Saliency Based Evaluation of Effect of Appearance on Motionese

Agrawal, Vivek
vagrawal39@gatech.edu

Das, Subhajit
das@gatech.edu

Nair, Lakshmi
lnair3@gatech.edu

***Abstract*— In the field of Human Robot Interaction, Learning from Demonstration (LfD) offers tremendous potential for human teachers to help social robots learn new skills by means of demonstration that is consistent with the task objective and solution approach. A critical success factor in LfD scenarios is the ability of the robot to identify salient points from the interaction that would be most useful in its replication of the task. Ability of humanoid robots to engage their human teachers in a more interactive manner that brings out natural *motionese* and related *scaffolding* actions can help robots in identifying the most important aspects of the task. In this paper, we explore the effect of appearance and infant like behavior on the human teachers in a setting similar to a LfD scenario. We hypothesize that infant likeness will induce *motionese* regardless of the appearance of the robot. Our results concur with our hypothesis indicating that infant likeness of the robot had a greater impact than the appearance in encouraging the human teachers to employ motionese. Additionally, as part of our experimentation, we made observation of a few key characteristics associated with motionese. These include the point of focus of the robot during the task, and the correlation between motionese and use of speech by the human teacher. Our results indicate higher amount of key-feature extraction by the Saliency Model when the focal point is the human teacher's hands rather than the objects being manipulated. Moreover, we observed a positive correlation between speech and motionese indicating that speech (*motherese*) is a common tool employed when the demonstration is naturally motionese oriented.**

Index Terms — motionese, motherese, infant directed actions, parental scaffolding, visual saliency, learning from demonstration

I. Introduction

As the robots gain more access to our homes and social environments, there is an opportunity for them to learn more about their surroundings and its inhabitants, by means of deriving functional knowledge about their world from their interactions with humans. Human teachers can help robots acquire social skills by teaching robots novel tasks by providing demonstrations that the robot can then observe and learn from. This ability to Learn from Demonstration (LfD) is an exciting field in study of Human Robot Interactions as it offers the potential for a robot to learn over and above the 'pre-installed' features it comes equipped with. Though very promising conceptually, the actual implementation of this idea is fraught with technical difficulties. When not explicitly encoded in machine language by a human programmer, the task of learning from demonstration assumes multidimensional structure that requires the robot to simultaneously process large volumes of data from visual and auditory sensors and select from the stream of binary encoded environment variables, the most important aspects of learning the task from the demonstration.

In many aspects, this difficulty that a robot faces when introduced to an 'alien' world, is similar to the cognitive load on an infant trying to learn the complexities of its social environment. Despite the sheer slope of the learning curve, human infants make tremendous progress during their early few months and display remarkable prowess in making sense of their physical and social world. A key factor that helps the human infants in this learning process is human parent's infant directed action (IDA) sequence, or *motionese*. In human parent-infant interactions, the naturally observed mode of parental action demonstration, motionese, helps the human infant to acquire new skills even when the infant's understanding of the task or its context is limited. By purposefully, yet unintentionally, emphasizing the most important aspects of the task; including the starting state, consequent state transitions, important landmarks in the demonstration process leading to the goal state, human parents help the infant tie the object and its characteristics with the possibilities of leveraging it to attain the task goal.

Some of the other ways in which human parents produce motionese induced demonstration is by scaffolding of demonstrations to draw the attention of the infant (*interactiveness of demonstration*) to guide their attention to the most important aspects of the demonstrated action (*simplification of demonstration*). This is typically achieved in turn by slowing down the task action, multiple iterations/repetition of the same task action and exaggerated, attention seeking movements. One of the ways in which the robots can leverage this 'natural' behaviour of its human teacher and successfully meet the need for identification of salient features of an interaction or demonstration is by taking advantage of motionese to direct its attention to focus its capabilities of observation to the most pertinent aspects of the action demonstration; thereby suppressing noise significantly, reducing the computational load on its computational machinery, yet in the process find meaningful sub-structures in the demonstrated action set.

To investigate how motionese can be induced in a human robot interaction, we propose to investigated the impact of appearance and the behavior of the robot on its human teachers. We intend to evaluate whether the appearance of the robot (infant or adult like features) causes people to employ motionese in the LfD scenario, or whether it is the infant like behavior and mannerisms that causes the human teacher to employ motionese during demonstrations irrespective of the appearance. Our model for evaluation is similar to that employed in [1] and does not assume any past knowledge about the task, objects or human actions used for the tasks by the robot. Our model is designed to be able to detect salient locations in a demonstration setup, that stand out because of the motion-based and visual features (color, intensity) in the environment. Taking cues from the human parent-infant interactions, we hypothesize that a robot equipped with infant-like abilities will induce motionese in the human teacher irrespective of its appearance. This paper presents our experiment with two virtual robots with appearance of an infant and an adult, and evaluation of whether parental/motionese based teaching demonstration is employed in the LfD context solely based on infant-like attention mechanism.

