Faculty of Humanities
Version September 2014

PLAGIARISM RULES AWARENESS STATEMENT

Fraud and Plagiarism
Scientific integrity is the foundation of academic life. Utrecht University considers any form of scientific deception to be an extremely serious infraction. Utrecht University therefore expects every student to be aware of, and to abide by, the norms and values regarding scientific integrity.

The most important forms of deception that affect this integrity are fraud and plagiarism. Plagiarism is the copying of another person's work without proper acknowledgement, and it is a form of fraud. The following is a detailed explanation of what is considered to be fraud and plagiarism, with a few concrete examples. Please note that this is not a comprehensive list!

If fraud or plagiarism is detected, the study programme's Examination Committee may decide to impose sanctions. The most serious sanction that the committee can impose is to submit a request to the Executive Board of the University to expel the student from the study programme.

Plagiarism
Plagiarism is the copying of another person's documents, ideas or lines of thought and presenting it as one's own work. You must always accurately indicate from whom you obtained ideas and insights, and you must constantly be aware of the difference between citing, paraphrasing and plagiarising. Students and staff must be very careful in citing sources; this concerns not only printed sources, but also information obtained from the Internet.

The following issues will always be considered to be plagiarism:

- cutting and pasting text from digital sources, such as an encyclopaedia or digital periodicals, without quotation marks and footnotes;
- cutting and pasting text from the Internet without quotation marks and footnotes;
- copying printed materials, such as books, magazines or encyclopaedias, without quotation marks or footnotes;
- including a translation of one of the sources named above without quotation marks or footnotes;
- paraphrasing (parts of) the texts listed above without proper references: paraphrasing must be marked as such, by expressly mentioning the original author in the text or in a footnote, so that you do not give the impression that it is your own idea;
- copying sound, video or test materials from others without references, and presenting it as one's own work;
- submitting work done previously by the student without reference to the original paper, and presenting it as original work done in the context of the course, without the express permission of the course lecturer;
- copying the work of another student and presenting it as one's own work. If this is done with the consent of the other student, then he or she is also complicit in the plagiarism;
- when one of the authors of a group paper commits plagiarism, then the other co-authors are also complicit in plagiarism if they could or should have known that the person was committing plagiarism;
- submitting papers acquired from a commercial institution, such as an Internet site with summaries or papers, that were written by another person, whether or not that other person received payment for the work.

The rules for plagiarism also apply to rough drafts of papers or (parts of) theses sent to a lecturer for feedback, to the extent that submitting rough drafts for feedback is mentioned in the course handbook or the thesis regulations.
The Education and Examination Regulations (Article 5.15) describe the formal procedure in case of suspicion of fraud and/or plagiarism, and the sanctions that can be imposed.

Ignorance of these rules is not an excuse. Each individual is responsible for their own behaviour. Utrecht University assumes that each student or staff member knows what fraud and plagiarism entail. For its part, Utrecht University works to ensure that students are informed of the principles of scientific practice, which are taught as early as possible in the curriculum, and that students are informed of the institution’s criteria for fraud and plagiarism, so that every student knows which norms they must abide by.

I hereby declare that I have read and understood the above.

Name: Bo Molenaar

Student number: 5981107

Date and signature: 7 September 2019

Submit this form to your supervisor when you begin writing your Bachelor’s final paper or your Master’s thesis.

Failure to submit or sign this form does not mean that no sanctions can be imposed if it appears that plagiarism has been committed in the paper.

Bo Molenaar
5981107
Willem Schuylenburglaan 60, 3571SK, Utrecht
BA Thesis
Supervisor: Aoju Chen
19 August 2019
5922 words

Acoustic-prosodic entrainment by social robots tutoring English vocabulary

Abstract

In recent years, a growing body of research has investigated social robots as tutors. Robot-assisted language learning (RALL) is a promising application of employing social robots to help both children and adults acquire a language, which is a fundamental domain of knowledge. The present study attempted to provide new insights into RALL by introducing acoustic-prosodic entrainment as a facilitative tool for tutoring. It was hypothesised that pitch-level entrainment by a Nao robot during a word learning task performed with school-aged children would result in increased learning. The results indicated entrainment had no significant effect on participants' learning. This unexpected finding is likely due to methodological shortcomings and an implementation of pitch-level entrainment that corresponds poorly to previous related studies. In conclusion, it was identified that acoustic-prosodic entrainment may facilitate RALL, but due to methodological limitations, this study did not find the expected evidence supporting that claim.

Key words: entrainment, acoustic-prosodic, robot-assisted language learning, word learning

Acknowledgements

I would like to thank Breixo Soliño Férnandez for helping me design and carry out this project and Aoju Chen for her advice and supervision throughout these months, as well as the funding provided by her Westerdijk Talent grant. Special thanks to Judith Brinksma for her invaluable assistance as an experimenter and providing us with excellent recordings, along with Nelleke Jansen. I am also grateful for the help provided by Dr. Emilia Barakova and her Social Robotics Lab as well as Mirjam de Haas, from TU Eindhoven. I would like to thank Kees de Korte and the teaching staff of the Noachschool in Schoonrewoerd for facilitating us with a great location and helping us find participants. I am very grateful for their help throughout the project. I would also like to extend my thanks to the children that participated and their parents. Finally, a thank you to Niki Liebregts for supporting me emotionally every step of the way.

