

An Exploratory Study to Determine the Effects Conversational Repetition Has on Perceived Workload and User Experience Quality in an Online Human-Robot Interaction

Curtis Gittens

The University of the West Indies Cave Hill Campus
curtis.gittens@cavehill.uwi.edu

Damian Garnes

The University of the West Indies Cave Hill Campus
damian.garnes@mycavehill.uwi.edu

Abstract

Human-robot interaction studies in the Caribbean currently face two challenges. First, the robots used in these studies have difficulty understanding many of the regional accents spoken by study participants. Secondly, the global pandemic has made in-person HRI studies in the Caribbean more challenging due to the physical and social distancing mandates. This paper reports on our exploratory study to determine what kind of impact these two challenges have on HRI by evaluating the effect conversational repetition has on a human-robot conversation done using video conferencing software. Using network analysis, the results obtained suggest that conversational repetition has several subtle relationships on perceived workload. One interesting finding is that frustration and effort are indirectly affected by conversational repetition. Results from the short User Experience Questionnaire indicate that the overall quality of the user experience is perceived as positive-neutral. This encouraging result indicates that video conferencing may be a suitable interaction modality for HRI studies in the Caribbean.

1. Introduction

The Caribbean is made up of more than 30 territories, which includes a combination of countries and dependencies. There are at least 44 million people in the region who speak English, Spanish, Chinese, Dutch, or French, as well as creoles such as Papiamentu and Caribbean Hindustani. In addition to these official languages, there are multiple accents and dialects spoken by the people in each Caribbean country.

Except for the data available through the World Bank and other United Nations entities, there is little formal documented use of technology – including social or service robots in the region. This does not imply technology has not been adopted; but it does indicate that there are few formal studies of technology adoption. Since there is a dearth of information on technology adoption, and the territories that make up the region are

largely middle-income economies with a broadly educated populace and high cell-phone network penetration, the Caribbean is an attractive place to undertake Human-Robot Interaction (HRI) research [1].

Unfortunately, the COVID-19 pandemic has restricted in-person HRI studies, and this has pushed researchers to adjust to a new normal [2]. In the Caribbean, undertaking HRI studies is even more challenging because the social robots used in our studies are made in Asia and their speech recognition systems are not trained on Caribbean dialects. As a result, there are always problems with such robots understanding users who speak with Caribbean accents.

1.1. The Barbadian Dialect

Linguists have concluded that everyone, regardless of social, geopolitical, or economic status, speaks a dialect [3]. For scholars, the notion of a dialect is key; it is specified as "any variation of a language used by a community of speakers." This means that anyone who speaks a language, speaks a dialect.

The Barbadian dialect, commonly referred to as Bajan, has an ambiguous origin. It is a Caribbean creole language that can be discerned in American English and Gullah — a language found in the Carolinas. It can also be detected in Guyanese and Belizean creole languages.

Bajan is mostly an oral language that lacks a regular written form. Consequently, the spelling of Bajan terms varies greatly between authors who write in the dialect. Bajan dialect words are generally spelled as they are spoken; for example, *dem* (them) or *wunna* (you all/your). The interdental *th* in words like *thing* and *them* is pronounced similarly to other Caribbean creoles in which the *th* is compressed into a single consonant, thus *thing* becomes *ting* and *them* becomes *dem*.

Given its close relationship to American and Mainstream English, the Bajan dialect is relatively easy to understand by non-speakers. Additionally, voice-assistants such as Alexa and Siri are used by Bajan speakers with no complaints. However, real-time captioning systems like those used in YouTube videos

and Google Meet sessions, have been observed to inaccurately translate words spoken in the Bajan dialect.

1.2. Motivation for the Study

Conducting HRI research in such a language- and culturally diverse environment, with the added complications of the pandemic, has also meant that HRI user studies must be conducted online. While there is general agreement that in-person (co-present) user studies are preferred to online (tele-present) studies [4], there is also consensus that not all in-person user studies yield better results [5]. Although online HRI studies are a viable alternative to in-person studies, they may cause higher levels of frustration in participants because of their accent or dialect and the inability of the robot to fully understand either [6]. The purpose of this exploratory study is to investigate whether conversational repetition (due to the robot's inability to understand the accent/dialect) affects either the perceived workload, or the quality of the experience, associated with an online human-robot conversation.

2. Related Work

In this section, we will investigate the work done in conversational repetition in four areas: (i) human-human conversational repetition; (ii) robot command repetition; (iii) human-robot conversational repetition and (iv) human-robot lexical entrainment. We will also discuss studies that have been undertaken to evaluate online HRI studies.

2.1. Human-Human Conversational Repetition

Repetition in conversation is a common phenomenon in human-human conversation. Studies have been done to create a taxonomy of repetition types of same-speaker and second-speaker repetition [7]. It has been proposed as a building block for discourse because it acts as form or poetry and provides a level of coherence and relationship building to the conversation [8], [9]. Tannen also indicated that repetition is a form of spontaneous pre-patterning in conversations that serves to help increase camaraderie and a sense of self [10]. More practically, repetition forms the basis for learners of second languages as a communicative, cognitive, and scaffolding function [11], [12]. It can also be used to detect Alzheimer disease by discerning changes in the words and topics repeated by the speaker [13]. Finally, repetition is used as a correction in conversations with a native and non-native speaker. The correction repetition varies in context as well as the nature of the asymmetrical relationship [14].