The rest of this paper is organized as follows. In Section 2, we summarize the related work from other researchers in the area of motionese from psychological and computational studies. In Section 3, we introduce the saliency-based visual attention model and describe the implementation details. We also discuss the implementation of the virtual human in this section. Next, we describe the experimental setup in Section 4, and discuss the results in Section 5. Finally, we conclude with discussion, future directions and acknowledgements in Sections 6 and 7.

II. Related Work

A. Robot Learning via Demonstration

Schaal [2] presents the idea of Learning from Demonstration as a key mechanism that helps humans extract initital biases and strategies for solving a complicated problem, or learning novel actions and object manipulations by observation of demonstation by experts. This method of learning has became a cornerstone of Robotic leaning systems that seek to imitiate or reproduce human-like actions. Choice of the demonstrator, problem state definition, policy derivation and performance have been evaluated as the foundations of the LfD research as reported by Argall wt al. [3]. Various aspects of the LfD domain and its application to Robotics has been extensively studied in areas such as bipedal walking by Nakanishi et al. [4], learning motor skills by Pastor et al. [5], interactive policy learning by Chernova and Veloso [6], human gesture imitation by Calinon and Billard [7] and in dogged learning by Grollman and Jenkins [8].

B. Infant Directed Interaction

The basic premise of motionese is based on the idea of observable human behavior and natural occurance of infant directed interactions in case of teaching by the human parents. Zukow-Goldring and Arbib [9] bring to light the fact that during the course of their early development, infants' discovery of a range of affordances and effectivities contributes to participating in a new activity. By leveraging the adaptation of demonstration by the human teacher to the specific needs of the learners, robots can benefit from similar behavior in results from Fernald [10] and Cooper et al. [11] as seen with motherese. Given the limited development of the cognitive capabilities of an infant, a reduction in the percieved complexity of the learning mechanism is most desirable as described by Horowitz [12]. By taking into account, what the infant lacks in context about the learning process, the human parent augments the demonstation process as note by Brand et al. [13] and Koterba and Iverson [14] to include additional details about the task and actions to successfully transfer the knowledge to the human infant as demonstrated by Dunst et al. [15] and Brand and Shallcross [16].

C. Application of Motionese in LfD

In the seminal paper by Nagai and Rohlfing [1] discuss the computational models of Motionese towards designing robots that can take advantage of this parental teaching method in a LfD setting. In separate papers the authors have also addressed the issues of applicability of motionese as a mechanism to help infants and robots with identifying "what to imitate" [17], the modification of the parental actions in highlighting the goal versus the means in relation to a task [18], and design aspects for building a robot that is able to learn novel actions from parental demonstrations by leveraging motionese [19].

D. Visual Saliency

An indepth view of Visual saliency methods is presented by Ciptadi et al. [20] with focus on generic object segmentation and detection capabilities. Borji and Itti have presented comprehensive review of the stimulus-driven, saliency-based attention models [21]. We have based our implementation of the Visual Saliency on the work by Hou et al. [22].

III. Approach

A. Saliency Model Implementation

1) Saliency Model: One of the prime challenges faced in learning from demonstration scenarios is the inability of the robot to accurately identify objects that are key to the task being taught. When introduced into a completely novel surrounding with unfamiliarity of the task it has to perform, the robot receives a lot of sensory information as input from the environment. Being able to direct its attention correctly to key aspects can go a long way in easing the process of teaching. A handy tool that can be utilized to achieve this is the Saliency based visual attention model. In this experiment, we use the saliency model as an indicator of whether the teacher employed Motionese while teaching the robot. A description of this model, its architecture and implementation is described in the following sections.

2) Saliency Model Architecture: The saliency model employed is the Image Signature Saliency model [22]. Saliency is defined as the difference between the intensity of the foreground pixels and the background pixels. The architecture of the model is indicated in the flowchart in Figure 1. This approach involves the separation of the background and foreground thus enabling identification of salient aspects. An image descriptor called the "Image Signature" is used. This descriptor is defined as the Discrete Cosine Transform (DCT) of an image. It contains information regarding the foreground of the image which is used to separate the background and foreground in order to isolate salient objects. As indicated in Figure 1, an input image is first decomposed into the three color channels, namely the red, blue and green channels and an image descriptor is generated for each of these channels. These descriptors are called the Channel maps. The channel maps are then summed up to produce the final saliency map. Further details regarding the implementation can be obtained from [22].

