Pointing Gestures for a Robot Mediated Communication Interface

John-John Cabibihan[1,2], Wing Chee So[3], Medi Nazar[2], and Shuzhi Sam Ge[1,2]

[1] Social Robotics Laboratory, Interactive and Digital Media Institute
[2] Department of Electrical and Computer Engineering
[3] Department of Psychology,
National University of Singapore, Singapore
{elecjj,psyswc,medinazar,elegesz}@nus.edu.sg

Abstract. This paper asked whether the pointing gesture accompanying with speech would facilitate comprehension of spatial information in the videoconference communication. Ten adults participated in our study and communicated with the experimenter over Skype (Skype Technologies, Luxembourg). The experimenter described the spatial layout of items in a room to the participants in two conditions – dynamic and static. In the static condition, the notebook was not moving; in the dynamic condition, the notebook moved around with the arms pointing to abstract spatial locations that represented the locations of items in the room. The movement was done by putting the notebook on the three-wheeled Wi-Fi enabled device that was equipped with two artificial arms and was controlled by the experimenter over the Internet. At the end of each description, the participants were asked to lay out the items properly. Reaction times and accuracy rate were recorded. The findings showed that the accuracy rate was higher in the dynamic condition than in the static condition. In addition, the response time was faster in the dynamic condition than in the static condition. It turned out that pointing gestures facilitated the speech comprehension of spatial information.

Keywords: Videoconferencing, pointing gesture, communication modality, robot-mediated interaction.

1 Introduction

Communication through Internet is the most common way of keeping in touch with our friends and family members from a distance. Email, chat, videoconference, and social networking sites (SNS) are becoming part of our daily lives. In face to face communication, it is very natural for us to combine our different senses to fully express ourselves and to fully understand what the other person is saying. We often use our audio-visual sensory system and incorporate non-verbal cues such as gestures to convey information using these multiple channels unconsciously to different degrees. In mediated communication such as videoconferencing, nonverbal cues are either limited or filtered thus the richness of information sent/received decreases.

There are various existing works focusing on different aspect of mediated communication. In [1], Hwang and Park mentioned that computer mediated communication

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 67–77, 2009.

(CMC) is a medium with low social presence especially text-based CMC environments since non-verbal cues like gestures, direction of gaze and posture are missing, where social presence here is defined as the ability of communicating parties to project themselves, socially and emotionally, as real people through a medium of communication [2]. In [3], the difference among communication modes (text chat, audio, audio–video, and avatar) in an Internet-based collaboration were investigated and significant differences were found between text-based chat and all other communication modalities. Although some researchers [4] argued that the quality of communication is based on the medium, Fish and colleagues [5] showed that video does not significantly improve telecommunication as compared to audio with respect to efficiency or user satisfaction. Experiments were conducted between human-robot interaction [6] to test the effect of social presence. These current works motivate us to find other means of increasing the level of social presence in existing communication through the Internet.

In this paper, we added pointing gesture to a typical videoconference over Skype. To produce such gesture, we thought of an embodiment of a person (a robot) on the other side of communication channel that can be controlled through the Internet. Furthermore, we used memory test in our experiment to investigate the effect of such communication compare to a typical videoconference. In our experiment, ten adults participated in our study where the spatial layout of items in a room is described by the experimenter with and without pointing gesture to each participant. Then they were asked to layout the items properly. Reaction times and accuracy rate were recorded.

In [7], Kita argued that pointing gesture lessens the confusion between the *left* and *right* concepts. Thus, we hypothesized that adding pointing gesture will increase the accuracy rate of the participants and also, their response will be faster.

The succeeding sections are organized as follows: Section 2 presents the methodology that we used in our experiment, Section 3 shows the results, Section 4 contains the discussion and analysis, and finally, the conclusion in presented in Section 5.

2 Methodology

2.1 The Robot Interface

The aim of the experiment is to find out if there will be a difference when we add pointing gesture in describing the spatial location of the objects seen by the person in a room over a Skype-based communication. In order to do this, we built a robot interface that consists of three components: (1) Rovio (WowWee Group Limited, Hong Kong), a three-wheeled Wi-Fi enabled robot; (2) mini notebook put on top of Rovio; and (3) a static pointing arm attached to the combined Rovio and mini notebook, see Figure 1a. The movement of the robot interface is controlled over the Internet through a software joystick installed in another Wi-Fi enabled computer (Computer A), see Figure 1b. During the experiment, the face of the experimenter who is communicating with the participant appears on the monitor of the notebook through a Skype videoconference.