Table of contents

1. Introduction

As children learn a language, it is crucial to their learning progress that they receive appropriate tutoring and feedback. The role of tutoring becomes especially relevant when children start learning a second language (L2). A child develops meta-linguistic awareness around the same time that they start to learn a second language, which allows L2 teaching to make use of explicit teaching methods such as focus-on-form(s) in grammar (Sheen, 2002). An effective language tutor should be able to make use of explicit teaching methods and tutor a learner as they engage with these methods.

Notice that the term ‘language tutor’ does not necessarily entail a person qualified to be a language tutor. After all, there are other ways of tutoring. An intelligent tutoring system (ITS) may be able to tutor a child’s language learning just as effectively. It is not difficult to imagine a computer programme or mobile application performing L2 tutoring – in fact, these already exist. A digital solution, however, may also take a more physical form to promote a sense of engagement and to more closely resemble the ‘gold’ standard of a human tutor. This can be achieved by using robots (especially social, humanoid robots). In a meta-analysis of the use of social robots for education, it was shown that social robots can indeed function as language tutors. However, the authors also emphasise the shortcomings of social robots for education. Due to technological constraints such as speech recognition, robots are not yet able to understand children well enough to react to the situation appropriately. Combined with a lacking ability to generate appropriate verbal and nonverbal output this considerably constrains their educational ability (Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018).

Within the field of second language acquisition (SLA) research, there is an increasingly large number of studies investigating how to improve robots as language tutors. A recent doctoral dissertation by van den Berghe (2019) set out to identify the state of robot-assisted language learning (RALL) and its challenges. While the reviewed literature suggested that the use of a robot peer in L2 learning activities would match the positive learning effect of a human peer, the author found no evidence supporting their hypothesis that “children would learn more in both peer conditions compared to when they performed the learning task without a peer” (van den Berghe, 2019, p. 105). It is worth investigating these inconsistent findings across studies to clarify the effects of robot peers in education.

The present study attempts to extend the body of literature on RALL by introducing entrainment, the event of interlocutors adapting to one another, as a tool to be used by a robot tutor to facilitate L2 learning. To closely relate to van den Berghe (2019) as well as other RALL and social robot tutoring studies (Belpaeme et al., 2018), the present study uses a Nao robot. The robot was named Robin for its gender neutrality, to prevent a gendered engagement bias, and after the Nao robot named Robin in Vogt et al. (2019). We decided to test school-aged children, as “school-aged children and adults demonstrate a more consistent picture [than young children], showing clear word learning across studies” (van den Berghe, 2019, p. 38).

The remaining part of this paper proceeds as follows: the second section presents previous research on social robots and entrainment in tutoring. The third section introduces the present study, its thesis and hypothesis. The fourth section presents the method. The fifth and sixth sections will present the results and a discussion thereof, followed by the conclusion in the seventh section.

2. Theoretical background

2.1 Robot assisted language learning (RALL)

Recent research on social robots has pointed to their usefulness in second language tutoring, namely through robot assisted language learning (RALL). RALL usually refers to a humanoid or animal-shaped social robot teaching language to an individual or a group of people. RALL has been found to benefit various types of language learning such as grammar, word learning, reading and speaking skills. Although studies have introduced RALL to various age groups, it appears to be most effective with school-aged students and adults (van den Berghe, 2019).

Vocabulary acquisition research has pointed out the benefits of RALL with students of various ages. Mazzoni & Benvenuti (2015) found that Italian preschool students improved more on an English vocabulary acquisition task when paired with a social robot partner than with a child partner. Eimler, von der Pütten, Schächtle, Carstens, and Krämer (2010) reported that German primary school students partaking in an English vocabulary acquisition task showed a greater learning effect when receiving tutoring from a social robot than those learning the German-English vocabulary without any tutoring. Lastly, Alemi, Meghdari, and Ghazisaedy (2014) found that Iranian junior high school students learned more from an English vocabulary acquisition task when receiving additional tutoring from a social robot during English lessons compared to those only receiving tutoring from their human teacher. Although these studies all focus on teaching English and thus offer a limited view of SLA, they clearly show that language tutoring by social robots has a positive effect on vocabulary acquisition. Moreover, it can be concluded that both preschool, primary school and junior high school students can benefit from RALL.

Several limitations of RALL have been identified in van den Berghe (2019). Most importantly, it is unclear whether interaction with robot peers or tutors adds the same learning gain as interaction with human peers or tutors. The present study should thus take into account that there is no clear relationship between having a robot tutor and L2 learning gain.

2.2 Entrainment

Robot-assisted learning can be improved by implementing mechanisms that are known to raise the efficacy of learning tasks performed by humans. Entrainment, the event of a speaker adapting to their interlocutor during an interaction, is one such mechanism. For example, we can entrain to an interlocutor’s syntactic structure (Reitter, Keller, & Moore, 2006, 2011), pronunciation (Pardo, 2006) or fundamental frequency (Levitan & Hirschberg, 2011).

2.2.1 Entrainment by human interlocutors

According to Gravano, Benus, Levitan, and Hirschberg (2014), there is a link between entrainment and “positive conversational attributes, including task success, smoothness of interaction, speaker attitude, cooperation, social attractiveness, and power relations, inter alia” (p. 1). This belief is motivated by Communication Accommodation Theory (Giles et al., 1991), which posits that “speakers converge their speech behaviour to that of their interlocutor in order to minimize social distance” (Gravano et al., 2014, p. 1). Moreover, Gravano et al. (2014) provide clear evidence that entrainment on the intonational contour level correlates with increased engagement, which in turn is known to promote learning (Carini, Kuh, & Klein, 2006).