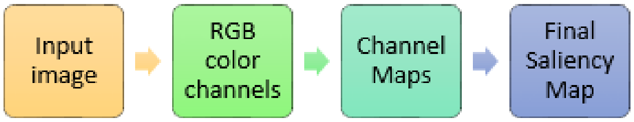
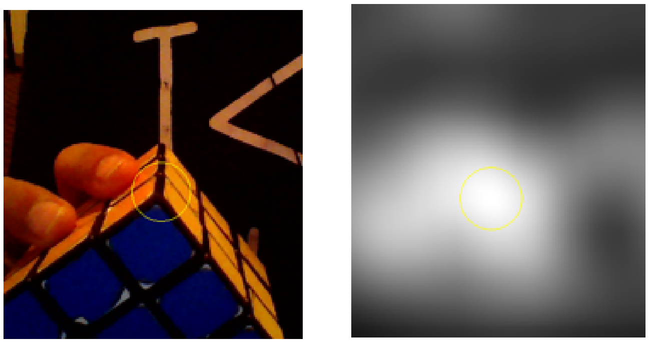


Fig. 1: Signature Saliency Architecture

3) Attention Model Implementation: Following the generation of a saliency map, the most salient regions (indicated by pixels with highest intensity in the saliency map) are sampled. Since there may be multiple points of interest, the robot is equally likely to attend to any of these key areas and hence, one of the regions is selected by uniform sampling. The sampled region is then treated as the attended location. The key aspects of the task thus derived form the Region of Interest (ROI).

Habituation is not implemented here since its purpose is to test for the use of motionese by the teacher. In lieu of our original hypothesis, if the teacher employs motionese, then the



(a) Original Image (b) Saliency Map

Fig. 2: Images With Yellow Marker Showing Region of Interest

saliency map generated will correctly capture the objects and the teacher's hands as they move while the task is being performed. If the rest of the teacher's body does not remain stable (teacher does not employ motionese), then multiple salient regions will be generated, for instance, the teacher's head or face, in which case the robot's attention would not be directed to the ROI. This aspect has been visualized in the images given in Figure 2. The number of times attention is correctly focused on the ROI is measured. If this value is larger than the number of times attention is focused elsewhere, based on the algorithm design, we can conclude that motionese was employed. When motionese is employed focus will only be on the ROI and this will increment its corresponding count.

In Figure 2, the Rubik's cube is detected as the most salient aspect since it was moved while all the other areas remained stable. Hence, maximum intensity was captured for the Rubik's cube on the Saliency map. This allows the Saliency model to be used as a reliable indicator of motionese.

4) Complete Algorithm: The algorithm returns the number of times the attention was correctly focused on the key objects and the number of times focus was diverted elsewhere. Based on these results, analysis of whether motionese was employed during the demonstration can be performed. The algorithm for the saliency based attention model is shown in Algorithm 1.

B. Virtual Human Implementation

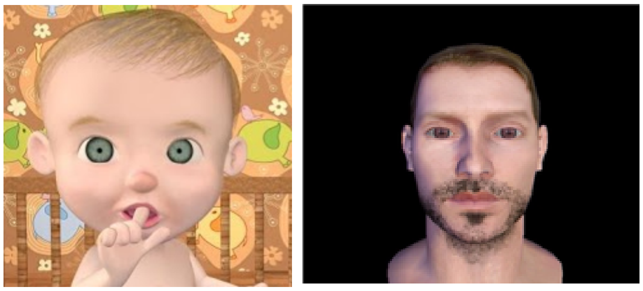
To study whether humans apply motionese or not to robots irrespective of their appearance, we created an opportunity for interaction between the human subject and a robot. For this experiment we implemented a virtual model to study the interaction. This would help us analyze the human behavior towards a robot which would be similar to what a human in the future might do, when social robots would be placed at a home setting or other social gatherings in order to assist humans. From our research on saliency model and infant directed action, we learned that parents are known to significantly alter their actions when they interact with infants as compared to interaction with other adults [1]. For instance, parents usually take a longer time to demonstrate tasks and additionally, they take more pauses between actions while exaggerating their body gestures, in order to help the infant better understand the task while demonstrating actions unfamiliar to them. This behavior, motionese, can be directly extrapolated to an LfD scenario when a robot is to be taught a series of actions unfamiliar to it. We proposed the experiment in an attempt to test if the subjects would employ a similar behaviour with adult-like and/or infant-like appearance of robots while demonstrating a Lego block stacking task. Hence, we decided to have two virtual humans as our robots, giving the user a virtual experience of almost teaching a real human. The experiment proceeds to understand whether or not the subjects implement motionese, depending on the robot's adult-like or infant-like appearance, while teaching the task.

```
Data: Video input
initialization;
Create video object;
while Key not pressed do
    Image = Retrieve frames;
    SaliencyMap = Signature Saliency(Image);
    MaxValue = Max(Saliency map);
    Rows columns = find(MaxValue);
    A = UniformSampling(rows);
    XCoordinate = Rows(A);
    YCoordinate = columns(A);
    Create circular mask for ROI;
    if color value greater than threshold then
        Increment ObjCount
    else
        Increment NullCount
    end
end
Return(ObjCount, NullCount);
```

