UTRECHT UNIVERSITY

HUMANITIES

Entrainment in Social Robots

THE INFLUENCE OF PROSODIC ENTRAINMENT BY SECOND LANGUAGE TUTOR ROBOTS ON STUDENT ENGAGEMENT

7.5 ECTS BA THESIS ENGLISH LANGUAGE AND CULTURE

Author:
Alessandra Polimeno

First assessor:
Prof. dr. Aoju Chen (UU)

Second assessor:
Dr. Emilia Barakova (TU/e)

June 25, 2020

Abstract

This thesis investigates the influence of prosodic entrainment by a social robot on student engagement in a second language learning setting. Two approaches are adopted to analyze engagement, namely by means of speech and behavior analysis. Speech analysis investigates pitch range and utterance duration, and behavior analysis is concerned with eye gaze and facial expressions. The results indicate that the entrainment group displays a significantly higher percentage of eye gaze towards robot, hinting at increased behavioral engagement compared to the control group. For the remaining three features, no significant difference between the groups was found. However, the absolute values of the features are generally higher in the entrainment group. Taking into account the small sample size, the results of this study cautiously suggest that entrainment in social robots can enhance student engagement.

Key words: student engagement, prosodic entrainment, robot-assisted language learning

Contents

1 Introduction

In the past few decades, robots have started to take a place in society where they can assist humans in a variety of ways. Where at first this mostly concerned the automation of relatively simple tasks, much research is currently carried out on how to design socially intelligent robots that can have meaningful interactions with people. For researchers, an interesting question to ask is how social robots can support humans, for example in pedagogical settings, where they may contribute to learning gains. Recently, studies have investigated robots as an educational tool in second language (L2) learning contexts. Robot-assisted language learning (RALL) is a promising method with which both children and adults can receive personalized feedback, which is essential in L2 learning (Okita & Schwartz, 2013). One of the factors that make robots a suitable language tutor is their physical presence. For example, they can use their body to manipulate objects, perform collaborative tasks with the learner, convey meaning with gestures and create joint attention (van den Berghe, 2019). This situatedness facilitates language learning (Barsalou, 2008; Hockema & Smith, 2009). However, studies that investigate learning gains resulting from a robot tutor yield mixed results, indicating that more research is needed to clarify the effects of social robots in education (van den Berghe, 2019).

A linguistic feature that has received attention because of its potential contribution to learning gains in the context of social robots is speech entrainment. This is the process in which the speech of the interlocutors becomes more similar to each other's during an interaction. Entrainment has been found to diminish social distance, often resulting in more positive interactions (Beňuš, 2014), and to increase feelings of liking between speakers (Chartrand & Bargh, 1999). Several studies report that entrainment can have positive effects on learning gains (Friedberg, Litman, & Paletz, 2012; Sinha & Cassell, 2015). Others report that entrainment can increase student engagement (Carini, Kuh, & Klein, 2006). For these reasons, designing a tutor robot that can entrain to students' speech could positively influence their academic performance, either by directly enhancing learning gains, or indirectly by increasing engagement.

While in some studies on robot-assisted learning in educational settings, the positive learning outcomes resulting from entrainment have been found to transfer to human-robot interaction (Thomason, Nguyen, & Litman, 2013), a direct effect is not always present. Soliño Fernández (2020) investigated whether implementation of speech entrainment in a tutor robot improved second language learning results of Dutch primary school children. Contrary to the expectations, the study found no difference in test scores between the control group and the group that interacted with an entraining robot. However, Soliño Fernández (2020) did not measure the degree to which the robot increased student engagement and motivation, while it is plausible that this aspect is positively affected by entrainment. For this reason, the current project aims to investigate whether speech entrainment in social robots can increase student engagement by reanalyzing the data collected by Soliño Fernández (2020).

2 Theoretical Background

2.1 Robot-assisted language learning

A classroom L2 learning context can potentially benefit from the use of technologies such as computers and intelligent software because they can support learners in ways that are impossible in traditional classrooms (van den Berghe, 2019). For example, learners can have one-to-one conversations with a well-designed chat program in their target language, which might provide both native-like input and personalized feedback. Technology-supported ed-

ucational tools such as chat bots can be especially helpful for children who are struggling with the target language, as the teacher might not always have enough time to provide the help the learner needs. Moreover, language learning technologies can theoretically provide input in any language. This can be useful in classrooms where learners with different language backgrounds are present, for example migrant children with a first language that is different from the language that is spoken in class. Balanced bilingual development can be achieved if both languages receive sufficient practice. Technology-based language tools might be able to enhance this process by providing input in the bilingual students' L1, and possibly by providing learning strategies for the target language that align with their L1. In this way, RALL can contribute to a balanced bilingual development (Blom, 2019, as cited in van den Berghe, 2019)

A relatively new technology that has entered the stage is social robotics. Social robots are designed to interact and communicate with humans in a lifelike manner, often being equipped with a body resembling a human or animal (van den Berghe, 2019). Herein lies one of their strengths compared to other technologies: their physical reality poses advantages for language learning (Barsalou, 2008; Hockema & Smith, 2009; van den Berghe, 2019). Robots can use gestures or the environment to convey word meanings in the same way as humans do when they communicate. The use of non-verbal cues such as pointing, eye gaze and gestures results in a more natural interaction compared to what is attainable for other technologies (van den Berghe, 2019).