2.2. Robot Command Repetition

The study of repetition in human-robot interactions focuses on different aspects of repetition from the work done in human-human conversation. In the area of robot command design for instance, work focused on stopping the robot from repeating the execution of commands [15]; or the robot itself has repetition commands that it uses to help it learn [16], [17]. Repetition in commands is also sometimes used as a grounding exercise when natural language commands are given [18].

2.3. Human-Robot Conversational Repetition

Research in human-robot conversations also focus on repetition as part of the conversation's structure. Work has been done that also focus on repetition within paralinguistic contexts like repeating gestures and non-verbal sounds [19], [20], repetition as part of the conversational structure [21], [22] and repetition to help robots detect uncertainty expressed by learners of a second language [23]. Conversational repetition has also been used to help robots determine the task they are to perform [24], [25].

2.4. Human-Robot Lexical Entrainment

Lexical entrainment as a phenomenon is a form of repetition that occurs when one party in a conversation adopts (repeats) the terminology used by their interlocutor. This is seen as a manifestation of the cooperative principle where participants develop a "conceptual pact" [26]. Lexical entrainment has also been observed in human-robot interactions, with the entrainment persisting after the interaction.[27], [28]

2.5. Evaluating Online HRI User Studies

Because HRI user studies examine how robots engage with humans, some researchers include an examination of the mode of interaction; however, the majority do not. This is because the HRI study incorporates all modes of engagement, whether they are tele-operative, in-person physical/tactile, in-person verbal, or issued online as verbal or gesticulated commands. The pandemic has made the question of interaction modality – and its quality – a fundamental concern for user study experts. Researchers conducting HRI user studies who planned to use a face-to-face modality had to switch to an online mode [29]. The effect of this change in interaction mode on the study's outcome could not be determined. This has rekindled interest in the effects of non-physical modalities on user studies, with academics intending to incorporate virtual

reality and online HRI interactions into future research [6], [30].

HRI user studies on interaction modalities explored whether online (virtual or tele-present) versus in-person (co-present) interactions influenced a participant's emotion, behavior, attitude, or perception [4].

According to Li et al. meta-analysis, co-present (in-person/physical) robot interactions are superior to tele-present (online/video) robot and virtual agent interactions [4]. The truth, however, is more complicated. As shown in the Li et al. study and later work, there are instances when tele-present interactions outperform co-present interactions, and in some instances, cross-effects are detected [31], [32].

2.6. Summary

Having to repeat oneself in a conversation with someone who is either hard of hearing, or whose who is a non-native speaker of the language is a common occurrence. The same occurs when interacting with robots.

A study by Mousalli and Cardoso investigated non-native speakers and their interaction with the Amazon Alexa Echo Dot [33]. This study did not investigate the effect of repetition, nor did it use an embodied robot. Another by Irfan et al. undertook an in-the-wild experiment using Pepper as a barista robot. While they observed that participants had to repeat their order. The effect of this repetition was not captured or analyzed [34].

Based on these studies we conclude that no work has been undertaken to date on the effects repetition has participant perception of the robot, or any negative attitudes or other outcomes associated with conversational repetition. Additionally, the quality of the user experience of interacting with a robot using video conferencing software must be examined.

We propose an exploratory study to investigate these two factors and believe this work will be applicable to other cultural contexts.

3. Research Questions and Hypotheses

We will answer the following two research questions:

- RQ1: Does repetition influence the perceived workload experienced by a user who must frequently repeat statements in their conversation with a robot?
- RQ2: What is the quality of the interaction a user experiences when having a conversation with a robot using video conferencing software.

Based on these research questions, we will investigate the following two hypotheses:

- H1₀: Conversational repetition effects perceived workload.
- H2₀: The quality of the user experience is positive.

The first hypothesis will be investigated using network analysis. We anticipate that, based on prior observations, that some effect will be detected. Since we are unsure of where the effect will manifest itself, we will use network analysis to identify potential relationships.

The second hypothesis will be investigated using a validated experience survey tool that can provide sufficient insight into how the participant perceived the experience.

4. The Method

4.1. Participants

The participants were students from a Caribbean university and consisted of 38 participants—26 males ($\mu=22.65$, $\sigma=3.97$) and 12 females ($\mu=21.6$, $\sigma=3.82$). A prerequisite for being a participant was that he or she must be taking at least one course in the Information Technology or Computer Science program. The number of participants represented approximately 8.5% of the total number of students registered in Computer Science and Information Technology programs at the university. Eighty-five percent of participants never interacted with a social robot, with 55% never hearing of, nor interacting with, a social robot before participating in the study.

The participants were randomly placed in one of two groups. The first group interacted with the robot that spoke using words predominantly taken from the Bajan dialect. The second group interacted with the same robot, however, in this group, the robot spoke using only words from Mainstream English.

4.2. The Experiment Conditions

We used the Zoom® video conferencing platform to conduct the experiment online. The online session was configured for one researcher to have two simultaneous Zoom sessions, to let participants see the robot's face in one Zoom panel and the full body of the robot in the other. The second researcher in the session was responsible for explaining the experiment to the participants through a brief tutorial on how to interact with the robot.