Assuming the positive effect that increased engagement has on learning and the increased engagement in the case of entrainment, the link between entrainment and learning can be evaluated more directly. Entrainment in interpersonal interaction has been found to correlate with learning in several studies. Friedberg, Litman, & Paletz (2012) analysed lexical entrainment by groups of undergraduate students working on a project and found that the groups who received high scores for said projects used entrainment significantly more than the low scoring groups. Another study by Sinha and Cassell (2015) on reciprocal peer tutoring by students also found a significant positive correlation between entrainment and learning. Considering these studies, a positive correlation between entrainment by human interlocutors and learning is to be expected.

2.2.2 Entrainment by social robots

Given what is known on entrainment between human interlocutors, let us now consider the behaviour of a social robot tutoring a student. Robot tutors, like their human counterparts, will likely speak as part of their tutoring effort. Speaking is especially important for language learning tasks because a robot may give spoken feedback to a student, and because a student should learn both the orthographic and phonetic representation of words.

Considering the various cases where a social robot may speak to a student, entrainment in RALL is expected to be effective on the acoustic-prosodic level. It is hard to imagine a different level that social robots may use to entrain to students. Speech contains several directly measurable features such as pitch and loudness, while other features such as speech rate or syntactic structure are more complex and harder to code online. Entraining to acoustic-prosodic features requires less computational power from the robot than entrainment on other levels and is thus easier to implement in a broad range of RALL research.

There is a relatively small body of literature on acoustic-prosodic entrainment by social robots. The authors of these studies have generally opted for entrainment on the pitch level. Sadoughi, Pereira, Jain, Leite and Lehman (2017) explored the effect of pitch-level entrainment on children's engagement when playing a fast-paced cooperative game with a robot. The experiment consisted of two rounds, with children divided over two conditions. In one condition, the first round involved entrainment and the second did not, and vice versa for the second condition. The results indicated significantly higher engagement for participants in the entrainment-first condition. Interestingly, these participants also retained more engagement over the course of the experiment than the entrainment-second condition. This implies that children are sensitive to the immediacy of entrainment by their robot partner, and that their resulting engagement may remain for the duration of their interaction, even if the robot has already stopped entraining.

Acoustic-prosodic entrainment by social robots also formed the central focus of a study by Lubold, Walker, Pon-Barry, and Ogan, which measured learning effects for middle school students performing a mathematics task with an entraining social robot. Lubold et al. employed a Nao robot with the ability to use social dialogue and entrain to the participant's pitch. They compared three conditions: a non-social (control) condition, a social condition where the robot used social dialogue, and a social + entrainment condition where the robot also used entrainment by convergent pitch. Results indicated a significant improvement in learning between the social + entrainment condition and the control condition and a significant improvement between the social and control condition. Notably, there was no non-social + entrainment condition, therefore this study does not measure the effect of entrainment alone. The authors found no significant difference in learning between the social + entrainment and social condition, meaning this study does not point towards an increased learning effect as a result of acoustic-prosodic entrainment alone. However, it may facilitate social dialogue during tutoring to result in more effective tutoring.

The studies reviewed here suggest that both tutoring by social robots and entrainment can facilitate language learning. The following section will introduce the present study, which explores the connection between robot-assisted language learning and the facilitative role of entrainment.

3. Present study

Previous studies have suggested that RALL is an effective method of tutoring school aged students, and that entrainment is an effective tool to facilitate engagement and learning. However, whether entrainment can facilitate RALL remains unknown. This study aims to contribute to research on social robot tutors and RALL by identifying whether entrainment can indeed facilitate RALL.

While entrainment may occur on several linguistic levels and may thus facilitate several types of language learning, this study focuses on acoustic-prosodic entrainment and its applicability in vocabulary tutoring. In order to more closely relate to the existing body of literature, and to bridge the gap between research on language-learning tasks and other learning tasks where entrainment has shown to improve learning, this study, too, employs pitch-level entrainment. The following research question was thus formulated:

RQ: Does pitch-level entrainment by social robots during L2 vocabulary tutoring improve children's learning?

This study tests the null hypothesis *H0: learning(entrainment*$^{+}$*) = learning(entrainment*$^{-}$ *)*. Learning is measured as the difference between post-test and pre-test scores on vocabulary tests. Based on previous research discussed here, the following hypothesis was proposed:

H1: Pitch-level entrainment by social robots during L2 vocabulary tutoring will result in an increased difference between children's scores on the post-test *relative to the pre-test*.

4. Method

4.1 Experimental design

The experiment followed a pre-test–training–post-test design, which allowed us to assess the effect entrainment had on learning for each participant individually. The pre-test served to measure the participant's prior knowledge of the words in the learning task. During the training phase, participants completed a word learning task with Robin. In the experimental, or entrainment condition, Robin entrained to the participant's mean pitch during this task. In the control condition, Robin did not entrain to the participant. Finally, during the post-test, the participant’s knowledge of the words in the learning task was measured again. The difference between pre-test and post-test scores indicated the participant’s learning.

4.2 Participants

For the present study, 35 participants between ages 8:10 and 11:5 were tested. The participant pool was split into two homogeneous groups; one comprising the experimental (entrainment) condition and the other the control (no entrainment) condition. The mean age for the control condition was 10:3 and 10:4 for the experimental condition. Both groups comprised 11 female and 5 male participants. All participants were students of the Noachschool primary school in Schoonrewoerd, the Netherlands. In order to reduce the novelty effect of playing with a humanoid robot, all participants had the chance to meet Robin during a 10 minute introductory session in a classroom at the Noachschool, which took place six weeks before the testing sessions.