Algorithm 1: Saliency Based Attention Model

As previously explained we relied on simulated human characters for the robot, to whom the human subjects would demonstrate the Lego block stacking task. For the experiment we used two virtual characters, one simulating an adult while the other simulated an infant. The adult model thus developed is named "Tony" and the infant is named "Eric". These virtual characters were developed with the HapFACS technology [23], which in effect is available as an open source software and API. With the HapFACS software, we could generate selections of realistic FACS-validated facial expressions. Evidently, with HapFACS API, we could animate virtual characters with real-time realistic facial expressions, without any prior experience in computer graphics and modelling. To implement the same we used HapFACS control over 49 FACS Action Units (AU) at all levels of intensity. It enabled the animation of faces with a single AU or a composition of AUs, activated unilaterally or bilaterally. The system came with its own set of virtual humans as an in house library from which we could make selection of expressions, which can be applied to any supported character in the underlying 3D-character system.

This basically generates facial expressions on 3D virtual characters based on FACS technology. Following this we use the software to create a series of pre coded facial expressions and gestures like smile, head nodding, head tilt left, right etc. We then built an animation out of these selected facial gestures by placing these snippets of facial expressions one after the other.



(a) Virtual Infant (b) Virtual Adult

Fig. 3: Virtual Human Implementation

Eventually we modeled a full length virtual character with the appeal of a real human. We refrained from adding any speech capabilities into either of the characters to avoid its potential impact on the judgement of the experimenter when teaching the block stacking task to the virtual humans.

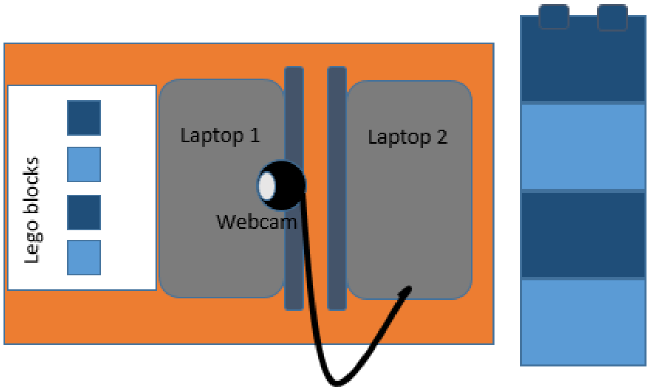
In order to maintain the integrity of our results, we ensure that the only modified variable in both the virtual humans is the appearance. As a result, both the virtual adult and the infant are equipped with infant-like capabilities like limited attention span and absence of speech capabilities. This helps prevent any bias in the experiment while confirming the influence of appearance alone, rather than behavior. The virtual humans thus developed are shown in Figure 3.

IV. Experiment

To validate our hypothesis, a within-subjects user study was conducted with 15 participants comprising of 5 females and 10 males. The independent variables include the adult-like and infant-like appearance of the two virtual humans. In order to study the effect of modification of the independent variables, all other variables were kept constant. Hence, both the virtual humans were equipped with infant-like capabilities. The dependent variables measured were the results obtained via the saliency based attention model which indicated the number of times attention was focused upon the blocks, the subjects hands and other locations. Following is a detailed description of our experimental setup.

