Universiteit Leiden
The Netherlands

Opleiding Data Science and Artificial Intelligence

A Study on the Effects of Speech Performance of a NAO Robot on Information Recall

Riham Mushmush

Supervisors:
Dr. Rob Saunders & Dr. Joost Broekens

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

13/01/2025

Abstract

In recent years, robots have increasingly become a subject of study and development. In the 1990s, social robots, as a new generation of robots, began to be studied and developed. One of the earliest examples in this field is Kismet, a robotic head created at MIT to mimic human expressions using a highly abstracted form of a creature's head [4]. Since the main purpose of social robots is to work in environments surrounded by people or in direct interaction with them, such as in education, healthcare, reception services, and more; social robots have become a common subject of study. These studies are aimed mostly at understanding the key factors for smooth interactions with humans and their surroundings, as well as achieving optimal performance in their tasks, such as teaching or other specific roles.

This thesis aims to investigate the effects of design in social robots on human-robot interactions and human cognitive tasks. Specifically, it focuses on the impact of voice design in educational robots, on information recall. The NAO robot was used in this study and programmed in two conditions: the first was a lecture without variability in voice, and the second involved giving a lecture while varying voice parameters for key and important words. The manipulated variables included pitch, speech rate for these important words, and a brief pause before them. The participants were divided into two groups, with each group assigned to one NAO robot. They were given multiple-choice questions to answer, including questions about the lecture given by the NAO and questions about their interactions with the robot. The results of the two groups were then studied and compared.

Contents

1 Introduction

Communication is very important in daily life, and it can occur in multiple ways. One of these is through talking, but communication does not just happen through words; it also depends on how these words are spoken, including the tone and sound of the speaker, such as pitch and speech rate. The voice is handled by the human brain in complex ways [21], and the effects of voice modulation on cognitive functions like attention and memory are profound [3]. Studies have demonstrated that the brain's response to changes in voice can significantly impact a listener's ability to focus on and retain information. For example, research by Saffran and colleagues [19] found that infants can use the statistical relationships between speech sounds to segment words in fluent speech, that suggest that we learn the tone and rhythm of language before the actual words themselves. This findings underscore the importance of voice design in social robots.

The tones in people voices determine our emotions and how we feel [5]. Furthermore, when hearing different tones in another person's voice, individuals can understand others' feelings and determine whether they are happy or sad, serious or joking. Even if we do not see their gestures or movements or do not understand what they are saying, we can still discern -through their voices- the impressions they want to convey, whether what they are saying is important or casual, as well as identify the key or important words in an entire speech we listen to [20]. Similarly, in written texts, words with typographical cues such as underlining, coloring, or other styles draw our attention and give us the impression that the underlined or highlighted word is important or more significant to remember than other words in the same text [13]. Based on the ideas mentioned previously, it can be concluded that "coloring" important words in speech can help draw attention to these words and improve their memorization or recall.

Sounds and voices are important aspects of social robots, particularly educational robots, as talking is not only a channel for communication but also a way to transfer experience and knowledge. This is why it is important to know how to modify sounds and voices and design an optimal voice technique to achieve the best performance and interactions between robots and humans. Additionally, this helps these robots afford the purpose of their existence, which is to be social, as well as the purpose of their programming, such as understanding and conveying information if the robot is an educational robot.

One example of voice variation in our daily speech is how we convey our intentions through interrogative, exclamatory, and neutral statements by simply changing our tone. For example, one can tell whether someone is telling something, asking something, or expressing an emotion by hearing their words. For example, if we take a simple sentence like "Ahmed plays tennis," the way it is said can convey different meanings. For example, if one emphasizes the word "Ahmed," this suggests surprise or a question, as if asking, "Is it Ahmed who plays tennis?" If the emphasis is on "tennis," this indicates surprise that Ahmed specifically plays tennis. If the sentence is said without emphasis, this is understood as the person merely stating that Ahmed plays tennis, which may imply a subsequent question about whether the other person also plays tennis [2].

1.1 Background Research

Speech performance has been studied and explained for a long time, with many scholars and philosophers shedding light on the subject over the years. Some have treated it as an explanatory subject, detailing what a speaker or lecturer should say and how to say it, while others have treated it as a subject of study, examining the characteristics that must be taken into account while speaking in front of an audience or communicating to convey an idea effectively.

An example of a philosopher who explained what a rhetorician should focus on is Aristotle, who set out the principles of effective rhetoric in his book "Rhetoric," which has been translated into many languages. Aristotle emphasized in "Rhetoric, Book III" that expressing ideas requires not only knowing what to say, but also knowing how to say it in the right way. Aristotle explains further that a speaker can engage listeners and make them feel more connected through the performance, as appropriate voice control can convey the intended emotion by adjusting volume and pitch. Further, he highlights that language should have a rhythm, with each type of rhythm appropriate to a particular subject; additionally, the manner an idea is expressed impacts its clarity and intelligibility [1].

Later, the scholars in the Middle Ages had translated and completed Aristotle works, developing the art of rhetoric and the qualities of an effective rhetoric. Among the scholars is Al-Jahiz, who emphasized in his book "Al-Bayan wa Al-Tabyin" the necessity for the orator to be characterized by "Fsaha" (eloquence). This means that the orator must speak in a clear and understandable manner, pronouncing letters distinctly so that the audience can easily comprehend. Al-Jahiz also highlighted that the orator should not have any lisp because clarity in speech is essential.