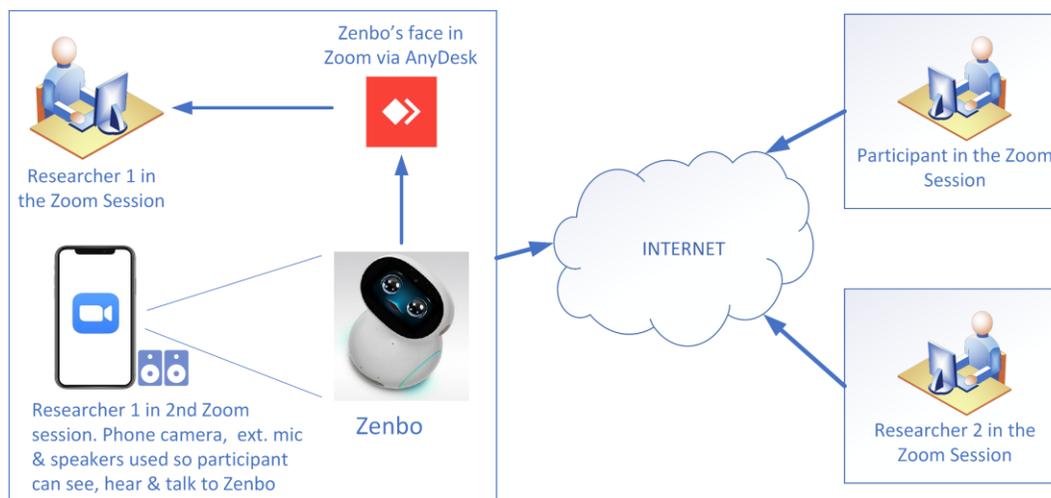


Figure 1 Online experiment set up showing how the participant could interact with the Zenbo robot

Displaying the robot’s face was achieved by installing AnyDesk® clients on both the laptop and the robot. The connection established between the laptop and the robot was then used to see the robot’s interface on the laptop – which for the Zenbo robot is its face. The screen on the laptop where the face was displayed was then shared with the Zoom session. A smartphone was used to log into a second Zoom session for the experimenter to display the entire robot in a separate panel. To allow the participant and the robot to communicate and hear each other speak, a Bluetooth external speaker with a built-in microphone was used. All sessions were recorded but the participants were asked to leave their video off, so only the voice of each participant was captured, see Figure 1.

To start the pre-scripted conversation, participants used two commands. The first was “Hey Zenbo.” This command placed the robot into listening mode awaiting further instructions. The participant then had to say “Talk to me” to start the pre-scripted conversation. The participants were instructed to repeat the statements in the conversation as needed if the robot did not respond.

4.3. The Robot

For this work the Asus Zenbo social robot, which is marketed as a companion robot, was employed. The robot can entertain children with songs and stories. It can dance and be programmed to perform other entertaining activities like interactive games [35].

The robot uses either Mainstream English or Chinese in its Speech Recognition System (SRS). However, there are no open APIs available to update the SRS so that it can recognize, or talk with, different accents. A previous pilot study indicated that the robot had some difficulty processing accents from Caribbean speakers. This could have caused greater frustration

among the users that interacted with the robot online [6]. This pilot study indicates that poor or slow robot reactions could negatively influence users' perceptions, as what is found in earlier surveys [36], [37]. Given that the intended uses for the robot is within family and close personal settings, we believe the Zenbo robot would make a good candidate for first-time users.

4.4. The Survey Instruments

The NASA Task Load Index (TLX) was used for this study because it has been validated to measure factors like Effort (E), Frustration (F) Mental Demand (MD), Physical Demand (PD), Temporary Demand (TD) and Performance (P). The overall score from this instrument measures the workload associated with the task under investigation – in this case interacting with a robot online [38].

In its simplified form, the User Experience Questionnaire [39] is a seven-point, eight-item questionnaire that lets users assess their experience with a technology. It has two subscales: Pragmatic Quality, which measures how efficient it was to perform a task using the product, and Hedonic Quality, which measures how enjoyable the product was to use. Positive experiences exceed a score of 0.8, while negative experiences fall below -0.8. Between -0.8 and 0.8 is a neutral experience.

4.5. The Conversation

The ASUS online dialogue editor was used to create the Barbadian dialect (Bajan) and Standard English conversations as a basic question/answer/follow-up exchange. For instance, posing the question, "What do you like to eat?" to the robot would elicit the pre-scripted response, "I enjoy ones and zeros," and the

follow-up question, "What do you like to eat?" (Table 1). To create the Bajan version of the Standard English conversation, we merged recommendations from a Barbadian focus group that suggested appropriate Bajan dialect responses.

4.6. The Experiment Design

For modeling interactions between many variables, network analysis is a somewhat new but effective tool. Estimation of the relationship between all variables is carried out explicitly, rather than trying to minimize the structure of the variables to their shared information, which is usually done in latent variable modeling.

We created a partial correlation network with the independent variables: Duration, which is the total session time spent talking to Zenbo in the pre-scripted conversation. "Hey Zenbo" Repetition, which was the number of times that command was repeated during the session. "Talk to Me" Repetition, which was the number of times the command was repeated to start the conversation; and Number of Moderator Interventions, which was the number of times one of the experimenters had to instruct the participant to repeat the command or provide clarification on what the robot was doing.