The participants were tested over the course of two weeks, either in a quiet classroom at the Noachschool during school time or in a private home near the school after school time. As the experiment setup was very similar in both locations, the testing location should not affect the participant’s performance. Moreover, the number of participants per condition was spread evenly across testing locations; effectively controlling for any differences in performance.

Two participants were excluded from the results because they were diagnosed as having a language impairment. One participant was excluded because they were Dutch-Italian simultaneous bilingual. Another participant was excluded from the results because they answered everything correctly on the pre-test, which means they did not learn any new words during the experiment and thus no learning effect could be measured. Lastly, another participant was excluded from the results because they had already been exposed to part of the experiment stimuli before partaking in it. Their increased exposure to these stimuli vis à vis other participants may have affected their ability to acquire certain words.

4.3 Materials

In addition to any mechanical equipment required, the two key parts to testing our research question were the implementation of an entrainment mechanism and word learning task in the robot.

4.3.1 Implementing entrainment in the robot

Robin entrained to the participant through the following steps. Firstly, the participant's speech was recorded using a dynamic microphone connected to a laptop. The recording started and stopped with one of the experimenters pressing a button on the laptop. These recordings were then passed through a programme that used a volume threshold to detect background noise at the peripheries of the recording and cropped the recording to remove this background noise, resulting in cleaner recordings. Secondly, the recorded speech sample was analysed in Praat (Boersma & Weenink, 2019) to get the mean pitch value for the utterance. Thirdly, the mean pitch was rounded to the nearest 5Hz value. Lastly, the rounded mean pitch was passed to Robin's audio player, together with the English translation of the word spoken by the participant, which allowed Robin to say the English word with a pitch value corresponding to the participant's pitch.

Robin's pool of English utterances was created by generating utterances using Robin's text-to-speech engine and using a Praat script to impose a mean pitch value on them. We opted for creating realisations at 5Hz intervals to ensure the minimal distance between two samples was smaller than the just noticeable difference between two pitch values, which is approximately 7Hz for English listeners (Jongman, Qin, Zhang, & Sereno, 2017).

In order to establish an expected pitch range for our participants, and thus find the pitch range Robin should be able to speak at, we analysed a set of speech data by children aged 8 and 10 years (balanced by gender). These data were recorded by Chen (2011), who elicited from their participants "[n]aturally spoken declarative sentences with either sentence-initial topic and sentence-final focus or sentence-initial focus and sentence-final topic" (p. 1055). We can expect declarative sentences by our participants of the same age range to have very similar prosodic features, and thus very similar mean pitch to these data. The mean pitch from 808 utterances ranged from 87Hz to 366Hz. We opted to use the 130Hz to 350Hz (or 3.05 SD) range to generate Robin's utterances. Robin's pool of English utterances thus consisted of 45 realisations of each English word appearing in the game with mean pitch values ranging from 130Hz to 350Hz at 5Hz intervals. Robin's English voice was set at 130Hz for the control condition in order to maximise the difference of Robin's pitch between conditions.

4.3.2 Word learning task design

The second key part to our experimental materials was designing a word learning task that allowed the robot to entrain to the participant's pitch. The word learning task was framed as a game in order to promote the student's engagement during the task. The game we designed introduced the task as an opportunity for the student to learn some of the words Robin had learned during its recent vacation to the UK, where it has had many conversations with local English speakers. This allowed us to present the student with a fixed selection of words in a fixed order whilst also making the game more interesting by adding a story element.

The selection of words consisted of 10 nouns and 10 verbs taken from PPVT-IV-EN (Dunn & Dunn, 2007) and can be found in Appendix A. All words were monosyllabic in both English and Dutch. The set of nouns and verbs each consisted of 3 target words and 7 filler words, which had been selected using two steps. Firstly, all selected words appear in PPVT-

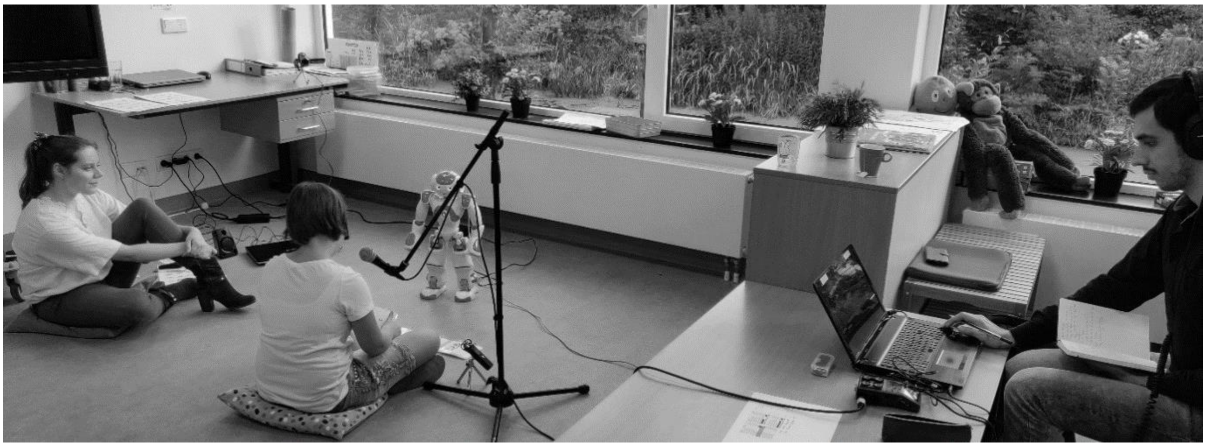
III-NL (Schlichting, 2005). The PPVT increases in difficulty from beginning to end, and a certain difficulty level corresponds to an estimated mental age – more specifically the age at which native speakers of the target language will typically be familiar with the words that appear at this difficulty level. Filler words were selected from difficulty levels corresponding to ages below and up to 8 years old, which is slightly less than the age of the youngest participants in order to maximise the possibility that all participants were familiar with the filler words in their L1. Target words were selected from difficulty levels corresponding to ages 10 to 12. In effect, this allowed us to expect participants to be familiar with the fillers in their L1 and not to be familiar with the target words in both their L1 and L2, since their L2 vocabulary will likely be less developed than their L1 vocabulary.

