
Features Extraction and Audio Analysis of Human-Robot Interaction

Author:
Léa Blinière

Supervisor(s):
Daniel Tozadore

Professor:
Pierre Dillenbourg

June 11, 2023

Contents

1	Introduction	1
2	Experiment Design	1
2.1	Participants	1
2.2	Interview protocol	2
2.3	Lesson design	2
2.4	Equipment	3
2.5	Audio Preprocessing	3
2.6	Extraction of audio features	4
3	Results	5
3.1	Data exploration	5
3.2	Impact of the order of passage on human-robot interaction differences . . .	5
3.3	Impact of Language on the Interaction Differences between the Robot and the Human	7
3.4	Comparison of Human-Robot Interaction Differences in the Features, Isolating Other Effects	8
4	Discussion	10
4.1	Impact of Technical Issues on Data	10
4.2	Relevance of Statistical Tests	10
4.3	Impact of Order of Interaction	11
4.4	Influence of Language on Interactions with the Robot and the Human . . .	12
4.5	Comparison of Human-Robot Interaction Differences in the Features, Isolating Other Effects	12
4.6	Participant perception through questionnaire analysis	13
5	Conclusion	15
6	Appendix	17
A	Lessons	17
B	Features Details	26
C	Post-experiment questionnaire	28
D	Box Plots of Significant Features by Starting Order and Agent	32
E	Box Plots of Significant Features by Language and Agent	37
F	Box Plots of Significant Different by Agent, Isolating Other Effects	39

1 Introduction

The field of human-robot interaction has been a subject of increasing interest in recent years, as advancements in technology have made it possible for robots to take on roles traditionally filled by humans. Some studies have already examined the potential of social robots in supporting individuals' psychological well-being through conversation by studying participants' audio features [1]. Others have evaluated the attraction and acceptance of a robot instructor in an educational and sports context[2] [3]. Despite the progress made, there are still gaps in our understanding of how people interact with robots and the factors that influence these interactions. This study aims to fill some of these gaps by examining the nuances of human-robot interactions.

Understanding how people interact with robots is crucial as it has implications for various sectors, including education, healthcare, and customer service. For instance, robots can be used as teaching aids in classrooms, assist in patient care in healthcare settings, or provide customer service in retail environments. However, for these applications to be effective, it is essential to understand how people interact with robots and how these interactions can be improved.

The objective of this study is to delve deeper into the dynamics of human-robot interactions. Specifically, we aim to answer the question: "To what extent do human responses to academic questions vary based on the embodiment of the agent (human or robot) asking the questions?" By exploring this question, we hope to gain insights that can help improve the design and implementation of robots in various settings. This study involves a series of experiments conducted in a controlled setting with a group of participants. The participants, all of whom are native French speakers, interact with both a human instructor and a humanoid conversational robot. The interactions are recorded and analyzed to identify differences in the participants' behavior and responses when interacting with the human and the robot. The study also includes a post-experiment questionnaire to gather participants' perceptions of their interactions with both the human and the robot.

In the following sections, we will provide a detailed description of the experiment design of the study, present the results of our analysis, and discuss the implications of our findings for the field of human-robot interaction.

2 Experiment Design

2.1 Participants

For this experiment, a recruitment process resulted in the participation of 33 individuals (M age = 27.03, SD age = 11.72, 14 female, 19 male). All participants voluntarily agreed to take part in this study and signed an informed consent form prior to their participation. Among them, 26 had previously interacted with a human who performed one part of the experiment, while only 3 had interacted with a humanoid conversational robot. Twenty-five participants hailed from a scientific background (engineers or researchers). All participants were native French speakers.

2.2 Interview protocol

First, participants were informed about the experiment’s procedures, which consisted of two parts: one with the social humanoid robot QT and one with a human agent. The order of interactions was randomly assigned to all participants and was evenly distributed among them. Participants were advised that the agent (human or robot) would give a series of three short lessons, interspersed with questions. Participants were not informed that the robot operated in "Wizard of Oz" mode [4] to ensure their interaction with the robot was as natural as possible. Participants were informed that they could not ask the human agent or the robot to repeat the lesson or question to maintain fairness among the participants. However, they were encouraged to indicate if they did not understand or could not answer a question. After each answer, brief feedback was provided by the instructor (robot or human) to minimize a "robotic" effect. Participants were also informed that their interactions would be recorded to collect audio data. Participants were then guided to a soundproof room, which already contained the first instructor, either the robot or the human. The instructor was switched between the two experiment phases, without moving the participant. At the end of the experiment, participants were asked to complete a post-experiment questionnaire to evaluate various aspects of their interaction with both the robot and the human.

Questions covered:

- *Comfort level:* Participants were asked to rate their comfort level when answering questions posed by the robot and the human, on a scale of 0 to 5. This could help understand how comfortable participants felt when interacting with the robot compared to the human.
- *Robot speech perception:* The questionnaire asked participants if they found the robot’s speech to be natural and understandable. This could provide information about the quality of the robot’s communication.
- *Differential behavior:* Participants were asked whether they perceived any differences in their behavior when interacting with the robot compared to the human. This could reveal any differences in how participants interact with robots versus humans.
- *Factors influencing responses:* The questionnaire also asked to what extent the robot’s alternating between English and French speech could have disrupted their interactions. They were also asked to evaluate the difficulty of the questions and to rank their preference for the topics covered in the lessons to assess if this influenced their responses.

2.3 Lesson design

Each session consisted of a series of three lessons delivered by the agent. These lessons covered either science or history (both in French) or English. The order of the lessons was

also randomly determined and varied between human and robotic sessions. We selected comparable lessons in terms of the questions asked and the topics covered. For example, in history, we covered "France on the eve of the French Revolution" and "Feudal Society." In science, we addressed "Insect Development" and "Animal Growth." In English, we covered "Symbols of the United States" and "Symbols of England." The lessons can be found in the appendix A. The order of lesson presentation was determined so that if a participant started with the robot, the order of lessons was randomly drawn for the subject and lesson number. For the session with the human agent, the order was also randomly determined, and the lesson number corresponded to the one that was not given in the previous session. Each lesson lasted between 2 and 3 minutes and included two questions, one in the middle and one at the end, related to the corresponding part of the lesson. In the experimental design, we considered the inclusion of "trigger questions." These questions were related to the previous question and aimed to encourage participants to provide more detailed responses when their initial answers were too brief. The use of question triggers allowed us to collect sufficiently long audio files to extract relevant audio characteristics. The triggers were activated based on the subjective judgment of the human agent or the person controlling the Wizard of Oz.

The lesson topics were chosen from the French educational program for 8-11-year-olds [5] to be accessible to all participant. All questions were designed so that participants could answer them by listening to the lesson, ensuring fairness in terms of knowledge. Moreover, to obtain participants with diverse backgrounds and interests, I chose to include lessons in history and science as they can appeal to different audiences. The inclusion of an English lesson aimed to analyze participant behavior when speaking in a language other than their mother tongue.

2.4 Equipment

During this experiment, a BLUE Yeti microphone was used. This device is ideal for podcasts and interviews, allowing for the capture of excellent quality sound. To collect the audio data and carry out pre-processing, including the cutting of audio segments of interest, Audacity software was utilized. The robot used in the experiment was a QTrobotV1, it is an interactive humanoid robot designed for use in educational and therapeutic contexts. It was developed by LuxAI, a company specializing in the creation of social robots for education and therapy. The robot operated using the Wizard of Oz technique to minimize potential errors. It was programmed entirely using the "QTrobot studio" application, which simplifies robot programming. The lesson scripts were programmed into the robot and were the same as those provided to the human agent. The QT robot was programmed to mimic lip movements while speaking and to display facial expressions such as smiling to enhance the naturalness of its interaction.

2.5 Audio Preprocessing

Each participant's entire experience was recorded in a single audio file. To conduct the analysis for each participant, two sub-audio files were created: one for interactions with

the robot and one for interactions with the human. Subsequently, audio segments corresponding to each lesson (history, science, or English) were extracted from these sub-audio files. Finally, specific parts of these lessons, where the participant responded to a specific question, were further extracted.

2.6 Extraction of audio features

Python librairies

For our study, we utilized two Python libraries to extract and analyze acoustic and prosodic characteristics from the vocal signals: Simple-Speech-Features [6] and My-Voice Analysis [7]. Simple-Speech-Features is a straightforward Python library for the extraction and manipulation of acoustic and prosodic features from audio signals. This library, based on Praat and Parselmouth, includes all the necessary utilities and functions for speech extraction and analysis. My-Voice Analysis, on the other hand, is a Python library designed for voice analysis without the need for transcription. It segments utterances, detects syllable boundaries, fundamental frequency contours, and formants. It also includes built-in functions for analyzing articulation rate, speech rate, and more. These tools were chosen for their ability to provide comprehensive quantitative and analytical analysis of acoustic speech features. They leverage Parselmouth, a Python interface for Praat, to extract relevant acoustic characteristics for our human-robot interaction analysis.

Choices of the features

The choice of these features was motivated by these papers [8] [1], and the details of how they can be calculated can be found here [9]

We can categorize the extracted features into five categories:

- *Temporal characteristics*: Features related to duration and rhythm of the audio, such as speaking duration, rate of speech, articulation rate, and balance.
- *Intensity characteristics*: Features related to the intensity or volume of the audio, including mean intensity, standard deviation of intensity, and voiced fraction.
- *Pitch characteristics*: Features related to voice pitch, which is associated with fundamental frequency, such as mean, maximum, minimum, standard deviation of pitch, and mean absolute pitch slope.
- *Voice quality characteristics*: Features related to voice clarity and stability, including mean harmonic-to-noise ratio, standard deviation of HNR, jitter, and shimmer.
- *Other features*: In the context of using the My Voice Analysis library, we implemented a function called "silence threshold modulation" to detect silent periods in a voice. Initially set to -20dB in the library, the threshold was automatically lowered until the library could detect relevant characteristics. The minimum threshold required to extract these characteristics was added as a new feature.

For a more detailed description of the features, please refer to the appendix B.