Al-Jahiz main focus was on using appropriate language-choosing words that are neither overly complex nor too colloquial. Furthermore, he stressed that an orator should carefully select both words and delivery based on the audience. For example, the way one addresses a king is different from the way one speaks to the king's subjects, and the language used with scholars should be different from that used with the uneducated. Thus, adapting speech to the context and audience is a crucial aspect of effective communication [12].

Speech performance has continued to be studied in various ways, examining how voice modulation affects listeners by influencing their emotions, attracting attention, changing feelings or persuading them about an idea. The impact of this topic spans multiple areas of life, including education and advertising, and has even extended into robotics and social robots.

A study conducted by Pell and colleagues examined the effect of voice modulation on emotional predictions. In this study, participants were tasked with predicting the emotions of speakers who spoke a language different from their own, often a language unfamiliar to them. The results were impressive, as most participants were able to accurately predict the emotions of the speakers based only on their vocal expressions. However, some differences were observed across languages and in the specific emotions detected [17].

In another study, voice modulation was shown to enhance attention and influence recognition memory. This was demonstrated in the study by Potter and his colleagues [18], where twelve audio

commercials were created using four different announcers. Each commercial incorporated a single voice change, categorized into Low-, Medium-, or High-Pitch Differences.

Substantially, grabbing someone's attention is important to make them listen to the idea and memorize it [15]. However, grabbing attention is not an easy task for lecturers. While some people succeed in this task, others cannot; voice pitch plays an important role in this process [6]. Additionally, speech rate also has an impact, as it can affect recalling and memorizing words. This was demonstrated in the study by Bradlow and her colleagues [3], where participants were asked to segment words marked as "new" or "old" from a list of words that had been previously spoken by a person at the same rate.

Recently, studying the social robots has become a common research topic, particularly with the rise of new generative tools and AI. Although, as previously mentioned, the study of voice effects is not a new subject, integrating it with social robots is relatively novel. In addition to what has been discussed earlier, researchers have also explored its impact on various areas not only in human-human interactions but also in human-robot interactions.

A study conducted on receptionist robots, using two robots with different voice pitches, found that people perceived the robot with a higher pitch as more attractive and assigned it better personality traits and behaviors [14]. Other studies have found that participants tend to prefer robots with voices matching their own gender [10].

Unfortunately, the research on the effects of voice modulations in social robots is somewhat limited, with relatively few references available. However, it is worth noting the advancements in tools used for text-to-speech (TTS) technology. These tools, such as ElevenLabs, utilize deep learning to achieve high-quality and natural-sounding voice synthesis, paving the way for more realistic interactions.

1.1.1 Important Definitions

In this section, we will explain some key definitions that will be used in the conclusion (Section 3). The first and most relevant to the study of social robots is anthropomorphism [9], which refers to attributing human characteristics to non-living things, such as saying "this computer is tired" or "my phone is dying." Another important concept is affordance [16], which describes how the design of an object reflects its function; for example, the type of door handle can indicate whether the door should be slid or pulled.

Another important term is the uncanny valley [8], which describes the phenomenon where a robot becomes unsettling or creepy if it appears nearly human but not entirely perfect. Finally, mirror neurons are neurons that fire both when a person performs a specific action and when they observe someone else performing the same action. In other words, if a person grasps an object, certain neurons fire, and the same neurons will fire if they watch someone else grasping an object [22].

1.2 Research Questions

1.2.1 Expectations and Hypotheses

Although social robots are a relatively new field, numerous studies have been conducted. However, fewer studies have explored the effect of varying voice. In human conversations, people naturally adjust their speech rhythms from the beginning to the end. Thus, incorporating this variation into robots could make them appear more approachable and relatable. Nevertheless, it remains uncertain whether such variations can directly enhance information recall or influence cognitive abilities. Moreover, as previously discussed, voice modulation in speech significantly affects people. This study aims to bridge this gap by investigating the effects of voice modulations in social robots, focusing on whether they can improve information recall and enhance interactions between humans and robots.
This is why the research question is:

How does the speech performance of social robots influence information recall, speech clarity and likability of the robot?

It is expected that the experiment will show a significant difference between the two groups, with better answers and interactions observed in the group exposed to voice modulations.

2 Experiment

2.1 Study Design

2.1.1 Method

In this paper, an experiment was conducted using the NAO robot (Figure 1). The NAO robot has an embodied form and can speak multiple languages, including English, which was used in this experiment. Although the NAO robot's settings, such as speech rate and pitch, can be adjusted, a Python library was used to generate the voice for this study. The NAO robot can also play pre-recorded audio files. The lecture delivered by the NAO robot was generated using ChatGPT, and the voice was generated using the Python library "pyttsx3", which offers multiple settings. A female voice was selected with a pitch of 1.0 and a speech rate of 150 words per minute.

Figure 1: A still from the video recording used in the experiments showing the NAO robot as presented to participants.

