UNIVERSITY OF TWENTE.

"Talk to me, Hal": A Study of Player Experience and Interaction in a Voice Interaction VR Game Featuring AI-driven Non-player Characters

Leandro Tonini

Psychology of Conflict, Risk and Safety

University of Twente

Master Thesis

1st Supervisor: Dr. Max A. Friehs

2nd Supervisor: Prof. Dr. Sven Zebel

22.04.2024

Abstract

Current trends in industry and research highlight the potential of implementing generative AI in non-player characters (NPCs) to enable natural interaction in future digital entertainment media. This paper aims to investigate how players experience a virtual reality (VR) game with AI-driven NPCs using voice interaction (VI) and explores players' perceptions and interactions with said NPCs. In an international study, participants primarily from Canada, the Netherlands, and Germany (N = 48) played a VR game with NPCs featuring large language model (LLM) implementation and provided feedback on their experience and game conceptions through a survey and subsequent interviews, with a following sentiment analysis of the player's voice input. The results indicate that participants generally found the game to be a positive and immersive experience, with aspects such as the entertainment value, immersion, and clarity of direction being highlighted. However, issues with speech recognition were also noted. Players perceived the NPCs as realistic and mostly believable conversation partners and utilised mostly polite and friendly language, although concerns were raised regarding the sometimes-artificial flow and structure of conversation, as well as frequent occurrence of erroneous or uninformative answers. Despite the challenges, the findings suggest numerous potential benefits of implementing AI-driven NPCs in VR games using VI for industry and future studies. Further research on NPCs utilizing LLMs and VI is necessary to overcome challenges and fully leverage the benefits for games and interventions.

Keywords: Non-player characters, virtual reality, video games, generative AI, LLM, voice interaction

Introduction

The digital entertainment industry continues to grow and evolve, lead to change not only through increasing consumer demand, but also innovations in media of entertainment, technology and continuously changing creative pursuits in the artistic vision of entertainment media. To name one of the many examples, the reveal of the Unreal Engine 5 in May of 2020 (Unreal Engine, 2020) showcased unprecedented levels of detail and realism and continues to make waves especially in the video game industry through continuous releases of visually striking games built in said engine. Scenes, models, movies, and avatars crafted in the engine can seemingly offer such a high degree of realism and immersion that one can sometimes struggle to tell, what is real and what is not (Unreal Engine, 2024).

Paralleling this push of innovation towards realism and authenticity in video games is specifically the creation and implementation of intelligent and authentic non-player characters (henceforth referred to as NPCs) in video games. Just recently, on the 19th of March 2024, Ubisoft presented “NEO NPC”, a project of their R&D team of Ubisoft Paris in collaboration with Nvidia’s Audio2Face application and Inworld’s Large Language Model (LLM) to create non-player characters that are poised to break away from dialogue trees and pre-determined conversation options. They aim to emulate authentic conversations with the player (O’Brian, 2024). The NEO NPC approach enables writers to train LLMs and employ generative AI to shape characters that engage the player just as the writers imagine they would – regardless of situation or question. The goal is thus to be able to introduce NPCs into games that players can have an authentic conversation with, just like they would with a real person, thereby eclipsing the realms of possibility set by previous games such as Façade (Mateas & Stern, 2005) attempting to simulate a similar experience. Inworld specifically has already shown its LLM’s potential for more authentic NPCs in its demo of Inworld Origins,

an immersive detective sim, that garnered substantial attention by gamer audiences upon its release (Inworld AI, 2023).

Higher sophistication of NPC conversation and increased realism through visual fidelity facilitating the experienced immersion by players is already observable in recent innovative enterprises to combine the two in virtual reality (VR) environments. For example, VR social platforms such as Riff XR have recently begun to build and support online environments that feature NPCs with full AI integration, as spotlighted recently in an online HTC VIVE showcase (HTC VIVE, 2023). The integration allows players to converse seamlessly and naturally with NPCs and players alike, addressing both through the use of voice interaction.

Yet, this combined implementation also introduces the risk of facing commonly occurring problems associated with the technology in addition to possible birthing pains of new technological advances. Specifically, voice interaction (henceforth denoted as VI) inhabits a somewhat obstruse position in the video games industry (cf. Allison et al., 2018), as it rarely experiences central implementation in games and often sees little commercial success. This is often due to typically hard to eliminate challenges such as inaccurate voice recognition or inefficient controls compared to traditional, mechanical input (Allison et al., 2020) Examples of different developers implementing voice interaction include games such as Lifeline (Sony Computer Entertainment, 2003), Odama (Vivarium, 2006), Phasmophobia (Kinetic Games, 2020), or, as mentioned previously, Inworld Origins (Inworld AI, 2023). In 2017, Electronic Arts also patented an interactive voice-controlled companion application to be used in future games (Sardari et al., 2017), which has yet to be seen implemented in a game published by the company. However, studies have noted a renewed growing interests in implementing voice interaction into their games (Waqar et al., 2021). In their paper, Allison et al. (2020) outline a “tidal” pattern, wherein new release platforms and increasingly capable

technology coincide with new and innovative iterations of voice interactions in video games. Extrapolating the insights by Allison et al. (2020) onto current trends, one could argue that with the previously discussed technological advancements, the interest in the application of voice interaction in video games is currently experiencing a high tide or possibly even breaking out of the pattern.