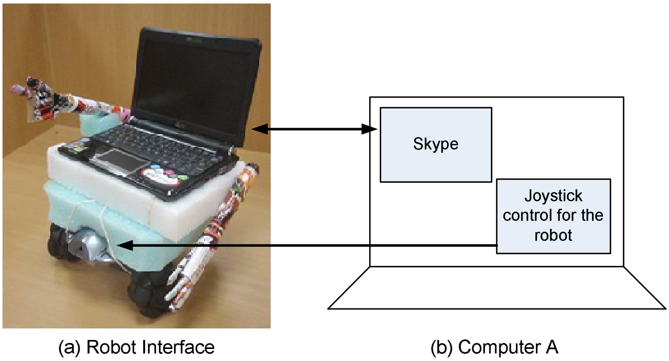


Fig. 1. The Robot-Mediated Communications Interface

2.2 Communication Set-Up

We used two communication set-ups in our experiment, first is the typical videoconference over the Internet which we called *static mode* (Figure 2a) and the second is the one with additional robot movement called *dynamic mode* (Figure 2b). In static mode, the experimenter and the research subject will communicate through a typical videoconference (without movement) using Computer A and the mini notebook while in dynamic mode the robot interface will be used to create the pointing gesture, the unidirectional arrow in Figure 2b illustrates this additional communication modality towards the research subject. In our experiment, we studied the difference of these two set-ups: static and dynamic modes.

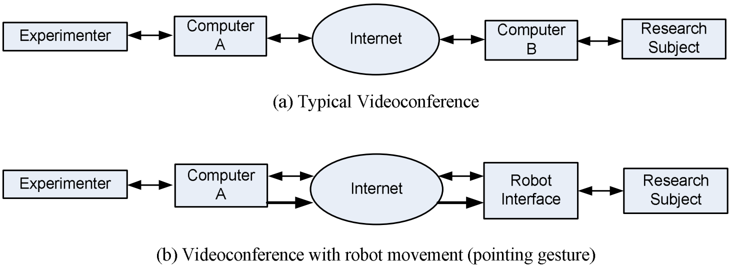


Fig. 2. Communication set-up for *static* and *dynamic* modes

2.3 Experimental Set-Up

Figure 3 shows the experimental set up in the laboratory room. The robot interface is placed on top of a wide table to give space for its movement in a dynamic condition facing the subject who sits on the chair. The dashed area that surrounds the subject's chair is the empty part of the room where the items described in the scripts are located. Ten layout sheets and ten sets of small cut pieces of papers needed for the memory test are provided on a separate table (not included in the figure).

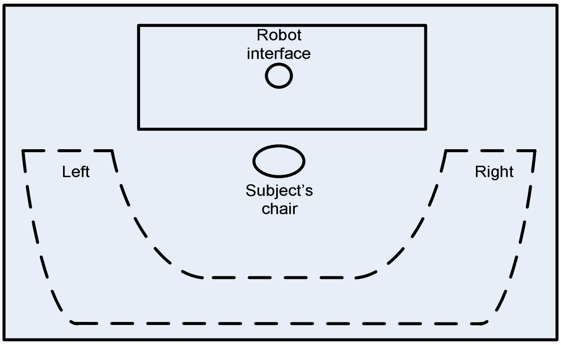


Fig. 3. Experimental Set-up

The layout sheet contains the layout of the room shown in Figure 3 which is basically empty. The items to be laid out on the sheet were written on the small cut pieces of papers. Each set of cut papers has its own color and has a label in a room-mode format (e.g. kitchen-dynamic). These ten sets were arranged in a right to left manner based on the random sequence of the script for each subject.

2.4 The Scripts

For the scripts, we considered five venues that can be found at home such as kitchen, living room, bedroom, study room and kid's playroom. We thought of five items that can be found in each room, then, prepared a description/script of the room with the five items and their specific location in a given room. There are ten scripts, five for the static condition and another five for the dynamic condition, see Appendix. One room is described in each script. Scripts 1 to 5 are the static mode description of the five given rooms whereas scripts 6 to 10 are the description for the dynamic mode, same rooms were used but with different list of items. The first statement in the script varies according to the sequence of delivery which is random for each subject.