Since the lecture content itself was not the primary focus of the study, ChatGPT was asked to generate three random texts:

- A short story
- An article about sharks
- An article about space

The lecture on sharks was chosen because it contains both quantitative and qualitative variables and includes some challenging and new vocabulary. In contrast, the text about space was considered far more challenging (a personal judgment). The shark text was then reviewed and revised to highlight the relevant information for the lecture, with unnecessary details removed. Key words were identified and emphasized. Finally, a Python script was used to generate two recordings: one with a monotonous tone and the other with voice modulation that highlighted the key words. The code, text, and questions used in this study are to be found here. Additionally, a loop of random movements is been used to give the robot a sense of liveliness while speaking.

The experiment involved two groups:

- A control group
- An experimental group

For the first group, the NAO robot was programmed with the monotonous lecture recording about sharks. For the second group, the NAO robot delivered the modulated recording.
The experiment consisted of the following stages:

1. **Initial Survey**: Participants were asked for their virtual names, age, knowledge about sharks, native language, proficiency in English (and its order of acquisition), and if they had ever interacted with a robot before.

2. **Lecture Viewing**: Participants watched a video of the NAO robot lecturing about sharks.

3. **Questionnaires**:

    o **First Questionnaire**: The main questionnaire that focuses on information recall from the lecture given by the NAO robot.

    o **Second Questionnaire**: This focused on participants' perceptions of the NAO robot, including how they felt about receiving a lesson from a robot and whether they found it interesting and clear.

    o **Additional Question**: An open question allow participants to write down any additional notes or feedback they wanted to deliever.

Additionally, the questions of the first questionnaire were categorized as follows:

- **Numbered Questions (NQ)**: Questions requiring participants to remember specific numbers (quantitative answers) where the answers were numbers e.g. question: 8, 10, 13).

- **Simple Questions (SQ)**: Basic recall questions where only one straightforward answer was correct. (qualitative answers). Where the questions were partially simple (e.g., 7, 9, 15), where participants chose the correct answer from multiple choices.

- **Complex Questions (CQ)**: Questions requiring participants to pay close attention to the lecture and analyze the options, as only one answer was correct, though all were derived from the lecture content (e.g., 11, 12, 14, 16), where all choices were derived from the text, but only one was correct.

In Figure 2 you can find the questions and their type.

Questions and there type	Type Questions
What is the text about?	SQ
What is the age of sharks on Earth?	NQ
To which class do sharks belong	SQ
How many shark species are there?	NQ
What allows the unique head shape of the hammerhead shark mentioned in the lecture?	CQ
Which of these sentences is correct?	CQ
How many rows of teeth in total does the bull shark have according to the lecture?	NQ
Which of these sentences is true according to the lecture?	CQ
According to the lecture, which of these is a reason for overfishing sharks?	SQ
Which of these types is often referred to as a "living fossil"?	CQ

Figure 2: Questions used in the experiment classified by type numeric question (NQ), simple question (SQ), and complex question (CQ).

The experiment involved a total of 21 participants: 11 in the control group and 10 in the experimental group. They were relatives, friends, and other students recruited through SurveySwap. To respect participants' privacy, real names were not requested; instead, participants were asked to provide pseudonyms of their choice. Due to the recruitment method, it was somewhat challenging to gather detailed information about the participants prior to the survey.

The experiment is conducted using online platforms. The NAO robot was pre-programmed with the recordings, and two videos were made – one for each recording. A survey was created using Microsoft Forms, containing the questions and the respective video for each group. The NAO robot was chosen due to its popularity in social robotics and its ease of transport, if needed. A female voice with a pitch of 1.0 and a speech rate of 150 was used for normal speech in experiment group

and whole speech in control group, while a pitch of 0.0 and a speech rate of 100 were used for key words, with a short pause before the key words. The female voice was chosen because some studies have found that female voices are generally more preferable [7]. The normal speech settings are the default in the Python library, while the settings for key words were specifically chosen to highlight them. These changes, including pauses and a slower speech rate, were made to provide participants with more time to process words that are often new or difficult, mimicking the natural human tendency to emphasize important words during speech. Both quantitative and qualitative key words were included based on the belief that not all verbal-types are equally influential in recall [23].

2.2 Results

At the beginning of the analysis, the time taken to complete the test was examined, and a time range was determined as follows: the minimum time was set as the video duration plus 1 minute, and the maximum time was set as the video duration plus 10 minutes. The results were as follows:

- **Control Group** (video time: 4 minutes and 40 seconds):
    - o Minimum Time: 5 minutes and 40 seconds
    - o Maximum Time: 14 minutes and 40 seconds
- **Experiment Group** (video time: 5 minutes and 5 seconds):
    - o Minimum Time: 6 minutes and 5 seconds
    - o Maximum Time: 15 minutes and 5 seconds

This time range was chosen because it is not reliable to watch nearly the entire video/lecture and answer the questions in less time than the minimum or to take more than 10 minutes to answer the questions.

Unfortunately, applying these criteria resulted in the removal of nearly half of the data. For more details on the time taken and the data removed, see the tables in the appendix Tables 1 & 2, where "TRUE" indicates retained data and "FALSE" indicates removed data. This adjustment was necessary to ensure the reliability of the remaining data, which resulted in 7 participants for the control group and 5 for the experiment group. The newly filtered data was then analyzed.

