Modeling and Evaluating Narrative Gestures for Humanlike Robots

Chien-Ming Huang, Bilge Mutlu
Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI 53706
E-mail: {cmhuang,bilge}@cs.wisc.edu

***Abstract*—Gestures support and enrich speech across various forms of communication. Effective use of gestures by speakers improves not only the listeners' comprehension of the spoken material but also their perceptions of the speaker. How might robots use gestures to improve human-robot interaction? What gestures are most effective in achieving such improvements? This paper seeks answers to these questions by presenting a model of human gestures and a system-level evaluation of how robots might selectively use different types of gestures to improve interaction outcomes, such as user task performance and perceptions of the robot, in a narrative performance scenario. The results show that robot deictic gestures consistently predict users' information recall, that all types of gestures affect user perceptions of the robot's performance as a narrator, and that males and females show significant differences in their responses to robot gestures. These results have strong implications for designing effective robot gestures that improve human-robot interaction.**

I. Introduction

People use gestures in different forms of communication, from narration [20] to conversations [20], across a wide range of social settings that include instruction [1, 19, 29] and spontaneous and rehearsed speech [16]. In these contexts, gestures communicate semantic information [14], visualize imagery [20], draw the attention of the recipients [12], regulate discourse [3], and enhance communication by disambiguating references and supplementing speech with additional information [14]. In return, gestures help recipients comprehend the presented information [10, 14] and form opinions about the speaker [16], thus serving as a key communicative mechanism for both speakers and recipients [14, 20].

How might robots use these communicative mechanisms to improve human-robot interaction? What gestures are most effective in achieving improvements in outcomes like user task performance and perceptions of the robot? Research in human-robot interaction has explored how robots might use gestures to give directions [25], to coordinate joint activities [30], and to persuade their users [7]. While these studies demonstrate the potential robot gestures hold in shaping human-robot interaction, research to date has not explored how robots might selectively use different types of gestures to target improvements in specific interaction outcomes such as user task performance, perceptions of the robot's effectiveness, and perceptions of the social and affective characteristics of the interaction. Such an exploration will not only provide robot designers with a better understanding of how to maximize certain outcomes, such as targeting improved student learning in the design of an instructional robot, but it will also inform researchers on how robots might go beyond human capabilities in effective communication. However, this exploration bears methodological challenges. It needs to consider interconnections between different types of gestures and between gestures and other channels of communication such as speech and gaze [34], thus requiring a system-level study instead of following conventional experimental paradigms, such as manipulating gestures in isolation and measuring their effects in interaction.

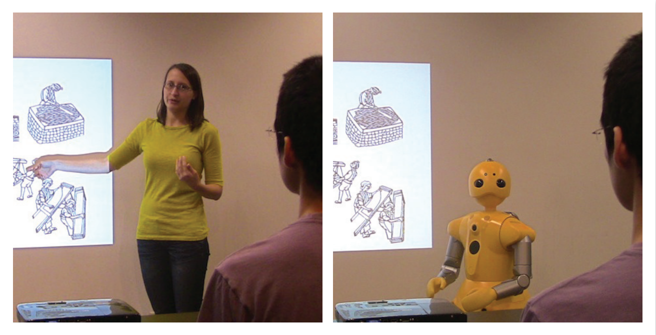

Fig. 1: Gestures of human narrators (left) were modeled, implemented into a humanlike robot, and evaluated in a human-robot interaction scenario (right).

To answer the research questions posited above and address these methodological challenges, this paper presents a *model* of how people coordinate their gestures with their speech and gaze behaviors and a *system-level evaluation* [26] of a robot's use of this model to display different types of gestures toward improving participants' recall and retelling of information and perceptions of the robot's effectiveness and the social and affective characteristics of the interaction (Figure 1). More specifically, this work makes two key contributions: (1) a model of how four key types of gestures—*deictic*, *iconic*, *metaphoric*, and *beat* gestures—temporally and semantically align with gaze and speech in humans and (2) a system-level evaluation that uncovers the predictive relationships between different types of gestures and specific outcomes in human-robot interaction.

The remainder of the paper is organized as follows. Section II reviews prior research on human and robot gestures and introduces the system-level evaluation research paradigm. Section III outlines our process for modeling human gestures and implementing the model to a humanlike robot. Section IV presents the methodology for and results from the system-level evaluation. Sections V and VI discuss the findings, design implications, and limitations of this work and provide a summary of its contributions.

II. Background

A. Gestures in Human Communication

Conversations across different tasks, cultures, and age groups involve people *gesturing* by moving their hands and arms [9]. For speakers, these gestures convey information in support of their speech, enabling them to reinforce or supplement spoken content [14, 20]. For recipients, the accompanying information facilitates the comprehension of the spoken content [10, 14]. These benefits are observed across a broad range of communicative settings including narrative [20], conversational [14, 20], and instructional [19] communication and across cultures [11, 15, 20]. The ability to interpret gestures also has developmental significance [10], making gestures particularly important for teaching and learning [29]. For example, education research has shown that gestures help young children recall information [36] and learn new algebraic concepts [1]. Moreover, gestures contribute to the shaping of students' perceptions of the teacher and keep them motivated and engaged in the classroom [27].

While people use gestures in a wide range of communicative settings and with different communicative goals [8], researchers have identified particular patterns in which people display gestures and have proposed classifications for these patterns [17, 20]. The majority of these classifications agree that human gesture is composed of four typical types of movements. Following the terminology used by McNeill [20], these movements include (1) *deictics*, (2) *beats*, (3) *iconics*, and (4) *metaphorics*. Deictics point toward concrete objects or abstract space in the environment to call attention to references. Beats include short, quick, and frequent up-and-down or back-and-forth movements of the hands and the arms that co-occur with and indicate significant points in speech, such as the introduction of a new topic and discontinuous parts in discourse. Iconics depict concrete objects or events in discourse, such as drawing a horizontal circle with the arms while uttering "a big basket." Finally, metaphorics visualize abstract concepts or objects through concrete metaphors, such as using one hand to motion forward to indicate future events and motion behind one's self to refer to past events. Also referred to as *representative* gestures, deictic, iconic, and metaphoric gestures are closely related to the semantics of speech. Figure 2 illustrates a human narrator and a robot performing these four types of gestures.