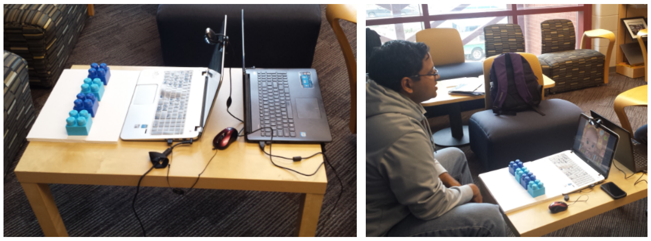
We conducted the experiment in the RIM Library at Georgia Tech, College of Computing Facility. The setup consisted of two laptops (hereafter referred to as Laptop 1 and Laptop 2) placed next to each other. Laptop 1 presented the two virtual human characters/robots who played a role similar to the learner in a LfD setting. Hence, the Lego block stacking task was performed by the subjects seated facing Laptop 1. Moreover Laptop 1 also had an independent USB webcam situated upon its screen facing the subject, in order to get a clear view of the experiment and the gesture adopted whilst teaching the task. This in effect, was used to record and computationally evaluate their body gestures including hand movements while they demonstrate the said task. Laptop 2 was placed behind Laptop 1. Laptop 2 is connected to the aforementioned webcam device with the aid of a USB connection. Laptop 2 thus, retrieved the video captured by the webcam placed on Laptop 1 and ran the computational model of the saliency map, and generated results. A top view of this setup is shown in Figure 4a.

Prior to beginning the experiment, its specifics were explained to the subjects. They were instructed to teach the virtual robots in Laptop 1, the task of stacking dark and light blue colored Lego blocks on top of each other in alternate colors to construct a tower of four blocks. The target structure is shown in Figure 4b. They were informed to teach the same task to both the adult and the infant virtual robots seen in Laptop 1. They were further instructed to teach the virtual robot in any manner they pleased. No specifics were given regarding the use of speech while executing the task. Up until this point, the subjects were not allowed to see the display of any of the laptop machines.



(a) Top View of Setup (b) Target Lego Structure

Fig. 4: Experimental Setup



(a) Object Placement (b) Subject Interaction

Fig. 5: Snapshots of Setup

Once the subjects were confident regarding the task execution, they were asked to begin demonstrating the task. On the back end i.e, on Laptop 2, the saliency based computational model was initialized to analyse and evaluate the subject's method of teaching and body gestures simultaneously. The subjects, in effect, taught Tony and Eric one after another. Steps were taken to ensure that each of the subjects would randomly start with either Tony or Eric such that the subjects do not get influenced by any one of the virtual robots while teaching the other. Additionally, while they were demonstrating the task, the procedure was recorded, in order to further study the body gesture they adopted to teach the robot the mentioned task. All of the subjects taught both Tony and Eric.

V. Results

A. Evaluation of Main Hypothesis

Before proceeding towards the evaluation section, based on the data obtained from the experiment, percentage of correct focus is calculated for both the virtual models. Correct focus implies focus on the ROI. This is computed as follows:

Percentage of correct focus = $(N_h + N_b) * 100/(N_h + N_b + Null)$

Where, N_h indicates the number of times focus was on hands; N_b indicates number of times focus was on blocks; Null indicates number of times focus was on other locations. N_h and N_b together indicate that focus of the robot was upon the key objects. This percentage is calculated for both the virtual models. These values are tabulated and the data obtained therein is subjected to further evaluation.

In order to evaluate the results, T-test was conducted since the experiment involved only two independent variables. T-test can be used to determine if two sets of data are significantly different from each other. However, in order to determine the equality / inequality of variance, first F-test is conducted. The F-test is used to assess whether the expected values of a quantitative variable within several pre-defined groups differ from each other. Hence, the results produced are indicative of the equality/inequality of variance within the data. Table 1 shows the results of F-test. Variable 1 refers to the infant-like model and variable 2 refers to the adult-like model.

1) F-test Analysis: Interpretation of the results obtained from the F-test is done with the help of the *F* and *F critical one-tail* values. If *F* is greater than *F critical one-tail* value, it indicates unequal variance. In table 1;

F = 0.922617

F critical one-tail = 0.402621

This indicates variance inequality.

2) T-test Analysis: Since unequal variances are indicated by the F-test, T-test two sample analysis with unequal variance is performed. The results are indicated in table 2. **The null hypothesis is that percentage of correct focus is similar for both infant-like (variable 1) and adult-like (variable 2) appearances.** In other words, the mean percentage of correct focus is same for both the independent variables indicating application of motionese in both the scenarios. We test this null hypothesis on the basis of the T-test results obtained. The null hypothesis is rejected if *t Stat* is lesser than *-t critical two-tail* or greater than *t critical two-tail*. From table 2;

T-stat = 0.079305

T critical two-tail = 2.048407

From these results, we conclude that the **null hypothesis is *not* rejected**. In other words, the percentage of correct focus is similar for both infant-like and adult-like appearance models. In the context of our hypothesis, this means that motionese was employed for *both* the models since based on the algorithm design, the percentage of correct focus will be higher when motionese is involved.