3 Results

3.1 Data exploration

During the preprocessing phase, the audio features of two participants had to be excluded due to technical issues during their experiments. An analysis of the distribution of features was also conducted, revealing that none of the features followed a normal distribution. Throughout the analysis, several effects that may impact the differences between the audio features of interactions with the robot and those with the human will be examined. Firstly, the impact of the order of interaction on the audio differences between human-robot interactions will be investigated. Next, the influence of languages on the differences between the interactions will be examined. Finally, the differences between the audio features of the robot and the human will be studied by controlling for these two factors.

3.2 Impact of the order of passage on human-robot interaction differences

An initial study was conducted to analyze whether starting with a specific agent had an impact on the audio features between the robot and the human. To eliminate language influence during interaction, only interactions in French were retained. The dataset was then divided into two groups: participants who interacted with the human first and those who interacted with the robot first. Finally, for each of these two datasets, we split the data into interactions with the human and interactions with the robot. We calculated the mean features per participant for each dataset and then calculated the difference by subtracting the features of the first interaction from the features of the second interaction. The diagram 1 illustrates this process.

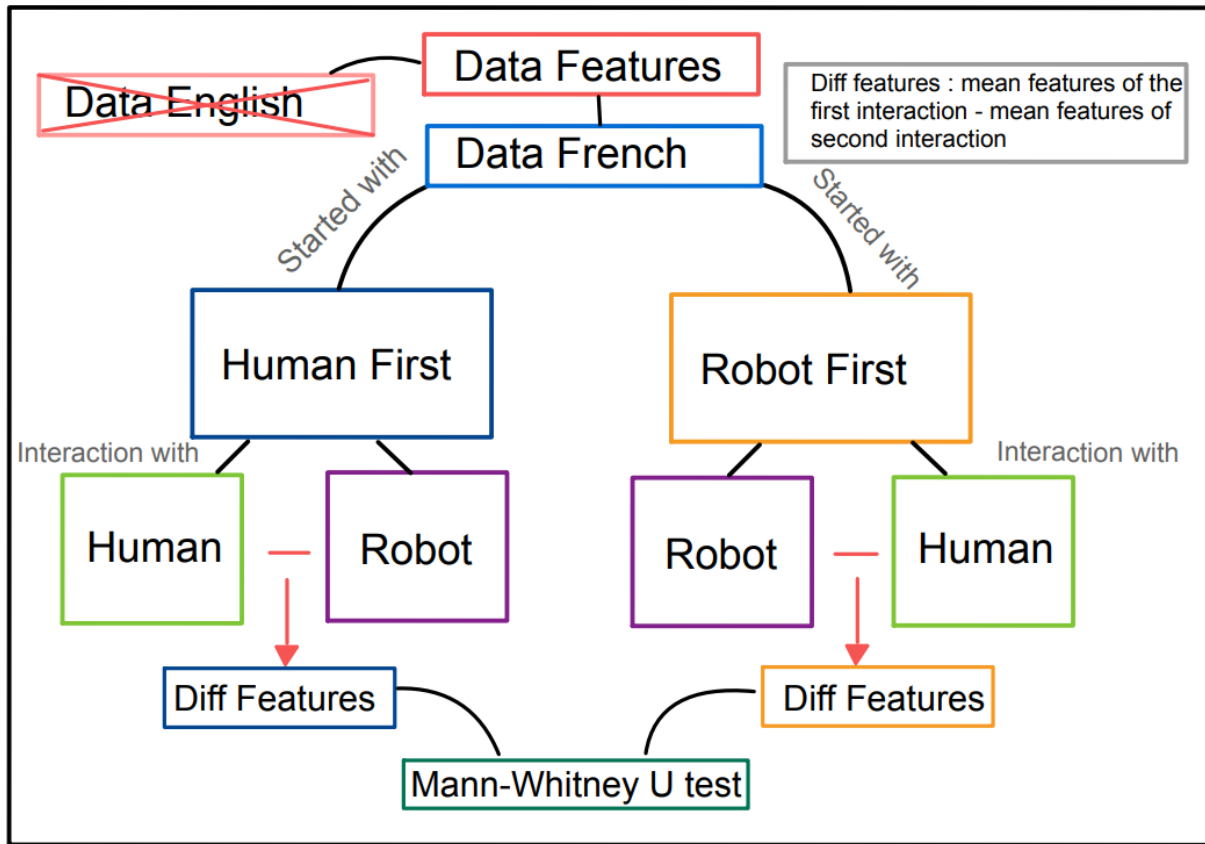


Figure 1: Diagram illustrating the process of selecting data to compare the impact of the order of interaction on the features

Choice of Statistical Test

To compare the distributions of the two samples, a Mann-Whitney U test was performed. The Mann-Whitney U test is a non-parametric statistical method used to compare the distributions of two independent samples when the data does not follow a normal distribution.

Results

The table 1 presents the features for which the Mann-Whitney U test yielded a p-value < 0.05 , indicating a statistically significant difference between the conditions.

Table 1: Table of Features Significantly Different ($p < 0.05$) based on Mann-Whitney U test

Features	p-value
stddev_hnr	0.0001
articulation_rate	0.0005
mean_intensity	0.0006
mean_absolute_pitch_slope	0.0007
voiced_fraction	0.0010
jitter	0.0044
mean_hnr	0.0096
mean_pitch	0.0135
duration	0.0151
min_pitch	0.0208
rate_of_speech	0.0229
shimmer	0.0312
speaking_duration	0.0459

The box plots of the features in both cases can be found in the appendix D

3.3 Impact of Language on the Interaction Differences between the Robot and the Human

During the experiment, two lessons were conducted in French (history and science), while one lesson was conducted in English. The participants were native French speakers with varying proficiency in English. This analysis aims to investigate the impact of language during interactions with an agent on the audio features. Two cases are considered: language change during interactions with the human and language change during interactions with the robot. The goal is to determine if there are differences in the features when interacting in different languages with a human compared to the robot. To address this, the dataset was divided into two groups: participants who interacted with the robot first and participants who interacted with the human first. The first interactions were selected for analysis. It is important to remind that each participant interacted with both the robot and the human in both languages. Therefore, the English interactions and the first French interaction for each participant were selected in both datasets. The diagram 2 illustrates the feature selection process.

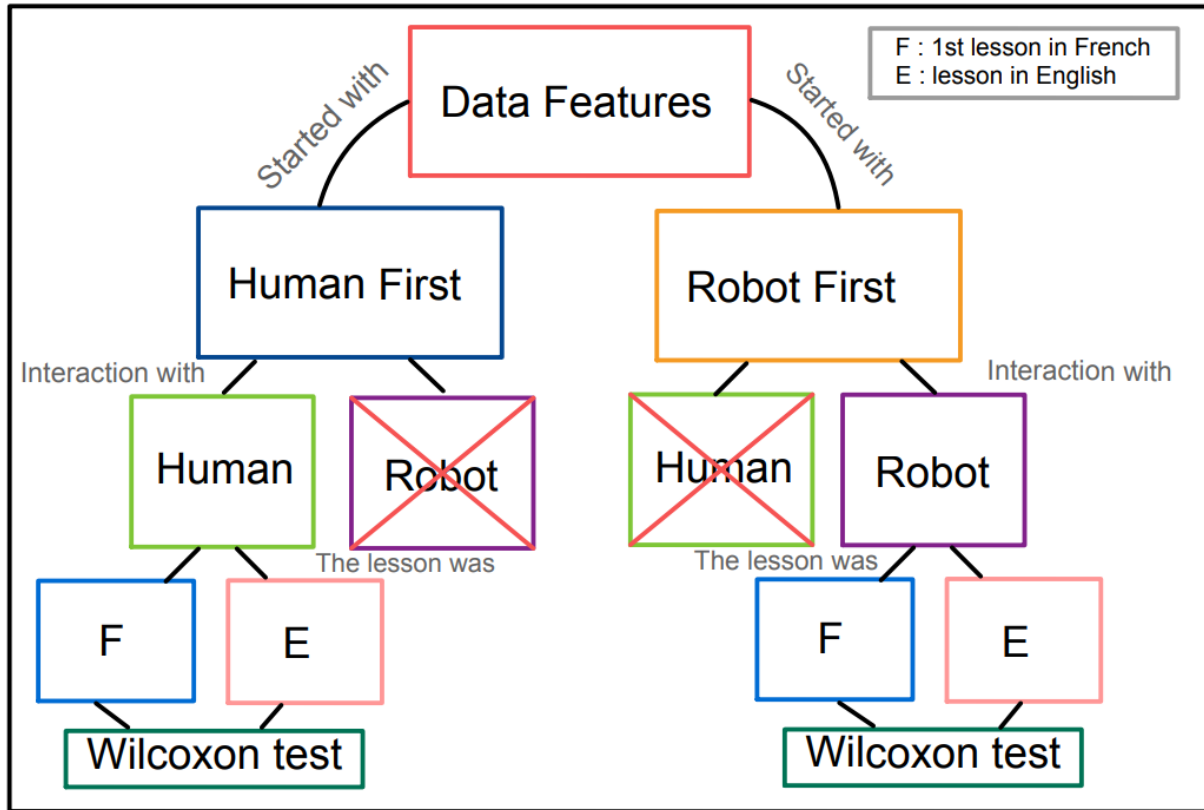


Figure 2: Diagram illustrating the process of selecting data to compare the impact of language on the features.

Choice of Statistical Test

For this analysis, the Wilcoxon test, also known as the signed-rank test, was chosen as the appropriate statistical test. This test is suitable for comparing paired samples, as we are comparing the same participants in each comparison. The Wilcoxon test is a non-parametric statistical method commonly used to assess differences between the distributions of two dependent or paired samples.

Results The table 2 and 3 presents the features for which the Wilcoxon test yielded a p-value < 0.05 in both cases, indicating a statistically significant difference between the conditions.

The box plots of the features in both cases can be found in the appendix E

3.4 Comparison of Human-Robot Interaction Differences in the Features, Isolating Other Effects

To ensure a strict comparison of differences between the robot and the human, only the first interactions of each participant were considered. This involved analyzing the interactions with the human for participants who started with the human, and vice versa. By

Table 2: Table of Features Significantly Different ($p < 0.05$) based on Wilcoxon Test for the human interaction in both languages

Features	p - value
shimmer	0.0185
mean hnr	0.0256
articulation rate	0.0439

Table 3: Table of Features Significantly Different ($p < 0.05$) based on Wilcoxon Test for the robot interaction in both languages

Features	p - value
stddev hnr	0.0015
voiced fraction	0.0054
rate of speech	0.0057
mean intensity	0.0353
articulation rate	0.0461

adopting this approach, the potential influence of the order of interaction was eliminated. Additionally, to mitigate the impact of language, only the interactions conducted in French were selected, as there were twice as many features available compared to English. Following this, a Mann-Whitney U test was conducted to determine the features that exhibited significant differences between these two scenarios. The diagram 3 illustrates the feature selection process.