B. Gestures in Human-Robot Interaction

Research in human-robot interaction has also recognized the importance of gestures as a key mechanism for human-robot communication, particularly exploring how robots might (1) recognize human gestures to help understand human intent and language, (2) display humanlike gestures to support their speech, and (3) use gestures in specific patterns to improve human-robot interaction. While recognition of human gestures has long been a research topic in robotics and computer vision, recent research has explored how gesture recognition might facilitate human-robot interaction [6, 24, 35].

Research on realizing gestures for robots has been focused primarily on a subset of the four typical types of gestures, such as how robots might use deictic gestures to give directions to their users [25] or to learn a task from their users [35]. This line of research also includes novel approaches to and control architectures for generating gestures [4, 23, 30]. For instance, Salem et al. [30] developed a control architecture for deictic and iconic gestures for a humanoid robot performing in a human-robot joint task. Similarly, Bremner et al. [4] introduced a gesture production approach based on actuator end-point and trajectory to generate open-hand gestures. Finally, Ng-Thow-Hing et al. [23] proposed a probabilistic model for synchronizing gestures and speech and conducted a video-based evaluation to explore how manipulating model parameters might produce gestures with different levels of expressivity.

Another body of work in human-robot interaction involves investigating how robot gestures affect people's perceptions of the robot and their experiences. This thread of research has shown that gestures positively shape participants' affective states [22], behavioral responses to the gestures [28], engagement in the interaction [5], and perceptions of the robot when gestures and speech are mismatched [30]. Such research involved laboratory studies in which participants performed a task with a physical robot [5, 30] or video-based studies in which participants observed robots performing gestures [22, 23, 28].

These different lines of research advance knowledge in how robots might use gestures toward improving human-robot interaction and highlight promise that gestures hold for shaping user experience and improving people's perceptions of robots. However, for robots to realize the full potential of using gestures, a better understanding of how they might selectively use different types of gestures to target improvements in specific outcomes in human-robot interaction is needed.

C. System-Level Evaluation

To understand the relationship between robots' use of different types of gestures and interaction outcomes in human-robot interaction, we followed the system-level evaluation paradigm [26]. This approach is based on multivariate regression analysis, in which predictor variables are randomly varied and the response variable is measured with the goal of understanding how different predictor variables are related to each other and how they contribute to the response variable. This approach was first introduced to evaluate dialogue systems [37] and has recently been used to study the effectiveness of an interactive robot system in an object learning scenario [26].

III. Gesture Modeling and Design

To answer our research questions, we *modeled* the four types of human gestures, *implemented* these gestures on a humanlike robot, and *evaluated* how robot gestures might affect specific measures in human-robot interaction. This section outlines the modeling and implementation steps in this process.

A. Modeling Human Gestures

To study how people use gestures, we developed a narration scenario in which a narrator described "the process of making paper" with the aid of a projected figure that depicted the

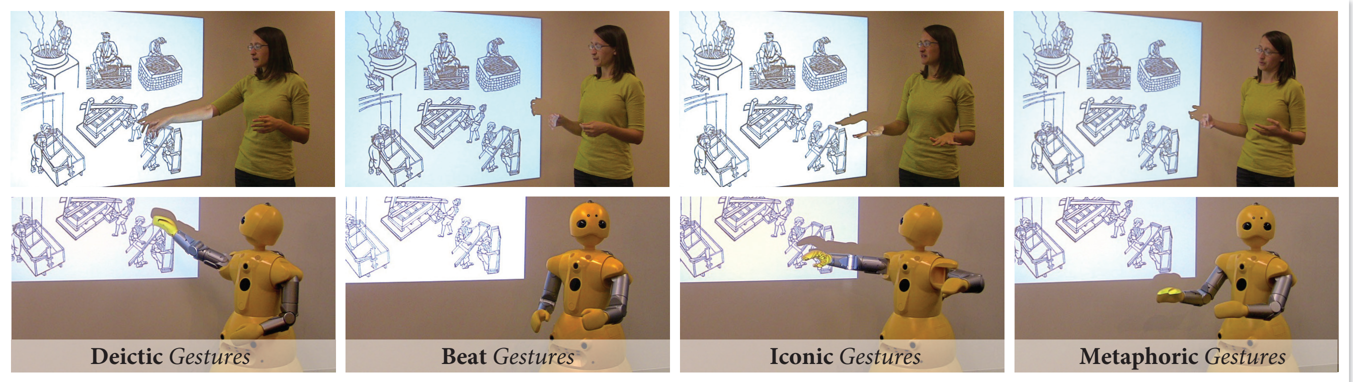

Fig. 2: Examples of the four common types of gestures—*deictic*, *beat*, *iconic*, and *metaphoric* gestures—observed in human storytellers (top) and implemented into the robot (bottom). The narrator uses deictic gestures to point toward an object of reference, beat gestures before introducing a new concept, iconic gestures to depict a concrete object such as "a flat wooden surface," and metaphoric gestures to visualize abstract concepts such as "about six hours."

process to a recipient, as shown in Figure 1. We chose narration as the setting for the study because narration naturally elicits the use of a rich set of gestures [20] and fully engages the observer with the narrator's behaviors. Narration also serves as an appropriate scenario for human-robot interaction, as robots are envisioned to provide their users with information following the norms and structure of this setting. Finally, narration provides us with a rich set of metrics to measure outcomes such as story recall and perceptions of the narrator.