Secondly, in order to validate these expectations, the selected fillers and target words were passed by the participants' English teacher, who had affirmed our expectations of the participants' familiarity with the selected words in English.

In order to present the selected words to the participant, the Dutch translation of the word and the corresponding image from PPVT-IV-EN[1] were printed on A6 paper cards. These cards were laminated and bound with rings into booklets to make them presentable as ordered sets of words and to make it easy for a participant to hold a full set of word cards. For each word, two different cards were created with each featuring a different image of the word. Thus, there were four unique ordered sets of cards in total: two sets of verbs and two sets of nouns with unique images on each card and cards appearing in different orders for each noun/verb set.

[1] The fourth edition of the PPVT was used for this because has coloured images that make them easier to identify, nicer to look at, and not come across as archaic as the black-and-white images in the third edition.

Figure 1. Overview of testing setting



A dynamic microphone was used to record the participant's speech. Experiment sessions were logged with a video recorder. A laptop with loudspeakers, also placed on the floor, was used to conduct the pre-test and post-test. Furthermore, a table featuring a laptop to control Robin (an experimenter was present for this purpose) was set up in the room, together with the technical equipment required to facilitate this. Figure 1 displays an overview of the testing setting.

The pre-test was an altered version of PPVT-IV-EN, featuring images of the 20 words from the learning task and recordings of these words by a speaker of British English. Like in the original PPVT, participants were asked to say or point at which of the four images presented to them corresponded to the word they heard. Both the visual and acoustic stimuli were presented using a PowerPoint slideshow. Answers were marked on an answer sheet. The post-test followed the same approach but had different images and recordings by a different speaker of British English, which was done to prevent participants from simply memorising the combination of acoustic signal and image.

The robot's voice was not used for the acoustic stimuli in the pre-test and post-test because a previous study on acoustic-prosodic entrainment by social robots indicated that a participant's engagement with the robot is significantly affected by whether or not the robot starts entraining to their voice from the start. In effect, using a different voice for the tests ensured that participants would not have altered engagement levels during the experiment.

Furthermore, since the participants receive their English lessons from a speaker of British English, the same accent was used for the test stimuli to as to not create any confusion or difficulty by having various accents. However, since it was only possible to generate Robin's utterances in a General American accent, this may have happened still.

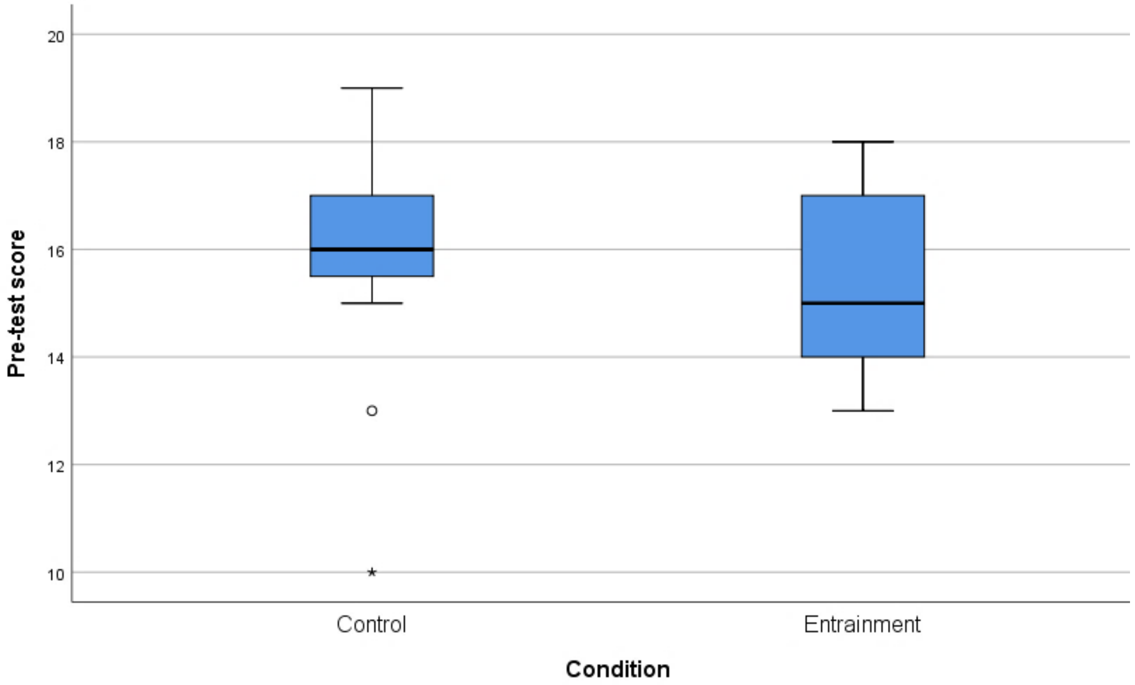
4.4 Procedure

A session started with the main experimenter leading the participant into the room, introducing themselves and the second experimenter, and sitting down the participant opposite Robin. The participant and Robin were seated on the floor (with a pillow for the participant) so that their heads were at about the same height, which enabled Robin track and look at the participant's face. When the participant sat down, the main experimenter explained to the participant that they were about to play a game with Robin that would allow them to learn some of the words Robin learned during its holiday. Robin then introduced themselves and explained its story before sending the participant to do the pre-test with the main experimenter.