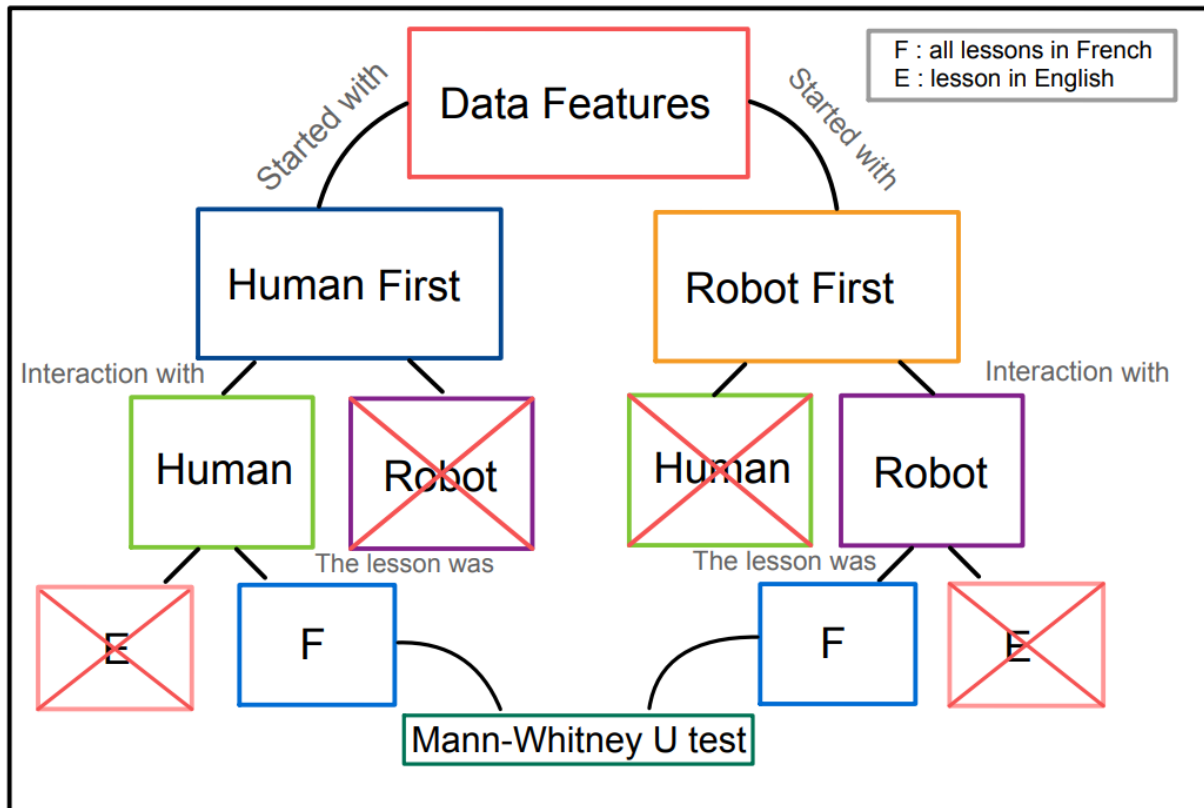


Figure 3: Diagram illustrating the process of selecting data to compare the features when interacting with a robot and when interacting with a human

Results The table 4 presents the features for which the Mann-Whitney U test yielded a p-value < 0.05 , indicating a statistically significant difference between the conditions.

Table 4: Table of Features Significantly Different ($p < 0.05$) based on Mann-Whitney U test

Features	p-value
mean_intensity	0.0019
mean_hnr	0.0037
stddev_hnr	0.0068
jitter	0.0125
rate_of_speech	0.0165

The box plots of the features in both cases can be found in the appendix F

4 Discussion

It is important to note that while I cannot provide a detailed explanation for all the results of each feature in this section of the analysis, I invite readers to refer to the appendix B, D, E; F for a complete overview of the results and a description of each feature.

4.1 Impact of Technical Issues on Data

Robot Noise

The robot inherently produces background noise due to its fan and joint movements, which could be captured during the experiments and potentially bias certain features, especially those related to HNR (Harmonic-to-Noise Ratio). Although noise filtering techniques were considered, participants' varying speech intensities made it challenging to effectively implement such filtering methods.

Participant Positioning

Unfortunately, it was not feasible to require participants to remain still during their experiments. It is important to note that some participants, particularly those who were stressed, exhibited movement during the recordings. This variation in distance from the microphone could introduce bias into the data. Furthermore, while all participants were placed in the same position, some individuals maintained different postures and distances from the microphone. These factors could introduce additional variations in the recordings and potentially influence the results.

4.2 Relevance of Statistical Tests

The non-normal distribution of audio features limited the choice of statistical tests for analyzing correlation effects. Parametric statistical tests such as ANOVA (Analysis of Variance) or Student's t-test typically assume a normal distribution of data. Since our

features did not follow this distribution, we had to resort to non-parametric statistical tests such as the Mann-Whitney U test and Wilcoxon test, which are more suitable for non-normal data. The choice to use the Mann-Whitney U test to compare the distributions of two independent samples and the Wilcoxon test to compare the distributions of two paired samples was justified by the non-normal nature of our data. These non-parametric tests do not rely on strict assumptions about data distribution, making them more appropriate for our situation. However, it is important to acknowledge that these tests may be less powerful than parametric tests in certain cases.

4.3 Impact of Order of Interaction

The results of the analysis on the impact of the order of interaction on the differences between human-robot interactions reveal that several features exhibit statistically significant differences. This suggests that the order in which participants interacted with the human and the robot had an impact on the audio characteristics of the interaction. An intriguing discovery has been made regarding the length of responses in relation to the two scenarios. Based on the data presented in table ??, it appears that the average duration of interviews for the group that initially interacted with a human is comparable for both their subsequent interaction with a robot and a human. On the other hand, for the group that began with the robot, there is a notable disparity in response length, with participants tending to provide briefer answers to the robot than to the human.

Table 5: Comparison of response duration between participants starting with Human and Robot

	Human first		Robot first	
	Human	Robot	Human	Robot
Mean	23.34	22.95	28.76	15.85
Standard Deviation	15.74	15.96	10.79	21.39

These differences could be attributed to two factors. First, participants tend to respond more concisely when interacting with the robot naturally. This tendency may be influenced by the participants' perception of the robot as a technological entity rather than a human. Second, the absence of these differences in the group that started with the human could be due to the loss of novelty effect [10]. Participants who had previously interacted with the human became more aware of what to expect during the experiment, leading them to feel more comfortable and adapt their discourse during the second interaction with the robot. This increased familiarity may have influenced their response style, resulting in similar response durations between the human and robot interactions.

4.4 Influence of Language on Interactions with the Robot and the Human

Language has a noticeable influence on the interactions with both the robot and the human, as there are significant differences in the audio characteristics in both cases. However, we observe more pronounced differences when it comes to transitioning between English and French in the context of the robot interactions. For example, in both scenarios, there are distinct variations in the rate of articulation when switching between languages as it can be observed in E. This finding aligns with expectations, considering that the rate of articulation is typically higher in one’s native language compared to a second language as stated in [11]. As all participants are native French speakers with English as a second language, it is expected that their rate of articulation would be higher in French. However, the box plots reveal that participants’ rates of articulation differ more significantly between languages in the context of the robot interactions compared to the human interactions. Additionally, participants have reported that the robot is less comprehensible in English compared to French. This observation suggests that language can impact the fluency and quality of responses during interactions with the robot. It is important to note that the groups of participants involved in the robot and human scenarios are not the same, and there may be variations in their English proficiency levels. This aspect should be considered when interpreting and comparing the results. Nevertheless, it is evident that language plays a role in shaping the interaction experience with both the robot and the human. In future experiments, taking participants’ English proficiency into account would provide deeper insights into the influence of language on interactions with the robot and the human.

4.5 Comparison of Human-Robot Interaction Differences in the Features, Isolating Other Effects

When comparing the differences in interactions between participants and humans or robots, while controlling for the order of interaction and language, several features still exhibit significant differences. This suggests that there are intrinsic differences in how individuals interact with humans and robots, which aligns with previous studies in the field of human-robot interaction. Among these differences, we find the average intensity, which reflects the average volume of the interaction. The results are shown in table 6

Table 6: Comparison of Mean and Standard Deviation of the Mean intensity between an interaction with the human and a robot

	Human	Robot
Mean	66.9	72.14
Standard deviation	10.08	1.68

Due to the complexity of the effects involved and the differences in participants, we cannot draw definitive conclusions about the higher average intensities during robot interaction.

It is important to note that the participants in both groups may speak at varying volumes. Nonetheless, we observe that the standard deviation of the intensity during interaction with the robot is very low compared to that of the interaction with humans, suggesting that participants tend to speak at a similar volume with the robot, and we may even assume that they align with its volume. Additionally, it is worth mentioning that the volume of the robot’s voice remained constant throughout the lesson and participants which could explain these results. These findings provide valuable insights into the nuanced differences in interaction patterns between humans and robots. However, further research is needed to better understand the underlying factors shaping these interactions.

4.6 Participant perception through questionnaire analysis

Report on comfort during interactions with the human and the robot

In a post-study questionnaire, we asked participants to rate their comfort level on a scale of 0 to 5 (with 5 being the highest) for interactions with the human and the robot. The results are show in tbale 7

Table 7: Comparison of Average and Standard Deviation between Human and Robot Comfort Ratings

	Humain	Robot
Mean	4.36	3.6
Standard Deviation	0.74	1.14

Participants reported feeling more at ease with the human compared to the robot. It is important to note that 81% of the participants were already familiar with the human interviewer prior to the study, and this familiarity influenced the results. Only 4 out of 33 participants rated their comfort level higher with the robot than with the human, and among them, only 1 had no prior interaction with the human interviewer, indicating that it was not solely due to unfamiliarity. None of these 4 participants had previously interacted with a humanoid conversational robot. Two of them mentioned feeling less judged by the robot, which made them more comfortable in the context of a lesson.