As mentioned above, the results of studies investigating the possible learning gains of RALL are mixed. Most of the research in this area is concerned with vocabulary learning (van den Berghe, 2019). In a review study, van den Berghe (2019) summarized several studies on social robots in language learning. While some research on robot-assisted vocabulary learning found substantial learning gains (de Wit et al., 2018), many others report more moderate learning gains (Movellan, Eckhardt, Virnes, & Rodriguez, 2009) or gains that only hold for a part of the participants (Gordon et al., 2016). Meanwhile, other studies found no learning gains (Soliño Fernández, 2020).

These mixed results can partially be explained by the complexity of both the effects that social robots can have on people and learning in general (van den Berghe, 2019). Another challenge in investigating RALL is the so-called novelty effect. This effect can occur when participants encounter a new technology for the first time. The newness sparks their excitement and motivation, resulting in higher learning outcomes. However, these gains reduce when the participants are more familiar with the technology, and the initial interest diminishes (van den Berghe, 2019). Another factor that makes researching RALL more difficult are technical limitations. Both speech recognition and generation are currently far from perfect, which can decrease the naturalness of the interaction (van den Berghe, 2019).

2.2 Entrainment

Entrainment is the process in which speakers become more similar to each other in terms of their communicative behavior (Beňuš, 2014). This convergence can be on linguistic features, such as syntactic structures, speaking rate and pitch, but also on non-verbal behavior, including gestures (Beňuš, 2014). Moreover, entrainment facilitates positive feelings amongst speakers (Chartrand & Bargh, 1999), and entraining speakers are regarded as more competent (Street Jr, 1984). Additionally, entrainment is often used to narrow social distance, i.e. to establish a more intimate relationship with the interlocutor. On the other hand, disentrainment is often perceived as signaling a negative attitude, thus impeding fruitful interactions and leading to a larger social distance between the speakers (Beňuš, 2014). Entrainment is not solely reserved for other human interlocutors, as humans have been found to entrain to robots or computers as well (Beňuš, 2014; Oviatt, Darves, & Coulston, 2004). For example, Breazeal (2002) found that participants entrained to a robot's speech intensity and

social cues.

A positive effect of entrainment in education has been proposed by several authors (Friedberg et al., 2012; Sinha & Cassell, 2015). Sinha & Cassell (2015) investigate whether speech alignment among students who tutored one another on algebraic concepts and procedures improved their performance on a related task. They found that a high degree of speech entrainment positively correlates with learning gains. Friedberg et al., (2012) compared the performance of groups on a project and found that the high performing groups displayed a significantly larger amount of lexical entrainment. Aiming to extend this effect to human-computer interaction, Thomason, Nguyen & Litman (2013) investigate the relationship between pitch entrainment and task performance of students interacting with a tutor dialogue system. They found that the amount of entrainment by students to the dialogue system was positively correlated with learning gains. These results support the claim that the relationship between positive learning outcomes and entrainment can be extended to human-computer interaction.

Most of the previously discussed research concerns the relationship between students' learning gains and the amount to which they entrain to another entity. However, robots or dialogue systems can be designed to entrain to the human interlocutor as well. This poses a technical advantage in addition to the social benefits of implementing entrainment, since entrainment can be used as a cognitive constraint for the design and behavior of the robots (Beňuš, 2014). Beňuš (2014) argues that entrainment can provide a constraint that limits the degrees of freedom that exists when modelling the behavior of automatic systems. Limiting the degrees of freedom is beneficial as it improves the naturalness and efficiency of the system. Moreover, the implementation of an entrainment feature can improve speech recognition. A robot that produces speech at a certain pitch height that is ideal for its speech recognition software might trigger a response that entrains to this ideal pitch height. This, in turn, can facilitate easier speech recognition by the robot (Beňuš, 2014).

Not all studies on the influence of robot entrainment on learning gains find affirmative evidence. Soliño Fernández (2020) investigated whether implementation of entrainment in a social robot could facilitate language learning gains. Contrary to expectations, no effect was found. The experiment took place at a primary school in the Netherlands, where children performed an English vocabulary learning task on a one-to-one basis with a teaching humanoid Nao robot named Robin. In the experimental group, Robin entrained to the participant's mean pitch. In the control group, Robin's entrainment feature was not activated. The experiment followed a pre-test-training-post-test design, where the participants interacted with the robot during the training phase. The task successfully introduced new words to the participants, but there was no difference between the entrainment and control group. In other words, entrainment was not found to have an effect on learning gains. Soliño Fernández (2020) discusses several methodological issues, including the difficulty of the vocabulary learning task, which is thought to be too easy and limited. Moreover, the way entrainment is implemented in the robot forms a limitation. Entrainment could only be activated during the learning task, because of a lack of input from the participants during the other parts of the experiment. However, as Soliño Fernández (2020) points out, it remains unclear whether the entrainment mechanism triggered other positive effects that have been found in the context of entrainment, such as increased engagement. Gravano, Beňuš, Levitan & Hirschberg (2014) found that entrainment can increase the degree to which speakers are engaged in a conversation. Others found a positive effect of engagement on learning gains (Dörnyei, 1994; Carini et al., 2006). Since this relationship is not yet understood, it might be interesting to investigate the influence of entrainment on the degree of engagement of the participants.