The Duration variable is in seconds and was obtained by calculating the difference between the start and end time for each participant's recorded session. The "Hey Zenbo" and "Talk to Me" repetition variables were calculated by reviewing each video and manually counting each time the user said the phrase after it was first uttered. The Number of Moderator Interventions variable was also manually counted. An intervention was considered as any instruction, clarification or prompting provided by Researcher 2 during the session.

Since repetition is a condition imposed by the robot, there is no control group used in this study since there is no way to control the robot's response regardless of whether the interaction is in-person or online as reported in a previous pilot study [6].

We then took these variables and combined them with the NASA TLX subscale scores as dependent variables. We used the EBICglasso estimator to eliminate spurious edges and then evaluated the strength of the network to determine the soundness of the network generated. Once the level of soundness was determined, we evaluated the connections/relationships that exist within the network to identify the independent variables that effect the NASA TLX factors.

We used an independent Welch t-test to determine if dialect affected user experience. The user experience was scored using the UEQ tool that calculates the average score of the Pragmatic and Hedonic subscales and the Overall experience score

Table 1 Part of the scripted conversation that occurs after the participant says "Hey, Zenbo". ME - Mainstream English; BD - Bajan Dialect

| User Questions / Responses | Zenbo Responses / Follow-ups |
|--|---|
| <i>Start the Conversation</i> | |
| Say: "Talk to me" | BD: "Ok, everything bless fam, what going on with you?" |
| Say: "I am good" / "I am ok" / "I am fine" | ME: "Ok, what is your name?" BD: "That easy, so wha you name?" |
| Say: "My name is..." | ME: "Ok, there are a lot of things that I like, what would you like to know about me?" BD: "Safe, dey got a lotta stuff dat I like, wha you wann know about me fam?" |
| Ask: "What food do you like?" | ME: "I eat ones and zeros. What do you like to eat?" BD: "I does eat ones and zeros, real talk. Wha you does eat?" |
| Resp: "I eat..." | |
| Ask: "Where are you from?" | ME: "I am from Taiwan, but I've been living here for a while. Where are you from?" BD: "I from Taiwan, but I did bout here for a good while now, which part you from?" |
| Resp: "I am from..." | |
| : | |
| : | |
| <i>End Conversation</i> | |
| Say: "Goodbye" / "Bye" | |

4.7. Procedure

Before beginning the experiment, participants were required to complete an online consent form as well as baseline surveys on attitudes towards, and perceptions of, social robots. Upon completion of the consent forms and surveys they were given a brief overview of the commands to issue to start the conversation and the expected response from the robot. After this brief orientation, the interactive session began. The interaction was a scripted conversation with the robot where participants asked the robot predefined questions and received scripted responses, some of which is shown in Table 1. These responses were given using mostly words from the local dialect or completely in Mainstream English – depending on the group in which the participant was randomly placed. After the

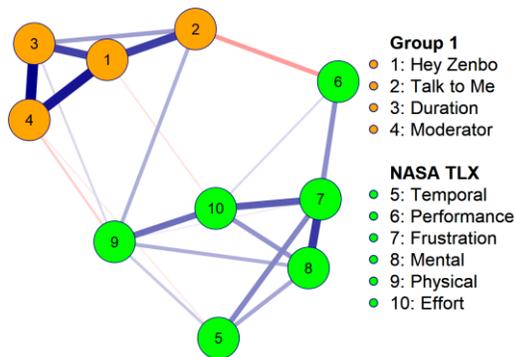


Figure 2: The partial correlation network with EBIC glasso hyperparameter $\gamma = 0.4$.

session ended, the participant completed the short User Experience Questionnaire as well as the attitude and perception surveys they completed pretest. Analysis of the attitude and perception surveys are beyond the scope of this paper.

5. Results

The network map, shown in Figure 2, was selected after evaluating the networks generated when the tuning parameter, γ was set to 0, 0.2, 0.4 and 0.5 respectively. The networks generated at $\gamma=0$ and $\gamma=0.2$ contained edges that were quickly removed as the threshold value increased. However, the edges for the networks generated at $\gamma=0.4$ and $\gamma=0.5$ were relatively consistent with small changes to the observed edge weights. We therefore selected the network generated with tuning parameter $\gamma=0.4$ since it had slightly stronger edge

weights and no edges were removed at a higher tuning parameter value.

The network map shows that Node 2, which represents the “Talk to Me” Repetition variable has a negative connection to Performance and a positive connection to Physical Demand. The variable “Hey Zenbo” Repetition has a weak negative connection to Effort, while Duration and Number of Moderator Interventions have weak positive and negative connections to Physical Demand respectively. There is also an imperceptible connection in the network of a connection between Number of Moderator Interventions and Temporal Demand. This relationship can be seen in the partial correlation (edge weights) matrix shown in Table 2.

5.1. The Centrality Plot

The centrality plot shown in Figure 3 indicates the importance of each node in the network using three measures: Betweenness, Closeness and Degree. The Betweenness graph shows that the node representing the “Talk to Me” Repetition variable has the highest Betweenness index. This means that it is frequently a part of the shortest path between other nodes in the network.

The Closeness graph indicates each node’s shortest average distance to other nodes in the network. The higher the index, the shorter the average distance. For this network, we see that the “Talk to Me” Repetition, the Overall Performance and the Frustration nodes have the three highest index values. This means that these nodes will quickly affect changes—and be affected the quickest by changes—in the network.