Report on perception of acting differently between the robot and the human

81% of the participants reported feeling that they acted differently, and among them, all participants who had never interacted with the human interviewer stated that they behaved differently. When asked why in the questionnaire, 48% attributed it to the lack of nonverbal cues from the robot. Many mentioned being unsettled by the absence of genuine eye contact with the robot. Overall, participants felt more at ease in responding to the human’s questions as they perceived approval or interest in the human’s gaze. 14% stated that they acted differently with the robot because they did not fear judgment. One participant mentioned that since the structure of both parts (robot and human) was similar, they quickly understood the setup, which made them more comfortable in the second part. 18% mentioned that difficulties in understanding what the robot was saying

and the lack of natural conversation contributed to their perceived differences in behavior.

Understanding and naturalness of language:

Prior to the experiments, we were aware that the robot was sometimes difficult to understand and lacked naturalness in its speech. Therefore, we included questions in the post-study questionnaire to assess participants' understanding and perception of naturalness in their interactions with the robot. The results are as follows:

Table 8: Mean and Standard Deviation of Comprehension and Naturalness

	Comprehension	Naturalness
Mean	2.84	2.51
Standard Deviation	1.06	1.22

The ratings are relatively low, which aligns with the feedback received from participants. It is important to note that several participants mentioned that this difficulty was partly due to the English language. Two possible explanations emerge: the fact that English was not their native language and the challenge of understanding a robot speaking in a language that did not adhere to natural intonations. Another hypothesis relates to the pre-programmed voices used on the robot, with different voices for English and French. The French voice resembled that of a young boy, while the English voice resembled that of a young girl. An infantile voice might be more challenging to comprehend and could emphasize the difficulty understand it. We also asked participants to rate on a scale of 1 to 5 how much the difference in voice between English and French perturbed them. The results are as show in table 9

Table 9: Note on the disturbance caused by the variation in voice between French and English

	Disturbances
Mean	2.09
Standard Deviation	1.58

24% of the participants rated a score of 4 or higher, with 5 indicating significant disturbance, indicating that the participants were divided on the issue, considering the average score remained relatively low.

5 Conclusion

The primary findings of this study underscore the multifaceted nature of human-robot interactions. The results indicate that factors such as language and order of interaction significantly influence the audio features of these interactions. Participants generally felt more at ease interacting with a human than a robot, and a majority reported a change in their behavior when interacting with the robot, primarily due to the absence of non-verbal cues. The study indicates that the lack of naturalness in the robot's speech can make understanding difficult, especially when it involves a language other than the participant's native language. This can have a negative impact on the quality of interactions. However, despite these challenges, some participants expressed feeling less judged by the robot. This suggests that there may be potential advantages to interacting with a robot in certain contexts. These findings have significant implications for the design and implementation of social robots. They highlight the need for improvements in the naturalness and comprehensibility of robot speech. They also emphasize the importance of considering the order of interactions and the potential influence of this factor on the outcomes of human-robot interactions.

However, this study is not without its limitations. The sample size was limited to 31 participants primarily from a scientific background, limiting the generalizability of the findings. Moreover, the study did not account for individual differences in participants' comfort and familiarity with technology, which could potentially influence their interactions with the robot.

Future research should aim to address these limitations by including a more diverse sample of participants and considering individual differences in technology use and comfort. Additionally, future studies could explore other factors that may influence human-robot interactions, such as the robot's physical appearance or the complexity of the tasks involved in the interaction. Lastly, research could also investigate the quality of responses from the participants.

References

- [1] G. Laban, J.-N. George, V. Morrison, and E. S. Cross, “Tell me more! assessing interactions with social robots from speech,” *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 136–159, 2021.
- [2] E. Park, K. J. Kim, and A. P. del Pobil, “The effects of a robot instructor’s positive vs. negative feedbacks on attraction and acceptance towards the robot in classroom,” in *Social Robotics* (B. Mutlu, C. Bartneck, J. Ham, V. Evers, and T. Kanda, eds.), (Berlin, Heidelberg), pp. 135–141, Springer Berlin Heidelberg, 2011.
- [3] N. Akalin, A. Kristoffersson, and A. Loutfi, “The influence of feedback type in robot-assisted training,” *Multimodal Technologies and Interaction*, vol. 3, no. 4, 2019.
- [4] “Wizard of Oz experiment.” Wikipedia. Consulté le 2023-06-11.
- [5] “Maxicours.”
- [6] “Simple speech features.” <https://github.com/uzaymacar/simple-speech-features>.
- [7] “My voice analysis.” <https://github.com/Shahabks/my-voice-analysis>.
- [8] E. Jacewicz, R. A. Fox, and J. Salmons, “Articulation rate across dialect, age, and gender,” *Language Variation and Change*, vol. 21, no. 2, pp. 233–256, 2009.
- [9] “Praat: Doing phonetics by computer.” <https://www.fon.hum.uva.nl/praat/manual/Manual.html>.
- [10] G. Hoffman and X. Zhao, “A primer for conducting experiments in human–robot interaction,” vol. 10, oct 2020.
- [11] H. Kallio, *The prosody underlying spoken language proficiency: Cross-lingual investigation of non-native fluency and syllable prominence*. PhD thesis, University of Helsinki, February 2022.

6 Appendix

A Lessons

Science 1 : Développement des insectes

Pour cette partie je vais te parler de science, est ce que tu es prêt ?

Partie 1

C'est parti ! Nous allons parler du développement des insectes. Tous les animaux changent au cours du temps. Certains êtres vivants ont un développement direct, c'est-à-dire que le petit ressemble déjà à l'adulte. Il subit simplement une augmentation de la taille, du poids et des changements dus à la maturité sexuelle. D'autres ont un développement indirect, c'est-à-dire que le petit ne ressemble pas du tout à l'adulte, on l'appelle une "larve". Pour atteindre le stade adulte, il subit plusieurs transformations qu'on appelle des "métamorphoses". Par exemple, la grenouille et la coccinelle ont un développement indirect.

Question

Commençons par une première question, prêt ?

1) Je t'ai parlé des deux types de développement des insectes, peut tu me rappeler de quoi s'agissait-il ?

Trigger

a) Peux tu m'en dire plus sur leurs différences ?

Partie 2

D'accord, merci pour ta réponse. Maintenant, reprenons la leçon avec le cas du papillon. Un papillon ne vit que quelques semaines. Certains papillons ne vivent parfois même que quelques jours seulement. Le cycle de vie du papillon comporte 4 stades : l'œuf, la larve, la nymphe et l'âge adulte. Dans le cas du papillon, la larve s'appelle une chenille, la nymphe s'appelle une chrysalide, et l'adulte s'appelle un papillon. Le stade de l'œuf dure entre 3 et 8 jours. Les chenilles se forment alors dans les œufs que la femelle a pondus. Le stade de la larve est le premier stade de développement après

l'éclosion de l'œuf. A ce stade, la chenille grandit très vite en mangeant une quantité considérable de nourriture. Après avoir bien grandie, la chenille trouve un endroit où se fixer. C'est alors que la phase suivante commence, c'est la nymphe. Durant cette phase, l'insecte se transforme totalement. A la sortie de la nymphe, le papillon possède alors sa forme adulte. Sa vie dure en général quelques semaines pendant lesquelles il recherche un partenaire afin de se reproduire.

Question

La leçon est maintenant terminée, tu es prêt pour la dernière question ?
2) La voici ! Peux-tu me parler des 4 stades de la vie du papillon ?

Trigger

b) Peux-tu m'en dire plus sur un des stades ?

Fin

Super ! Merci pour avoir répondu à mes questions !

Science 2 : La croissance des animaux

Pour cette partie je vais te parler de science, est ce que tu es prêt ?

Partie 1

C'est parti ! Nous allons parler de la croissance des animaux. Le phénomène par lequel le végétal ou l'animal grandit s'appelle la croissance. La croissance peut être continue tout au long de la vie, comme pour les arbres, ou bien être limitée à la jeunesse de l'être vivant comme chez la plupart des animaux. Certains animaux changent presque entièrement d'aspect au cours de leur croissance. C'est le cas des animaux qui subissent une mue ou de ceux qui subissent une métamorphose. La mue, c'est le fait pour un animal d'abandonner sa carapace ou sa peau lorsqu'elles sont devenues trop petites. Lorsque l'animal mue, une nouvelle enveloppe molle s'est déjà formée sous l'ancienne. L'ancienne enveloppe se fend, l'animal en sort et l'abandonne. Le crabe et le serpent sont des animaux qui muent. La métamorphose, elle, indique un changement complet d'aspect de l'animal

au cours de son développement. Lors de la métamorphose, de nouveaux organes apparaissent, d'autres disparaissent et parfois même le milieu de vie change. Par exemple, de l'eau à la terre pour la grenouille et de la terre au ciel pour le papillon.

Question

Commençons par une première question, prêt ?

1) Je t'ai parlé des phénomènes de la mue et de la métamorphose, de quoi s'agit-il ?

Trigger

a) Peux tu me donner un exemple d'animal qui mue et d'un subissant une métamorphose ?

Partie 2

D'accord, merci pour ta réponse. Cependant, la plupart des animaux ne muent pas et ne subissent pas de métamorphose. Parmi eux, on distingue deux types d'animaux : les ovipares et les vivipares. Chez les animaux ovipares : La femelle pond des œufs contenant le nouvel être vivant pas complètement formé et des réserves pour son développement. Une fois sorti de l'œuf, l'animal ressemble à peu près à l'adulte. Ses membres et ses organes vont ensuite évoluer en taille et en proportion. Et enfin, chez les animaux vivipares : Le jeune se développe à l'intérieur du corps maternel. À sa naissance, il est déjà complètement formé et a lui aussi l'aspect de l'adulte avec des proportions différentes. Dans le cas des mammifères, dont fait partie l'homme, le début de la croissance est assuré par l'allaitement maternel. C'est pendant cette période que le développement est le plus rapide.

Question

La leçon est maintenant terminée, tu es prêt pour la dernière question ?