In the *static* mode, the locations of the imaginary items in a given room were described using the following keywords: "on your left hand side", "on the right corner of the room". In the dynamic mode, since the robot is moving and pointing towards the direction of its movement, the scripts were shortened and simplified to "on that side" and "on that corner". The word "that" is associated with the location of the item where the robot is facing and pointing at, thus, no need for long description. Cues such as "behind you" and "at your back" remain to avoid ambiguity while the robot is facing to the subject.

2.5 Experimental Procedure

Participants

We invited ten participants in our pilot test, eight were female and two were male whose age is from 22 to 39 with different nationality and field of studies.

Procedure

Upon arrival of the participant, the communication set-up was explained and instructions were given by the experimenter. The experimenter set up the Skype connection between the mini notebook and Computer A which is in the adjacent room. Then, the experimenter entered to the adjacent room leaving the participant and the robot interface facing each other. The experimenter started the communication by giving introductory statements; consequently, the first script was recited. Each script contains the spatial description of five imaginary items memorized by the participant.

After the first script, the participant stood, picked the first set of small cut pieces of paper where the mentioned items were written, and laid out the items in their proper location on the provided layout sheet. After arranging the items on the layout sheet, he/she went back to his/her sit for the next script. The procedure was repeated for ten scripts. Different sequence of the scripts was given to each participant which are all randomized.

We used a timer written in C programming language (ran in Dev-C++ [8]) to measure the *response time* of the participant in each script. The timer starts when the participant gives a signal that he/she will start to lay out the items and it ends when a finish signal is sent by the participant.

After the ten scripts, the experimenter went out from the adjacent room to meet the subject in the experimental room. Then, checked the *accuracy* of the laid out items while having a conversation with the participant about what he/she's thinking regarding the location and the items and why they have mistakes on the item's location. Each script has five items and we have five scripts for each mode (see Appendix), thus, the perfect score per script is 5. The accuracy is calculated separately for static and dynamic mode.

3 Results

Figure 4 shows the response time of each subject in static and dynamic modes. It is evident that subject number 4 and 7 were faster in answering the memory test in dynamic compare to static mode. All subjects were relatively faster in dynamic mode except subject number 2.

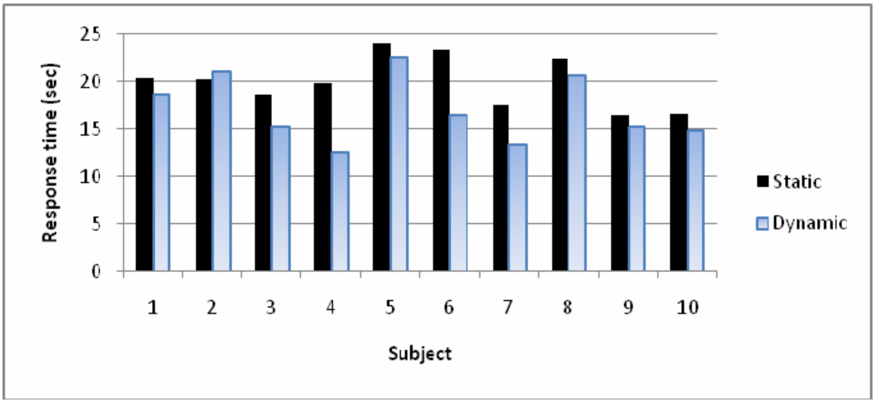


Fig. 4. Subjects' average response time

Note that the average response time varies per subject, which may be due to various factors such as alertness, current mental state of the subject and how they process data in their brain. These factors are out of the scope of this study.

Figure 5 shows the result of the memory test, the accuracy on the vertical axis shows the average correct laid out items of each subject with maximum accuracy level of five. The accuracy level of subject 1, 4 and 7 are much higher in dynamic compare to static mode. All the subjects got higher accuracy in dynamic mode except subject number 6 who got the same accuracy in both modes.