To model how human narrators employ gestures, we recruited four dyads of participants, aged 21.6 years on average ($SD = 1.77$), and matched them to represent all gender combinations (i.e., MM, MF, FM, and FF). For each dyad, one participant acted as the narrator, while the second acted as the recipient. The narrator was given text, pictures, and videos on the topic of the process of making paper and asked to review them for approximately 25 minutes. After the preparation phase, the narrator presented the process to the recipient. The average time for narrations across trials was 4.49 minutes ($SD = 1.13$). The trials were videotaped for behavioral coding and analysis. A primary rater coded all of the data, and a secondary rater coded 10% of the data to assess inter-rater reliability. The reliability analysis showed perfect agreement for gesture type (Cohen's $\kappa = .845$) and gaze target (Cohen's $\kappa = .916$) based on guidelines suggested by Landis and Koch [18].

B. Developing a Gesture Model for Robots

Gesture and speech are co-expressive channels in human communication [14, 20]. In this work, we modeled two particular aspects of gesture with respect to speech: *gesture points*, the points in speech where the speaker displays gestures, and *gesture timing*, the times when a gesture begins and ends. We also modeled *gesture-contingent gaze cues*, gaze cues displayed during gesturing, based on research that has identified interdependencies between gestures, speech, and gaze [32, 33].

1) Gesture Points: Utterances involve *lexical affiliates*—words and phrases that co-express meaning with representative gestures, including deictic, iconic, and metaphoric gestures [31]—that inform us on when a robot might need to gesture and what type of gesture it might perform. To capture this relationship, we identified lexical affiliates associated with representational gestures and used affinity diagramming to group them into categories of gesture points for each type of gesture (Figure 3). These categorizations showed that deictic gestures were frequently used to describe references, including visual representations of steps and objects involved in the process of making paper that appeared on the projected figure, and in conjunction with pronouns. Iconic gestures were mostly observed during descriptions of actions and concrete objects, while metaphoric gestures were generally observed when the speaker described abstract concepts involving actions, relative quantities, or time. Because beat gestures connect an utterance not at the semantic level but rather at the structural level [20], we did not empirically identify lexical affiliates for beat gestures. Instead, discontinuities in speech, such as introducing a new concept, served as gesture points for these gestures [20].

2) Gesture Timing: Gestures and utterances are closely related in the temporal domain [20]. The timing of a gesture might affect how people perceive and interpret it. While research has suggested that the *stroke* of a gesture ends before the end of its lexical affiliate [20], when a complete gestural phrase should begin and end relative to its lexical affiliate has not been specified. In this work, we empirically obtained these temporal parameters for representative gestures, as summarized in Figure 4, which confirmed that gesture initiation typically precedes the onset of lexical affiliates [20, 31].

3) Gesture-Contingent Gaze Cues: The distribution of gaze cues during each type of gesture was modeled in order to coordinate the production of gaze, gesture, and speech. We

Gesture	Categories	#	%	Example
Deictics *54 items*	Concrete references	30 items	55.6 %	*"the stamper"*
	Abstract references	18 items	33.3 %	*"the cooking process"*
	Pronouns	10 items	18.5 %	*"this person here"*
Iconics *105 items*	Action verbs	45 items	42.9 %	*"peel it off"*
	Nouns	44 items	41.9 %	*"a big basket"*
	Descriptors	8 items	7.6 %	*"the thickness"*
Metaphorics *51 items*	Actions	12 items	23.5 %	*"the order that we went in"*
	Relative quantities	8 items	15.7 %	*"more weight"*
	Time	7 items	13.7 %	*"the next day"*

Fig. 3: Top lexical affiliate categories for each type of representative gesture. Percentages represent the amount by which categories of lexical affiliates co-occur with each gesture type. For example, 42.9% of the lexical affiliates for iconic gestures are "action verbs." Note that categories might overlap.

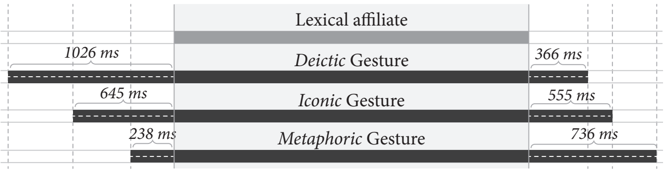

Fig. 4: A schematic of temporal alignment between gestures and lexical affiliates (not to scale). For instance, iconic gestures began on average 645 ms before and ended 555 ms after their corresponding lexical affiliates.

categorized gesture-contingent gaze cues into four gaze targets—*recipient*, narrator's *own gesture*, *reference*, and all *other* places. An additional category for *traveling*, the time spent transitioning between the four categories, was also created (Figure 5).

People tend to look toward shared references during conversation [2]. Our data showed a similar trend, as narrators looked toward references most of the time while gesturing. This behavior was observed during deictic, beat, and iconic gestures and was particularly notable for deictic gestures, during which speakers looked toward the reference approximately 83% of the time. During metaphoric gestures, narrators looked toward references and recipients at equal rates. Interestingly, approximately 10% of narrator gaze during iconic gestures was directed toward the gestures, which narrators might display to direct the recipient's attention to the gesture and to signal that the current gesture is relevant to the ongoing utterance [34]. Narrators did not display this behavior during metaphoric gestures, potentially because the abstract lexical affiliates with which they are associated might require speakers to look toward the recipient to affirm mutual understanding of their speech.

C. Implementation

The robot's gestures were designed based on our observations of the human narrators' gestures in the modeling study. While different narrators varied slightly in how they performed gestures at a given gesture point, they displayed semantically common elements. For example, when describing "beating (paper) with a stick," participants displayed one-handed or two-handed up and down movements at different speeds and with different degrees of tilt. For each unique gesture point, we created one robot gesture that captured the common elements that we observed from the human narrators displayed at that gesture point. Robot gestures were created through puppeteering, which involved manually moving the robot's arms while recording key frames of the gestural trajectories.