2) Et enfin, dernière question qu'est-ce qui distingue la croissance chez les ovipares et chez les vivipares ?

Fin

Super ! Merci pour avoir répondu à mes questions !

Histoire 1: La France a la veille de la Révolution française

Le sujet dont nous allons parler, est : l'histoire, tu es prêt ?

Partie 1

Allons-y ! Pour cette leçon, je vais te parler de la France à la veille de la Révolution française. Au XVIII^e siècle, la société française est inégale et divisée en trois ordres distincts. C'est une monarchie absolue, de droit divin, avec le Roi à la tête du pouvoir. Les trois ordres sont : le clergé, la noblesse et le Tiers-État. Je vais te parler de ces trois ordres. Commençons par le clergé. Le clergé est composé d'hommes d'église, certains très riches, tandis que les simples curés sont aussi pauvres que les paysans. Toutefois, ils partagent tous la capacité de lire et d'écrire. La noblesse, quant à elle, est constituée de ducs, comtes, vicomtes et autres descendants de familles de chevaliers. Les nobles occupent des postes importants dans l'armée et possèdent des terres. La noblesse et le clergé ont de nombreux privilèges : ils ne paient pas d'impôts lourds et ne sont pas soumis à la même justice que le Tiers-État. Et enfin, le dernier ordre est le Tiers-État, c'est le plus important en nombre. Il regroupe toutes les classes populaires, comme les paysans, les marchands, les écrivains, les avocats, les artisans et les ouvriers. Certains sont instruits et cultivés, tandis que d'autres vivent dans la misère. Ils paient tous beaucoup d'impôts.

Question

Tu devrais être prêt pour répondre à ma première question. C'est parti !

1) Je t'ai parlé des trois ordres qui régissent la société française de l'époque, peux-tu me les rappeler ?

Trigger

a) Quels sont les privilèges du clergé et de la noblesse ?

Partie 2

Ok, c'est noté ! Maintenant, tu connais les trois ordres et les inégalités qui existent entre eux. Au 18^{ème} siècle, le désir de liberté se fait de plus en plus présent en France, et une partie de la population, la plus instruite, réclame une constitution qui limiterait les pouvoirs du Roi. De plus, l'État

a besoin de beaucoup d'argent pour financer les guerres et les pensions distribuées aux nobles, mais le clergé et la noblesse s'opposent vigoureusement à l'idée d'être soumis à l'impôt. Finalement, les mauvaises récoltes et l'augmentation du prix du pain aggravent la situation de misère de la population, qui devient de plus en plus mécontente. Face à la crise financière et sociale, les trois ordres rédigent des cahiers de doléances pour faire part de leurs plaintes et de leurs revendications. Le Roi Louis XVI finit par accepter de réunir les États Généraux, une assemblée représentant les trois ordres, pour trouver une solution à la crise.

Question

Ça y est, maintenant tu devrais pouvoir répondre à ma dernière question ! On y va ?

2) Quels sont les facteurs déclencheur de la Révolution française ?

Fin

Trop chouette, merci d'avoir répondu !

Histoire 2: La société féodale.

Le sujet dont nous allons parler, est : l'histoire, tu es prêt ?

Partie 1

Allons-y ! Pour cette leçon, je vais te parler du Moyen Âge. Le Moyen Âge est marqué par une période d'insécurité : les guerres étaient très nombreuses et les invasions faisaient des ravages. La population terrorisée se dirige alors vers de riches seigneurs. La société féodale repose sur trois classes qui dépendent les unes des autres. Ces classes sont composées des seigneurs et chevaliers, du clergé et des paysans. Parlons du premier ordre, les seigneurs et les chevaliers, ce sont ceux qui font la guerre. Le seigneur possède un château fort et des terres sur lesquelles il exerce son pouvoir. Il est riche et occupe son temps à chasser, à participer à des tournois ou à faire la guerre à d'autres seigneurs pour agrandir son territoire. En temps de guerre, le seigneur protège les paysans qu'il accueille dans l'enceinte de son château. Le seigneur s'entoure également de fidèles chevaliers qui l'aident

à se défendre. Pour devenir chevalier, il faut passer par l'adoubement et l'hommage. Lors de l'adoubement, les guerriers deviennent chevaliers et reçoivent une épée et le droit de combattre à cheval. L'hommage est la cérémonie où le chevalier jure fidélité au seigneur et devient son vassal. En échange, le seigneur offre un fief à son vassal.

Question

Tu devrais être prêt pour répondre à ma première question. C'est parti !

1) Quels rôles ont les chevaliers et seigneurs dans la société féodale ?

Trigger

a) Quelles cérémonies les relient ?

Partie 2

Une autre classe de la société est le clergé, ce sont ceux qui prient. Au Moyen Âge, la population montre une grande ferveur religieuse. C'est durant cette période que de nombreuses cathédrales ont été construites et que les croisades furent lancées. Et enfin, nous avons les paysans, ce sont ceux qui travaillent. Les paysans doivent payer le prix de la sécurité que leur apporte le seigneur. Ils mènent une vie dure et misérable. En échange du droit de cultiver les terres du seigneur, les paysans doivent également lui reverser une grande partie de ce qu'ils produisent sous diverses formes. Ils doivent donner au seigneur des produits agricoles, des tissus et de l'argent. Ils doivent aussi effectuer des corvées, il s'agit de travaux effectués gratuitement pour le seigneur.

Question

Ça y est, maintenant tu devrais pouvoir répondre à ma dernière question ! On y va ?

2) Outre les chevaliers et seigneurs, quels sont les deux derniers ordres et quelle place occupent-ils dans la société ?

Trigger

a) Comment les paysans achètent-ils la protection du seigneur ?

Fin

Trop chouette, merci d'avoir répondu !

Anglais 1: Symbols of the United States.

Now, let's talk in English, are you ready?

Partie 1

I am going to tell you about some symbols of the United States. The American flag has a name: The Star-Spangled Banner. It is composed of 50 stars and 13 horizontal red and white stripes. The American flag was adopted in 1777, less than a year after the War of Independence and the drafting of the Constitution, formalizing the creation of the new state. At that time, there were only 13 stars, symbolizing the 13 states already conquered on the east coast. The stars were added gradually as new states were formed until 1959 with the integration of Hawaii as the fiftieth state of the Union (hence the 50 stars today). The 13 stripes still symbolize the 13 secessionist states that led the War of Independence against the United Kingdom.

Question

Are you ready for the first question ?

1) Can you describe me the American flag?

Trigger

a) What does represent the 50 stars? and the 13 stripes?

Partie 2

Great, thank you for responding. Another strong symbol of the United States is the Statue of Liberty! Famous worldwide, the Statue of Liberty welcomes travelers who have crossed the Atlantic to New York. It was the first thing that immigrants seeking work in the New World would see (until the 19th century). It was given to the United States by France in 1886 to celebrate the 100th anniversary of independence and the friendship between the two peoples. It is the symbol of liberty and democracy. You may have already seen that the American Eagle is also a strong symbol of

the United States! It is called "The Bald Eagle" and has been the emblem of the United States since 1882. It symbolizes sovereignty with its power, longevity, and majestic appearance.

Question

The lesson is finish, are you ready for the last question?

2) The next question is why are the Statue of Liberty and the Bald Eagle symbols of the United States?

Fin

Great! That's it for this part.

Anglais 2: Symbols of England.

Now, let's talk in English, are you ready?

Partie 1

The flag of England, also known as the flag of Saint George, is an important symbol of England's history and culture. The red cross on a white background represents order and purity, while Saint George, the patron saint of England, is associated with bravery and courage. According to legend, Saint George killed a dragon to save a princess, and this story became a symbol of the struggle against evil and injustice. During the War of Welsh Independence in the 13th century, English soldiers began using the red cross of Saint George as a symbol of their faith and loyalty to their country.

Question

Are you ready for the first question ?

1) What are the origins of the flag of England?

Trigger

a) Why Saint George is the saint patron of England?

Partie 2

Great, thank you for responding. The rose is also an important symbol of England. In the 15th century, the War of the Roses pitted the House of Lancaster, whose emblem was the red rose, against the House of York, whose emblem was the white rose. Henry VII, belonging to the House of Lancaster, ended the war and married Elizabeth of York. He then created a new dynasty, the Tudors, and an emblem: the Tudor rose, which was red with a white center. The rose became a symbol of unity and peace in England and is considered the national flower of England. It is often used as a symbol of English identity and love for the country.

Question

The lesson is finish, are you ready for the last question?

2) Why is the rose a symbol of England ?

Trigger

b) And how did it became such a symbol?

Fin

Great! That's it for this part.

B Features Details

Temporal Features:

Speech duration: It is the total time during which a person speaks, usually measured in seconds. This can indicate how much time a person is engaged in the conversation.

Speech rate: It is the number of words spoken per minute. A faster speech rate may indicate excitement or anxiety, while a slower rate may indicate reflection or uncertainty.

Articulation rate: It is the number of phonemes pronounced per second. A higher articulation rate may indicate more fluent speech.

Balance: It is the ratio between the user's speech time and the total speech time. It can indicate the level of participation of the user in the conversation.

Intensity Features:

Average intensity: It is the average volume of the user's voice, usually measured in decibels (dB). A higher volume may indicate excitement or aggressiveness, while a lower volume may indicate shyness or uncertainty.

Intensity standard deviation: It is the variability in the volume of the user's voice. Greater variability may indicate a wider range of expressed emotions.

Vocalized fraction: It is the percentage of total speech time during which the user produces vocalized sounds. A higher percentage may indicate greater participation in the conversation.

Pitch Features:

Average pitch: It is the average fundamental frequency of the user's voice, usually measured in hertz (Hz). A higher frequency may indicate excitement or stress, while a lower frequency may indicate calmness or relaxation.

Maximum, minimum, and standard deviation of pitch: These are respectively the highest, lowest, and variability of the fundamental frequency of the user's voice. These features can indicate the range of emotions expressed by the user.

Average absolute pitch slope: It is the average rate of change of the fundamental frequency of the user's voice. A higher rate of change may indicate greater emotional variability.