2.3 Engagement

Since engagement has been found to positively correlate with learning gains (Dörnyei, 1994), investigating the relationship between entrainment and engagement might yield helpful insights into how educational tools, and more specifically social robots, can support both learners and teachers. Several studies have reported that the presence of a tutor robot can have a positive effect on student motivation and engagement. Children were reported to be less self-conscious and anxious about making mistakes in the presence of a robot than in the presence of a human teacher (Alemi, Meghdari, & Ghazisaedy, 2015). In the context of L2 learning classes, positive effects on engagement and motivation have been found in the domain of speaking, vocabulary, and overall learning experience. See van den Berghe (2019) for an overview of literature on this topic. Outside the domain of RALL, motivation has been positively influenced by the presence of a robot as well (Chin, Hong, & Chen, 2014).

2.3.1 Defining and measuring engagement

Engagement is an often ill-defined concept, as it overlaps with other constructs such as commitment and involvement. For this reason, Henrie, Halverson, & Graham (2015) emphasize the importance of clearly defining the concept when investigating engagement. In the current study, engagement is defined as "the psychological state of well-being, enjoyment and active involvement that is triggered by meaningful activities causing [participants] to be absorbed by the task, more energetic and in a more positive mood" (Perugia et al., 2018). This broad definition covers the three types of engagement as defined by Fredricks, Blumenfeld & Paris (2004): behavioral, emotional and cognitive engagement. These types are specifically developed for analyzing engagement in school settings. As argued by Fredricks, engagement unites these components in a meaningful way, resulting in a meta-construct which fuses behavioral, emotional and cognitive elements. Behavioral engagement includes actions that facilitate active participation in tasks, such as putting an effort into work, and following the rules. Emotional engagement entails the reactions of the student to the task, teachers or learning setting in general. This includes emotions such as boredom, happiness, interest and anxiety. Cognitive engagement encompasses intrinsic motivation and psychological investment in learning (Fredricks et al., 2004).

A second defining description of engagement is that the concept is a compound of objective elements (e.e. observable behavior) and subjective elements (e.g.self-reports) (Perugia, Díaz-Boladeras, Català-Mallofré, Barakova, & Rauterberg, 2020). In this thesis, only the observable dimension of engagement is investigated, as the subjective elements are impossible to measure in this case since no information on the internal state of the participants is available. The same reason prevents cognitive engagement from being examined in this thesis.

The distinction of the three types of engagement can result in a more detailed picture of how entrainment might influence engagement, but they are not measured individually. According to Fredricks et al.(Fredricks et al., 2004), operationalizing each type separately is not practical, and often not necessary, since they partly overlap. Additionally, developing reliable measures for each of the types requires more time than is available for this project. The aim of distinguishing different types of engagement is to be able to talk about it in a more specific and detailed manner.

There is no standardized method to measure engagement. Common methods include self-reports, observational checklists or rating scales and automated measurements using, for example, computer vision (Monk et al., 2003), and physiological measurements ((Perugia et al., 2020). Since automated measurements are beyond the scope of this project, and obtaining self-reports or physiological measurements of the participant are impossible as the data set used was collected almost a year prior to this analysis, this thesis uses observational features to measure engagement. This results in a measurement of perceived engagement, which can be defined as the degree of the participant's engagement as judged by an exter-

nal observer (Monk et al., 2003). Two types of features will be used. Firstly, the participants' speech is analyzed by extracting information on pitch range and utterance duration. This mainly covers behavioral engagement, because pitch range and duration reflect the effort of the participant (see section The Effort Code below). Secondly, participants' physical behavior, more specifically eye gaze and facial expressions, are analyzed in the light of both behavioral and emotional engagement respectively. Facial expressions can be regarded as a cue for emotional engagement, and eye gaze as a cue for behavioral engagement. The method section elaborates on these measurements.

2.3.2 The Effort Code

Two linguistic features will be used as cues for increased engagement, one of which is pitch range. The reason that this feature can signal engagement lies in one of the biological codes as proposed by Gussenhoven (2002). Intonation often expresses universal form-meaning relations, which derive from the physiological properties of the speech process. Based on these relations, three biological codes that account for the universal nature of paralinguistic meanings have been formulated, among which is the Effort Code. The Effort Code states that increased articulatory effort of an utterance results in a wider pitch range (Gussenhoven, 2002). In other words, pitch range reflects the articulatory effort of a speaker and the reason why the speaker increases their articulatory effort, such as engagement. A speaker who displays a low degree of engagement on a task is likely to exhibit evenly low articulatory effort. For engaged speakers, the reversed is true. This makes pitch range a relevant feature for analysing engagement.