After the pre-test, the participant returned to Robin, who explained the game and played a practice round with them using a set of cards that did not appear in the game. Two rounds were then played. During a round, the participant told all the Dutch words in a card set to Robin, who replied to each Dutch word with the English translation of that word. The first two rounds were followed by a 'break' where Robin told about its vacation and played some fun animations. Afterwards, the final two rounds were played. Robin then said goodbye to the participant and invited them to do the post-test with the experimenter, which concluded the session.

5. Analysis and Results

Figure 2. Boxplot diagram of pre-test scores per condition



The results of the pre-test and post-test are shown in Table 1. Figure 2 shows the distribution of pre-test scores per condition. An independent samples t-test showed no significant difference between pre-test scores in the control condition and the entrainment condition (t (28) = .653; *p*= .834), meaning the two groups of participants could be treated as equivalents for further analysis.

Table 1. Mean scores (and standard deviations) per test phase per condition

	Control (N=15)	Entrainment (N=15)
Pre-test score	15.87 (2.13)	15.40 (1.77)
Post-test score	18.07 (1.28)	17.00 (2.27)

Figure 3 shows the distribution of post-test scores per condition. According to the hypothesis of this study, the use of entrainment during the word learning task should improve learning by the participant between test phases. In other words, there is an expected positive interaction effect by condition and test phase on word recognition. A mixed effects binary logistic regression was used to find evidence for both the main effects (condition, test phase) and the expected interaction effect (condition × test phase). Results of the model are listed in Table 2. The model indicated a significant but weak positive effect for test phase. Figure 4 shows the estimated means of the effect for test phase. The model indicated a negligible, non-significant effect for condition. Notably, there was also no significant interaction effect for condition × test phase.