Voice Quality Features:

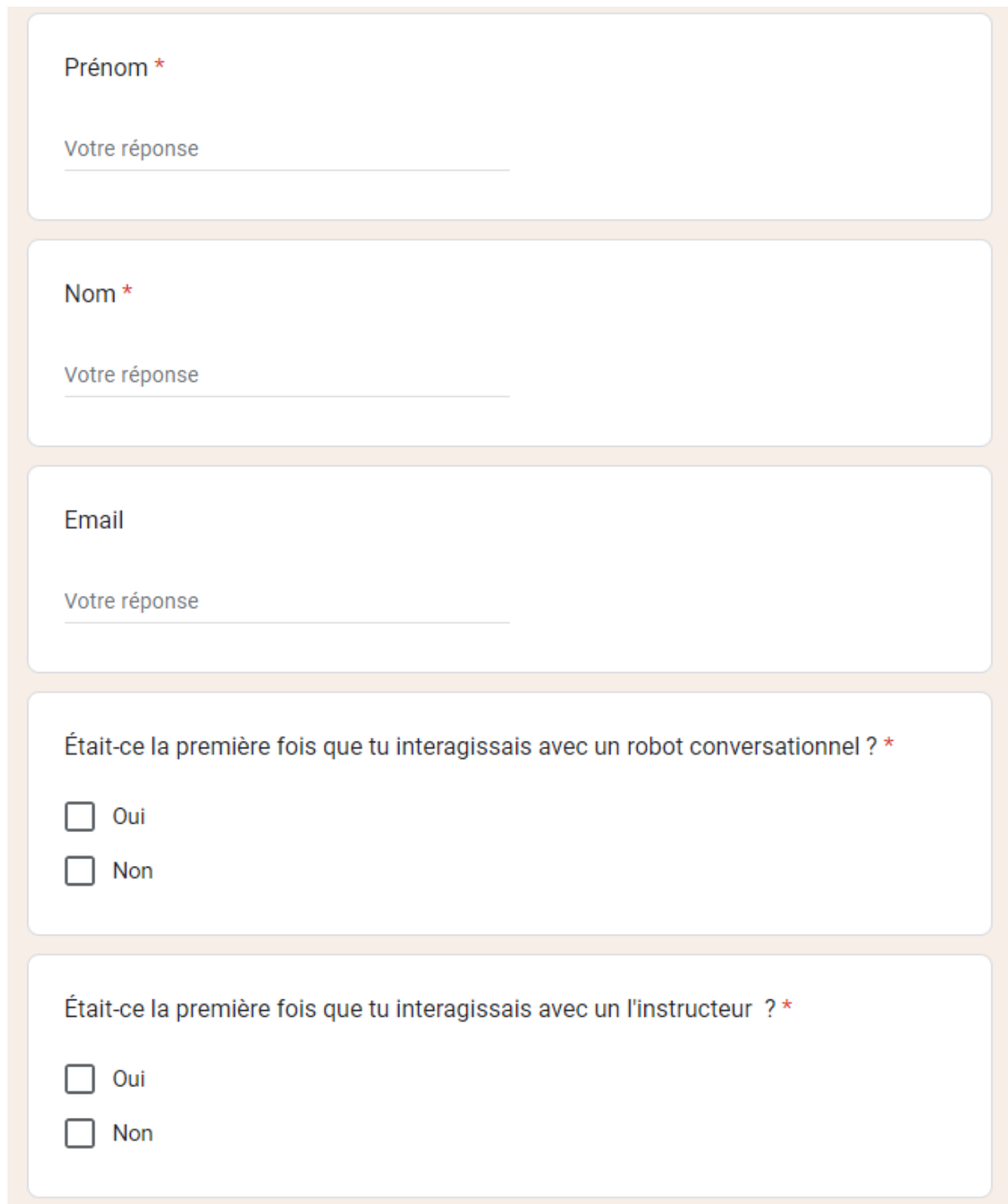
Harmonic-to-noise ratio (HNR): It is the ratio between the energy of harmonic components (voice) and the energy of non-harmonic components (noise) in the user's voice. A higher HNR may indicate a clearer and more stable voice.

HNR standard deviation: It is the variability of the HNR. Greater variability may indicate greater variability in voice clarity and stability.

Jitter: It is a measure of the variability of the fundamental frequency of the user's voice. Higher jitter may indicate a less stable voice and may be associated with anxiety or stress.

Shimmer: It is a measure of the variability of the amplitude of the user's voice. Higher shimmer may also indicate a less stable voice and may be associated with anxiety or stress.

C Post-experiment questionnaire



The form consists of five vertically stacked sections, each with a light beige background and rounded corners. The first three sections are for text input, while the last two are for multiple-choice questions.

Prénom *

Votre réponse

Nom *

Votre réponse

Email

Votre réponse

Était-ce la première fois que tu interagissais avec un robot conversationnel ? *

☐ Oui

☐ Non

Était-ce la première fois que tu interagissais avec un l'instructeur ? *

☐ Oui

☐ Non

Figure 4: Post experiment questionnaire

Te sentais-tu à l'aise pour répondre aux questions du robot ? *

0 1 2 3 4 5

Pas du tout ☐ ☐ ☐ ☐ ☐ ☐ Très à l'aise

Te sentais-tu à l'aise pour répondre aux questions de l'instructeur ? *

0 1 2 3 4 5

Pas du tout ☐ ☐ ☐ ☐ ☐ ☐ Très à l'aise

As-tu trouvé que le robot parlait de manière naturelle ? *

0 1 2 3 4 5

Pas du tout ☐ ☐ ☐ ☐ ☐ ☐ Oui, très

As-tu trouvé que le robot parlait de manière compréhensible ?

0 1 2 3 4 5

Pas du tout ☐ ☐ ☐ ☐ ☐ ☐ Oui, très

Figure 5: Post experiment questionnaire

Estimes-tu avoir agi différemment envers le robot par rapport à l'instructeur ? *

☐ Oui

☐ Non

Si oui, pourquoi ?

Votre réponse

Dans quelle mesure le fait que la voix du robot soit différente en anglais et en français t'a perturbé ? *

0 1 2 3 4 5

Je n'ai pas été perturbé ☐ ☐ ☐ ☐ ☐ ☐ Très perturbant

Comment évaluerais-tu la difficulté des questions posées ? *

0 1 2 3 4 5

Très facile ☐ ☐ ☐ ☐ ☐ ☐ Très difficile

Figure 6: Post experiment questionnaire

Ordre de préférence : classe les sujets que tu as préféré de 1 (le meilleur) à 3 (le moins bien). *

	1	2	3
Histoire	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Science	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anglais	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Pour quelle raisons as-tu classé les sujets dans cet ordre ? *

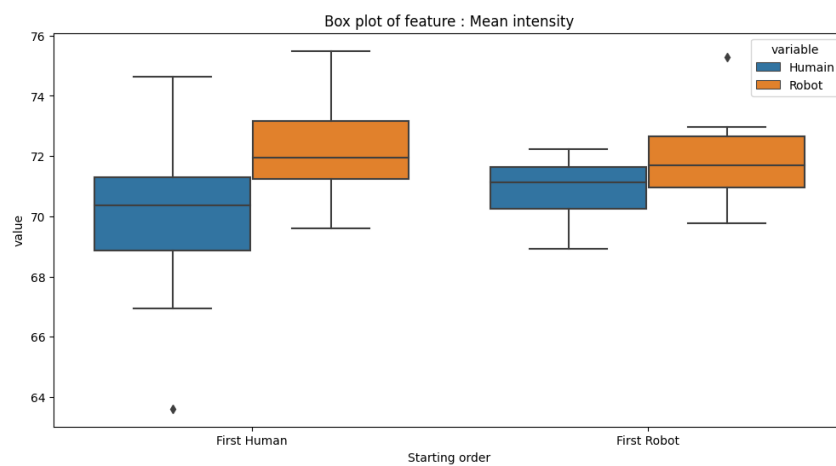
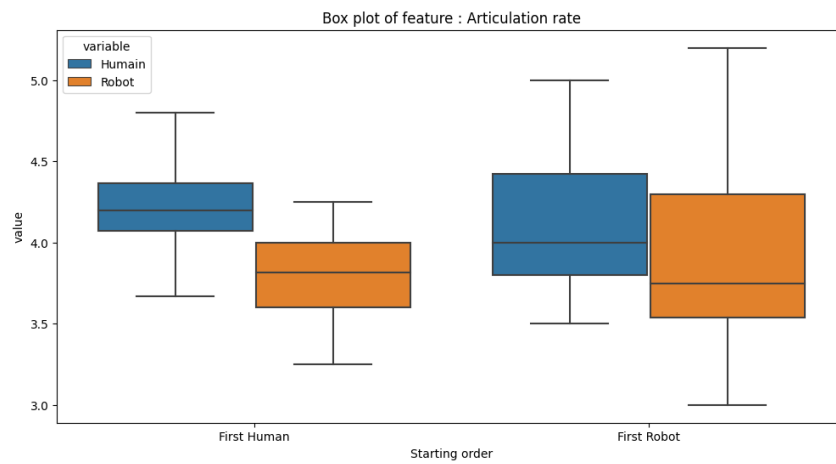
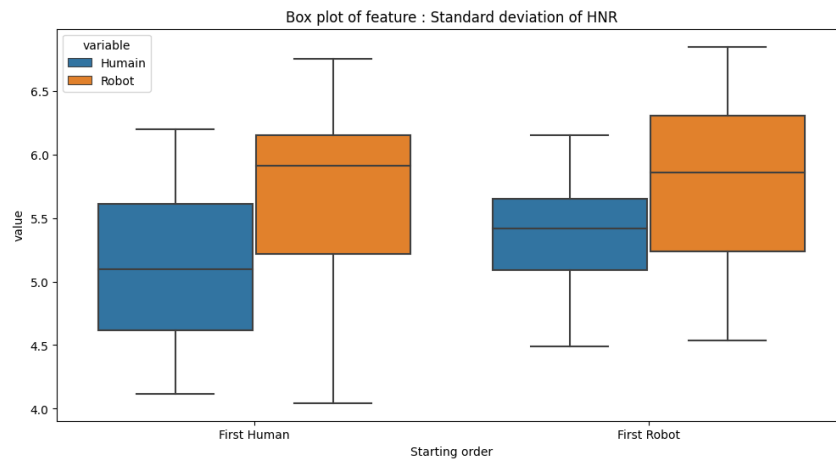
Votre réponse