The Degree graph indicates the number of edges incident to each node. The “Hey Zenbo” Repetition and Frustration nodes have the degree in the network.

Table 2: The partial correlation matrix generated by the relationships between the independent and dependent variables.

| Variable | Network | | | | | | | | | |
|-------------|-----------|------------|----------|-----------|----------|-------------|-------------|--------|----------|--------|
| | Hey Zenbo | Talk to Me | Duration | Moderator | Temporal | Performance | Frustration | Mental | Physical | Effort |
| Hey Zenbo | 0 | 0.334 | 0.325 | 0.405 | 0 | 0 | 0 | 0 | 0 | -0.048 |
| Talk to Me | 0.334 | 0 | 0.168 | 0 | 0 | -0.171 | 0 | 0 | 0.139 | 0 |
| Duration | 0.325 | 0.168 | 0 | 0.436 | 0 | 0 | 0 | 0 | 0.066 | 0 |
| Moderator | 0.405 | 0 | 0.436 | 0 | -0.036 | 0 | 0 | 0 | -0.078 | 0 |
| Temporal | 0 | 0 | 0 | -0.036 | 0 | 0 | 0.211 | 0.156 | 0.095 | 0 |
| Performance | 0 | -0.171 | 0 | 0 | 0 | 0 | 0.195 | 0.16 | 0 | 0.065 |
| Frustration | 0 | 0 | 0 | 0 | 0.211 | 0.195 | 0 | 0.341 | 0.031 | 0.28 |
| Mental | 0 | 0 | 0 | 0 | 0.156 | 0.16 | 0.341 | 0 | 0.139 | 0.184 |
| Physical | 0 | 0.139 | 0.066 | -0.078 | 0.095 | 0 | 0.031 | 0.139 | 0 | 0.252 |
| Effort | -0.048 | 0 | 0 | 0 | 0 | 0.065 | 0.28 | 0.184 | 0.252 | 0 |

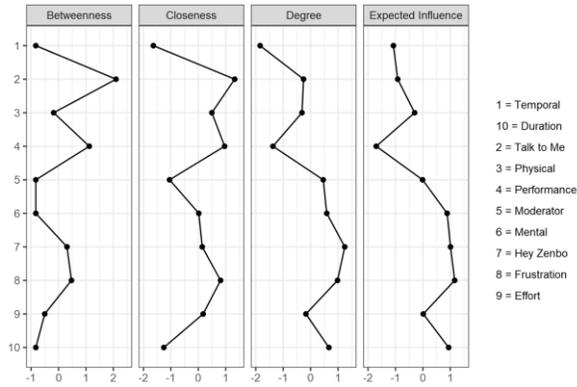


Figure 3 The network centrality plot showing that frustration has the highest expected influence followed by the repetition of “Hey Zenbo”.

Finally, the Expected Influence graph indicates which nodes have a combination of strongest connections, degree connections. This graph indicates that the *Frustration*, “Hey Zenbo” *Repetition* and *Duration* nodes have the greatest influence in the network.

5.2. Edge Stability

This graph, Figure 4, evaluates the robustness of the edges in the network by bootstrapping samples to test the correlation between subsample estimates and the original entire sample. Based on this graph, the correlation remains relatively stable above 0.75, down to 40% of the population sample. This indicates that

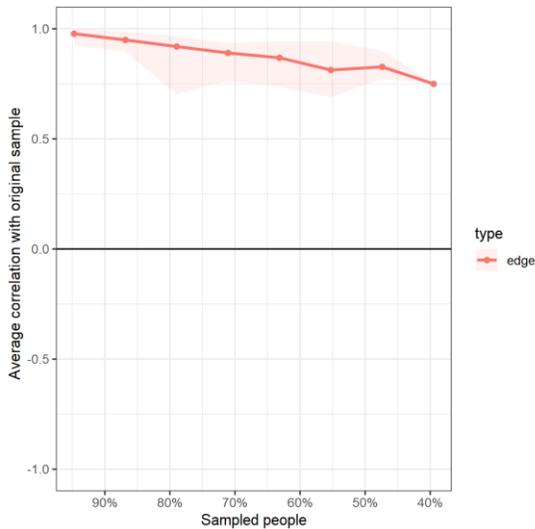


Figure 4: The edge stability graph generated from 500 bootstrapped samples, shows a correlation of over 0.75 at 40% of the population.

there is some stability in the edge values present in the network.

5.3. Centrality Stability

The centrality stability graph, Figure 5, shows that the betweenness and closeness indices are not reliable because their correlation values are below 0.7 at 40% of the population. However, the strength index, which relates to expected influence is 0.75 at 40% of the sample and indicates this index has some significance. It must be noted that the betweenness and closeness indices typically are not reliable for small samples sizes, such as what is used in this study, so this result is expected.