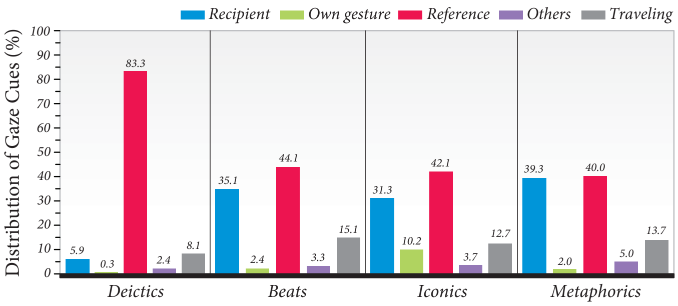

Fig. 5: Distributions of targets for gesture-contingent gaze for each type of gesture. The human data showed four main gaze targets: the *recipient*, the narrator's *own gesture*, the *reference*, and *other*, non-task-relevant targets. *Traveling* represents transitions between these targets. Narrators gazed most toward references while displaying deictic gestures but split their gaze evenly between the recipient and references while displaying other types of gestures.

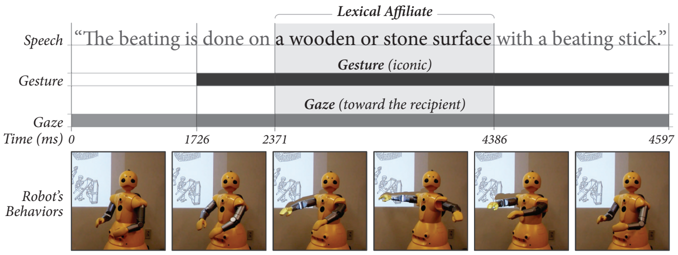
Lexical Affiliate
Speech
"The beating is done on a wooden or stone surface with a beating stick."
Gesture
Gesture (iconic)
Gaze (toward the recipient)
Gaze
Time (ms)
0
1726
2371
4386
4597
Robot's Behaviors

Fig. 6: An example utterance from the robot's narration, including the lexical affiliate "a wooden or stone surface," its corresponding iconic gesture, and the gesture-contingent gaze behavior of looking toward the recipient.

For implementation and evaluation, we used the narrative script that the participants used for preparation in the modeling study. We manually marked all possible gesture points in the script based on the lexical affiliate categories shown in Figure 3, identifying 30, 30, and 25 points for deictic, iconic, and metaphoric gestures, respectively. Twenty-four points for beat gestures were marked based on heuristic principles [20]. Gestures created for the robot were manually assigned to these gesture points. The gesture-contingent gaze cues were produced by mirroring the distributions shown in Figure 5.

We implemented the gesture model and its contingent gaze cues on a Wakamaru humanlike robot shown in Figure 1 using the Robot Behavior Toolkit, a Robot Operating System (ROS) module for controlling multiple channels of robot behavior [13]. The Appendix illustrates the algorithm for synchronizing gestures and gaze with the robot's speech. Figure 6 illustrates an example synchronization of speech, gesture, and gaze cues.

IV. Evaluation

A. Study Design and Procedure

Following a system-level evaluation paradigm, we manipulated the amount by which the robot displayed each type of gesture and measured how this variability affected interaction outcomes. For each gesture type, we marked all possible gesture points at which human narrators displayed gestures of that type. To manipulate the *amount* by which the robot would display gestures of that type, a random number between 0 and the number of possible points (i.e., 30, 30, 25, and 24 for deictic, iconic, metaphoric, and beat gestures, respectively) was drawn from a uniform distribution, which served as the amount by which the robot would display gestures of that type during its narration. To manipulate the gesture *points* at which the robot would display gestures, a subset of gesture points that matched this percentage amount was randomly selected from the set of all possible gesture points. For each participant, this process was repeated for all gesture types, producing the necessary variability in the robot's gestures to investigate how well different gestures predicted interaction outcomes. Section IV-C outlines how we modeled this predictive relationship.

The evaluation study started when the experimenter obtained the participant's consent to taking part in the study. Following this step, the experimenter asked the participant to be seated across from the humanlike robot and to listen to the robot tell a story on the process of making paper, as shown in the right image in Figure 1. The robot's narration lasted approximately six minutes. After the story, the participant was asked to complete a five-minute distractor task followed by a quiz on the topic of the narration. The participant was then asked to retell the process of making paper in the same experimental setting, which was videotaped for later analysis. The experiment concluded with a post-experiment questionnaire for evaluating the participant's perceptions of the robot. For each participant, the experimental protocol took approximately 30 minutes. Each participant received $5 for taking part in the study.

B. Measures

We developed objective, subjective, and behavioral measures to understand how the robot's use of different types of gestures affected the participants' information recall, their perceptions of the robot, and their ability to retell the robot's story.

The primary objective measure was how accurately the participants recalled the information presented by the robot, measured by a quiz consisting of 11 multiple-choice or true-or-false questions. The subjective measures sought to capture the participants's perceptions of the robot in terms of naturalness of behavior (5 items; Cronbach's $\alpha = .78$), competence (8 items; Cronbach's $\alpha = .82$), and effective use of gestures (2 items; Cronbach's $\alpha = .81$). We also measured their evaluations of their engagement (8 items; Cronbach's $\alpha = .82$) and rapport with the robot (6 items; Cronbach's $\alpha = .83$). Participants responded to these measures using seven-point rating scales.

In addition to the objective and subjective measures, we measured how the robot's gestures affected the participants' ability to retell the robot's story, particularly the participants' articulation of the process of making paper, as indicated by story length, and use of gestures during narration.

C. Analysis Method

We used a backward stepwise multivariate linear regression to analyze the data. The linear regression method is used to model a response variable y by a linear combination of predictor variables x_i, represented as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + e$$

In the model, β_0 is a constant, and β_i is a regression coefficient for corresponding predictor variable x_i, indicating to what extent the variable predicts the response variable. e denotes the residual error of the model. The backward stepwise regression process starts with all predictors included in the model. The process iteratively removes the predictor with the highest p-value as a relatively less powerful predictor of the response variable. The iterations continue until certain stop criteria are met. We used a stop criterion of $p < .10$. The remaining predictors construct the final model for the response variable.