2.3 Analyses

2.3.1 First Questionnaire

Using t-test method there was no significant difference between the control group and the experiment group for recalling information as shown in the table below:

	Total	NQ	SQ	CQ
T-Test	-1.490028015	0.538494321	-1.090768247	-1.89514702
df	9.898771089	7.250641868	9.859863945	7.943591965
p-value	0.1674	0.6064	0.3013	0.09493
Confidence interval - 95%	[-3.9245525,0.7816954]	[-0.8641524, 1.3784381]	[-1.3927618, 0.4784761]	[-3.0422400, 0.2993828]
Minimum Value	0	0	0	0
Maximum Value	10	3	3	4

Figure 3: Summary of the results of the t-test on information recall for different types of questions NQ, SQ and CQ, which do not show any significant difference between the experimental group and the control group.

But on the other hand, after reviewing the scores out of 10 (Figure 4), we could see that, although there was no notable difference in simple questions and a negative difference in numeric questions, the complex questions showed the most noticeable difference between the control group and the experiment group.

Figure 4: The mean scores of the experimental group and the control group for the different types of questions, and the greatest difference was shown in the complex questions.

Additionally, the difference in scores between participants who had previously met a robot and those who had never met a robot is also noticeable, as shown in Figure 5, the participants who had interacted with a robot before in the experiment group achieved the highest average grade. Furthermore, in both conditions, the experiment group had better scores compared to the control group. Also in Figure 6, which lists the robots participants mentioned they had interacted with before, we can see that some participants included chatbots in their definition of robots. This is a very interesting observation and should be taken into consideration for future studies.

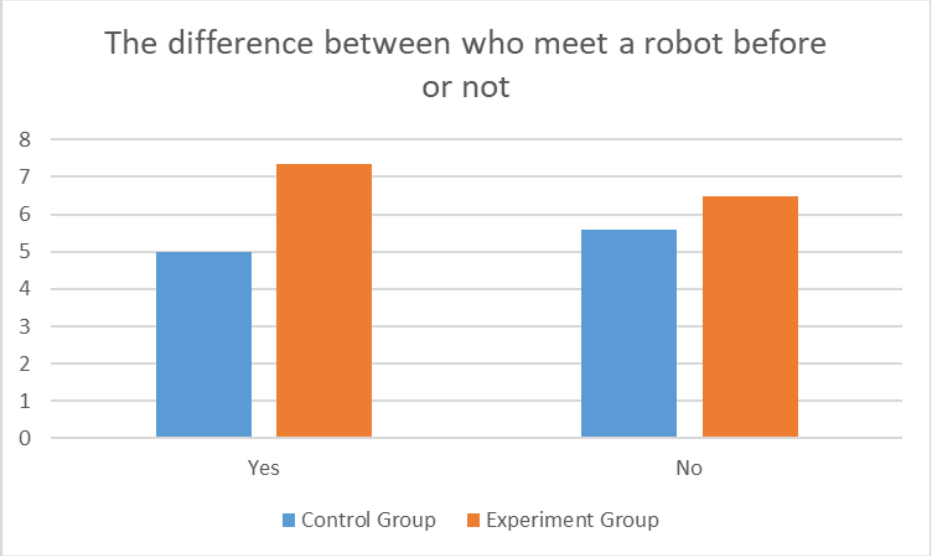
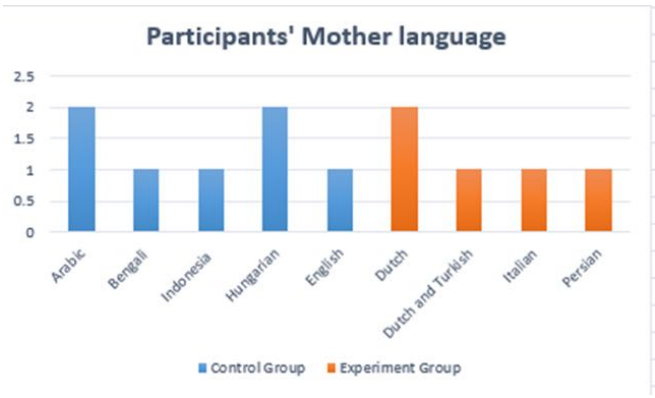


Figure 5: Mean scores of the experimental group and control group for participants who had previously met a robot and those who had not.

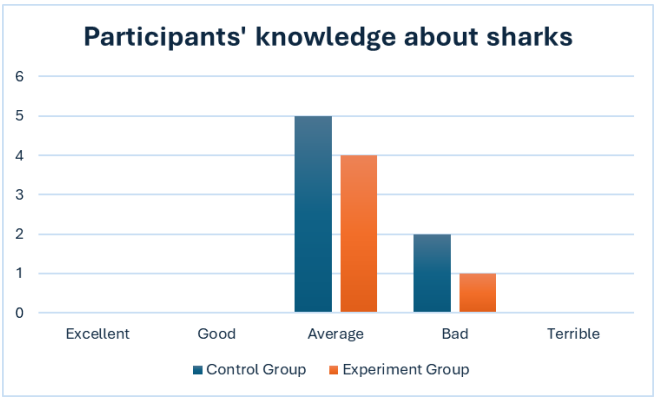
Did you ever meet/have a conversation with a robot? If yes who	Yes	No	Who
Control Group	2	5	ChatGPT, Hotel Robot
Experiment Group	3	2	ChatGPT, NAO robot, restaurant robot, AI chatbots, Google Brard

Figure 6: Names of robots mentioned by participants who answered that they had previously met a robot in both groups.