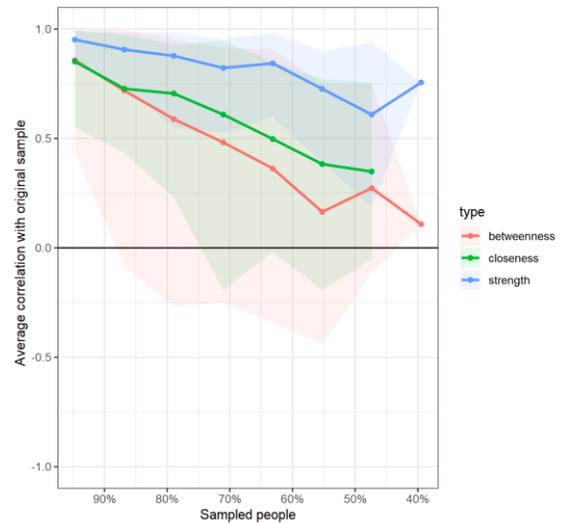


Figure 5: The centrality stability graph. Centrality strength is over 0.75 at 40% of the sample population

5.4. The User Experience Quality

Dialect had no effect on user experience quality $F_{Hedonic} (1, 37.764) = 0.447, p = .658$; $F_{Pragmatic} (1, 35.198) = -0.375, p = .71$ and $F_{Overall} (1, 37.827) = 0.074, p = .941$ and the variance was equal, so we pooled the dialect group data to evaluate the quality of the user experience. The graph in Figure 6 shows that the quality of the experience was positive-neutral.

6. Discussion

In this paper we accepted our null hypotheses—repetition affects perceived workload, and the user experience is positive-neutral in nature.

The first feature of the network is that two discernable clusters are present. This shows that the

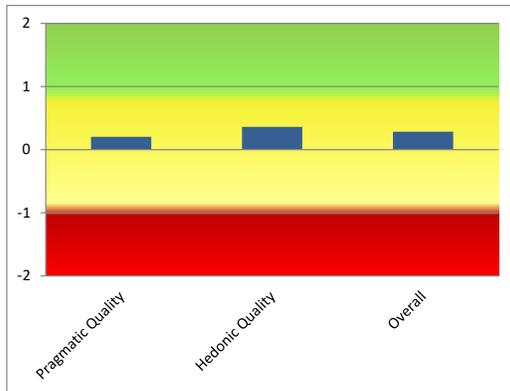


Figure 6 The UEQ score showing that the subscales and overall score are all positive-neutral

relationships within the input variables are stronger than the relationships these variables have with the NASA TLX scores; the same applies to the NASA TLX scores.

The network appears to have stable edge and centrality strength. This allows us to draw some conclusions about the relationships present in the network.

The possibility of frustration being triggered by interacting with a robot online as discussed in [6] appears to have some credence. While there is no direct relationship between the repetition variables and the Frustration score; there several mediation pathways. The strongest mediation pathway exists between the “Talk to Me” Repetition, Performance and Frustration nodes. Repetition of the command “Talk to Me” influenced Overall Performance, and Overall Performance influenced Frustration, therefore performance mediated the relationship between “Talk to Me” repetition and the level of frustration reported by the participant – nodes 2, 6 and 7. There is a longer mediation path where “Talk to Me” repetition also influences Frustration and that is where Physical Demand and Effort are mediators – nodes 2, 9, 10 and 7. This also a relatively stable pathway where all paths have a positive influence. This pathway shows that as repetition of the “Talk to Me” command increases, it increases the effect of Physical Demand, Effort and Frustration. This indirect relationship is an important finding that confirms our intuitive belief that having to repeat oneself can become frustrating. It is well to note that even though the correlation values are small, our Edge and Centrality stability measures are valid, so we can draw conclusions about these correlations since they are part of the model. The other mediation pathways that exist between the repetition variables and Frustration are weaker. The edges between the “Hey Zenbo” and Effort nodes as well as the Moderator and Physical Demand nodes are both less than 0.1. Therefore, although these

pathways exist, their influence on Frustration would not be as significant as the other mediation pathways.

We now examine the implications of some of the main direct connections in the network. The relationship between repetition of the “Talk to Me” command and Physical Demand is an interesting one. We say this because there was no physical interaction with the robot since the experiment was held online. The interesting outcome here is that there is no connection between “Talk to Me” and Mental Demand, even though the Mental Demand scores are significantly higher than the Physical Demand scores ($t(37) = 4.378, p < .001$). The existence of this relationship demonstrates how network analysis can uncover relationships that are not readily discernable using other methods. Such a relationship has no readily explainable reason to exist and would therefore merit deeper study.

Duration has a similar but positive relationship with Physical Demand as does Moderator Interaction. These relationships reflect expected outcomes: (i) the longer the session last, the more physically demanding it becomes; (ii) the more the moderator intervenes, the less physically demanding the session becomes. The second observation reflects what happened during the experiment. In sessions where the moderator intervened more, there was less doubt about what was expected of them and less doubt about how the robot is expected to respond. In sessions where the robot did not give a timely response and the moderator did not intervene, but waited for the participant to solve the problem, the participant was less sure and then asked questions or made verbal comments expressing their confusion.

The value of the network demonstrates that if we had chosen to undertake this study using a linear regression approach, these subtle relationships would be missed for two reasons: (i) lack of direct correlation between the factors under examination and (ii) low correlation values. Using network analysis helped identify valid unexpected relationships since the edge stability and centrality stability evaluation indicated that for at least the strength of the nodes (their influence) we can be confident that these relationships exist. We also have been able to see the relationships that exist within the NASA TLX tool itself. This provides additional insight on the factors that can affect the results we observe in other HRI studies that use this tool.

6.1. Limitations

The major limitation of this study is the sample size. Given that the participants were drawn from computer science and IT students from a small Caribbean university, we still believe that there is value in the study given that they account for just over eight percent of the students enrolled in the program.