The second linguistic feature that serves as a cue for engagement is utterance duration. In line with the Effort Code, a longer utterance duration can flag increased articulatory effort, which in turn signals increased engagement. Another interpretation of a long utterance duration is that the participant might speak more slowly in order to be more comprehensible if they are unsure of the robot's proficiency in their language. Lower speech rate and thus longer utterance duration can in this case be used by the participant to be helpful and facilitate successful communication. While trying to be helpful does not directly indicate a higher degree of engagement, it nevertheless signals that a participant is putting effort into successful communication, which justifies the use of longer utterance duration as a cue for engagement.

3 Current Study

This study aims to gain new insights into the potentially positive effects of robot entrainment in educational contexts. Specifically, it examines whether entrainment in the speech of a social robot can lead to increased student engagement in language learning settings. The following research question is formulated:

> *RQ: Does implementation of entrainment on mean pitch in social robots improve student engagement in L2 tutoring?*

As previous studies have suggested, entrainment can have a positive effect on engagement. The following hypothesis is consequently proposed:

> *H1: Entrainment on mean pitch by a social robot during a L2 vocabulary training task will result in increased student engagement during the task.*

To address the research question, the effects of prosodic entrainment on four types of engagement measurements were analyzed. Subsequently, the following predictions could be made:

> *P1: Pitch range will be larger in the entrainment group.*
>
> *P2: Utterance duration will be longer in the entrainment group.*

P3: The eye gaze of participants in the entrainment group will be focused on the robot for a longer time.
P4: The facial expressions of participants in the entrainment group will convey a higher percentage of positive emotions.

4 Methodology

4.1 Data

The data was taken from the study by Soliño Fernandez (2020). This data set originally contained recordings of 35 monolingual Dutch-speaking children, ages 8:10 to 11:5, interacting with tutor robot Robin on a one-to-one basis. Robin is NAO V4 humanoid robot, running the NAOqi OS version 2.1.4.13. For the current project, a total of seven participants were excluded. The reasons for their exclusion are the following: missing or incomplete sound or video files, being bilingual, having a language impairment, and being exposed to the experiment before taking part. In total, the number of participants in the current project is 28. Both the control group and the entrainment group consist of 14 participants.

Each session follows the same structure, as outlined in Soliño Fernandez (2020):

1. Introduction, where the participant was introduced to the robot and the task. In this phase, the robot presented themselves and asked a few introductory questions to the participant.
2. Testing (pre-test), where the familiarity of the participant with the English words from the learning task was tested.
3. Practice round, to prepare the participant for the learning task.
4. Training (part 1), in which the first half of the learning task was done, reviewing once each of both categories of words.
5. Break, where the robot and participant had a fun interaction. The robot told a story about their recent trip to England, asked the participant questions, and performed three entertaining tricks.
6. Training (part 2), the second half of the learning task.
7. Testing (post-test), which measured the number of new words that the participant had obtained as a result of the training phase.

The current study is only concerned with the interaction between the participant and the robot. Hence the testing phases, which were carried out by a human experimenter, are not analyzed here. The training task consisted of a word learning exercise. The participant was instructed to read a Dutch word from a booklet, after which the robot responded with the corresponding English translation. Two categories of words were tested: nouns and verbs, each consisting of ten different words. All words contained one syllable. In total, each word appeared in the experiment twice: once in the first training phase and once in the second. Each session lasted approximately 20 to 25 minutes. During the word learning exercise, the robot spoke in English and the participant in Dutch. Outside the task, the conversations were entirely in Dutch.

During the two parts of the training, the robot entrained to the mean pitch of the participants in the entrainment group only. Outside the training phase, the entrainment feature was disabled for both the entrainment and control group. This was necessary since it was the robot who was mostly leading the conversation, leaving little room for the participant to speak. As a consequence, the robot could not entrain to the participant's last utterance without the risk of sounding unnatural.

4.2 Data annotation and analysis

The influence of entrainment on engagement is investigated in two ways: by analysis of the participants' speech, more specifically their pitch range and utterance duration, and by analysis of their observable behavior. The latter includes eye gaze and facial expressions. This section describes the details of both analyses.

4.2.1 Speech analysis

As explained above, pitch range and utterance duration can be regarded as cues for engagement, where a wide pitch range and longer utterance duration signal a higher degree of engagement. The pitch range of the utterances was compared between the two groups. For utterance duration, only the duration of utterances during the word learning exercise was compared. The analysis was carried out in three steps: obtaining pitch and duration values, preprocessing the data, and statistical analysis utilizing mixed linear regression modelling.

Step 1: Obtaining pitch values in Praat (Boersma & Weenink, 2009). Firstly, each utterance received a label containing relevant information. Each label contained the following information: participant id, group (control or entrainment), phase, number of the utterance, content of the utterance. The words in the learning task received an additional tag for whether they had been a filler or target word in the original experiment. An example of a labelled utterance would look as follows: p1_c_intro_ut1_hello. The corresponding pitch values in Hz were obtained for each labeled utterance using a script. Utterance duration in seconds was extracted only for the utterances in the learning task. The reason for this is that the participants' utterances outside the learning task displayed too much variation in terms of content and total number to be compared. In the learning task, all participants read the same words from the booklets, so the duration of each word could be compared between the groups. Pitch values included mean, minimum and maximum pitch. This step resulted in a file containing raw, unarranged data.