For the current work, we constructed linear models for nine interaction outcomes. Each model involved eight predictor variables—the four types of gestures and the four gesture-contingent gaze targets—and a response variable that represented an interaction outcome. Each predictor variable represented the percentage amount by which the robot displayed that behavior during its narration, which varied by a small margin from the amount randomly generated for the manipulation due to overlaps and conflicts in gesture production. Only the amounts by which the robot displayed gestures were manipulated. Gesture-contingent gaze cues were drawn from the distributions in the modeling study and were included in the analysis as covariates. We constructed separate models for females and males based on findings from research in human-robot interaction that show strong differences in how females and males perceive and respond to robot behaviors (e.g., [21]), producing a total of 18 models for nine interaction outcomes.

All predictor variables were standardized. A *log* transformation was applied to predictor and response variables to obtain linearity for each constructed model. A small number, 0.000001, was added to all predictor variables to avoid the singularity of taking *log* of values of zero.

D. Participants

A total of 32 participants, including 16 females and 16 males, were recruited from the University of Wisconsin–Madison campus community. The average age of the participants was 24.34 ($SD = 8.64$), ranging from 18 to 55. Based on seven-point rating scales, the participants reported that their familiarity with robots ($M = 2.47$, $SD = 1.34$) and their familiarity with the task ($M = 1.78$, $SD = 1.21$) were fairly low.

E. Results

The regression analysis yielded 15 significant models, categorized into four groups: *task performance*, *perceived performance*, *social and affective evaluation*, and *narration behavior*. These models are shown in Figure 7 and illustrated in Figure 8. The statistical tests for the models are provided in Figure 7, and the paragraphs below provide only a textual description of the results for readability. In each model, the adjusted R-squared score, R^2_A, shows the degree to which the predictor variables account for the variance in the response variable. The standardized β coefficient for each predictor represents the individual effect of the predictor on this variance, and the t-test and p-value summarize the significance of this effect. In this work, $p < .05$ and $p < .10$ were considered as significant and marginal effects, respectively. Note that, because we cannot directly manipulate gesture-contingent gaze cues, we cannot draw conclusions on how they affect interaction outcomes. Hereafter, we only highlight results on gestures.

1) Task Performance: The robot's use of *deictic* gestures significantly predicted information recall for both female and male participants, suggesting that it is important for the robot to point toward references to help the participants ground their understanding of the narration in the references, which in this particular scenario included illustrations of the steps involved in making paper. Additionally, *metaphoric* gestures significantly predicted male participants' recall performance, indicating that they might have leveraged the robot's visualization of abstract concepts such as actions and processes involved in making paper to reinforce their understanding of these concepts.

	Measure (y)	Gender	Function ($\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon$)	$R_A{}^2$	Predictor	β	T-test	Significance
Task Performance	Information Recall	**F**	$-.45 + .123 \times \mathit{Deictic} + .202$	.232	*Deictic*	*.123*	$t(14)$=2.35	$p = .034^{*}$
		M	$-.739 + .623 \times \mathit{Deictic} + .335 \times \mathit{Metaphoric} - .549 \times \mathit{Gaze}_{\mathit{Reference}} - .314 \times \mathit{Gaze}_{\mathit{Gesture}} - .195 \times \mathit{Gaze}_{\mathit{Other}} + .393$	.364	*Deictic* *Metaphoric*	*.623* *.335*	$t(10)$=3.08 $t(10)$=2.36	$p = .012^{*}$ $p = .040^{*}$
Perceived Performance	Gesture Effectiveness	**F**	$1.193 + .452 \times \mathit{Deictic} + .252 \times \mathit{Beat} + .464 \times \mathit{Metaphoric} + .224 \times \mathit{Gaze}_{\mathit{Gesture}} - .455 \times \mathit{Gaze}_{\mathit{Reference}} + .373$	.575	*Deictic* *Beat* *Metaphoric*	*.452* *.252* *.464*	$t(10)$=3.28 $t(10)$=2.10 $t(10)$=3.53	$p = .008^{**}$ $p = .062^{\dagger}$ $p = .006^{**}$
		M	$1.215 + .414 \times \mathit{Beat} - .957 \times \mathit{Gaze}_{\mathit{Listener}} - .371 \times \mathit{Gaze}_{\mathit{Gesture}} - 1.05 \times \mathit{Gaze}_{\mathit{Reference}} - .386 \times \mathit{Gaze}_{\mathit{Other}} + .356$	.640	*Beat*	*.414*	$t(10)$=4.02	$p = .002^{**}$
	Competence	**F**	$1.600 + .122 \times \mathit{Gaze}_{\mathit{Other}} + .132$	.440	*No gesture predictor*			
		M	$1.698 + .127 \times \mathit{Iconic} + .117$	.527	*Iconic*	*.127*	$t(14)$=4.21	$p < .001^{***}$
	Naturalness	**F**	$1.454 + .198 \times \mathit{Metaphoric} - .453 \times \mathit{Gaze}_{\mathit{Listener}} - .580 \times \mathit{Gaze}_{\mathit{Reference}} + .183$	.554	*Metaphoric*	*.198*	$t(12)$=3.02	$p = .011^{*}$
		M	$1.529 + .146 \times \mathit{Iconic} - .117 \times \mathit{Metaphoric} + .190$	.333	*Iconic* *Metaphoric*	*.146* *−.117*	$t(13)$=2.78 $t(13)$=−2.23	$p = .016^{*}$ $p = .044^{*}$
Social & Affective Evaluation	Rapport	**F**	$1.212 - .673 \times \mathit{Gaze}_{\mathit{Listener}} - .198 \times \mathit{Gaze}_{\mathit{Gesture}} - .689 \times \mathit{Gaze}_{\mathit{Reference}} + .330$	.121	*No gesture predictor*			
		M	$1.179 + .281 \times \mathit{Deictic} - .608 \times \mathit{Gaze}_{\mathit{Listener}} - .246 \times \mathit{Gaze}_{\mathit{Gesture}} - .718 \times \mathit{Gaze}_{\mathit{Reference}} + .250$	.345	*Deictic*	*.281*	$t(11)$=2.54	$p = .028^{*}$
	Engagement	**F**	$1.575 - .127 \times \mathit{Metaphoric} - .418 \times \mathit{Gaze}_{\mathit{Listener}} - .347 \times \mathit{Gaze}_{\mathit{Reference}} + .174$	.279	*Metaphoric*	*−.127*	$t(12)$=−2.04	$p = .064^{\dagger}$
		M	$1.589 - .120 \times \mathit{Metaphoric} + .242$	.152	*Metaphoric*	*−.120*	$t(14)$=−1.92	$p = .076^{\dagger}$
Narration Behavior	Gesture Use	**F**	$-1.229 + .830 \times \mathit{Metaphoric} + .621 \times \mathit{Gaze}_{\mathit{Listener}} + .698 \times \mathit{Gaze}_{\mathit{Gesture}} + .742$	.482	*Metaphoric*	*.830*	$t(12)$=3.02	$p = .011^{*}$
		M	$-.800 + .289 \times \mathit{Iconic} + 1.300 \times \mathit{Gaze}_{\mathit{Listener}} + 1.209 \times \mathit{Gaze}_{\mathit{Reference}} + .244 \times \mathit{Gaze}_{\mathit{Others}} + .468$	.409	*Iconic*	*.289*	$t(11)$=2.11	$p = .059^{\dagger}$
	Narration Duration	**F**	*No significant model*	*N/A*	*N/A*			
		M	$11.740 + .511 \times \mathit{Deictic} + .204 \times \mathit{Beat} + .331 \times \mathit{Metaphoric} - .305 \times \mathit{Gaze}_{\mathit{Listener}} - .346 \times \mathit{Gaze}_{\mathit{Gesture}} - .819 \times \mathit{Gaze}_{\mathit{Reference}} - .109 \times \mathit{Gaze}_{\mathit{Others}} + .117$	.840	*Deictic* *Beat* *Metaphoric*	*.511* *.204* *.331*	$t(8)$=7.50 $t(8)$=5.38 $t(8)$=7.06	$p < .001^{***}$ $p < .001^{***}$ $p < .001^{***}$