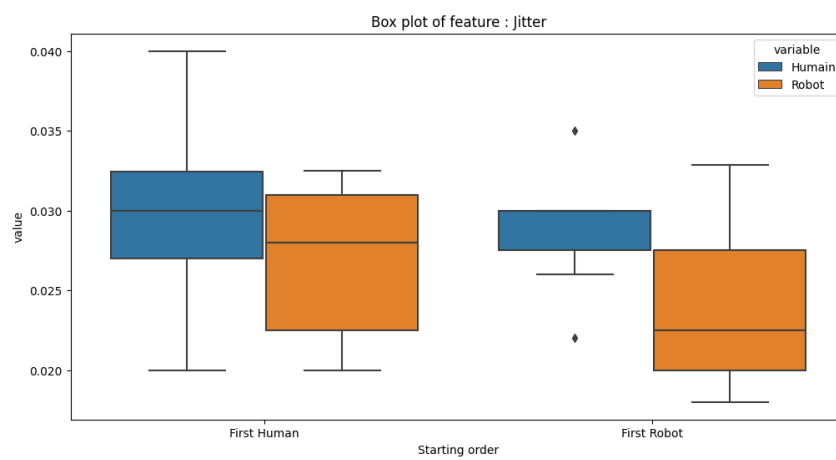
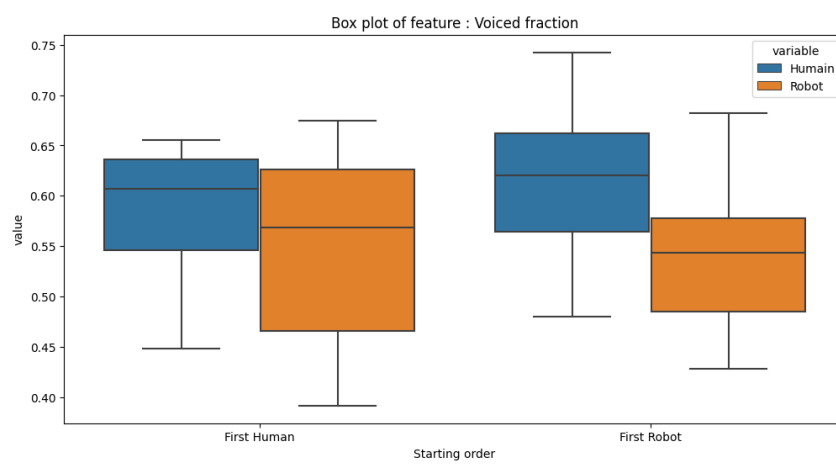
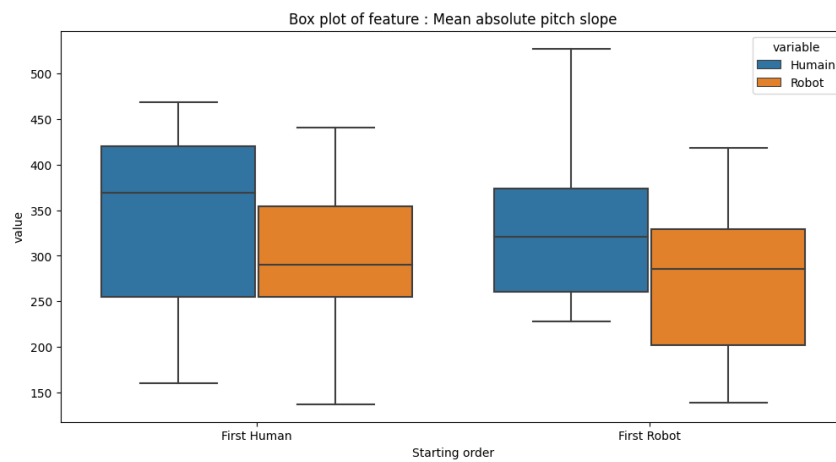
Voudrais-tu ajouter un commentaire ?

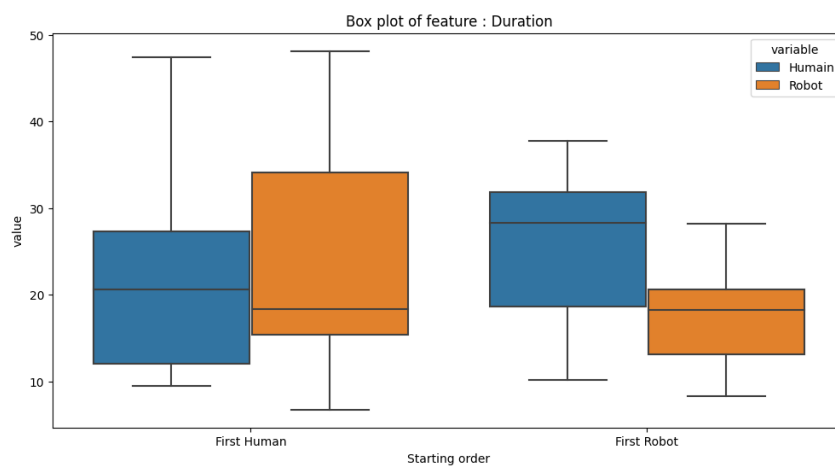
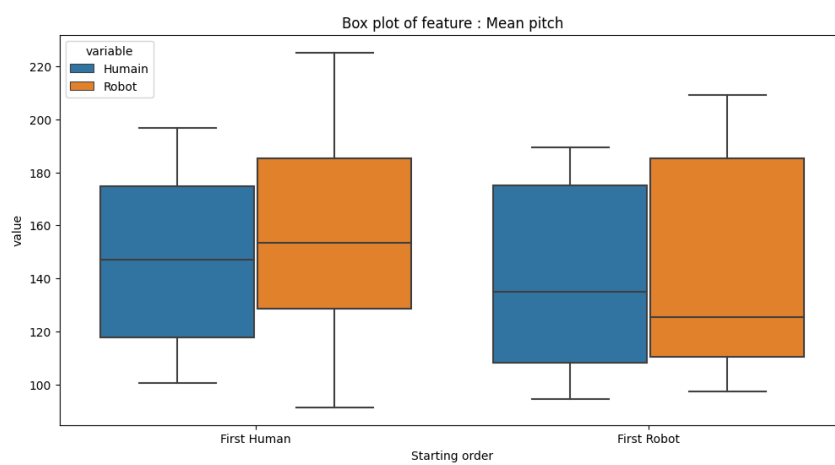
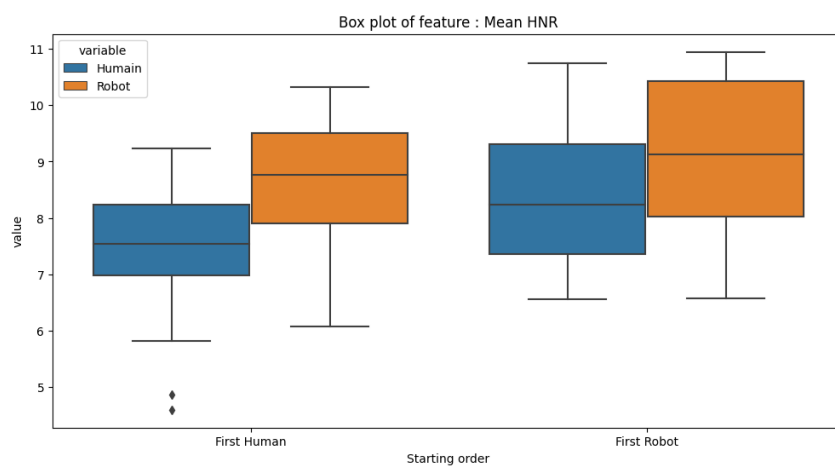
Votre réponse

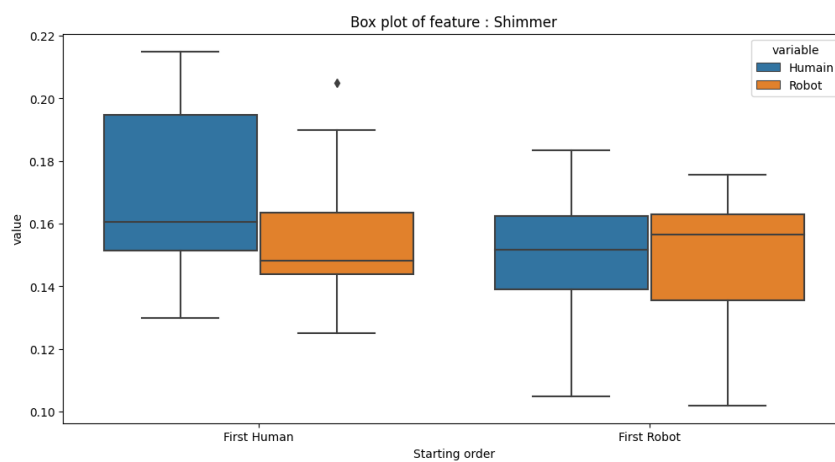
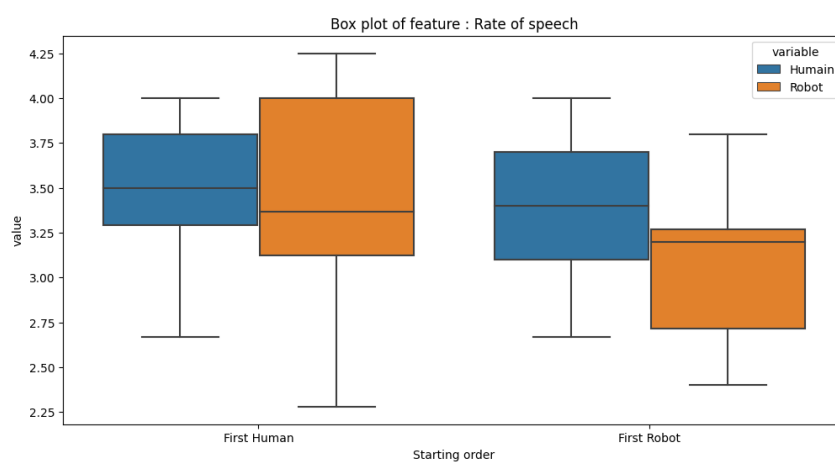
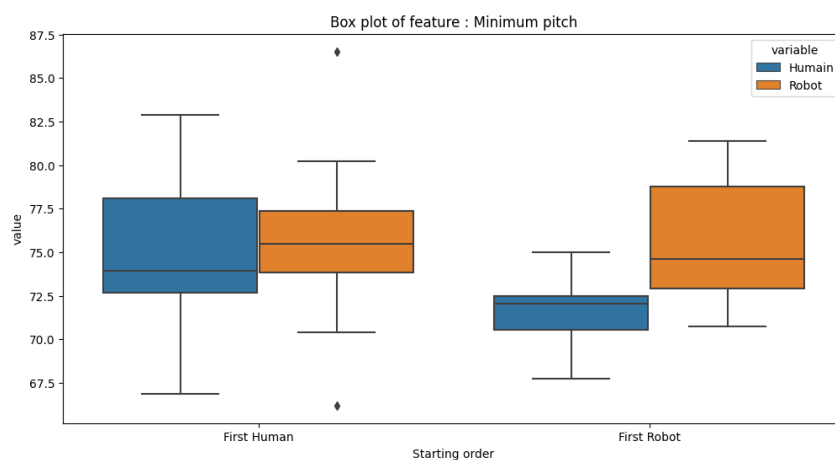
Figure 7: Post experiment questionnaire