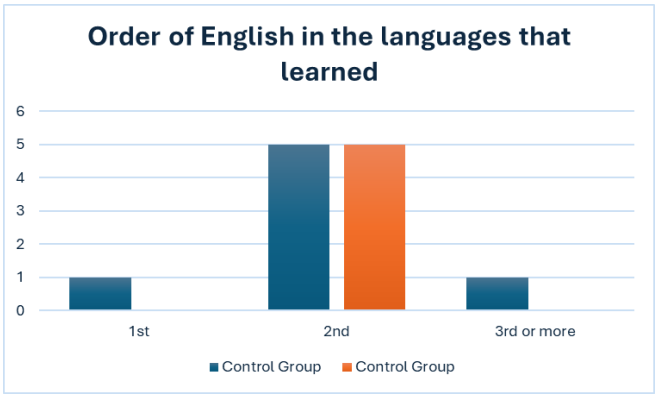
After analyzing the participants' information, we observed a variety in their mother languages, as shown in Figure 7a. The participants ranged in age from 12 to 59, distributed as shown in Figure 7d. They also had limited prior experience with sharks, as indicated in Figure 7b. Additionally, for most participants, English was their second language, as shown in Figure 7c.



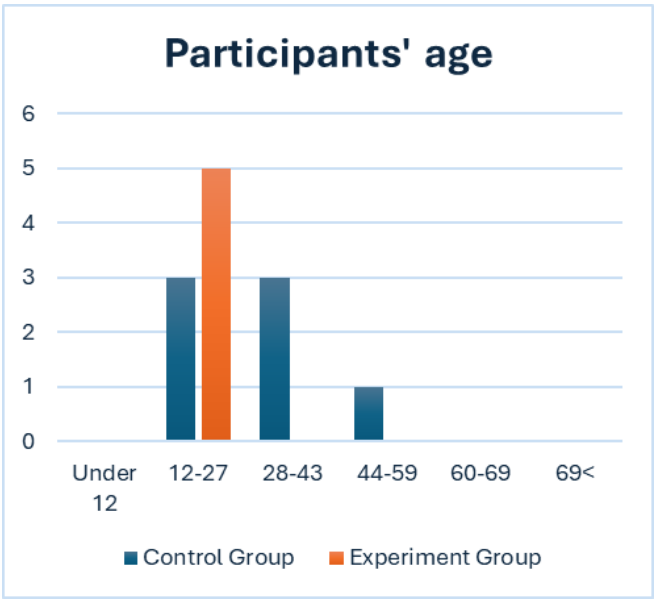
(a) The mother language of the participants, indicated a great diversity of languages and no common mother language between the two groups.



(b) Participants' knowledge about sharks before the experiment. Most participants in both groups had average knowledge.



(c) English language order by languages learned by participants in their lifetime, with most participants learned English as a second language.



(d) Ages of participants in the two groups: All participants in the experimental group were between 12-27 years old, while most participants in the control group were equally divided between 12-27 and 28-43 years old.

Figure 7: General information collected from participants prior to the experiment.

2.3.2 Second Questionnaire

Analyzing the second questionnaire to examine the human-robot interactions, which includes questions about how participants found the NAO robot, as shown in Figure 8.

How did you feel about taking a lesson from a robot
Is it clear what NAO says?
Are you interested in trying this more in the future
Would you choose to take a lesson from a robot over a human teacher

Figure 8: Questions used in the experiment to measure likability and speech clarity of the robot.

We see from Figure 9 that, for the experimental group, participants who liked the NAO's lecture scored higher than the participants who did not find it interesting. For the control group, it was the other way around: participants who did not find it interesting to take a lecture from the NAO scored higher than those who found it more interesting.

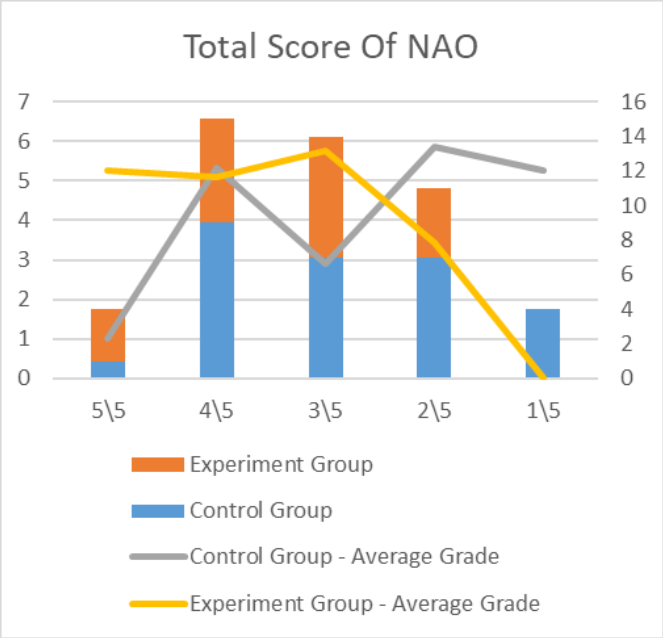


Figure 9: Total Scores for likability and speech clarity compared to the average grades of the participants.

The Figure 10 show how the participants responded to the second questionnaire, and we could see that the experiment group is biased more to like the NAO, and the control group is biased more to not like it.

