**Patricc, Child, Parent interaction**

**Roi Hezkiyahu**

[roihezkiyahu@mail.tau.ac.il](mailto:roihezkiyahu@mail.tau.ac.il)

Sagol School of Neuroscience ,Tel-Aviv University

Tel-Aviv, Israel

Under the supervision of:

**Omer Gvirsman,**

omergvir@mail.tau.ac.il

Curiosity Lab, Department of Industrial Engineering,

Tel-Aviv University

Tel Aviv, Israel

**Goren Gordon**

goren@gorengordon.com

Curiosity Lab, Department of Industrial Engineering

Sagol School of Neuroscience, Tel-Aviv University

Tel Aviv, Israel

**Abstract (1)**

There has been an increasing amount of research conducted to determine how a robot tutor influences language learning in children, most of this research was conducted on children aged 3-12, still the use case for robots for toddlers (ages 1-4) in their homes hasn’t gained similar attention. We investigate the influence of Patricc, a specially designed robot for toddler-robot interaction, on the parent toddler interaction with comparison to a tablet interaction. We conducted a follow up study for the triadic interaction (with 18 parent-toddler pairs) in order to better understand patricc’s role in the social interaction and to better understand how we can utilize it. We compared the key features of the robot interaction versus the tablet interaction to understand how they differ, also we used ganger causality test in order to understand which interactions are linked be a cause-and-effect relation. We saw that: results… This suggest that: Conclusion

**Introduction (1-2)**

Social robots are slowly taking place in our everyday activities, in our workplace [1], in agriculture [2], in medicine [3] and in educational systems [4]. the use of social robots has recently been explored in the educational system, with the expectation of positive feedback on learners [5]. Childhood education is an important subject and many researchers are seeking a way to improve current methods, especially Robotics researchers have been attempting to improve children education by introducing robots [9-11]. Most of the educational application are targeting children from preschool to adolescence, only a few studies have targeted toddlers from ages 1-4. These age group require a different approach as they require a triadic (Parent, Robot, Child) interaction as opposed to higher age groups which do not require help in handling the robot. They also require a different design for the robot because toddlers tend to physically interact with the robot and their attention span is much shorter, therefore a Platform for Triadic interaction (Patricc) has been presented [6]. Patricc is a 3D printed robot with changeable puppet-like characters, it lays on a platform which can be placed with physical educational props and can be easily integrated with new content via a specific made authoring tool for non-programmers.

There has been some research suggesting that a HRI (human-robot interaction) may increase the parent-toddler interaction and enjoyment of the experience [7], also a robot touter might be better at teaching new words then humans [8]. Although it remains unclear whether their learning improved due to the presence of the robot or because children taught the words to another agent, the study demonstrated the unique role robots can play in vocabulary learning.

In this paper we present a comparison between the patricc interaction versus a tablet interaction and investigate how does patricc interaction effect and contributes to the parent child interaction in English learning session for toddlers.

**Research question (0.5)**

We sought to investigate 3 main research questions:

Q1: How does the interaction with a robot differ from the interaction with a tablet?

We addressed this question by comparing different features of the interactions using paired t-tests.

Q2: What patterns of the robot can contribute to the parent-toddler interaction?

We tried to address this question by asking which features granger cause others.

Q3: What are the main patterns for the parent- toddler interaction?

We also tried to address this question with granger causality tests.

**Methods (2-3):**

**Data collection**

Audio and video data was collected for 18 pairs of parents and toddlers engaged in a hands-on activity with both the robot and tablet, during each activity the parent and toddlers participated in three learning sessions ordered RRT(robot, robot, tablet) or TRR (tablet, robot, robot). Each session had its own vocabulary set and was taught by a specific puppet.

Each session began with the puppet introducing itself followed by a song which included the vocabulary of the session, each time the robot sang about one of the vocabulary words he pointed and gazed towards the correct object. After the song the robot pointed and gazed towards each of the objects, this action was accompanied by saying the object’s name and asking the toddler to pick it up. After teaching all five objects the robot asked the toddler once again to pick up each object, this time without pointing and gazing at it. If the toddler picked up the correct object then the robot gave positive feedback such as waving his hands and saying: “Very good”. If the wrong object was picked up then the robot told the toddler the name of the object that was picked up, asked the toddler to put it back and then repeat the name of the correct object to pick up. If the toddler did not pick up anything the robot would encourage the toddler by saying “Try again, you can do it!”. Only after the correct object was picked up the robot continued to the next word. After finishing all the words the robot waved his hand and said “Goodbye, see you soon”.