Building on the by now well-known risks of voice interaction, an integration of AI into non-player characters may also introduce new and potentially compounding risks and unintended consequences into video games and other digital entertainment products. Documented risks and problems when interacting with generative AI can extend to speech that may be harmful, misleading, discriminatory, deceptive and/or psychologically manipulative (Fischer, 2023; Henderson et al., 2023). Infamously, Replika, a therapy AI chatbot app marketed as a digital partner to talk to and form an emotional connection with, has been spotlighted in a collection of studies as an avenue to reinforce harmful gender paradigms (Depounti, 2022), place emotional distress on its users that is suggestive of an abusive relationship (Laestadius et al., 2022), or possibly reinforce a relationship based on, essentially, manipulation (Brandtzaeg, 2022). The discussion around the app highlights the potential risks for users in terms of psychological exploitation of users or the generation of harmful content through generative AI.

Given these current developments in technology and market in addition to the many potential pitfalls, questions now arise of how players would perceive and experience a video game combining virtual reality and voice interaction and how players would interact with this new iteration of authentic digital conversation partners. However, current research and literature on the user experience of virtual reality games, principally featuring AI backed NPCs or employing verbal communication as the main tool of interaction, is sparse given its novel nature and specificity. As such, further academic ventures investigating potential

application of such mechanics require additional research to build research endeavours on a fundament of empirical findings – even more so given the previously outlined risks associated with these technologies. Realising this imperative, this paper poses the following research questions:

RQ 1: How do players experience a virtual reality game featuring artificial intelligence non-player characters using their voice as the main tool of interaction?

RQ 2: How do players perceive and interact with artificial intelligence non-player characters in a virtual reality game using their voice as the main tool of interaction?

Before further investigating these research questions, however, key topics of relevant academic literature must be further examined to adequately embed the research presented in this paper in a scientific fundament.

Literature Review

Voice Interaction

Notable breakthroughs in the complexity and dependability of natural language processing (NLP) technology over the past decade (Cambria & White, 2014) has led to speech technology being implemented into various technologically suffused areas, extending to telecommunication, consumer electronics products, automotive electronics, and many more (Ning et al., 2019). Here, studies have indicated that voice interaction holds significant potential for enhancing machine interfacing and improving workspace convenience (Li et al., 2023). Voice interaction generally serves as a valuable tool for ensuring accessibility in human interfacing, particularly for individuals with disabilities or impairments (Pradhan et al., 2018). Further, Kowalski et al. (2019) concluded in their pilot study, examining the potential

of voice assistants for older adults, that voice assistance systems in combination with IoT devices were well adjusted to the needs of many older adults through accessibility benefits, while discerning a potential in the technology to empower selective groups of older adults even further.

According to McLean and Osei-Frimpong (2019), the growing popularity of voice assistants could potentially be attributed to a combination of utilitarian, symbolic, and social benefits the technology offers. According to their findings, voice assistants may symbolize an individual's openness to embracing innovative technologies and potentially provide social benefits, including an enhanced social presence and attractiveness, beyond the pragmatic offer of hands-free and intuitive control. Voice assistants display humanlike attributes in their voice communication, further fostering rapport, engagement, and empathy among users (McLean & Osei-Frimpong, 2019; Pousneh, 2021). However, it is worth noting that voice interaction systems may also, even under the premise of potential benefits, introduce additional accessibility challenges for individuals with various disabilities impeding their use of voice (Pradhan et al., 2018) and thereby potentially offset accessibility benefits. Additionally, Porcheron et al. (2018) underline in their paper that denoting of voice user interfaces platforming voice interaction as full conversational agents or carriers should be considered a misnomer, as accessible home devices providing voice interfaces are rather embedded into conversations instead of being an active conversation partner.