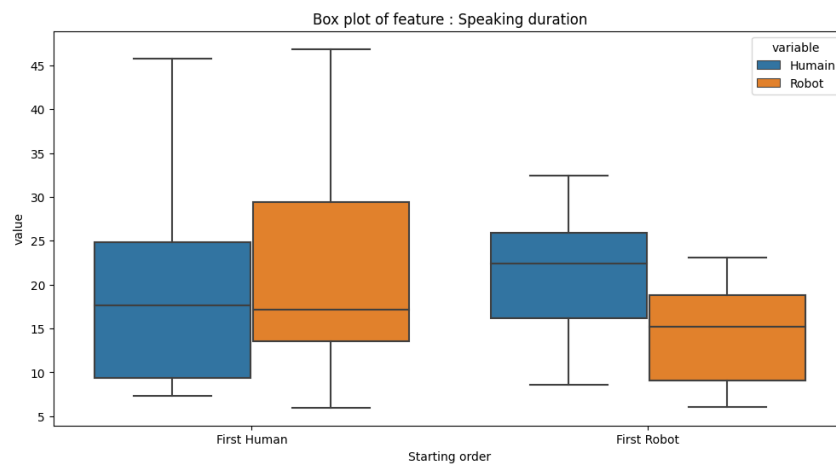
D Box Plots of Significant Features by Starting Order and Agent



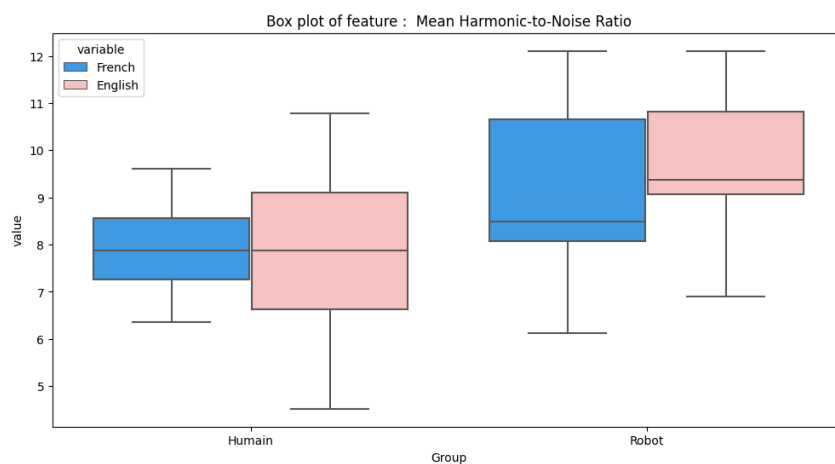
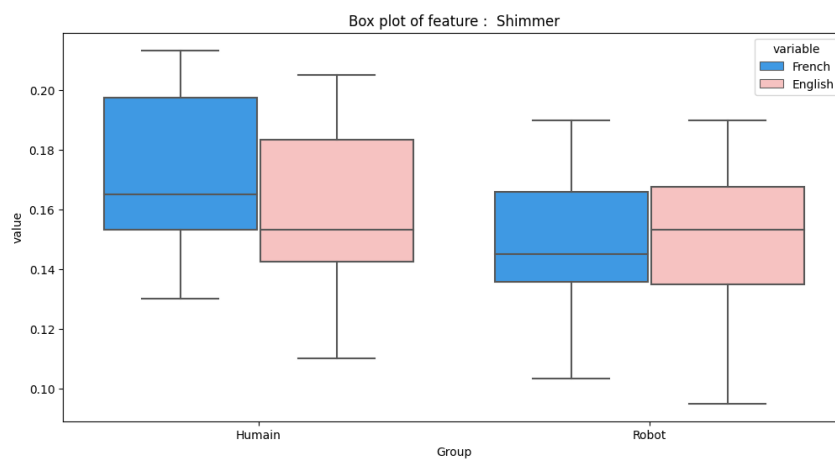
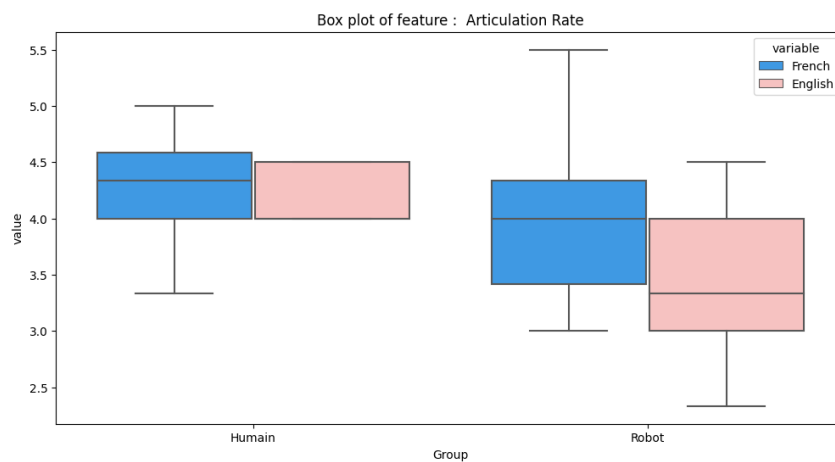


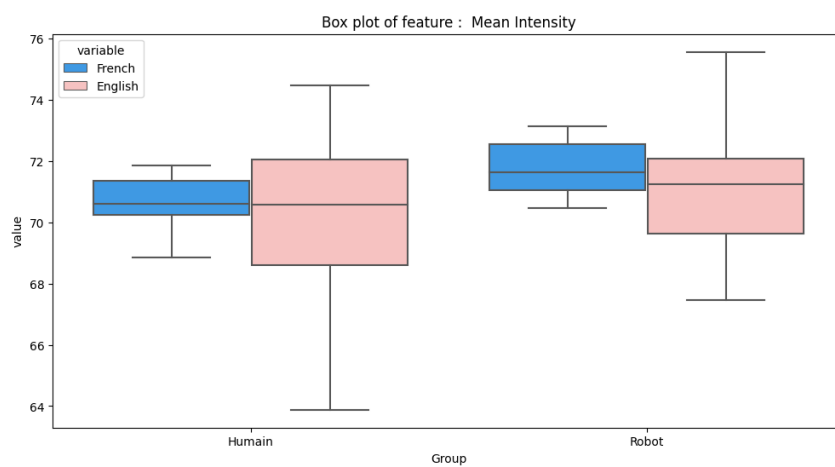
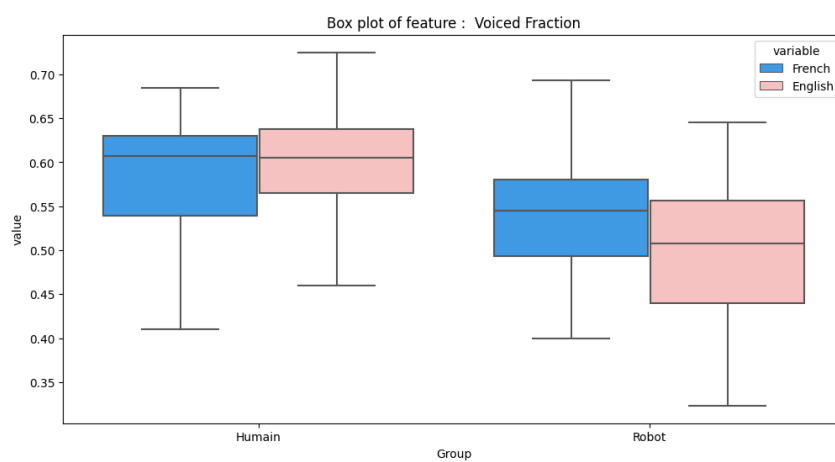
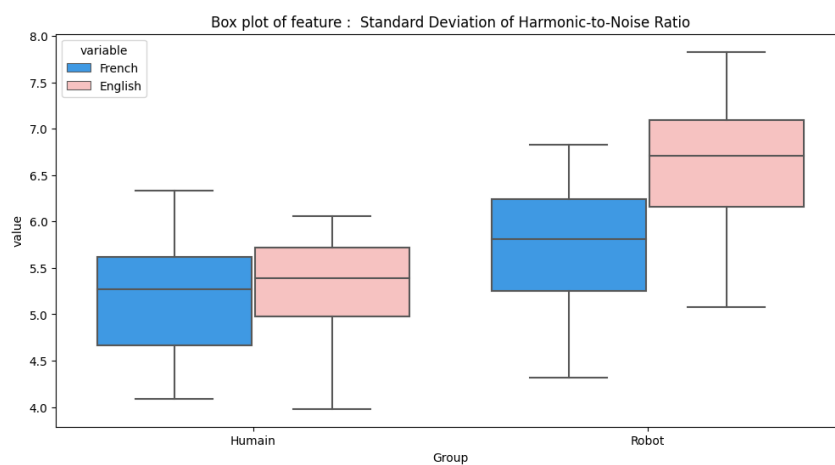






E Box Plots of Significant Features by Language and Agent





F Box Plots of Significant Different by Agent, Isolating Other Effects