TABLE I: F-test Results

	Variable 1	Variable 2
Mean	49.76667	49.27333
Variance	278.5467	301.9092
Observations	15	15
df	14	14
F	0.922617	
P one-tail	0.441174	
F critical one-tail	0.402621	

Figure 6 indicates the box plot for the data obtained. The boxplot shows similarities for both the virtual human implementations and further confirms the hypothesis. From the boxplot we observe that in case of the infant implementation, the outliers are at 82 percent correctly focused for cases where motionese was employed and only 16 percent in the case of subjects who did not employ motionese. Similarily for the adult implementation we observe the outliers to be at 80 percent and 16 percent. The median for both the implementations lie around 50 percent of correct focus. The quartile ranges from about 45 - 59 percent for the infant model and about 44 - 55 percent for the adult implementation.

TABLE II: T-test Results

	Variable 1	Variable 2
Mean	49.76667	49.27333
Variance	278.5467	301.9092
Observations	15	15
df	28	
T stat	0.079305	
P one-tail	0.468677	
T critical one-tail	1.701131	
P two-tail	0.937354	
T critical two-tail	2.048407	

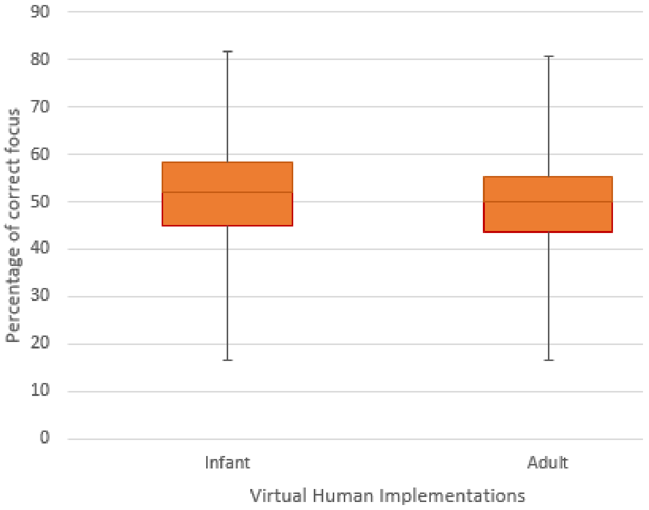


Fig. 6: Boxplot for the Implementations

B. Evaluation of Characteristics of Motionese

Apart from the main hypothesis, we also analyzed the data obtained in order to understand some of the characteristics associated with motionese. Some of our observations have been summarized below.

We evaluated the proportion of attention focused on three different areas: The hands of the subject, the Lego blocks and other locations. Based on our evaluations, we plotted the bar chart shown in Figure 7 which indicates that majority of the focus (50 percent) was on the hands of the subjects than the blocks.

In addition to this, we also observed that speech is an integral component of motionese. As indicated in Figure 8, most of the subjects that employed motionese also used speech cues when demostrating the task. According to Figure 8, 60 percent of the individuals used speech cues to supplement motionese while 6.7 percent used motionese without the aid of speech cues. 6.7 percent used only speech to articulate their gestures without actually using motionese.