Voice Interaction in Video Games

In the realm of video games, the utilization of voice interaction has demonstrated its capacity to enhance players' engagement and immersion and lead to a more profound learning experience within single player environments (Andrus & Fulda, 2020; Jung et al., 2019; Zhao et al., 2018, as referenced in Zargham, 2023). The potential of a more immersive experience

is also suggested to manifest in multiplayer scenarios (Fan et al., 2011), alongside the potential benefits of fostering connections among players in multiplayer games when utilizing voice communication, while also shielding against adverse effects that may arise from text-based interactions, such as increased uncertainty and anxiety due to reduced verbal cues (Williams et al., 2007). However, Wadley et al. (2005) observe a tendency among players to shy away from voice communication if they perceive it as cumbersome to navigate – a viewpoint that is corroborated by Allison et al. (2018), who also highlight concerns regarding speech recognition technology limitations and the perceived discomfort associated with executing voice commands, potentially deterring players from engaging in games incorporating voice interaction. In their explorative expert study on speech interaction with NPCs, Zargham et al. (2024) elaborated an extensive overview of potential advantages and challenges, as well as tips for regulation and implementation for speech interaction with NPCs in video games. Summarised, voice interaction with NPCs according to Zargham et al. (2024) can offer a natural and realistic mode of interaction and immersive experience with accessibility advantages, while however also being prone to disruptive factors such as higher cognitive demand, low recognition accuracy, and potential privacy issues.

Another notable risk that occurs when players engage with other players in a multiplayer setting and sometimes extends to player and NPC interaction is online toxicity. Toxic behaviour in gaming, describing offensive communication and/or intentional griefing, can have a strong and negative impact on players and the gaming community at large (Kwak et al., 2015; Neto et al., 2017, as cited in Liu & Agur, 2023). Competitive gaming environments are particularly susceptible to toxicity, with studies linking competitive drive to an increased likelihood of toxic behaviour (Türkay et al., 2020). Toxic behaviour can cause player disengagement and ultimately lead to financial losses for game creators (Beres et al., 2021; Kordyaka et al., 2020). Thus, addressing the underlying factors driving toxic behaviour,

such as online disinhibition, player traits, and motivators like emotional catharsis and egocentrism, is crucial for developing effective preventive measures (Suler, 2004; Beres et al., 2021; Liu & Agur, 2023).

An opportunity to combat toxicity through thoughtful game design may lie in integrating elements into the gameplay that actively promote positive behaviour. Social regulatory mechanics in voice interaction, such as a courtesy-based mechanic that checks for appropriate language and conduct when interacting with NPCs, could help nurture a more welcoming gaming environment (Zargham et al., 2023). Preliminary findings from a study featuring a game prototype with polite interaction mechanics suggest that players prefer games that prioritize courteous interactions, highlighting the potential effectiveness of incorporating such features (Zargham et al., 2023).

Non-Player Characters

As the term suggests, non-player characters (NPCs) are characters in games controlled by the computer rather than the player, often appearing in narrative-driven or single-player experiences for player interaction. The concept of NPCs predates digital games, originating from older formats like tabletop Role-Playing Games (Warpefelt, 2016). Examining the role of NPCs in video games, Warpfelt (2016) introduces a typology for NPCs, categorizing them into functions (e.g. sellers, adversaries, or friends) based on an earlier typology of NPCs by Bartles (2004, as cited in Warpfelt, 2016). In an effort to explore features constituting compelling NPC companions, Emmerich et al. (2018) provide an extensive investigation of NPCs with a focus on characters that are commonly placed in closer narrative or gameplay proximity to the player character and can thereby exert a greater influence on the player's experience and enjoyment. Notably, both authors underline the importance of believability for NPCs in order to engage players and immerse them in the

game – a sentiment that other authors such as Lim et al. (2012) previously highlighted in their papers.

Believability is tied to meeting player expectations and the NPC's role as a social agent, affecting the feeling of togetherness within the game world (Warpfelt, 2016; Gilbert, 1996, as cited in Warpfelt, 2016). Chowanda et al. (2016) also explore the impact of social dimensions in NPCs, wherein an implementation of higher social and emotional awareness through for example facial recognition and enhanced emotional state interpretation using their framework "ERiSA" (cf. Chowanda et al., 2014) resulted in enhanced emotional attachment and significant positive changes in player engagement and immersion. Emmerich et al. (2018) stress the importance of social competences for NPCs, emphasizing social relations and contextually dependent communication particularly for more present NPCs such as player companions, which can foster relationship building and increase player engagement, especially in single-player games.

Non-Player Characters and Artificial Intelligence

Artificial Intelligence in video games has experienced a remarkable improvement over the last 15 years and games shape a well-suited domain for its implementation (Fan et al., 2020). Especially in dialog, generative AI can be employed to create conversations using flexible natural language that are context aware created by artificial intelligence (Bubeck et al., 2023; Eysenbach, 2023). According to Lv (2023), this technology can enhance NPCs in games to display genuine emotions and varied responses to player actions, enhancing game immersion. Existing LLMs such as Inworld or ChatGPT by OpenAI are already able to provide such an experience (Huang & Huang, 2023; Zeng, 2023). ChatGPT, powered by generative AI based on the GPT architecture, responds to user inputs in conversations and is a neural network pre-trained on immensely vast text data (Dwivedi et al., 2023, as cited in Lv.