Step 2: Preprocessing of the data. This was necessary to frame the data in a convenient way and to exclude or convert data, and was carried out in Python. Although these steps could also have been carried out in R, the reason that Python was used is the comfort and familiarity that the author has with Python but not with R. Firstly, information that was irrelevant for this thesis was removed from the labels. This concerned information on whether a word in the learning task was a filler or target word in the original study. Secondly, the data was arranged neatly into columns using the Pandas tool. Thirdly, some of the pitch values could not be extracted by Praat and thus were removed from the data. This only concerned seven instances out of the total of 5904. Fourthly, two extra variables were added, one that displayed the pitch range of each utterance, and one that indicated whether an utterance had taken place during the introduction phase of the experiment or not. Pitch range was calculated by subtracting the min pitch value from the max pitch. Lastly, pitch and duration values were converted to numbers, as the output of the Praat script contained strings.

Step 3: Statistical analysis. The analysis was done using a linear mixed-effects model in R with the lme4 package. The dependent variables were pitch range and utterance duration. The independent variables for the analysis of pitch range were group (entrainment or control) and phase (intro or other). The latter variable specifies whether an utterance took place during the introduction phase of the experiment. This information is relevant because this analysis investigated whether the entrainment group became more engaged during the experiment. Since the entrainment feature was disabled during the introduction phase for both groups, this phase was regarded as the base line for pitch height, against which the pitch values in the other phases were compared. For the analysis of utterance duration, the independent variable was group (entrainment or control). Random variables included the participant and the content of the utterances. The effect of group or phase on the independent variables was investigated by adding them to an empty model in a step-wise fashion.

If the hypothesis is correct, we expect an interaction effect of phase and group on pitch range. That is, we expect the entrainment group to exhibit a larger increase in pitch range group in the training phase compared to that in the introduction phase than the control group. Moreover, a main effect of group on duration is expected, where duration will be longer in the entrainment group compared to the control group.

4.2.2 Behavior analysis

The second type of feature that was used to analyze degree of engagement concerned the behavior of the participant. Two observable features functioned as engagement indicators: eye gaze and facial expressions. Degree of engagement is expressed here as the percentage of the engaged behavior relative to the total observation time. Eye gaze towards the robot and a positive facial expression were regarded as cues for engaged behavior. This process again consisted of three steps: annotating the behavior in the Noldus Observer XT software, preprocessing the data in Python, and statistical analysis in R.

Table 1: The coding scheme for behavior analysis.

Type	**Behavior**
Eye gaze	Towards robot, not towards robot
Facial expression	Positive, neutral, negative

Step 1: Annotating the behaviors in the Noldus Observer XT software (version 15). Annotation was done by means of a coding scheme, which is displayed in Table 1. The scheme is largely based on the paper by Perugia et al. (2018), in which the authors develop a reliable and extensive coding scheme for investigating engagement in activities of people with dementia. While it was not investigated whether the coding scheme proposed by Perugia et al. (2018) is as accurate outside the context of dementia, it nevertheless provides a useful guideline for engagement analysis in general. They state that eye gaze is a relevant indicator of engagement, alongside various categories of body movements and gestures. Unfortunately, the current experiment was set up in such a way that the participants display almost no body movements, because they were sitting in front of the robot while mainly reading from a booklet. The experiment did not stimulate active manipulation of the context. For this reason, the features relating to body movements could not be adopted in the coding scheme, as there would be almost none to annotate.

Since positive attitudes towards the stimulus can be regarded as a cue for engagement, the analysis of facial expressions was included (Olsen, Pedersen, Bergland, Enders-Slegers, & Ihlebæk, 2019). This is in line with the working definition of engagement in the current study, which relates engagement to 'the psychological state of well-being, enjoyment and active involvement' (Perugia et al., 2018) of the participant. For these reasons, facial expressions were included in the coding scheme, referring to either positive, negative or neutral emotions. Positive facial expressions include smiles, laughs and affirming head nods. Negative facial expressions include boredom and agitation. Neutral was assumed to be the default state, i.e. if no obviously positive or negative emotion was expressed, the participant was considered to have a neutral facial expression. Both facial expressions and eye gaze were scored in duration. Afterwards, the proportion of the presence of the features was calculated as percentages of the total observation time.

Step 2: Preprocessing in Python. This step facilitates readability of the data both by R and the human eye. Firstly, the data was conveniently framed using the Pandas package. Secondly, the percentages were converted from strings to numbers. All code, including the R script, can be found at github.com/aapolimeno/thesis

Step 3: Statistical analysis. Linear regression modelling was performed in R using the lm function. In this case, mixed linear regression could not be used because the only random variable in this analysis, namely participant, corresponds to one data point for each observation. As a result, participant cannot be regarded as a random variable. Mixed linear regression requires at least one random factor, which the current data could not provide. Thus linear regression modelling was used. The independent variables are the same as before, namely group (entrainment or control) and phase (intro or other). Again, the introduction phase was used as a baseline. The dependent variables are eye gaze towards robot and positive facial expression, both expressed in percentages of the total observation time.