**Data annotation:**

To analyse the videos we used the thin slice approach [12] with a 15-second slices. The interactions were manually coded by 2 coders. Each coder coded Y% of the videos. Inter-rater reliability (Cohen’s kappa) was assessed on a set of Z randomly selected videos (% of the data). The average kappa score Z with the lowest Z , indicating ? reliability for all codes [13]. the behaviours that were coded are:

|  |  |  |
| --- | --- | --- |
| Action | Sub actions | Who |
| Gaze | At: Parent/Toddler, props, robot, tablet | Parent, Toddler |
| Gesture | point at prop | Parent, Toddler |
| prop manipulation | - | Parent, Toddler |
| utterance |  | Parent, Toddler |
| affective touch |  | Parent |
| affect | positive1, positive2, positive3 | Parent |
| pointing |  | robot |
| robot text | pick up, positive feedback, put down, song, teach, try again | robot |
| Verbal scaffolding | affective, cognitive, technical | Parent |
| Conversational turns\* | PT (Parent, Toddler), TP, TPT, PTP | - |
| Joint attention\* | At props, robot, tablet, mutual gaze | Parent, Toddler |

Table 1: Specification of the coded features

\*features that were created from the annotated data

**Data processing:**

We pre-processed the data before the analysis, first we cut the video to contain only the interaction, afterwards we created the conversational turns and joint attention features from the annotated data. Conversational turns were counted as PT if the parent talked and the toddler started talking 1 second after the parent started talking and 5 seconds after he had finished talking, TP was counted the same, PTP and TPT where counted the same way. The conversational turns were timed from the second that the first talker started until talking the time the second (or third) talker ended talking. Joint attention were counted if both parent and toddler gazed at the same thing (or at each other), it was timed from the moment they both gazed together until one of them has finished gazing. The data was then transformed into a time series, the rows were time stamps from the video with a window width of 0.5 seconds, the columns are the annotated features. Each cell in the matrix was given the value 1 if the event of the column occurred at the time window, otherwise it was 0.

**Analysis:**

To better understand cause and effect relations between 2 distinct features we build a GUI to present a time window histogram. Our time window histogram was created by fixating one feature and counting how many events occurred several seconds before and several seconds after that said feature occurred. For example: if a child pointed at the props in second 10 we counted the number of times the parent gazed at the props in seconds 7-,8,8-9,9-10,10-11,11-12,12-13,13-14,14-15. We counted the number of events for each second that the child pointed at the prop and made a histogram. This GUI helped us to better understand the changes within the interaction.

Chart, treemap chart

Description automatically generated

Figure 1: GUI example

Counted events histogram for events accruing before and after child pointed at prop at one of the trails

**Comparison of interactions:**

In order to compare how does the interaction with a robot differ from the interaction with a tablet we conducted 13 paired t-tests to compare the mean of each group for several features of the interaction. The multiple testing was addressed with Benjamini and Hochberg FDR procedure [14] in order to keep the alpha level at 5%.

The pair t-test assumptions are:

1. The dependent variable must be continuous – our variables were the percentage of time each feature took place in the interaction, therefor continuous.
2. Observations are independent of one another – each interaction is independent of the other interactions.
3. The measured differences should be approximately normally distributed – due to a small dataset available at the analysis we could not check this assumption with certainty.
4. The compared conditions are dependent for each observation – each observation is an interaction with both the tablet and robot for the same participant and therefor dependent.

Our null hypothesis was that there is no difference in the percentage of time each feature took place in the interaction

**Granger causality:**

We used the granger causality test to figure out which features of the interaction has cause and effect relations. The granger causality was introduced by Clive Granger [15], he argues that if a time series X improves the prediction of time series Y compared to only lagged values of Y then X granger causes Y. It can also be treated as predictive causality meaning X forecasts Y [16,17].

The tested hypothesis is that:

Where P is the probability function, A is an arbitrary non-empty set, I, is the information available with or without X respectively, if the hypothesis is rejected, we say that X granger cause Y.

In order to test if X can help predict Y we use VAR (vector auto regression) to predict Y only based on lagged values of Y and predict Y also with past values of X. Then we can use a chi-squared test or F-test (whose null hypothesis is: no explanatory power jointly added by X values) if this hypothesis is rejected, we can conclude that X granger cause Y.

Because we use autoregression the following assumptions must be met:

has a mean of 0, they are independent, they are homoscedastic and distributed normally. Also, the time series Y must be stationary.

We had a few limitations with the granger tests:

1. Stationary, our data was not stationary so we used the diff method in order to stationaries the data.
2. We had multiple samples of Y and X. A participant interaction is a sample of a time series, we could not find related work with multiple samples, our solution to this problem was to “stitch” the data meaning we took the values from Y1 and added to them the values of Y2 and so on, between 2 times series we added a vector of 0 at length of double the max lagged used for the granger causality test.
3. Our time series is binary. The regression model assumes normality, because the data we use is binary it can’t distribute normally. In order to address this limitation one can use a logistic autoregressive model, but it requires further research which is out of scope for this project.