2023). Referring to an already existing implementation (cf. Bankhurst, 2023, as cited in Huang & Huang, 2023), Huang and Huang (2023) outline the potential use of ChatGPT to evolve NPC conversations with players into realistic and diverse speech interactions, as opposed to limited and repetitive dialogue seen in traditional NPC models. Additionally, Cox and Ooi (2024) highlight in their analysis on user reviews of a game featuring LLM implementation that players enjoyed the flexibility and open-endedness of conversations with NPCs, providing an immersive experience that is furthermore enhanced with the use of speech interaction.

Huang and Huang (2023) stress, however, potential challenges when implementing generative AI, such as data privacy and the possibility of harmful content generation through user interactions on the ethical side, and latency or performance issues on the technical side, which may detract from the overall player experience. Cabe et al. (2024) expand on this list of challenges, by highlighting specific points of difficulty for AI NPCs, listing concerns such as “being tricked into forgetting their character by the user” or “lacking character motivation or agency within the world or story they inhabit.” Cox and Ooi (2024) add in their review that players also often negatively experience an absence of conversational memory of the NPCs, coupled with instances of inappropriate responses and easily disprovable lies.

The Present Study

The present study aims to investigate player experience of virtual reality games utilising voice interaction as the central mode of communication by letting participants engage with a VR game that tasks them to complete objectives and interact with AI supported non-player characters. Participant gameplay experience is then assessed through survey data employing quantitative measures and player interviews. Additionally, this study aims to investigate player perception of and interaction with AI supported non-player characters

through quantitative sentiment analysis of verbal player utterances during gameplay, player interviews and additive survey data.

Methods

Participants

Participants were recruited in two different locations as part of an internationally joined research effort between the University of Twente (Netherlands) and the University of Waterloo (Canada). Research subjects were recruited via convenience and snowball sampling through word-of-mouth or through intranet structures of university facility-wide shared research subject pools. Participation was on a voluntary basis, with either no compensation or a small compensation in the form of university course credits. Additionally, one recruitment location offered the chance of winning a coupon for an online store in the amount of 25€. Requirements for inclusion in the study were an age above 18 years and at least B2 level English language proficiency. Sufficient language proficiency was not tested due to the high general English proficiency in the population sampled on both research sites. Exclusion criteria extended to denied consent. The final sample consisted of 48 respondents. The average age of the sample was 23.58 years (SD = 6.15), ranging from a minimum of 18 to a maximum of 52 years. Gender distribution of the sample extended to 21 male participants (43.8%), 26 female participants (54.2%),, and 1 diverse participant (2.0%). Regarding residency, 17 participants (35.4%) were from the Netherlands, 20 (41.7%) from Canada, 9 (18.8%) from Germany, and 2 (4.2%) from other countries. In terms of gaming frequency, the sample entailed 20 participants who indicated that they played less than once per week (41.7%), followed by 9 participants that indicated once per week (18.8%), 7 participants that indicated 2-3 times per week (14.6%), 5 participants that indicated every other day (10.4%), 3

participants that indicated every day (6.25%), 2 participants that indicated multiple hours a day (4.2%), and 2 participants that indicated more than 4 hours a day (4.2%).

Materials and Procedure

The collection of data proceeded from August 16, 2023, to October 6, 2023. The chosen platform for the VR study was the Meta Oculus Quest 2, while the VR game was built in Unity (2022) and employed through the Meta Quest Developer's Hub (v3.0) and SideQuest (v0.10.39). Before partaking in the study, respondents were handed a consent form that informed them about their rights as participants, the research purpose, the content of the study, possible risks, and potential compensation (see Appendix A). Ethical approval for the study was received from the University of Twente. Participants were furthermore informed that they could withdraw from the study and/or playing the game at any point in time.

After giving their consent, participants were asked to read through a briefing detailing the scenario and narrative of the VR game present in the study, as well as listing the central goals and possibilities in the game. As additional material, respondents were also provided a list of NPCs they might encounter in the game with pictures. After the briefing, participants were shown the controls of the Meta Oculus Quest 2, and how they were implemented in the game, as well as given the opportunity to ask questions if they had any. Additionally, participants were also given a short period of time to familiarise themselves with the virtual reality environment before proceeding to the game.

Upon introducing the participants to the game, they were again informed that they would have up to a maximum of 20 minutes to play the game as well as some final remarks pertaining to potential bugs and/or visual features. After introducing participants into the game, the entire gameplay session was recorded. Once participants had finished all the goals in the game and wished to stop playing, ran out of time, or wished to quit playing due to other

factors, e.g. motions sickness, the recording was stopped, and participants were guided toward exiting the game. Proceeding further, after a quick break for participants to acclimate again, respondents were directed to fill out a survey on the online survey platform Qualtrics. Here, participants were asked about their demographical data, i.e. age, gender, and residence. Following demographical data, respondents were inquired about their experience with voice interaction systems, video games and a combination of both. Continuing, participants were additionally asked to answer questions pertaining to their player experience of the game using the Player Experience Inventory (PXI), followed by questions connected to their perception of the NPCs in the game, their interaction with the player and questions concerning player performance and satisfaction. Questions regarding the featured NPCs in the game were supplemented with pictures of matching NPCs to facilitate recall. All questions featured in the survey are detailed in Appendix B.