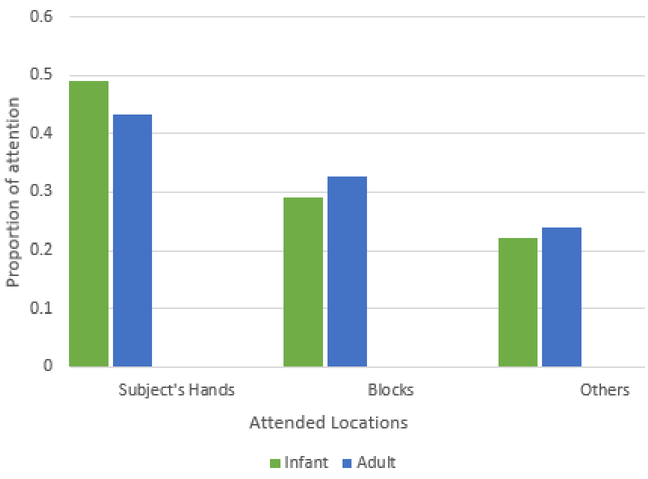


Fig. 7: Proportions of Attended Locations

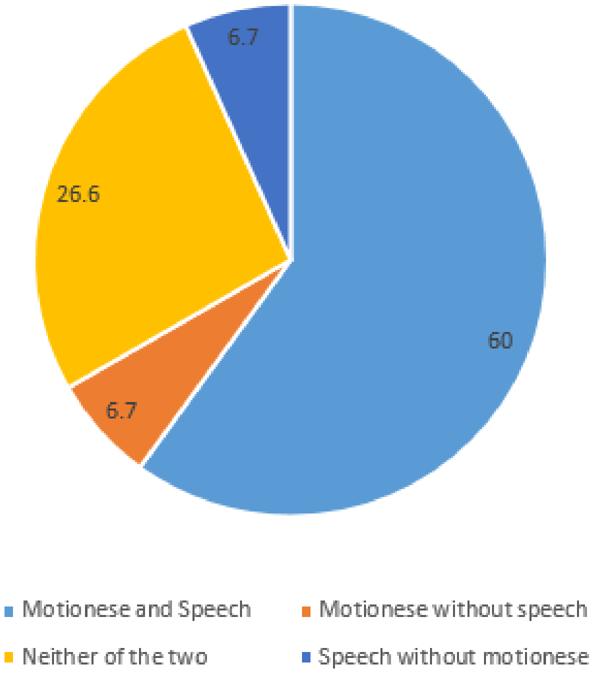


Fig. 8: Percentage of Motionese With and Without Speech

VI. Discussion

Our evaluations returned promising results that potentially influence design of robots for deployment into a social environment involving LfD. Our results indicate that regardless of the appearance of the robot, the infant-likeness played a more significant role in harnessing motionese on the part of the teacher. The virtual adult implementation, which in its design lacked the use of speech, successfully encouraged motionese in the subjects indicating that speech capabilities could be a key factor causing individuals to attribute intelligence. In its absence, subjects treated the adult like an infant. We also observed that subjects who generally employed motionese supplemented their actions with motherese. Additionally, there were also subjects who employed motionese without the use of speech. For the subjects that employed motionese we further evaluated the attention points and the results revealed that the attention was mainly focused on the subjects' hands rather than the blocks. Using a backend habituation algorithm in conjunction with the saliency based learning model can help ameliorate this. However, we have avoided implementing a habituation model since it would bias the results of the experiment.

The results obtained can be intuitively justified in terms of interactions observed between humans themselves. In the case of individuals who are mentally challenged, teachers take much patience and care when teaching them to perform tasks. This is also an example of motionese. This leads us to understand that motionese is in no way limited to infant directed actions. In fact, motionese is a behavioral mode adopted by teachers who feel the need to emphasize their gestures and actions in order to make the student understand the task regardless of whether the student is an infant. Teachers feel the need to employ motionese depending on the kind of behavior exhibited by the learner. A bright and keen student would warrant less motionese as compared to a student exhibiting a less inquistive behavior. However, we ignore the idiosyncracies that may be involved, for instance, impatient nature.

A possible shortcoming of the evaluation presented in previous sections would be the small sample size (number of subjects who participated) in the experiment. Also, the experimental study group consisted only of students drawn at random from the pool in the graduate college and all have background in Robotics research. It is possible that the study participants may have prior interaction with virtual human agents and could have preformed views and biases. Given the background of the study participants, it is highly likely that very few (if any) of the study participants have parenting experience with infants. Nevertheless, the findings of the experiment are interesting as motionese was still observed as being expressed in their interaction with the virtual agents. In our opinion, this finding is encouraging since this opens doors for robots to learn not just from humans with parental instincts, but from people from all backgrounds and age groups since motionese appears to be universally prevalent.

However, we would like to further affirm / validate our results by conducting the study with a much wider group of participants, those from different backgrounds and age groups, and also evaluate cultural variations in the expression of motionese as part of the experiment.

VII. Future Work

One of the areas that warrants more investigation is the use of verbalization of tasks or actions, as part of the demonstration process. We found a high correspondence between the performance of an act, its vocalization and emergence of motionese-like patterns. By vocalizing their actions, human teachers tend to slow down, reflect on their actions and become more sensitive to the effect of the demonstration on the learner.

We would also like to explore other avenues in the robot physical/form design and interaction planning that hold the promise of engaging the human teacher in a more mutually reciprocative manner so as to induce motionese more often, and thereby help social robots improve their chances at maximizing the benefits from the interaction with their human counterparts.

VIII. Acknowledgment

The authors would like to thank our instructor, Dr. Andrea Thomaz, for pointing us to relevant research in this area, and for her valuable guidance throughout the course of this project.

REFERENCES

[1] Nagai, Yukie, and Katharina J. Rohlfing. ”Computational analysis of motionese toward scaffolding robot action learning.” Autonomous Mental Development, IEEE Transactions on 1, no. 1 (2009): 44-54.

[2] Schaal, Stefan. ”Learning from demonstration.” Advances in neural information processing systems (1997): 1040-1046.