**Results (with graphs, statistics, legend, and description)**

Comparison of interactions results:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| feature | Mean robot | Std robot | Mean tablet | Std tablet | T score | Adjusted p values |
| Child gesture | 0.021 | 0.022 | 0.042 | 0.039 | -1.698 |  |
| Conversational turns | 2.501 | 3.266 | 1.048 | 0.458 | 0.732 |  |
| Mutual gaze | 0.055 | 0.012 | 0.011 | 0.013 | 19495.53 |  |
| Parent gesture | 0.056 | 0.016 | 0.049 | 0.018 | 4.907 |  |
| Child gaze at parent | 0.085 | 0.026 | 0.017 | 0.021 | 19.915 |  |
| Parent gaze at child | 0.246 | 0.044 | 0.139 | 0.011 | 4.536 |  |
| Child gaze at props, robot or tablet | 0.624 | 0.24 | 0.93 | 0.023 | -1.642 |  |
| Parent gaze at props, robot or tablet | 0.48 | 0.129 | 0.759 | 0.093 | -1.783 |  |
| Verbal affective scaffolding | 0.023 | 0.002 | 0.015 | 0.009 | 1.094 |  |
| Verbal cognitive scaffolding | 0.088 | 0.052 | 0.129 | 0.092 | -1.454 |  |
| Verbal technical scaffolding | 0.015 | 0.021 | 0.014 | 0.014 | 0.051 |  |
| Parent affective touch | 0.002 | 0.003 | 0.02 | 0.021 | -1.366 |  |

Table 2: results for paired t-test on all tested features

Since we had only 2 fully annotated videos we can’t conclude significant results except for much higher Mutual gaze in the robot condition, even though it is not significant (due to low number of observations) we can see a big difference in the following features:

Higher at robot condition: Conversational turns, Child gaze at parent, Parent gaze at child.

Lower at robot condition: Child gaze at props, robot or tablet, Parent gaze at props, robot or tablet, Parent affective touch.

Granger causality results:

**Discussion(3-4 including the results)**

**References**

1. Jilles Smids, Sven Nyholm, Hannah Berkers,2020, Robots in the Workplace: a Threat to—or Opportunity for—Meaningful Work?. Philosophy & Technology <https://doi.org/10.1007/s13347-019-00377-4>
2. Luis Emmi , Mariano Gonzalez-de-Soto, Gonzalo Pajares, and Pablo Gonzalez-de-Santos 2014, New Trends in Robotics for Agriculture: Integration and Assessment of a Real Fleet of Robots, The Scientific World Journal
3. Satwant Kaur (2012) How Medical Robots are going to Affect Our Lives, IETE Technical Review, 29:3, 184-187, DOI: 10.4103/0256-4602.98859
4. Jacqueline M. Kory-Westlund and Cynthia Breazeal. 2019. A Long-Term Study of Young Children’s Rapport, Social Emulation, and Language Learning With a Peer-Like Robot Playmate in Preschool. Frontiers in Robotics and AI 6 (2019).https://doi.org/10.3389/frobt.2019.00081
5. F. Tanaka, S. Matsuzoe, Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. J. Hum. Robot Interact. 1, 78–95 (2012).
6. Omer Gvirsman, Yaacov Koren, Tal Norman, and Goren Gordon. 2020. Patricc: A Platform for Triadic Interaction with Changeable Characters. In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20). Association for Computing Machinery, New York, NY, USA, 399–407. DOI:https://doi.org/10.1145/3319502.3374792
7. Strouse, G. A., O'Doherty, K., & Troseth, G. L. (2013). Effective coviewing: Preschoolers' learning from video after a dialogic questioning 252 EWIN ET AL. intervention. Developmental Psychology, 49(12), 2368–2382. <https://doi.org/10.1037/a0032463>
8. Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. Journal of Human–Robot Interaction, 1, 78–95. <https://doi.org/10.5898/JHRI.1.1.Tanaka>
9. Han, J., Jo, M., Jones, V., & Jo, J. H. (2008). Comparative study on the educational use of home robots for children. Journal of Information Processing Systems, 4(4), 159-168, <http://dx.doi.org/10.3745/JIPS.2008.4.4.159>.
10. Kanda, T., Sato, R., Saiwaki, N., & Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction
11. Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., et al. (2011). On the effectiveness of robot-assisted language learning. ReCALL, 23(1), 25-58, <http://dx.doi.org/10.1017/S0958344010000273>.
12. Nalini Ambady and Robert Rosenthal. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological bulletin 111, 2 (1992), 256.
13. J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. biometrics (1977), 159–174.
14. Benjamini, Y. and Hochberg, Y. (1995), Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57: 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
15. Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". *Econometrica*. **37** (3): 424–438. [doi](https://en.wikipedia.org/wiki/Doi_(identifier)):[10.2307/1912791](https://doi.org/10.2307%2F1912791). [JSTOR](https://en.wikipedia.org/wiki/JSTOR_(identifier)) [1912791](https://www.jstor.org/stable/1912791).
16. Diebold, Francis X. (2007). [*Elements of Forecasting*](https://www.sas.upenn.edu/~fdiebold/Teaching221/FullBook.pdf) (PDF) (4th ed.). Thomson South-Western. pp. 230–231. [ISBN](https://en.wikipedia.org/wiki/ISBN_(identifier)) [978-0324359046](https://en.wikipedia.org/wiki/Special:BookSources/978-0324359046).
17. Hamilton, James D. (1994). [*Time Series Analysis*](http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_02_Hamilton_Time-Series-Analysis.pdf) (PDF). Princeton University Press. pp. 306–308. [ISBN](https://en.wikipedia.org/wiki/ISBN_(identifier)) [0-691-04289-6](https://en.wikipedia.org/wiki/Special:BookSources/0-691-04289-6).