Again, an interaction effect of phase and group on eye gaze and facial expression is expected. For eye gaze, this means that for the entrainment group, the increase in the amount of time that the participants' gaze is pointed towards robot is higher in the training phase compared to that in the introduction phase than for the control group. Similarly, positive facial expressions should increase to a larger extend for the entrainment group in the training phase compared to that in the introduction phase than for the control group.

5 Results

5.1 Speech analysis

5.1.1 Pitch range

Pitch range was slightly higher in the entrainment group (M = 119.1) compared to the control group (M = 115.0), but this difference was not significant (p = .733). In other words, there was no main effect of group on pitch range.

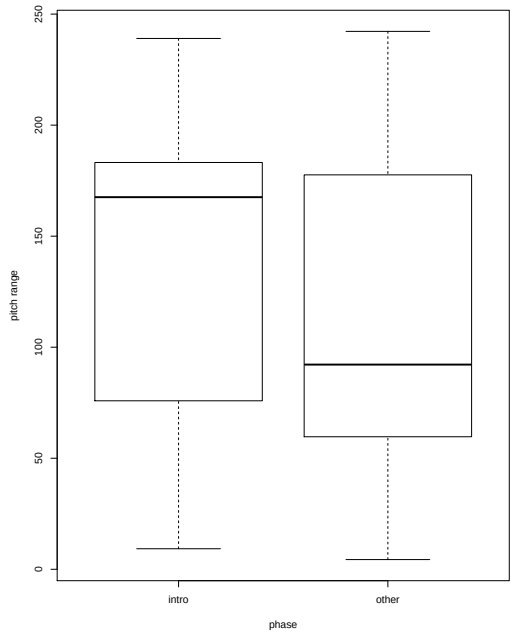
A highly significant main effect of phase on pitch range was found, where pitch range in the introduction phase differed from the pitch range in the other phases (β = 20.36, SD = 5.89, t = 3.46, p < .001), see Figure 1 below. Contrary to the expectations, pitch range was significantly wider in the introduction (M = 135.72) than in the other parts of the experiment (M = 115.93).

Table 2 displays the mean pitch range per phase per group. The interaction effect of phase and group was not significant (p = 0.247).

Table 2: Mean pitch range (and standard deviations) per phase per condition

	Control (n=14)	Entrainment (n=14)	Total (N=28)
Introduction	144.5 (60.1)	128.0 (71.5)	135.72 (66.6)
Other	113.5 (67.2)	118.5 (67.7)	115.93 (64.7)

Figure 1: Boxplot diagram of mean pitch range per phase



5.1.2 Utterance duration

Although the mean utterance duration was slightly higher in the entrainment group (M = 0.859 s) than in the control group (M = 0.833 s), this difference was not significant (p = 0.255).

5.2 Behavior analysis

5.2.1 Eye gaze

A significant main effect of phase on eye gaze towards robot was found (β = 62.8, SD = 3.36, t = 18.7, p < .001), where the percentage of eye gaze towards the robot was higher in the introduction phase (M = 83.0) than in the rest of the experiment (M = 59.9). However, a marginally significant interaction effect of group and phase on eye gaze was found (β = 13.9, SD = 3.81, t = 3.65, p = 0.052), overruling the main effect of phase on eye gaze. Subsequent analysis of the interaction effect pointed out that the entrainment group displayed a marginally significant higher value of eye gaze towards robot compared to the control group, but only in the non-intro phases (p = 0.052). The mean percentage of gaze towards robot was 63.9 for the entrainment group and 55.9 for the control group, see Table 3.

Table 3: Mean percentage of gaze towards robot (and standard deviations) per phase per condition

	Control (n=14)	Entrainment (n=14)	Total (N=28)
Introduction	84.0 (8.05)	82.0 (8.94)	83.0 (8.41)
Other	55.9 (10.1)	63.9 (10.6)	59.9 (10.9)

5.2.2 Facial expression

A significant main effect of phase on positive facial expression was found (β = 58.9, SD = 3.66, t = 16.1, p < .001). In the introduction phase, the mean percentage of positive facial expressions was 15.7, while this was 7.06 for the rest of the experiment. No main effect of group was found, nor was there a significant interaction effect between phase and group. See Table 4 for the means (and standard deviations).

Table 4: Mean percentage of positive facial expression (and standard deviations) per phase per condition

	Control (n=14)	Entrainment (n=14)	Total (N=28)
Introduction	15.1 (15.6)	16.4 (16.4)	15.7 (15.2)
Other	6.32 (4.63)	7.80 (6.46)	7.06 (5.56)

6 Discussion

The behavior analysis has shown that entrainment can enhance student engagement, while the speech analysis has not yielded evidence for this. The participants displayed a marginally significantly higher percentage of eye gaze towards robot in the entrainment condition than those in the control condition. However, no effect of group was found on facial expressions, pitch range and utterance duration. One of the four predictions has been proven to be true. The eye gaze of participants in the entrainment condition was more often pointed towards the robot, indicating increased behavioral engagement. No effect of entrainment on emotional engagement was found. All in all, the hypothesis that entrainment significantly improves student engagement is partly supported.