[3] Argall, Brenna D., Sonia Chernova, Manuela Veloso, and Brett Browning. ”A survey of robot learning from demonstration.” Robotics and autonomous systems 57, no. 5 (2009): 469-483.

[4] Nakanishi, Jun, Jun Morimoto, Gen Endo, Gordon Cheng, Stefan Schaal, and Mitsuo Kawato. ”Learning from demonstration and adaptation of biped locomotion.” Robotics and Autonomous Systems 47, no. 2 (2004): 79-91.

[5] Pastor, Peter, Heiko Hoffmann, Tamim Asfour, and Stefan Schaal. ”Learning and generalization of motor skills by learning from demonstration.” In Robotics and Automation, 2009. ICRA’09. IEEE International Conference on, pp. 763-768. IEEE, 2009.

[6] Chernova, Sonia, and Manuela Veloso. ”Confidence-based policy learning from demonstration using gaussian mixture models.” In Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, p. 233. ACM, 2007.

[7] Calinon, Sylvain, and Aude Billard. ”Incremental learning of gestures by imitation in a humanoid robot.” In Proceedings of the ACM/IEEE international conference on Human-robot interaction, pp. 255-262. ACM, 2007.

[8] Grollman, Daniel H., and Odest Chadwicke Jenkins. ”Dogged Learning for Robots.” In ICRA, pp. 2483-2488. 2007.

[9] Zukow-Goldring, Patricia, and Michael A. Arbib. ”Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention.” Neurocomputing 70, no. 13 (2007): 2181-2193.

[10] Fernald, Anne. ”Four-month-old infants prefer to listen to motherese.” Infant behavior and development 8, no. 2 (1985): 181-195.

[11] Cooper, Robin Panneton, Jane Abraham, Sheryl Berman, and Margaret Staska. ”The development of infants’ preference for motherese.” Infant Behavior and Development 20, no. 4 (1997): 477-488.

[12] Horowitz, Frances Degen. ”Infant learning and development: Retrospect and prospect.” Merrill-Palmer Quarterly of Behavior and Development (1968): 101-120.

[13] Brand, Rebecca J., Dare A. Baldwin, and Leslie A. Ashburn. ”Evidence for ‘motionese’: modifications in mothers’ infant-directed action.” Developmental Science 5, no. 1 (2002): 72-83.

[14] Koterba, Erin A., and Jana M. Iverson. ”Investigating motionese: The effect of infant-directed action on infants’ attention and object exploration.” Infant Behavior and Development 32, no. 4 (2009): 437-444.

[15] Dunst, Carl J., Ellen Gorman, and Deborah W. Hamby. ”Effects of motionese on infant and toddler visual attention and behavioral responsiveness.” Center for Early Literacy Learning 5, no. 9 (2012).

[16] Brand, Rebecca J., and Wendy L. Shallcross. ”Infants prefer motionese to adult-directed action.” Developmental science 11, no. 6 (2008): 853-861.

[17] Nagai, Yukie, and Katharina J. Rohlfing. ”Can motionese tell infants and robots” what to imitate”.” In Proceedings of the 4th International Symposium on Imitation in Animals and Artifacts, pp. 299-306. 2007.

[18] Nagai, Yukie, and Katharina J. Rohlfing. ”Parental action modification highlighting the goal versus the means.” In Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on, pp. 1-6. IEEE, 2008.

[19] Nagai, Yukie, Claudia Muhl, and Katharina J. Rohlfing. ”Toward designing a robot that learns actions from parental demonstrations.” In Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on, pp. 3545-3550. IEEE, 2008.

[20] Ciptadi, Arridhana, Tucker Hermans, and James M. Rehg. ”An In Depth View of Saliency.” In Eds: T. Burghardt, D. Damen, W. Mayol-Cuevas, M. Mirmehdi, In Proceedings of the British Machine Vision Conference (BMVC 2013), pp. 9-13. 2013.

[21] Borji, Ali, and Laurent Itti. ”State-of-the-art in visual attention modeling.” Pattern Analysis and Machine Intelligence, IEEE Transactions on 35, no. 1 (2013): 185-207.

[22] Hou, Xiaodi, Jonathan Harel, and Christof Koch. ”Image signature: Highlighting sparse salient regions.” Pattern Analysis and Machine Intelligence, IEEE Transactions on 34, no. 1 (2012): 194-201.

[23] Amini, Reza, and Christine Lisetti. ”HapFACS: An open source API/software to generate FACS-based expressions for ECAs animation and for corpus generation.” In Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on, pp. 270-275. IEEE, 2013.