7. Conclusions and Future Work

This exploratory study determined that when conversational repetition occurs, due to the speech recognition system, it directly affects Overall Performance, Physical Demand and Effort, while there are significant indirect effects on Frustration. Using network analysis, we have shown that there are subtle relationships both between our independent and subscales, as well as within the NASA TLX tool itself. We have also shown that the quality of the online experience was positive-neutral with the Overall score being 0.363 in the UEQ survey.

We intend to undertake a large study to verify and investigate the relationships that may exist between selected independent variables and the psychometric tools and scales used in HRI. This form of investigation is important and, as has been shown here, can shed light on important relationships that may be overlooked using other methods.

8. References

- [1] 'World Bank Country', World Bank Country and Lending Groups – World Bank Data Help Desk, 2021. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519> (accessed Apr. 30, 2021).
- [2] D. Feil-Seifer, K. S. Haring, S. Rossi, A. R. Wagner, and T. Williams, 'Where to Next? The Impact of COVID-19 on Human-Robot Interaction Research', *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 1, pp. 1–7, Jul. 2020, DOI: 10.1145/3405450.
- [3] W. Wolfram and N. Schilling-Estes, *American English: dialects and variation*, Third edition. Chichester, West Sussex [England]; Malden, MA: Wiley Blackwell, 2016.
- [4] J. Li, 'The benefit of being physically present: A survey of experimental works comparing co-present robots, tele-present robots and virtual agents', *International Journal of Human-Computer Studies*, vol. 77, pp. 23–37, May 2015, doi: 10.1016/j.ijhcs.2015.01.001.
- [5] S. Thellman, A. Silvervarg, A. Gulz, and T. Ziemke, 'Physical vs. Virtual Agent Embodiment and Effects on Social Interaction', in *Intelligent Virtual Agents*, Cham, 2016, pp. 412–415. doi: 10.1007/978-3-319-47665-0_44.
- [6] C. Gittens, 'Remote-HRI: A Pilot Study to Evaluate a Methodology for Performing HRI Research During the COVID-19 Pandemic', in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021, p. 1878.
- [7] N. R. Norrick, 'Functions of repetition in conversation', *Text-Interdisciplinary Journal for the Study of Discourse*, vol. 7, no. 3, pp. 245–264, 1987.
- [8] D. Tannen, 'Repetition in conversation: Toward a poetics of talk', *Language*, pp. 574–605, 1987.
- [9] S. Machi, 'Managing relationships through repetition: How repetition creates ever-shifting relationships in Japanese conversation', *Pragmatics*, vol. 29, no. 1, pp. 57–82, Mar. 2019, doi: 10.1075/rag.17021.mac.
- [10] D. Tannen, 'Repetition in conversation as spontaneous formulaicity', *Text-Interdisciplinary Journal for the Study of Discourse*, vol. 7, no. 3, pp. 215–244, 1987.
- [11] R. F. Roebuck and L. C. Wagner, 'Teaching repetition as a communicative and cognitive tool: evidence from a Spanish conversation class', *International Journal of Applied Linguistics*, vol. 14, no. 1, pp. 70–89, 2004, doi: 10.1111/j.1473-4192.2004.00054.x.
- [12] F. J. DiCamilla and M. Anton, 'Repetition in the Collaborative Discourse of L2 Learners: A Vygotskian Perspective', *The Canadian Modern Language Review*, vol. 53, no. 4, pp. 609–633, Jun. 1997, doi: 10.3138/cmlr.53.4.609.
- [13] Y. Yamada, K. Shinkawa, and K. Shimmei, 'Atypical Repetition in Daily Conversation on Different Days for Detecting Alzheimer Disease: Evaluation of Phone-Call Data From a Regular Monitoring Service', *JMIR Mental Health*, vol. 7, no. 1, p. e16790, Jan. 2020, doi: 10.2196/16790.
- [14] S. Kurhila, 'Correction in talk between native and non-native speaker', *Journal of Pragmatics*, vol. 33, no. 7, pp. 1083–1110, 2001.
- [15] T. Oka, T. Abe, K. Sugita, and M. Yokota, 'RUNA: a multimodal command language for home robot users', *Artificial Life and Robotics*, vol. 13, no. 2, pp. 455–459, 2009.
- [16] H. Elci, R. W. Longman, M. Q. Phan, J.-N. Juang, and R. Ugoletti, 'Simple learning control made practical by zero-phase filtering: Applications to robotics', *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 6, pp. 753–767, 2002.
- [17] T. W. Kim and J. Yuh, 'Task description language for underwater robots', in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No. 03CH37453)*, 2003, vol. 1, pp. 565–570.
- [18] T. Kollar, S. Tellex, D. Roy, and N. Roy, 'Grounding Verbs of Motion in Natural Language Commands to Robots', in *Experimental Robotics: The 12th International Symposium on Experimental Robotics*, O. Khatib, V. Kumar, and G. Sukhatme, Eds. Berlin, Heidelberg: Springer, 2014, pp. 31–47. doi: 10.1007/978-3-642-28572-1_3.
- [19] S. Fujie, Y. Ejiri, K. Nakajima, Y. Matsusaka, and T. Kobayashi, 'A conversation robot using head gesture recognition as para-linguistic information', in *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, Sep. 2004, pp. 159–164. doi: 10.1109/ROMAN.2004.1374748.
- [20] T. Kobayashi and S. Fujie, 'Conversational robots: An approach to conversation protocol issues that utilizes the paralinguistic information available in a robot-human setting', *Acoustical Science and Technology*, vol. 34, no. 2, pp. 64–72, 2013, doi: 10.1250/ast.34.64.
- [21] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, 'How quickly should communication robots respond?', in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, New York, NY,