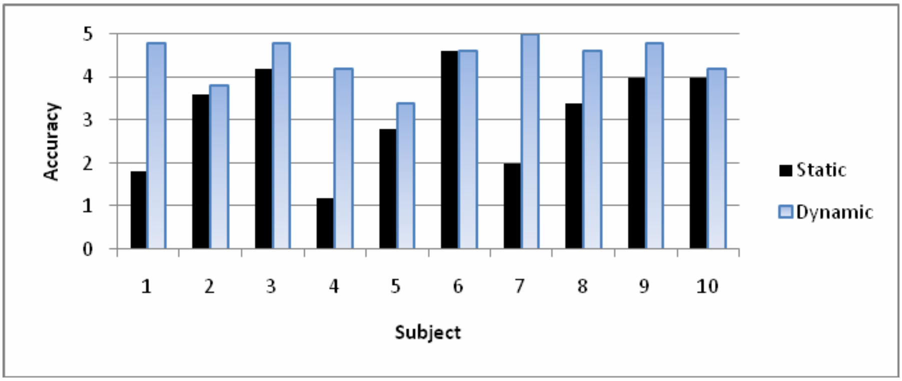


Fig. 5. Subjects'Average Accuracy

The overall response time and accuracy rate for all the subjects are shown in Figure 6. The accuracy rates in static and dynamic modes are 63.2% and 88.4% respectively. It clearly shows that the response time in dynamic mode is faster than in static mode, moreover, the accuracy rate is higher in dynamic mode.

A paired-samples t-test was conducted to compare the response time of the research subject in static and dynamic conditions. The static condition makes use of the typical conversations while the dynamic condition has a robot moving and pointing directions. There was a significant difference at the 5% level in the scores for the static (mean = 19.94, sd = 2.70) and dynamic (mean = 17.04, sd = 3.45) conditions; t (3.56, p = 0.006).

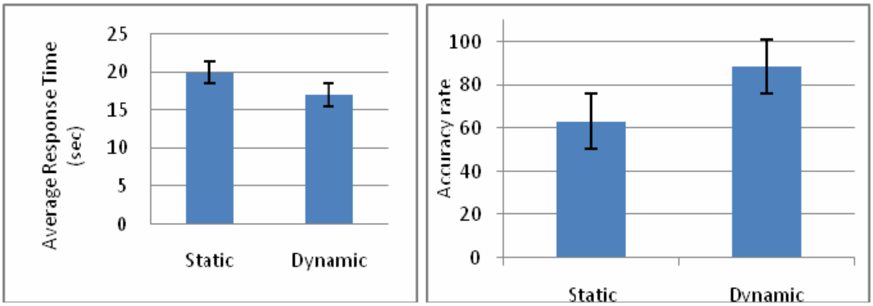


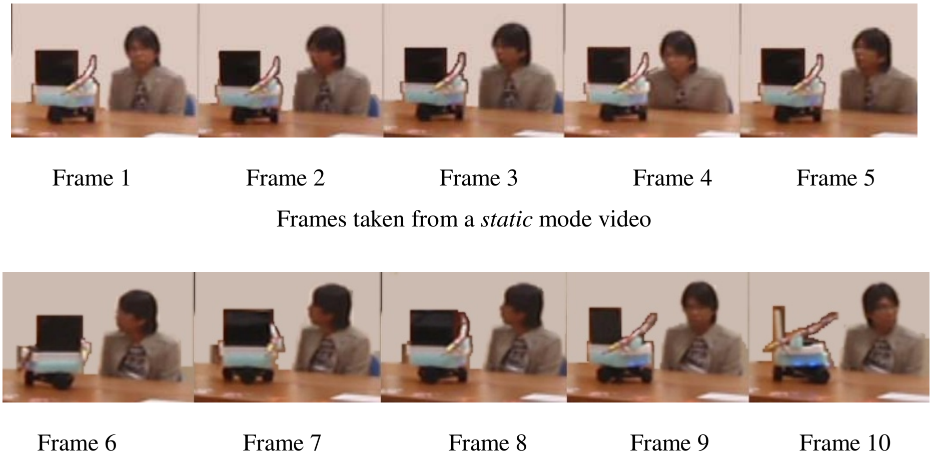
Fig. 6. Overall average response time and accuracy rate

Paired-samples t-test was also conducted to compare the accuracy of the research subject in the memory test for static and dynamic conditions. Likewise, there was a significant difference at the 5% level in the scores for the static (mean = 3.16, sd = 1.16) and dynamic (mean = 4.42, sd = 0.51) conditions; t (-3.194, p = 0.011).