More specifically, we observe that the most significant differences lie in the questions about the clarity of what the NAO robot says and whether the participant is interested in trying this again in the future. In these areas, the experiment group shows more positive responses compared to the control group. However, regarding the questions about how participants feel about taking a lesson from a robot and whether they would replace a human teacher with a robot, both groups show approximately the same responses.

How did you feel about taking a lesson from a robot?	Control Group	Experiment Group
Loved it	0	1
Liked it	2	0
It was okay	2	3
Disliked it	2	1
Hated it	1	0

Is it clear what NAO says	Control Group	Experiment Group
Yes, the entire lecture is clear	1	1
Yes, most of the lecture is clear	2	3
It is okay	3	1
No, most of the lecture is not clear	1	0
No, the entire lecture is not clear	0	0

Are you interested in trying this more in the future	Control Group	Experiment Group
Definitely	0	1
Probably	2	1
Not sure	2	3
Probably not	2	0
Definitely not	1	0

Would you choose to take a lesson from a robot over a human teacher	Control Group	Experiment Group
Yes, definitely	0	0
Yes, probably	3	2
Not sure	0	0
No, probably not	2	3
No, definitely not	2	0

Figure 10: Participants' responses to the second questionnaire. Each number represents the number of participants who selected each answer.

A significance test for this questionnaire was also conducted, and no significant result was found either (as shown in Figure 11 marked "total"); this test was conducted by converting the possible answers that fall within the qualitative scales to quantitative scales where the best answer is 5 and the worst answer is 1 and then the mean for each question individually was calculated by dividing the sum of the answers by the number of participants using the following equation:

$$\text{average} = \frac{\sum \text{converted}}{n}$$

Where *converted* is the value after the conversion, and n is the number of participants.
After that, a t-test was performed on the means of the four questions.

	Total	First Q	Second Q	Third Q	Fourth Q
T-Test	-1.6827	-0.75228	-1.1761	-1.526	-0.3173
df	5.5608	8.8601	9.9768	9.7712	9.8398
p-value	0.1473	0.4714	0.2668	0.1587	0.7576
Confidence interval - 95%	[-1.3476360,0.2619217]	[-1.9498107,0.9783821]	[-1.6543210,0.5114638]	[-2.183055,0.411626]	[-1.837172,1.380029]
Minimum Value	0	0	0	0	0
Maximum Value	5	5	5	5	5

Figure 11: Summary of the results of the t-test on likability and speech clarity of the robot.

In addition, making a t-test on each question apart (without taking the mean) does not show any significant result as well (Figure 11).

2.3.3 Additional Question

For this section, the removed participants are involved again to gather more information. In Figure 12a, you can see the current participants, and in Figure 12b, you can find the notes of the removed participants.

Control Group	Experiment Group
Robot lecture is not engaging. It's just like reading with monoton voice. I couldn't finish it, too boring.	The movements of the robot disturb the lecture. It makes it hard to understand what the robot says.
THis robot guy was so life less, most boring talk, hands were totally randomly moving not appropriate fo what was being said, and the whoel lecture was so monotonous no person could pay attention to this	I enjoyed
It is a fun video	

(a) Additional notes for participants who were taken into the experiment.

Control Group	Experiment Group
The video voice is a little bit difficult to hear	No. Thank you and good luck!
The robot's motors (servo joints, I think?) were too loud compared to the robot's voice audio, which made it hard to understand the lecture. Also, I generally prefer a human teacher because I don't think a robot, as it stands with current technology, is be able to imitate the cadence and countenance of a human, which, in my opinion, are important factors in differentiating between a good human teacher and a bad one. Therefore, a robot may be good for relaying tidbits of information here and there, but I would not personally like to be taught a full lecture by one. Nonetheless, this was a very interesting survey. Thank you for the experience :)	There was too much noise.
no, I just dont like the monotonous voice	

(b) Additional notes for participants who have been removed from the experiment due to timing.

Figure 12: Participants Additional Notes

3 Conclusion

Previous research has overlooked an important aspect: humans are naturally not monotonous, and we should consider that the brain tends to notice and retain differences more than similarities [11]. By considering mirror neurons [22], we can conclude that people interpret others' actions as if they were their own, helping them understand how others might think or feel in similar situations besides translating the speaker's thoughts and intentions.

Since people often emphasize important words in conversations, we conducted this study to measure participants' ability to recall information, as well as the speech clarity and likability of the robot. We did this by having a group of participants listens to a robot speak in a varied, non-monotonous manner and comparing it to a control group that listened to a robot speaking monotonously.

Although the experiment did not yield statistically significant results neither in the first nor in the second questionnaire, this may be due to several factors, and it does not necessarily mean that the experiment failed. The number of participants was too small, and after excluding those who did not complete the experiment within the designated time range, the sample size became even smaller. However, by comparing the average scores, we observed that the group exposed to the NAO robot with varied speech performed slightly better on complex questions than the other participants. Additionally, it was noted that participants who were not particularly fond of the experiment performed slightly better than those who were more engaged.

Moreover, the sounds of NAO's movements were loud, mostly because it was a recorded video, making it more difficult for the participants to hear what was being said. This issue was mentioned in the additional notes provided by the participants at the end of the experiment (Figure 12).