- USA, Mar. 2008, pp. 153–160. doi: 10.1145/1349822.1349843.
- [22] H. Minami, H. Kawanami, M. Kanbara, and N. Hagita, ‘Chat robot coupling machine responses and social media comments for continuous conversation’, in 2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW), Jul. 2016, pp. 1–6. doi: 10.1109/ICMEW.2016.7574758.
- [23] R. Cumbal, J. Lopes, and O. Engwall, ‘Detection of Listener Uncertainty in Robot-Led Second Language Conversation Practice’, in Proceedings of the 2020 International Conference on Multimodal Interaction, New York, NY, USA, Oct. 2020, pp. 625–629. doi: 10.1145/3382507.3418873.
- [24] P. E. Rybski, J. Stolarz, K. Yoon, and M. Veloso, ‘Using dialog and human observations to dictate tasks to a learning robot assistant’, *Intel Serv Robotics*, vol. 1, no. 2, pp. 159–167, Apr. 2008, doi: 10.1007/s11370-008-0016-5.
- [25] R. Cantrell, K. Talamadupula, P. Schermerhorn, J. Benton, S. Kambhampati, and M. Scheutz, ‘Tell me when and why to do it! run-time planner model updates via natural language instruction’, in Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, New York, NY, USA, Mar. 2012, pp. 471–478. doi: 10.1145/2157689.2157840.
- [26] S. E. Brennan, ‘Lexical entrainment in spontaneous dialog’, *Proceedings of ISSD*, vol. 96, pp. 41–44, 1996.
- [27] T. Iio, M. Shiomi, K. Shinozawa, K. Shimohara, M. Miki, and N. Hagita, ‘Lexical Entrainment in Human Robot Interaction’, *Int J of Soc Robotics*, vol. 7, no. 2, pp. 253–263, Apr. 2015, doi: 10.1007/s12369-014-0255-x.
- [28] J. Brandstetter, C. Beckner, E. B. Sandoval, and C. Bartneck, ‘Persistent Lexical Entrainment in HRI’, in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, New York, NY, USA, Mar. 2017, pp. 63–72. doi: 10.1145/2909824.3020257.
- [29] L. Chan, B. J. Zhang, and N. T. Fitter, ‘Designing and Validating Expressive Cozmo Behaviors for Accurately Conveying Emotions’, in RO-MAN Virtual 2021, pp. 1037–1044.
- [30] G. Nejat, ‘Growing Old with Socially Assistive Robots: Intelligent Robots that can Help with Everyday Tasks’, presented at the IEEE RO-MAN Virtual 2021, Aug. 11, 2021.
- [31] F. Correia, S. Gomes, S. Mascarenhas, F. S. Melo, and A. Paiva, ‘The Dark Side of Embodiment-Teaming Up With Robots vs Disembodied Agents’, *Proceedings of Robotics: Science and Systems*. Corvallis, Oregon, USA, 2020.
- [32] C. Straßmann, A. Grewe, C. Kowalczyk, A. Arntz, and S. C. Eimler, ‘Moral Robots? How Uncertainty and Presence Affect Humans’ Moral Decision Making’, in *HCI International 2020 - Posters*, Cham, 2020, pp. 488–495. doi: 10.1007/978-3-030-50726-8_64.
- [33] S. Moussalli and W. Cardoso, ‘Can you understand me? Speaking robots and accented speech’, *CALL in a Climate of Change: Adapting to Turbulent Global Conditions—Short Papers from EuroCALL*, pp. 217–221, 2017.
- [34] B. Irfan, M. Hellou, A. Mazel, and T. Belpaeme, ‘Challenges of a Real-World HRI Study with Non-Native English Speakers: Can Personalisation Save the Day?’, in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, Mar. 2020, pp. 272–274. doi: 10.1145/3371382.3378278.
- [35] ‘ASUS Zenbo - Your Smart Little Companion’, Zenbo Global. <https://zenbo.asus.com/> (accessed Oct. 10, 2019).
- [36] S. Andrist, M. Ziadee, H. Boukaram, B. Mutlu, and M. Sakr, ‘Effects of Culture on the Credibility of Robot Speech: A Comparison between English and Arabic’, in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, Mar. 2015, pp. 157–164. doi: 10.1145/2696454.2696464.
- [37] B. Lugrin, E. Strole, D. Obremski, F. Schwab, and B. Lange, ‘What if it speaks like it was from the village? Effects of a Robot speaking in Regional Language Variations on Users’ Evaluations’, in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Naples, Italy, Aug. 2020, pp. 1315–1320. doi: 10.1109/RO-MAN47096.2020.9223432.
- [38] S. G. Hart, ‘NASA-task load index (NASA-TLX); 20 years later’, in *Proceedings of the human factors and ergonomics society annual meeting*, 2006, vol. 50, no. 9, pp. 904–908.
- [39] M. Schrepp, A. Hinderks, and J. Thomaschewski, ‘Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)’, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, p. 103, 2017, doi: 10.9781/ijimai.2017.09.001.