4 Discussion

According to most of the participants, the movement/pointing gesture helped them in recalling the location of the item and the only thing they were memorizing during the *dynamic* condition are the items being mentioned whereas in static condition they have to remember the items at the same time. Subjects no.2 and no.9 mentioned that the movement distracted them. If we look on to the average response time they got, it is indeed confirmed that the movement/pointing gesture did not help subject no.2 in memorizing the items but we cannot say that it has the same effect with subject no. 9 because she had a higher response time during static than dynamic mode (see Figure 4). Most of the remaining subjects mentioned that they were having a hard time memorizing the items in *static* mode and it is confirmed by the results we just have presented in Section 3.

Figure 7 shows the frames from the video during the experiment. Figure 7a shows five frames while the scene is in the kitchen *static* mode description. Frames 1 to 5 were taken after each item was mentioned in the script. From the frames shown, it is apparent that the participant is trying to remember the items and the location being described by merely listening to the experimenter, notice that the position of the robot interface is not changing. In Figure 7b, during the *dynamic* mode description of the kitchen, the participant is following the movement of the robot interface, looking towards the direction where it is pointing while listening to the item being mentioned by the experimenter.



Frame 1 Frame 2 Frame 3 Frame 4 Frame 5

Frames taken from a *static* mode video

Frame 6 Frame 7 Frame 8 Frame 9 Frame 10

(b) Frames taken from a *dynamic* mode video

Fig. 7. Frames taken from the video during the experiment

In terms of accuracy, most of the wrong answers were due to swapped items. One participant was confused about the corner of the room though the layout of the room in Figure 3 was presented to them before the experiment started.

Unfamiliar words were compensated because they were written on the cut papers, so even if the item name is not that familiar to some of the subjects they were able to figure out what it is upon reading the words during the test.

During the experiment, the experimenter used to make a forward movement towards the subject in between scripts to make the interaction interactive and to take away the boredom from the subjects in case there is and to thank them at the end of the tenth script. Most of them smiled and responded positively and some of them felt anxious at the same time. It indicates that the presence of the embodiment strengthens the level of social presence of the person communicating on the other side of communication channel.

5 Conclusion

One of the most commonly used way of communication over the Internet today is videoconferencing, a combination of audio and video. In our experiment, we added another modality to this current set-up. We used an embodiment to add pointing gesture to the typical Skype-based communication. We used memory test to measure the effect of this embodiment.

From the result of our experiment, it appears that the dynamic mode significantly decreases the response time of the subjects and increases their accuracy rate in answering the memory test.

In the memory test, the sense of hearing was used by the subjects during the static mode while in dynamic mode two senses were used, sense of hearing and sight that is hearing the name of the items from the script and seeing the robot interface movement to the direction of the items. The results indicate that using these two senses at the same time to recall object's specific location is better than just using the sense of hearing alone. Thus, the addition of modality which is the pointing gesture movement in our experiment made a significant contribution in current Skype-based communication.

6 Limitations and Future Directions

There are several limitations to our current work. First, we utilized static arms and used the moving robot to point one arm to the direction of the items being described in the scripts. In our future work, we plan to develop moving arms that would point without the robot body's movement. Second, we used the available three-wheeled WiFi enabled device (Rovio) as our robot mediated interface, mounted a mini notebook above it and attached static arms. For the arm to move independently, this must be part of the interface that can be manipulated through the network, thus, the need to develop a robot interface with moving arms.

For our future studies, extending the number of non-verbal cues such as waving of the robot's arm and handshaking will be investigated. This requires us to modify our current experimental procedure or design a new one since our current memory test might not be applicable to these new gestures. It is also interesting to explore the effect of proximity.

Acknowledgments. The authors would like to thank Dr. Marcelo Ang, Department of Mechanical Engineering-NUS, for lending us the robot Rovio from the Control & Mechatronics Lab. This research was funded by the Social Robots: Breathing Life Into Machine Project (NRF-R705-000-017-279), Media Development Authority of Singapore.