After completion of the survey, participants were further asked to partake in a short semi-structured interview entailing 12 questions. The interview questions were drafted and developed in-house by the research team specifically for this study and aimed at capturing the participant player experience in the game with the central element of voice interaction (see Appendix B). The interview questions touch on different areas, such as aspects of the game the player may have liked or disliked and found interesting or uninteresting, as well as thematizing the voice interaction with the game's NPCs. After the interview, participants were debriefed and informed of their compensation possibilities and given room for further questions pertaining to the research topic.

Virtual Reality Game

The virtual reality game employed in this study was developed specifically for research purposes, incorporating voice interaction as the main mode of communication in a

virtual reality environment mimicking a classical office setting and featuring four non-player character archetypes in the form of an intern, a manager, a mechanic, and a designer. The designed game was called “Office Whispers”.

Most objects in the environment were programmed to be interactable, meaning that players had the option to not only verbally interact with the non-player character cast, but also physically interact with a wide array objects. Additionally, for the purpose of this study, the game featured two different versions, wherein the visually discernible sex of each character had been swapped to their binary counterpart to potentially test for differences in NPC interaction or perception based on gender and appearance. An overview of the game’s featured NPCs can be found in the Appendix C.

Using the Oculus Quest 2, participants were introduced to the virtual office space (see Figure 1), wherein players could adjust their view displayed in the headset through intuitive and natural head movement. In other words, respondents could look around as they naturally would to in turn visually look around in the virtual reality space. Alternatively, players could also use the controllers of the VR system to adjust their view vertically or horizontally through quick jumps or “snap-movements” by using a joystick. Spatial movement was also controlled through the Oculus controllers, likewise offering two modes of movement – either fluid directional movement using joysticks or quick and aimed teleporting (“blinking”) using the index trigger buttons (Figure 2). Additionally, items could be grabbed through a grabbing motion by using palm triggers or levitated in front of the player by pointing at them with the index triggers (Figure 3).

Figure 1

VR In-Game Environment

Figure 2

Moving through the Environment by Teleporting

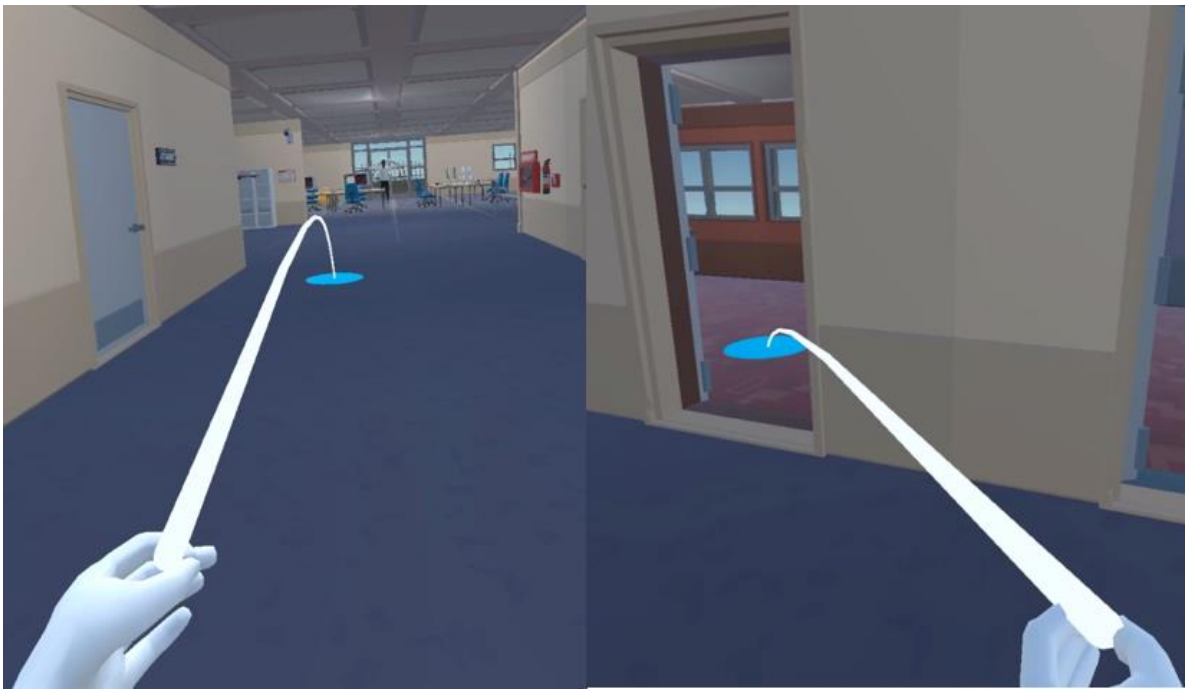


Figure 3

Interacting with Items in the Game

Further, conversations with NPCs could be initiated by simply approaching the digital character and speaking any inquiry out loud. As visible in Figure 4, audio input would be live transcribed and displayed in a chat-like textbox above the character during the conversation detailing both participant input and NPC output, meant as visual aid and documentation of