However, the second table in Figure 10 shows that participants in the experimental group mostly found what the NAO robot was saying to be clear, which is not the case in the control group, which was more widely distributed. While the same figure in the rest of the tables shows that there is only a slight difference in likability between the experimental and control groups.

We conclude our study with Carl Sagan's words, "The brain is a very big place in a very small space", and studying any effect on people can be challenging due to the complexity of the human brain. As seen in Figure 5, participants in the experiment group who had interacted with a robot before showed better results compared to the experiment group, who did not interact with a robot before. This could lead us to conclude that participants who had not met a robot before faced additional challenges in accepting the robot, or the interaction had an extra effect on them.

In conclusion, this study does not have any significant results and its impact in reality is weak, but there is much learned from participants' reactions. People naturally tend to see anthropomorphism in things. Moreover, based on some reactions that consider chatbots as robots, we could emphasize that small cues can *afford* sociality, and social robots do not necessarily need a physical body.

In some cases, there is no need to over-design social robots to avoid the uncanny valley and overestimating, which is partially supported by figures 12 and 6. Results show that the worst

average was from the control group who had interacted with a robot before. Additionally, the robot's movements could have a negative effect, especially if the sound is too loud, which could break attention. Furthermore, it is essential to build on previous research, incorporate insights from philosophers, and program robots that choose the best words and deliver them in the best way. Parallelly avoiding monotonous speech, as it can bore the listener (Figure 12), potentially affecting their focus.

4 Limitations and Further Work

The study has certain limitations that were not fully addressed. First, a notable point is that the experiment relied on multiple-choice questions, which could introduce a degree of guessing. However, we assumed that participants did not guess, as this would complicate the analysis. This aspect should perhaps be taken into consideration in future work, with a focus on open-ended questions to minimize the possibility of guessing.

Second, there were challenges in managing participants before conducting the experiment, primarily due to the use of an online platform to recruit them. This had an impact, as there was a noticeable variety in the participants' mother languages and ages, with no equivalent distribution across the two groups.

Although this study, in particular, does not show a significant effect, in my opinion, using robots in further research on the effects of voice on information recall or memorization could be very helpful. Studying the effect of a robot's voice may be more effective than studying the voice of people because the robot's voice can be preprogrammed and maintain consistent settings, whereas it is difficult for humans to replicate the exact same changes each time.

Additionally, by integrating techniques like facial expression recognition with self-learning and deep learning, we could develop a program that, based on the user's expressions, predicts whether the robot has chosen an appropriate tone and manner of speaking. This approach could enhance human-robot interaction by enabling the robot to adapt its communication style in real-time, ensuring better engagement and understanding.

References

[1] Aristotle. *Rhetoric*. Trans. by W. Rhys Roberts. Written 350 B.C.E, Translated by W. Rhys Roberts. MIT Classics, 350 B.C.E. URL: http://classics.mit.edu/Aristotle/rhetoric.html.

[2] Christoph Bartneck et al. *Verbal Interaction*. 2024. URL: https://www.human-robot-interaction.org.

[3] Ann R. Bradlow, Lynne C. Nygaard, and David B. Pisoni. "Effects of talker, rate, and amplitude variation on recognition memory for spoken words". In: *Perception & Psychophysics* 61.1 (1999), pp. 206–219. DOI: 10.3758/BF03206883.

[4] Cynthia Breazeal. "Toward sociable robots". In: *Robotics and Autonomous Systems* 3 (2003). Socially Interactive Robots, pp. 167–175. ISSN: 0921-8890. DOI: https://doi.org/10.1016/S0921-8890(02)00373-1. URL: https://www.sciencedirect.com/science/article/pii/S0921889002003731.

[5] Caterina Breitenstein, Diana Van Lancker Sidtis, and Irene Daum. "The Contribution of Speech Rate and Pitch Variation to the Perception of Vocal Emotions in a German and an American Sample". In: *Cognition and Emotion* 15.1 (2001), pp. 57–79. DOI: 10.1080/0269993004200114.

[6] Daphne Blunt Bugental and Eta K. Lin. "Attention-Grabbing Vocal Signals: Impact on Information Processing and Expectations". In: *Personality and Social Psychology Bulletin* 23.9 (1997). PMID: 29506447, pp. 965–973. DOI: 10.1177/0146167297239006. eprint: https://doi.org/10.1177/0146167297239006. URL: https://doi.org/10.1177/0146167297239006.

[7] Rebecca Cherng-Shiow Chang, Hsi-Peng Lu, and Peishan Yang. "Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan". In: *Computers in Human Behavior* 84 (2018), pp. 194–210. ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2018.02.025. URL: https://www.sciencedirect.com/science/article/pii/S0747563218300839.

[8] Marcus Cheetham. *The uncanny valley hypothesis and beyond*. eng. Frontiers Research Topics. Frontiers Media SA, 2018. ISBN: 2-88945-443-6.

[9] Nicholas Epley et al. "On Seeing Human: A Three-Factor Theory of Anthropomorphism". eng. In: *Psychological review* 114.4 (2007), pp. 864–886. ISSN: 0033-295X.

[10] Friederike Eyssel et al. "'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism". In: (Mar. 2012). DOI: 10.1145/2157689.2157717.

[11] R. Reed Hunt. "The subtlety of distinctiveness: What von Restorff really did". eng. In: *Psychonomic bulletin review* 2.1 (1995), pp. 105–112. ISSN: 1069-9384.