The main effect of phase on pitch range, facial expression and eye gaze, where all features had higher values in the introduction phase, suggests that the participants in both groups became less engaged as the experiment progressed. A few factors could have played a role in this pattern of disengagement, among which is the novelty effect. Although the participants were familiarized with the robot a few weeks prior to the experiment, it was not tested whether this method successfully reduces the effects of the novelty effect (Soliño Fernández, 2020). It is plausible that the one-to-one interaction with the robot at first sparked the participants' excitement, which is noticeable in the introduction phase, but reduced gradually as the experiment progressed. Another factor that might account for a reduction of engagement is that the task was rather easy for the participants, as Soliño Fernández (2020) already pointed out. The number of words in the learning task that were familiar to the participants was larger than intended in the design of the experiment. The limited amount of challenging content could have made the participants more bored, thus reducing their engagement to the task.

For all four engagement measures, it is true that the entrainment group displayed a higher absolute value in the non-introduction phase. This effect reached statistical significance only for eye gaze, while for the other features the difference between the groups was statistically negligible. However, this pattern should be regarded in the light of the small sample size of this study, which forms a considerable limitation. The sample size of this study consists of 14 participants per group. The results of this study may not be interpreted as strong evidence for the hypothesis that entrainment improves engagement, but the pattern that is found here nevertheless suggests that further research might be able to find affirmative evidence. However small the absolute difference in mean between the two groups, these results

do not exclude the possibility that entrainment can stimulate engagement, or alternatively reduce disengagement. Further research is needed to validate the current findings.

An interesting question to ask in the light of these results would be whether the two groups are equally disengaged, or whether differences in this pattern can be found. Inspection of Table 2, which shows the mean pitch per phase per group, allows a preliminary answer to be suggested. The distance between the pitch range in the introduction phase and the rest of the experiment is smaller for the entrainment group than for the control group. The relatively smaller drop in mean pitch range compared to the base line might carefully imply that the entrainment group became less disengaged during the experiment. To put it differently, while the entrainment feature in the robot did not result in an increase of engagement, it might be possible that it prevented the experimental group from becoming as disengaged as the control group. More research could be done to confirm this claim.

A methodological shortcoming that should be discussed is the fact that the body movements of the participants could not be studied. The reason for this is that the participants were seated on the floor in front of the robot and generally kept their body still. This excluded the possibility to analyze highly relevant engagement cues such as body movement. Since body movements are an informative cue for measuring engagement (Perugia et al., 2018), a study that investigates student engagement in the context of human-robot interaction should ideally adopt an experiment that includes elicitation of data on body movements. For example, a more interactive context could be included, where the participant can manipulate a stimulus using their hands.

Relatedly, the coding scheme for the behavior analysis might not have yielded a sufficient measure of engagement. Firstly, relevant information on physical behavior was absent. Secondly, facial expressions that obviously conveyed emotions, either positive or negative, were relatively rare. An explanation could be that the participants took the task seriously, similar to exercises in class, and thus generally conveyed comparatively neutral facial expressions. This was true for most of the participants, while the few participants who portrayed positive facial expressions did so more frequently, and for a longer duration. This explains the high values of standard deviations in Table 4.

Another issue with the behavior analysis concerns the reliability of the coding. Due to limited time and resources, the coding of eye gaze and facial expressions in Observer XT was carried out by one person. Having more than one coder would have resulted in a higher reliability. Furthermore, a more inclusive operationalization of engagement should cover a measure for cognitive engagement, for example by collecting data on the participants' intrinsic motivation and relationship with learning through self-report questionnaires. This would also yield more data on emotional engagement.

Lastly, utterance duration might not have been a good feature to measure engagement, because it is difficult to unambiguously interpret the meaning of a longer utterance duration. On the one hand, longer utterance duration might have been a result of the effort of the participant to make the robot understand what they said, as argued above. This is a plausible interpretation because the participants in this study are likely to have little to no experience with interacting with a robot, and thus might not have been convinced by the robot's language proficiency. On the other hand, longer utterance duration can signal boredom, as argued by Paeschke & Sendlmeier (2000). In this case, longer utterance duration would signal a lesser degree of engagement. Since both perspectives result in contradictory predictions, and it is not straightforward which one is more realistic, utterance duration might not have been a useful feature to analyze engagement.

The inconclusive answer to the research question indicates that it might be interesting for future research to further investigate the influence of entrainment on student engagement. Especially as speech technologies improve, allowing robots to interact more convincingly and naturally, their ability to become a helpful tool in language learning settings is promising. Since the implementation of entrainment in a social robot can be beneficial from both a

technical and a pedagogical perspective, the effects of entrainment on student engagement or learning in general remains an interesting topic for investigation.