Fig. 7: Summary of significant models. For each interaction outcome, models for both genders, $R_A{}^2$ values, and details of statistical tests for significant gesture predictors are presented. There was no significant model for narration duration for females. (†), (*), (**), and (***) denote $p < .10$, $p < .050$, $p < .010$, and, $p < .001$, respectively. Significance was assessed using two-tailed t-tests. β coefficients are standardized and therefore comparable across gestures.

2) Perceived Performance: Perceived performance involved three aspects: effectiveness of the robot's gestures, perception of competence, and naturalness of the robot's behavior. Females' ratings of the effectiveness of the robot's gestures were significantly predicted by the robot's use of *deictic* and *metaphoric* gestures and marginally predicted by its use of *beat* gestures, while males' ratings were significantly predicted only by *beat* gestures. The robot's use of *iconic* gestures significantly predicted males' perceptions of the robot's competence, while no gestures predicted female perceptions. The robot's use of *metaphoric* gestures positively predicted female participants' perceptions of the naturalness of the robot's behaviors while negatively predicting those of males. On the other hand, *iconic* gestures positively predicted male participants' perceptions of the naturalness of the robot's behaviors. These results suggest that a rich use of different types of gestures might positively shape the participants' overall perceptions of the robot's performance and that employing different gesture strategies might be beneficial to maximize perceived performance outcomes for females and males.

3) Social and Affective Evaluation: The robot's use of *deictic* gestures positively predicted male participants' rapport with the robot, while no particular type of gesture predicted this outcome for females. Surprisingly, *metaphoric* gestures marginally but negatively predicted how engaged with the robot males and females reported themselves. These findings indicate that the robot's gestures were less influential on the participants' evaluations of the social and affective characteristics of the interaction, which might be due to the limited interaction afforded by the narrative performance scenario.

4) Narration Behavior: The robot's use of *deictic*, *beat*, and *metaphoric* gestures significantly predicted the length of male participants' retelling of the robot's story. Moreover, male participants who retold the story longer also performed significantly better in the recall test, $\beta = 0.272$, $t(14) = 2.47$, $p = .027$, suggesting that they might have had better recall and thus provided more detail. However, the analysis did not show these relationships for female participants. The robot's use of *metaphoric* gestures significantly predicted how much female participants used gestures during their retelling of the robot's story, while the robot's use of *iconic* gestures marginally predicted males' use of gestures during their retelling. This finding is consistent with the differential effects of metaphoric and iconic gestures on females' and males' perceptions of the robot's performance, suggesting that participants might have employed gestures that they thought were effective.

V. Discussion

Understanding the relationship between robot gestures and interaction outcomes, particularly how robots might selectively use different types of gesture to improve specific interaction outcomes, promises significant implications for designing robotic systems that not only communicate effectively with their users to maximize certain outcomes, but also go beyond human capabilities in effective communication. This work serves as a first attempt toward building such an understanding in a narration scenario. To this end, we studied the gestures of human narrators, developed a model for controlling the gestures of narrative robots, and followed a system-level evaluation paradigm to investigate how the robot's use of different types of gestures affected participants' information recall, perceptions of the robot, and ability to retell the robot's story.