Figure 3. Boxplot diagram of post-test scores per condition

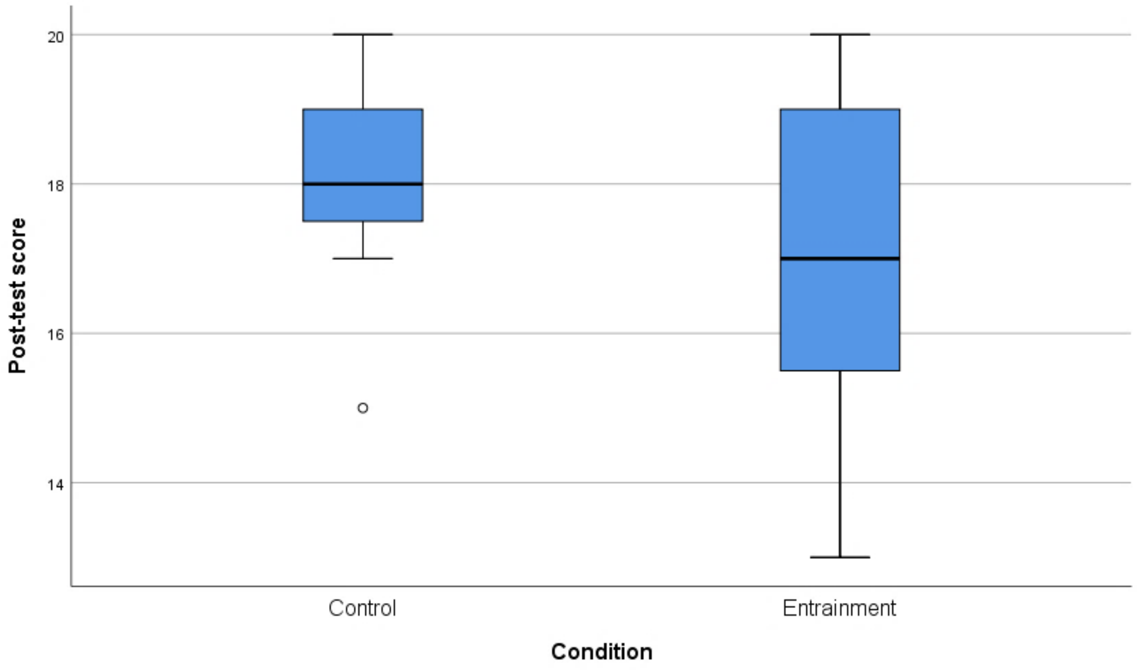


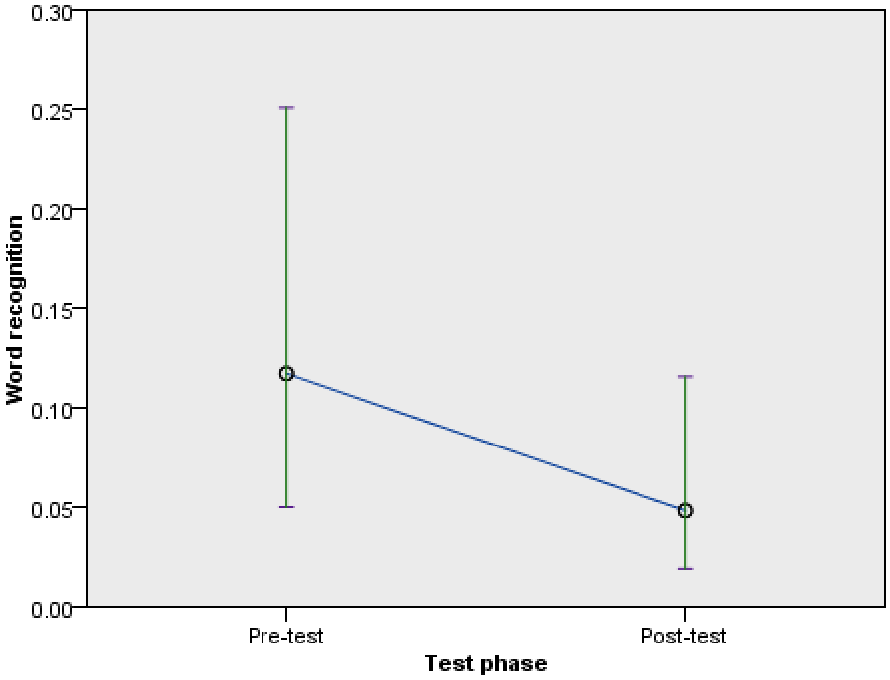
Table 2. Effect size estimates, exponential coefficients and confidence intervals for model predicting word recognition

	Effect size	Exponential coefficient	95% Confidence Interval for exponential coefficient	
			Lower	Upper
Fixed effects				
Intercept	3.31*	.04	.01	.10
Condition	.64	1.90	.91	3.97

Test phase	1.20*	3.31	1.91	5.75
Condition × Test phase	-.44	.65	.31	1.36
			95% Confidence Interval	
			Lower	Upper
Random effects				
Within-person variability	.42*		.18	1.02
Word variability	3.57*		1.65	7.74

* *p < .05*

Figure 4. Estimated means (including 95% CI) for test phase effect



6. Discussion

This study failed to find a significant main effect for condition on word recognition and a significant interaction effect for condition × test phase. Hence, the null hypothesis that learning in the entrainment condition equals learning in the control condition cannot be rejected. One interesting finding is the significant main effect for test phase on word recognition, which indicates that participants indeed learned some words as a result of the task performed with Robin. This finding was expected and in accordance with previous RALL research (Alemi et al., 2014; Eimler et al., 2010; Elvis Mazzoni & Martina Benvenuti, 2015). However, it cannot be considered evidence for RALL as an effective method of language tutoring because there was no condition controlling for the presence of the robot. Although it is expected to be true, the results do not indicate whether the robot's presence was of any effect because none of the participants performed the learning task without Robin and thus no effect could be measured.

Several methodological shortcomings may have contributed to this unexpected result. Firstly, the pre-test and post-test scores contained little intrapersonal variation. Each test was comprised of 14 filler words, which participants were expected to be familiar with, and 6 target words, which were expected to be new to them. However, most participants (83.3% on the pre-test, 93.3% on the post-test) were already familiar with one of the target words, 'brain'. It was therefore considered a filler word, increasing the ratio of filler versus target words to 15:5. The large amount of familiar words not only left little room for intrapersonal variation, but also skewed each individual result towards the maximum of 20 correct answers. A smaller range of variation between scores makes it more difficult to measure any effects.

Secondly, although informally assessed by the researchers, engagement by the participants was not measured. Hence, no relationship between engagement and test scores could be defined. It is possible that participants in the control condition already reached such a high level of engagement due to the robot's novelty effect that there was little to no room for increased engagement, and thus increased learning, in the entrainment condition. This concern is also found in van den Berghe (2019). Future work could improve upon this study by adding a survey inquiring about the participant's opinion towards the experiment and the robot, or by assessing perceived engagement using video footage of the participant interacting with the robot.

Thirdly, the method of entrainment in the robot was not consistent throughout a test session. The robot's utterances in Dutch were far longer than those in English, which made implementing entrainment in the same way for each language technically challenging. Pitch proximity on the word level was used for English, meaning the robot attempted to match the mean pitch of the participant's pronunciation of the Dutch word directly. However, this method could not be used for the Dutch utterances. Imposing a mean pitch on full sentences or even multiple sentences in a row led to unnatural sounding speech, which would be decremental to the robot's intelligibility and ability to increase engagement and learning using entrainment. We thus opted for entrainment by pitch convergence on the session level for Dutch by raising the robot's pitch by 7.5% between the introduction and the start of the game. By using two different methods of entrainment, the robot's entrainment might not have been as consistent as intended and may not have had the expected effect of facilitating learning.

Importantly, the method of entrainment in this study also differs from the method used in Lubold et al. (2018), which is the only study so far to find an effect for entrainment facilitating learning. Whereas the robot in Lubold et al. entrained to the participant by converging their pitch to the participant's over time, the present study used pitch proximity. Convergence may be more effective than proximity because it generates more rapport: "a feeling of connection, harmony and friendship", which represents the same values as the engagement the present study attempted to raise (Lubold et al., 2018, p. 283). Indeed, in accordance with the expected relationship between entrainment, engagement and learning presented in this study, Lubold et al. suggest that "[a]n agent which converges may build more rapport and a partner who feels more rapport may learn more" (p. 285). The use of proximity rather than convergence in the present study may thus explain the lack of an effect for entrainment on learning.