the conversation. Players were free to continue moving around or interacting with the environment during the conversation and similarly able to leave whenever they wanted. The game utilized the Inworld AI system (2024) for the featured NPCs, which are able to understand natural language, enabling seamless communication with players. Inworld employs multiple machine-learning models to create NPCs with human-like communication abilities. These NPCs can interact with players through speech using generative AI technology and can remember interactions and evolve over time. Inworld seamlessly integrates AI characters with popular game engines like Unity or Unreal, allowing developers to import them into specific scenes and set additional features such as animations and rigging systems. Each NPC in Inworld is generally defined by a 'Core Description' detailing their motivations, flaws, and actions. Developers can thus assign unique identities to NPCs, including names, pronouns, roles, and interests. During the NPC design process of the game, consistency was maintained across core character traits such as interests and hobbies for comparability. Players could engage with characters through speech interactions at any given moment, with a minimum proximity distance set, requiring players to be within close range of each character to initiate interaction. Players had the freedom to exit conversations at any time.

Figure 4

Conversations with NPCs in the Game

Measures

Player Experience Inventory

The Player Experience Inventory (PXI) describes a reliable and theoretically founded tool to measure player experience in games, based on the 'Mechanics - Dynamics - Aesthetics' (MDA) framework (Hunicke et al., 2004) and the Means-End theory (Gutman, 1982) as a theoretical foundation (Abeele et al., 2020). Player experience is measured at the level of functional or psychological, meaning a combination of immediate, practical consequences and experiences evoked by the game's design choices, and psychological and emotional experiences as a second-order response to the game design (cf. Abeele et al., 2020). Here, constructs of functional consequences extend to the ease of control, feedback on progress, the audiovisual appeal, goal clarity, and perceived challenge posed by the game.

Psycho-social constructs envelop mastery of the game, curiosity about it, immersion, personal meaning, and perceived autonomy. Lastly, enjoyment is proposed as the consequential instrumental value. In total, the 10 constructs that are measured with 3 items each, with an additional instrumental value measuring player enjoyment, also measured with 3 items. Each item offered answering options on a seven-point Likert scale ranging from “Strongly disagree” (1) to “Strongly agree” (7). The comprehensive list of constructs and connected items can be found in Appendix B

Non-Player Character Interaction & Player Performance

Additional questions pertaining to the likeability, attractiveness and interaction enjoyment between the player and NPCs as well as questions regarding the willingness to play similar games, a self-assessment of performance and the overall game experience of players were furthermore added to the survey to enrich the data obtained by the PXI. The resulting list of questions can be found in Appendix B. Generally, apart from question 2 and question 4 that were operationalised as open questions, each question prompted the participants to answer on a seven-point Likert-scale scale ranging from 1 to 7 with contextually corresponding statements such as “Not at all enjoyable” or “Extremely enjoyable” for question SQ6, or “Extremely bad” to “Extremely good” for question SQ9.

Data Analysis

Survey Data Analysis

Statistical analyses of survey data were conducted via the IBM Statistical Package for Social Sciences (SPSS Version 27) and RStudio (2024.04). Descriptive statistics (frequencies, means, and standard deviations) were used to explore demographic and survey data. Here, PXI item scores were computed into index mean scores for each construct for purposes of

further descriptive statistics such as means and standard deviations. Additionally, correlational analysis of PXI measures, additive items and demographic variables was conducted to further explore potential underlying factors and correlations.

Gameplay Analysis

Further, gameplay recordings of the VR game were screened, audio-transcribed, and prepared for sentiment analysis of participant input. For audio transcription, Amberscript (2024) was employed for primary and supplementary transcription, the results of which were further refined in accuracy and screened for errors. For purposes of focused analysis of participant input in the gameplay, transcripts were scrubbed of vocal fragments deemed unimportant for data analysis (e.g. vocalised inner monologue, questions toward the researcher and filler words or hesitation markers). Resulting transcripts of voice interactions between participants and NPCs were further filtered to separate NPC output from participant input to crystalise all participant utterances directed at NPCs over the course of their gameplay. Participant voice input was investigated via sentiment analysis through analysis tools and lexica of the Linguistic Inquiry and Word Count (LIWC) tool, as well as the RStudio packages containing AFINN and NRC lexica. The LIWC tool entails 100 built-in dictionaries created to capture people's social and psychological states, as it analyses and compares each word in a piece of text to the list of dictionary words and calculates the percentage of total words in the text that match each of the dictionary categories (*LIWC — How It Works*, n.d.; cf. Boyd & Schwartz, 2021). The AFINN Lexicon consists of 2,477 words split into 878 positive and 1,598 negative words arranged on a scale of +5 to -5, originally developed to conduct sentiment analysis of short posts and statements in social media platforms such as X, formerly known as Twitter (Nielsen, 2011, as cited in Yan, 2021; Yan, 2021). The NRC Emotion Lexicon on the other hand lists 5,636 words associated with a