The results showed that the robot's use of *deictic* gestures helped improve information recall for both female and male participants, perceived effectiveness of the robot's gestures for females, and ratings of rapport with the robot for males. *Beat*

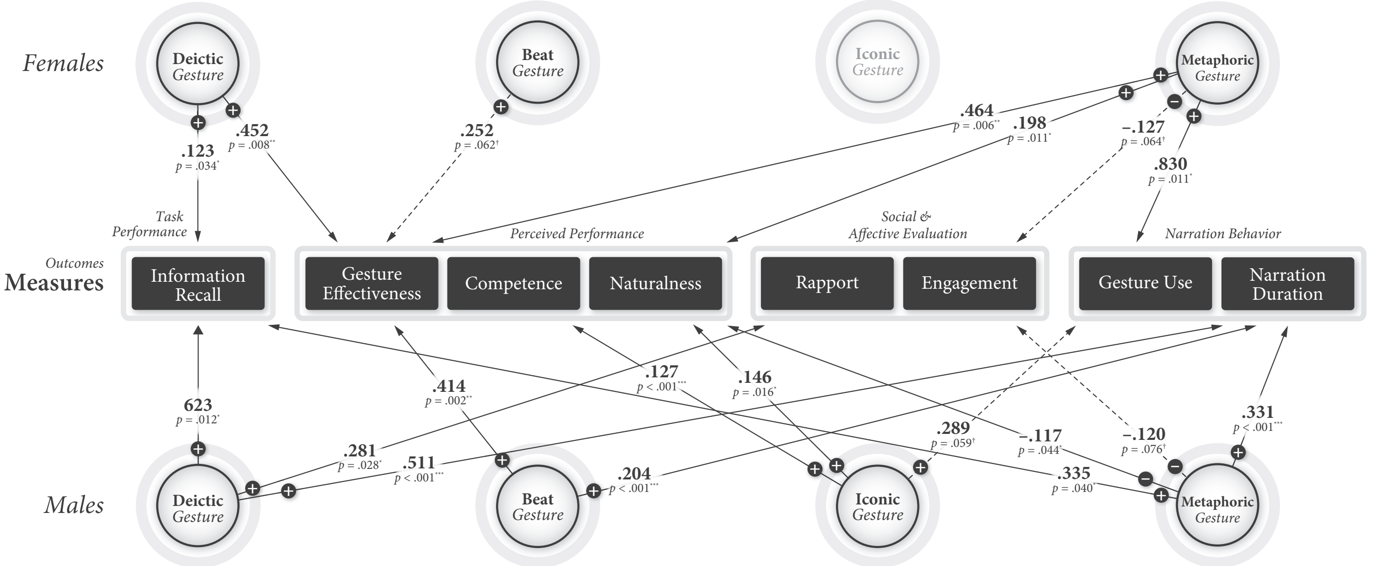

Fig. 8: A visual summary of the results, highlighting the predictive relationships between gestures and outcomes for females and males. Solid and dashed lines represent significant and marginal effects, respectively. Numbers on lines represent standardized β coefficients and p-values for predictors.

gestures positively contributed to the perceived effectiveness of the robot's gestures for both female and male participants, potentially due to their role in signaling key structural information on the discourse. *Iconic* gestures predicted male participants' perceptions of the robot's competence and the naturalness of the robot's behavior and gesture use during their retelling of the robot's story, while not predicting any outcomes for females. *Metaphoric* gestures predicted information recall for males and perceptions of the naturalness of the robot's behavior and the effectiveness of its gestures for females. Interestingly, metaphoric gestures negatively predicted the participants' engagement with the robot, indicating that more gestures do not necessarily mean better outcomes. We speculate that the abstract content and the large number of arm motions involved in this type of gesture might have been a distraction for the participants. On the other hand, these gestures contributed to the participants' ability to retell the robot's story, positively affecting narration length for males and gesture use for females.

A. Design Implications

These findings highlight the importance of robot gestures in shaping key outcomes in human-robot interaction, such as the ability to recall and retell information and perceptions of the performance of the robot and the social and affective characteristics of the interaction. Interaction designers and roboticists must leverage the design space for gestures to develop applications that maximize desired outcomes, such as increasing the use of deictic gestures by an instructional robot to improve student learning. Designers might also adapt the robot's use of the different types of gestures to the specific goals of the interaction and to user gender. For instance, the use of metaphoric gestures might be decreased if the application seeks to increase user engagement with the robot and increased if user ability to recall and retell information is important. Similarly, the robot might employ a different balance of metaphoric and iconic gestures in its interactions with females and males.

B. Limitations

The work presented here has two main limitations. First, our findings, such as the relative effects of different types of gestures on the measured outcomes, might have limited generalizability beyond the specific context of the study, requiring further work to establish the extent to which they generalize to other forms of interaction, cultural settings, and individuals with varying abilities to perceive and interpret nonverbal social cues. We expect the research approach presented here to provide the methodological basis for such future work and other research into understanding complex behavior-outcome relationships. Second, while the robot platform used in this work offered the expressivity needed to achieve research goals, hardware platforms that afford articulate hands and higher degrees of freedom would enable richer and more finely controlled gestures and thus a more thorough understanding of how robot gestures shape human-robot interaction.

VI. Conclusion

Gestures play a key role in human communication, supporting and enriching speech across various forms of interaction. While the rich space for designing robot gestures holds great potential for improving human-robot interaction, a deeper understanding the relationship between the different types of gestures and specific interaction outcomes is necessary to enable effective use of gestures by robots toward maximizing targeted outcomes. This paper sought to take a step toward building such an understanding in a narration scenario, following a process that involved modeling the gestures of human narrators, implementing these gestures into a humanlike robot, and following a system-level evaluation approach to evaluate how the robot's use of different types of gestures shaped interaction outcomes. The results provided several insights into the relative contribution of each type of gesture into different interaction outcomes. For example, deictic gestures were particularly effective in improving information recall in participants, and beat gestures contributed more to the perceived effectiveness of the robot's gestures. These gesture-outcome relationships serve as design guidelines for maximizing targeted outcomes, such as increasing an educational robot's use of deictic gestures to improve student learning. The research approach presented here might also inform future robotics research into the rich space for robot design to improve human-robot interaction.

Acknowledgments

The authors would like to thank Jingjing Du for her early help with this work and Jilana Boston for her help with behavioral coding. National Science Foundation awards 1017952 and 1149970 and and equipment loan from Mitsubishi Heavy Industries, Ltd. provided support for this work.

References

[1] M.W. Alibali and M.J. Nathan. *Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson*, pages 349–365. Cambridge U Press, 2007.