[12] Al-Jahiz. *Al Bayan Wal Tabain*. Accessed: 2024-10-09. Dar Al-Fikr Al-Arabi, 1921. URL: https://archive.org/details/in.ernet.dli.2015.322902/page/n232/mode/1up.

[13] Robert F. Lorch. "Text-signaling devices and their effects on reading and memory processes". In: *Educational Psychology Review* 3 (1989), pp. 209–234. ISSN: 1573-336X. DOI: 10.1007/BF01320135. URL: https://doi.org/10.1007/BF01320135.

[14] Andreea Niculescu et al. “The influence of voice pitch on the evaluation of a social robot receptionist”. In: *2011 International Conference on User Science and Engineering (i-USEr)*. 2011, pp. 18–23. DOI: 10.1109/iUSEr.2011.6150529.

[15] Klaus Oberauer. “Working Memory and Attention – A Conceptual Analysis and Review”. In: *Journal of Cognition* (2019). DOI: https://doi.org/10.5334/joc.58. URL: https://psycnet.apa.org/record/2019-52349-001.

[16] François Osiurak, Yves Rossetti, and Arnaud Badets. “What is an affordance? 40 years later”. eng. In: *Neuroscience and biobehavioral reviews* 77 (2017), pp. 403–417. ISSN: 0149-7634.

[17] Marc D. Pell, Silke Paulmann Laura Monetta, and Sonja A. Kotz. “Recognizing Emotions in a Foreign Language”. In: *Journal of Nonverbal Behavior* 33.2 (2009), pp. 107–120. DOI: 10.1007/s10919-008-0065-7. URL: https://doi.org/10.1007/s10919-008-0065-7.

[18] Robert Potter et al. “Effect of Vocal Pitch Difference on Automatic Attention to Voice Changes in Audio Messages”. In: *Communication Research* 46 (Dec. 2016). DOI: 10.1177/0093650215623835.

[19] Jenny R. Saffran, Richard N. Aslin, and Elissa L. Newport. “Statistical learning by 8-month-old infants”. In: *Science* 5294 (1996), pp. 1926–1928. DOI: 10.1126/science.274.5294.1926.

[20] Klaus R Scherer. “Vocal communication of emotion: A review of research paradigms”. In: *Speech Communication* 1 (2003), pp. 227–256. ISSN: 0167-6393. DOI: https://doi.org/10.1016/S0167-6393(02)00084-5. URL: https://www.sciencedirect.com/science/article/pii/S0167639302000845.

[21] Jamie Ward. *The Student’s Guide to Cognitive Neuroscience*. Fourth edition. Routledge, 2020, pp. 175–201. ISBN: 9781351035170.

[22] Jamie Ward. *The Student’s Guide to Cognitive Neuroscience*. Fourth edition. Routledge, 2020, pp. 247–258. ISBN: 9781351035170.

[23] Jan Zirk-Sadowski, Denes Szucs, and Joni Holmes. “Content-specificity in verbal recall: a randomized controlled study”. In: *PLoS One* (2013). DOI: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0079528.

5 Appendix

Control Group:

Id	Starting Time	Complete Time	Time in Minutes	Time in Seconds	Taken
1	5/31/2024 9:01	5/31/2024 9:03	1	46	FALSE
2	6/9/2024 13:22	6/9/2024 13:26	3	22	FALSE
3	6/10/2024 18:46	6/10/2024 18:53	7	2	TRUE
4	6/10/2024 22:35	6/10/2024 22:47	12	0	TRUE
5	6/11/2024 6:21	6/11/2024 6:27	5	47	TRUE
6	6/11/2024 11:54	6/11/2024 12:02	7	55	TRUE
7	6/11/2024 18:10	6/11/2024 18:51	41	0	FALSE
8	6/12/2024 1:55	6/12/2024 2:05	10	35	TRUE
9	6/19/2024 14:41	6/19/2024 14:43	1	13	FALSE
10	6/20/2024 14:53	6/20/2024 15:00	7	24	TRUE
11	6/21/2024 2:17	6/21/2024 2:30	13	18	TRUE

Table 1: Task completion times and details (Control Group).

Experiment group:

Id	Starting Time	Complete Time	Time in Minutes	Time in Seconds	Taken
1	5/29/2024 19:35	5/29/2024 19:42	7	41	TRUE
2	5/30/2024 12:07	5/30/2024 12:14	7	33	TRUE
3	6/9/2024 14:28	6/9/2024 14:31	2	49	FALSE
4	6/11/2024 18:51	6/11/2024 20:03	71	36	FALSE
5	6/12/2024 17:44	6/12/2024 17:51	7	0	TRUE
6	6/14/2024 20:48	6/14/2024 20:53	4	48	FALSE
7	6/28/2024 14:19	6/28/2024 14:28	9	2	TRUE
8	6/30/2024 0:37	6/30/2024 0:41	3	34	FALSE
9	6/30/2024 15:59	6/30/2024 16:08	9	28	TRUE
10	7/1/2024 21:52	7/1/2024 22:36	43	43	FALSE

Table 2: Task completion times and details (Experiment Group).

6 Usage of ChatGPT

- Helping with finding old references I read it before.
- Helping with some codes
- Helping with writing the sharks text
- Helping with grammars and choosed words
- Helping with rephrasing some sentences.