7 Conclusion

This study investigated the influence of entrainment on student engagement by means of speech and behavior analysis. Four features were compared between an entrainment and control group, namely pitch range, utterance duration, eye gaze and facial expressions. Only for eye gaze there was a significant difference between the two groups, where eye gaze in the entrainment group was significantly more often pointed towards the robot. This suggests that behavioral engagement can be affected by entrainment to some extent. For the other features (pitch range, utterance duration and facial expression), no significant difference between the groups was found, but their absolute values generally indicate higher engagement in the entrainment group. The lack of statistically significant results can partly be attributed to the small sample size. All in all, the hypothesis that entrainment positively influences student engagement is partially confirmed. Thus, this study provides modest support for existing claims that implementation of entrainment in social robots can indeed be beneficial from a pedagogical perspective.

8 Acknowledgements

I want to express my gratitude to prof. dr. Aoju Chen for supervising me throughout the past few months. By including me in this project, she allowed me to unite my interests in linguistics and artificial intelligence, for which I am very grateful. Secondly, I want to thank dr. Emilia Barakova, both for being my second reader and for providing access to the Noldus Observer XT software. Moreover, I want to thank the members of the Prosody and Language Learning research group for their input and advice. More specifically, I want to thank Na Hu for providing much appreciated help with the Praat scripts. Furthermore, I want to thank Cyril de Kock for his help with Python, as well as his emotional support. Lastly, I want to thank my dear friends Dinte and Pauw for their insightful feedback.

References

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on l2 learners' anxiety and attitude in english vocabulary acquisition. *International Journal of Social Robotics*, *7*(4), 523–535.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation*, *6*(4), 802–813.

Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer (version 5.1.13)*. Retrieved from `http://www.praat.org`

Breazeal, C. (2002). Regulation and entrainment in human—robot interaction. *The International Journal of Robotics Research*, *21*(10-11), 883–902.

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.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893.

Chin, K.-Y., Hong, Z.-W., & Chen, Y.-L. (2014). Impact of using an educational robot-based learning system on students' motivation in elementary education. *IEEE Transactions on learning technologies*, *7*(4), 333–345.

de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., ... Vogt, P. (2018). The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction* (pp. 50–58).

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The modern language journal*, *78*(3), 273–284.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59–109.

Friedberg, H., Litman, D., & Paletz, S. B. (2012). Lexical entrainment and success in student engineering groups. In *2012 ieee spoken language technology workshop (slt)* (pp. 404–409).

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., ... Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Thirtieth aaai conference on artificial intelligence*.

Gravano, A., Beňuš, Š., Levitan, R., & Hirschberg, J. (2014). Three tobi-based measures of prosodic entrainment and their correlations with speaker engagement. In *2014 ieee spoken language technology workshop (slt)* (pp. 578–583).

Gussenhoven, C. (2002). Intonation and biology. *Liber Amicorum Bernard Bichakjian (Festschrift for Bernard Bichakjian)*, 59–82.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53.

Hockema, S. A., & Smith, L. B. (2009). Learning your language, outside-in and inside-out. *Linguistics*, *47*(2), 453–479.

Monk, C. S., McClure, E. B., Nelson, E. E., Zarahn, E., Bilder, R. M., Leibenluft, E., ... Pine, D. S. (2003). Adolescent immaturity in attention-related brain engagement to emotional facial expressions. *Neuroimage*, *20*(1), 420–428.

Movellan, J., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th acm/ieee international conference on human robot interaction* (pp. 307–308).

Okita, S. Y., & Schwartz, D. L. (2013). Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *Journal of the Learning Sciences*, *22*(3), 375–412.

Olsen, C., Pedersen, I., Bergland, A., Enders-Slegers, M.-J., & Ihlebæk, C. (2019). Engagement in elderly persons with dementia attending animal-assisted group activity. *Dementia*, *18*(1), 245–261.

Oviatt, S., Darves, C., & Coulston, R. (2004). Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *11*(3), 300–328.

Paeschke, A., & Sendlmeier, W. F. (2000). Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements. In *Isca tutorial and research workshop (itrw) on speech and emotion*.

Perugia, G., Díaz-Boladeras, M., Català-Mallofré, A., Barakova, E. I., & Rauterberg, M. (2020). Engage-dem: a model of engagement of people with dementia. *IEEE Transactions on Affective Computing*.

Perugia, G., van Berkel, R., Díaz-Boladeras, M., Català-Mallofré, A., Rauterberg, M., & Barakova, E. (2018). Understanding engagement in dementia through behavior. the ethographic and laban-inspired coding system of engagement (elicse) and the evidence-based model of engagement-related behavior (emodeb). *Frontiers in psychology*, *9*, 690.

Sinha, T., & Cassell, J. (2015). Fine-grained analyses of interpersonal processes and their effect on learning. In *International conference on artificial intelligence in education* (pp. 781–785).

Soliño Fernández, B. (2020). *Using speech entrainment to improve second language tutoring by pedagogical robots*.

Street Jr, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139–169.

Thomason, J., Nguyen, H. V., & Litman, D. (2013). Prosodic entrainment and tutoring dialogue success. In *International conference on artificial intelligence in education* (pp. 750–753).

van den Berghe, M. A. J. (2019). *Social robots as second-language tutors for young children: Challenges and opportunities*.