body of eight emotions, namely anger, anticipation, disgust, fear, joy, surprise, sadness, and denoted as either negative or positive (Mohammad and Turney, 2013, as cited in Yan, 2021).

Post-Study Interview Analysis

Continuing, post-study interviews were transcribed for audio transcription, reiteratively employing the online transcription service Amberscript (2024) and prepared for coding. Coding of the interview responses was conducted with the use of the coding program ATLAS.ti (24.1.0). Analysis of the responses employed an inductive coding scheme, as the central aim of this study is to gather and condense key opinions, views and sentiments concerning voice interaction as the central tool of communication in the game, the interaction with NPCs and its combination with VR technology. The coding process was spearheaded by one researcher and informed by relevant literature and key insights of earlier studies (cf. Zargham et al., 2024). Following the first iteration of codes, the preliminary codebook was subject to review by co-researchers and other experts in the field. To ensure high analysis quality, recoding and reviewing were repeated until a consensus between the researchers was reached. Both latent (i.e., interpreting the commenters' intended meaning) and semantic (i.e., the participant's words, verbatim) codes were generated. Responses could be assigned multiple codes and adhered to the question structure as informed by the interview itself, denoting that codes were only coded as occurring once per answer to a question and coded as an additional occurrence if the sentiment was highlighted again by the participant in a response to a later question. This process of inductive coding was applied to yield a codebook that represents participants not only in frequency, but also in salience, present if participants deemed an aspect worth mentioning multiple times in connection to different questions.

