "optimal" α for any motion is preferred. However, we believe that optimality is context-dependent (e.g. we provide optimal values that correspond to robot-like, human-like, and cartoon-like). Since we do not believe that one α value is optimal for all situations, α can be exploited to control the amount of exaggeration at any given time based on whether a robot needs to attract attention, increase engagement of human partners, or simply move in a more human-like manner.

9. CONCLUSIONS

We presented an algorithm that creates exaggerated variants of an input motion in real-time. Our experimental data confirmed that exaggerated motion is perceptibly different than the input motion, provided that they motion is sufficiently exaggerated. We found that various levels of exaggeration in motion correlate to human expectations of robot-like, human-like, and cartoon-like motion. Use of exaggerated motion enhances the interaction through the benefits of increased partner engagement and perceived entertainment value. We provided statistical evidence that exaggerated motion also increases retention of interaction details. Exaggerated motion changes the salient body parts and viewing durations when motion is executing so that observer gaze is directed toward exaggerated body parts for longer time periods and away from body parts with less energy.

10. ACKNOWLEDGMENTS

This research is supported by NSF CAREER grant IIS-1032254.

11. REFERENCES

[1] A. Bruderlin and L. Williams. Motion signal processing. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 97–104, 1995.

[2] A. Bur et al. Dynamic visual attention: Motion direction versus motion magnitude. *Proceedings of the SPIE 20th Annual Symposium*, 6806, January 2008.

[3] M. Gielniak et al. Generalized motion through adaptation of velocity profiles. *ROMAN*, 2010.

[4] M. Gielniak et al. Secondary action in robot motion. *ROMAN*, 2010.

[5] M. Gielniak et al. Anticipation in robot motion. *ROMAN*, 2011.

[6] C. Howard and A. Holcombe. Unexpected changes in direction of motion attract attention. *Attention, Perception, and Psychophysics*, 72(8), 2010.

[7] O. Johnston and F. Thomas. *The illusion of life: Disney animation*. Disney, 1995.

[8] S. Lang et al. Providing the basis for human-robot interaction: A multi-modal attention system for a mobile robot. In *ICMI*, 2003.

[9] C. Liu et al. Motion magnification. In *SIGGRAPH*, 2005.

[10] B. Mutlu et al. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proc. Humanoids*, 2006.

[11] L. Takayama et al. Expressing thought: Improving robot readability with animation principles. *Proc. of HRI*, 2011.

[12] M. Unuma et al. Fourier principles for emotion-based human figure animation. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 91–96, 1995.

[13] J. Wang et al. The cartoon animation filter. In *SIGGRAPH*, 2006.

[14] Y. Ye and C. Liu. Animating responsive characters with dynamic constraints in near-unactuated coordinates. *ACM Trans. Graph.*, 27(5), 2008.