References

1. Short, J.A., Williams, E., Christie, B.: The social psychology of telecommunications. John Wiley & Sons, Ltd., London (1976)
2. Garrison, D.R., Anderson, T.D.: Learning in the 21st Century, p. 49. RoutledgeFalmer, London (2003)
3. Bente, G., Rüggenberg, S., Krämer, N.C., Eschenburg, F.: Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. Human Communication Research 34(2), 287–318 (2008)
4. Hwang, H.S., Park, S.: Being together: User's subjective experience of social presence in CMC environments, Beijing. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 844–853 (2007)
5. Fish, R.S., Kraut, R.E., Chalfonte, B.: The VideoWindow system in informal communications. Paper presented at the Proceedings of the Third Conference on Computer-Supported Cooperative Work Los Angeles, CA (1990)
6. Bainbridge, W.A., Hart, J., Kim, E.S., Scassellati, B.: The Effect of Presence on Human-Robot Interaction. In: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany (2008)
7. Kita, S.: Pointing: Where Language, Culture, and Cognition Meet. Lawrence Erlbaum Associates, Inc., New Jersey (2008)
8. Bloodshed Software, `http://www.bloodshed.net/devcpp.html`

Appendix: Scripts

Five scenes were used in our scripts namely the kitchen, living room, bed room, study room and kid's playroom. One room is described in each script. Scripts 1 to 5 are the static mode description of the five given rooms whereas scripts 6 to 10 are the description for the dynamic mode.

Script 1: Scene - Kitchen

Hi, ____(<subject's name>), I would like to introduce to you my kitchen.
There is a stove on your right side.
The sink is adjacent to the stove.
There are piles of dishes beside the sink.
The refrigerator is just right behind you and there are notes posted on its door.
And on your left side, there is the dining table surrounded by chairs.

Script 2: Scene - Living Room
Now, let me describe my living room to you.
The *door* is on your left hand side.
Next to the door, I can see an *aquarium* on top of wooden shoe rack.
At your back, there is a *small table with a dvd player* put on its top.
There are *two couches* adjacent to each other on the right side wall.
There is a *small table* right in front of the couches.

Script 3: Scene - Bed Room
Let's go to the bedroom
There is an *exit door* at your left.
At the left corner there's a *computer desktop* on top of the computer table.
There are *five paintings* hanging on the wall behind you.
Next to the paintings, there's a *window* with open horizontal blinds near the right corner of the room.
And there's a *bed* opposite to the window.

Script 4: Scene - Study Room
Let's go to my study room.
The *entrance door* is on your right.
There's a *computer and a printer* on the right corner of the room.
I can see an *office table* with two chairs in front of it behind you.
On the left corner, there is a *wooden shelf* full of books and documents.
Next to the shelf, there is a *glass wall* with vertical blinds.

Script 5: Scene - Kid's Playroom
Finally, let's go to the kid's playroom.
There is a *small table* surrounded by small chairs on your left.
At the left corner there is a *shelf* with colourful books in different sizes.
Adjacent to the shelf, behind you, there is a *stackable bin* full of toys.
A *painting on a canvas* stands on the right corner.
There is an *inflatable couch* on your right.

Script 6: Scene - Kitchen
Hi, ____(<subject's name>), I would like to introduce to you my kitchen.
There is a *trashcan* on that side.
Then, there is the *barbecue stand* beside.
Next to that is a *water dispenser*.
The *cabinet* is just right behind you
On this side, there is a *breakfast nook* with chairs beside it.

Script 7: Scene - Living Room
Now, let me describe my living room to you.
The *fireplace* is on this side.
I can see a *piano* on that corner.
At your back, there is a *window* with colourful curtain.
On that corner, I can see *two couches* adjacent to each other.
And there is a *coffee table* right in front of the couches.

Script 8: Scene - Bed Room
Let's go to the bedroom.
The *bed* is on this side.
Then, on that corner I can see *clothes* hanging.
There is a *lampshade* behind you.
There are *candles* on top of a rectangular table on that corner.
And there's a *television* on this side.

Script 9: Scene - Study Room
Let’s go to my study room.
There is the *study table*!
There’s a *bookshelf* on that corner.
At your back, I can see *medals & certificates* hanging on the wall.
On that corner, there is an *opened dictionary* on its stand.
I can see the *exit door* on this side.

Script 10: Scene - Kid’s Playroom
Finally, let’s go to the kid’s playroom.
Oh, I can see different sizes of *dolls* on the floor.
And at that corner, there are *coloring books* on top of a small table.
Behind you, there are *artworks* posted on the wall.
A *drawing board* stands on that corner.
On this side I can see a *toy house*.