[2] M. Argyle and M. Cook. *Gaze and mutual gaze*. Cambridge U Press, 1976.

[3] K. Bergmann, H. Rieser, and S. Kopp. Regulating dialogue with gestures: towards an empirically grounded simulation with conversational agents. In *Proc. SIGDIAL'11*, 2011.

[4] P. Bremner, A. Pipe, C. Melhuish, M. Fraser, and S. Subramanian. Conversational gestures in human-robot interaction. In *Proc. SMC'11*, pages 1645–1649, 2009.

[5] P. Bremner, A. Pipe, C. Melhuish, M. Fraser, and S. Subramanian. The effects of robot-performed co-verbal gesture on listener behaviour. In *Proc. HUMANOIDS'11*, pages 458–465, 2011.

[6] L. Brethes, P. Menezes, F. Lerasle, and J. Hayet. Face tracking and hand gesture recognition for human-robot interaction. In *Proc. ICRA'04*, pages 1901–1906, 2004.

[7] V. Chidambaram, Y.-H. Chiang, and B. Mutlu. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *Proc. HRI'12*, 2012.

[8] P. Ekman and W. Friesen. The repertoire of nonverbal behavioral categories. *Semiotica*, pages 49–98, 1969.

[9] S. Goldin-Meadow. The role of gesture in communication and thinking. *Trends in cognitive sciences*, 3(11):419–429, 1999.

[10] S. Goldin-Meadow. *Hearing Gesture: How our hands help up think*. Harvard U Press, 2003.

[11] J.A. Graham. A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, 10:57–67, 1975.

[12] C.C. Heath. *Gesture's discrete tasks: Multiple relevancies in visual conduct in the contextualization of language*, pages 102–127. John Benjamins, 1992.

[13] C.-M. Huang and B. Mutlu. Robot behavior toolkit: Generating effective social behaviors for robots. In *Proc. HRI'12*, pages 25–32, 2012.

[14] A. Kendon. Do gestures communicate? a review. *Research on Language and Social Interaction*, pages 175–200, 1994.

[15] A. Kendon. *Gesture: Visible action as utterance*. Cambridge U Press, 2004.

[16] R.M. Krauss. Why do we gesture when we speak? *Current Directions in Psychological Science*, 7(2):54–60, 1998.

[17] R.M. Krauss, Y. Chen, and R.F. Gottesman. *Lexical gestures and lexical access: A process model*, pages 261–283. Cambridge U Press, 2000.

[18] J.R. Landis and G.G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[19] S.C. Lozano and B. Tversky. Communicative gestures facilitate problem solving for both communicators and recipients. *Journal of Memory and Language*, 55(1):47–63, 2006.

[20] D. McNeill. *Hand and Mind*. U of Chicago Press, 1992.

[21] B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: Modeling and evaluation of humanlike gaze behavior. In *Proc. HUMANOIDS'06*, pages 518–523, 2006.

[22] H. Narahara and T. Maeno. Factors of gestures of robots for smooth communication with humans. In *Proc. RoboComm'07*, pages 44:1–4, 2007.

[23] V. Ng-Thow-Hing, P. Luo, and S. Okita. Synchronized gesture and speech production for humanoid robots. In *Proc. IROS'10*, pages 4617–4624, 2010.

[24] K. Nickel and R. Stiefelhagen. Visual recognition of pointing gestures for human-robot interaction. *Image and Vision Computing*, 25(12):1875–1884, 2007.

[25] Y. Okuno, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Providing route directions: design of robot's utterance, gesture, and timing. In *Proc. HRI'09*, pages 53–60, 2009.

[26] J. Peltason, N. Riether, B. Wrede, and I. Lütkebohle. Talking with robots about objects: a system-level evaluation in hri. In *Proc. HRI'12*, pages 479–486, 2012.

[27] V. Richmond. *Teacher Nonverbal Immediacy: Use and Outcomes*, pages 65–82. Allyn and Bacon, 2002.

[28] L. Riek, T.-C. Rabinowitch, P. Bremner, A. Pipe, M. Fraser, and P. Robinson. Cooperative gestures: Effective signaling for humanoid robots. In *Proc. HRI'10*, pages 61–68, 2010.

[29] W.-M. Roth. Gestures in teaching and learning. *Review of Educational Research*, 71(3):365–392, 2001.

[30] M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joublin. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics*, pages 201–217, 2012.

[31] E. Schegloff. *On some gestures' relation to speech*, pages 266–296. Cambridge U Press, 1984.

[32] J. Sidnell. Coordinating gesture, talk, and gaze in reenactments. *Research on Language and Social Interaction*, 39(4):377–409, 2006.

[33] J. Streeck. The significance of gestures: How it is established. *Papers in Pragmatics*, 2(1/2):60–83, 1988.

[34] J. Streeck. Gesture as communication I: Its coordination with gaze and speech. *Communications Monographs*, 60(4):275–299, 1993.

[35] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Natural deictic communication with humanoid robots. In *Proc. IROS'07*, 2007.

[36] L.A. Thompson, D. Driscoll, and L. Markson. Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, 22:167–187, 1998.

[37] M.A. Walker, D.J. Litman, C.A. Kamm, and A. Abella. Paradise: A framework for evaluating spoken dialogue agents. In *Proc. EACL'97*, pages 271–280, 1997.

Appendix
Algorithm for Synchronizing Robot Behaviors

```
Require: Speech utterances, timestamp[onset,end]-ID pairs for lexical affiliates and beat points
for each utterance do
    new nonverbalBehavior
    for each timestamp[onset,end]-ID pair do
        gesture.selectGestureFromLibrary(ID)
        gesture.setGestureDuration(duration(timestamp[onset,end]))
        gaze.setGazeTarget(sampleFromGazeDistribution(gesture.getType()))
        gaze.setGazeDuration(gesture.getDuration())
        nonverbalBehavior.append(gesture,gaze)
    end for
    nonverbalBehavior.executeBehavior()
end for
```