7. Conclusion

This study explored the effect of pitch-level entrainment by a social robot in an L2 vocabulary learning task. A learning task was designed to allow the robot to teach English words to a group of 34 monolingual Dutch participants. Using a combination of pitch proximity on the word level and pitch convergence on the session level, this study attempted to find increased learning in the entrainment condition vis à vis the control condition. The expected interaction effect for condition × test phase was not found. This may be due to any single or a combination of several methodological problems, most notably a lack of engagement assessment and not implementing entrainment in accordance with previous studies that found a positive effect for entrainment on learning. A significant effect was found for test phase on word recognition, indicating that participants acquired some words during the test session. However, this result cannot be used as evidence for RALL as an effective method of language tutoring because the present study did not control for the robot's presence.

Despite the unexpected findings reported here, this study may be of value to future RALL research. The identified methodological shortcomings can easily be overcome in a future study, and considering the body of literature pointing towards a positive effect of acoustic-prosodic entrainment on learning, such a study is likely to find evidence supporting the hypothesis tested here.

8. References

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2014). Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools. *International Journal of Humanoid Robotics*, *11*(03), 1450022. https://doi.org/10.1142/S0219843614500224

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (Version 6.1). Retrieved from http://www.praat.org/

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student Engagement and Student Learning: Testing the Linkages*. *Research in Higher Education*, *47*(1), 1–32. https://doi.org/10.1007/s11162-005-8150-9

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, *38*(5), 1055–1083. https://doi.org/10.1017/S0305000910000541

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test IV*. Circle Pines, MN: American Guidance Service.

Eimler, S., von der Pütten, A., Schächtle, U., Carstens, L., & Krämer, N. (2010). Following the White Rabbit – A Robot Rabbit as Vocabulary Trainer for Beginners of English. In G. Leitner, M. Hitz, & A. Holzinger (Eds.), *HCI in Work and Learning, Life and Leisure* (Vol. 6389, pp. 322–339). https://doi.org/10.1007/978-3-642-16607-5_22

Elvis Mazzoni, & Martina Benvenuti. (2015). A Robot-Partner for Preschool Children Learning English Using Socio-Cognitive Conflict. *Journal of Educational Technology & Society*, *18*(4), 474–485. Retrieved from JSTOR.

Friedberg, H., Litman, D., & Paletz, S. B. F. (2012). Lexical entrainment and success in student engineering groups. *2012 IEEE Spoken Language Technology Workshop (SLT)*, 404–409. https://doi.org/10.1109/SLT.2012.6424258

Giles, P. H., Giles, H., Coupland, J., Coupland, N., Oatley, K., Oatley, P. E. D. of H. D. & A. P. K., … Press, C. U. (1991). *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge University Press.

Gravano, A., Benus, S., Levitan, R., & Hirschberg, J. (2014). Three ToBI-based measures of prosodic entrainment and their correlations with speaker engagement. *2014 IEEE Spoken Language Technology Workshop (SLT)*, 578–583. https://doi.org/10.1109/SLT.2014.7078638

Jongman, A., Qin, Z., Zhang, J., & Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America*, *142*(2), EL163–EL169. https://doi.org/10.1121/1.4995526

Levitan, R., & Hirschberg, J. (2011). *Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.* 4. Florence, Italy.

Lubold, N., Walker, E., Pon-Barry, H., & Ogan, A. (2018). Automated Pitch Convergence Improves Learning in a Social, Teachable Robot for Middle School Mathematics. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, … B. du Boulay (Eds.), *Artificial Intelligence in Education* (Vol. 10947, pp. 282–296). https://doi.org/10.1007/978-3-319-93843-1_21

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Reitter, D., Keller, F., & Moore, J. D. (2006). Computational modelling of structural priming in dialogue. *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers on XX - NAACL '06*, 121–124. https://doi.org/10.3115/1614049.1614080

Reitter, D., Keller, F., & Moore, J. D. (2011). A Computational Cognitive Model of Syntactic Priming. *Cognitive Science*, *35*(4), 587–637. https://doi.org/10.1111/j.1551-6709.2010.01165.x

Sadoughi, N., Pereira, A., Jain, R., Leite, I., & Lehman, J. F. (2017). Creating Prosodic Synchrony for a Robot Co-player in a Speech-controlled Game for Children. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, 91–99. https://doi.org/10.1145/2909824.3020244

Schlichting, L. (2005). *Peabody picture vocabulary test-III-NL*. Amsterdam, the Netherlands: Hartcourt Assessment BV.

Sheen, R. (2002). 'Focus on form' and 'focus on forms'. *ELT Journal*, *56*(3), 303–305. https://doi.org/10.1093/elt/56.3.303

Sinha, T., & Cassell, J. (2015). Fine-Grained Analyses of Interpersonal Processes and Their Effect on Learning. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education* (pp. 781–785). Cham: Springer International Publishing.

van den Berghe, R. (2019). *Social robots as second-language tutors for young children: Challenges and opportunities* (PhD Thesis). Universiteit Utrecht.

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., … Pandey, A. K. (2019). Second Language Tutoring Using Social Robots: A Large-Scale Study. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 497–505. https://doi.org/10.1109/HRI.2019.8673077

Appendix A: Selected words

The words above the horizontal line are filler words. Those below it are target words.

Category 1: Nouns	Category 2: Verbs
boat	eat
dog	read
cat	walk
mouth	dance
shoe	drink
ship	swim
ear	wash
rope	dig
brain	lift
tin	pour

Practice words
Ball
Eye
Hand
Sheep
Sleep