Results

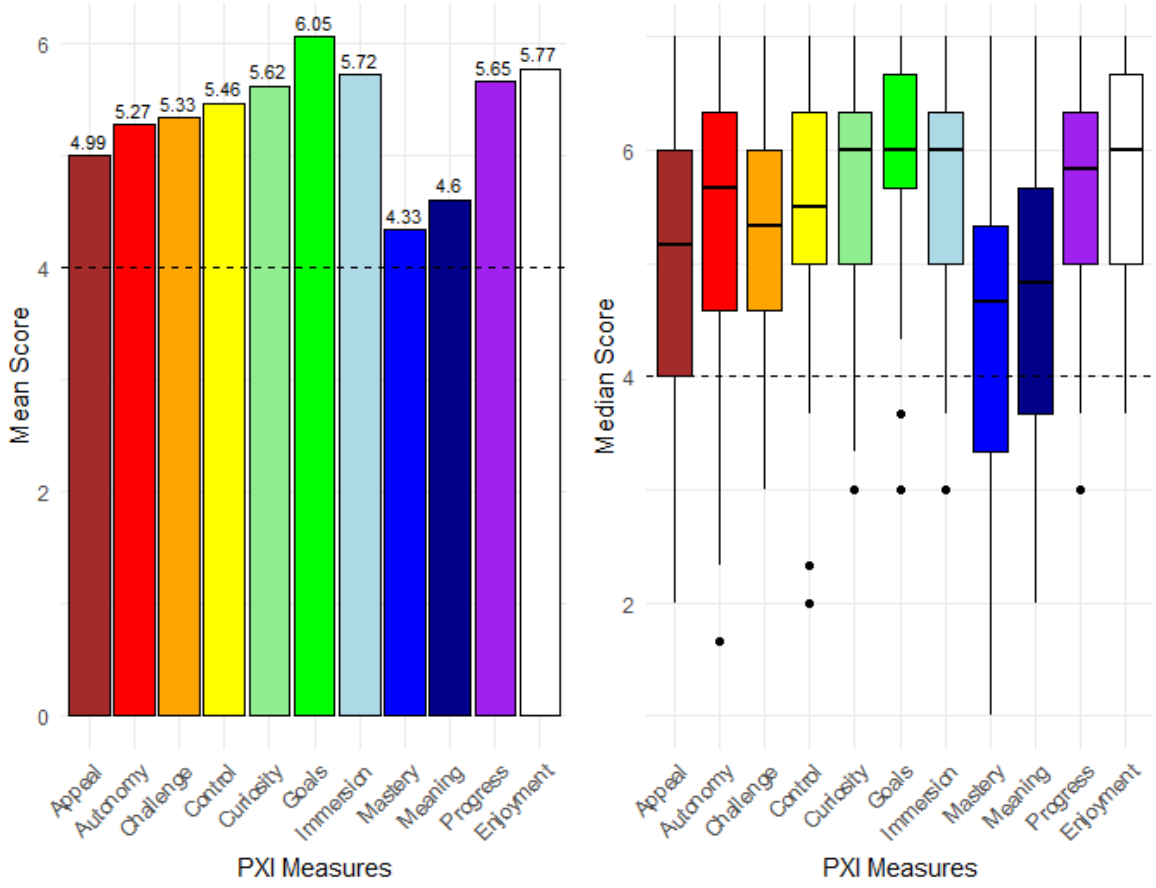
Survey Results

PXI Measures

The overall mean scores of the PXI measures were all above average (see Figure 5), with the highest score found in the "Goals" construct (M = 6.05, SD = 0.88) and the lowest score visible in the "Mastery" construct (M = 4.60, SD = 1.20). On average, participants rated the clarity and visibility of goals in the game the highest, while comparatively assessing the general personal meaningfulness of the game lower. Overall, mean scores are all above average, which accordingly indicates that player experience as captured by the PXI was a generally more positive one.

Figure 5

PXI Mean and Median Scores



PXI Correlations

Demographics and PXI Measures. To further investigate player experience, explorative data analysis in form of correlational tests was conducted, using an alpha level of .05 for all statistical tests. Correlational tests of demographic data and PXI inventory item mean scores resulted in a statistically significant negative correlation between participant age and mean scores of PXI Autonomy items with medium effect size, $r(46) = -.39$, $p < 0.01$. Similarly, mean scores of PXI Control items negatively correlated with participant age in likewise medium effect size , $r(46) = -.30$, $p = 0.04$. Said correlations indicate lower autonomy or control assessments with increased participant age, implying that participants of higher age felt less autonomous and a lower sense of control during the gameplay of the VR game featured in the study. Correlational tests of other PXI mean scores and demographical data yielded no statistically significant results; effect sizes ranged from $r = -.5$ to $r = -.003$ with $p \geq 0.1$.

PXI Measures. Following correlational tests with demographic variables, PXI item mean scores were investigated on internal correlations to enrich insights into player experience. Table 1 shows a correlation matrix between mean scores of all ten PXI key constructs. As shown, multiple construct mean scores correlated in a statistically significant manner with medium to large effect sizes. Most notably, PXI item mean scores of Meaning and Enjoyment correlate in a statistically significant manner with the largest effect size, $r(46) = .72$, $p < .001$. Further large effect sizes with statistical significance were found in correlations of Appeal and Enjoyment, $r(46) = .60$, $p < .001$; Curiosity and Enjoyment, $r(46) = .62$, $p < .001$; and Meaning and Curiosity, $r(46) = .62$, $p < .001$. Additionally, medium effect sizes with statsitical significance were found in correlations of Control and Autonomy, $r(46) = .51$, $p = .008$; as well as Appeal and Meaning, $r(46) = .51$, $p = .01$; and Control and Mastery, $r(46) = .55$, $p = .003$.

These correlations indicate that, with statistical significance, participants with higher scores in personal connection to and/or curiosity about the game, as represented by the constructs, also tended to have higher scores in measured items of the game enjoyment. Additionally, participants that reported a higher score of the overall audiovisual appeal of the game, also tended to show higher scores on the mean score of game enjoyment and personal connection, as represented by Enjoyment and Meaning, respectively. Lastly, as indicated by the correlational analysis, respondents with a higher score in mean scores of the PXI Control measure, representative of the extent to which a player finds the actions to control the game clear and intuitive, also tended to have higher scores on items related to a sense of competence, denoted as Mastery, and a sense of freedom, denoted as Autonomy, and vice versa.

Table 1

Correlation Matrix of PXI Mean Scores

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Appeal	-										
2. Autonomy	.34	-									
3. Challenge	.26	.28	-								
4. Control	.06	.51**	.28	-							
5. Curiosity	.49*	.48*	.25	.10	-						
6. Goals	.13	.14	-.01	.38	.08	-					
7. Immersion	.04	.35	.20	.38	.26	.24	-				
8. Meaning	.51*	.44	.28	.37	.62***	.30	.31	-			
9. Mastery	.18	.48*	.22	.55**	.29	.36	.31	.46*	-		
10. Progress	.41	.13	.31	.18	.37	.38	.16	.14	.24	-	
11. Enjoyment	.60***	.48*	.31	.32	.62***	.45	.29	.72***	.44	.38	-

. Correlation is significant at the 0.05 level (2-tailed), ** Ibid. at the 0.01 level, * Ibid. at the < 0.001 level.*

Performance & Experience Items

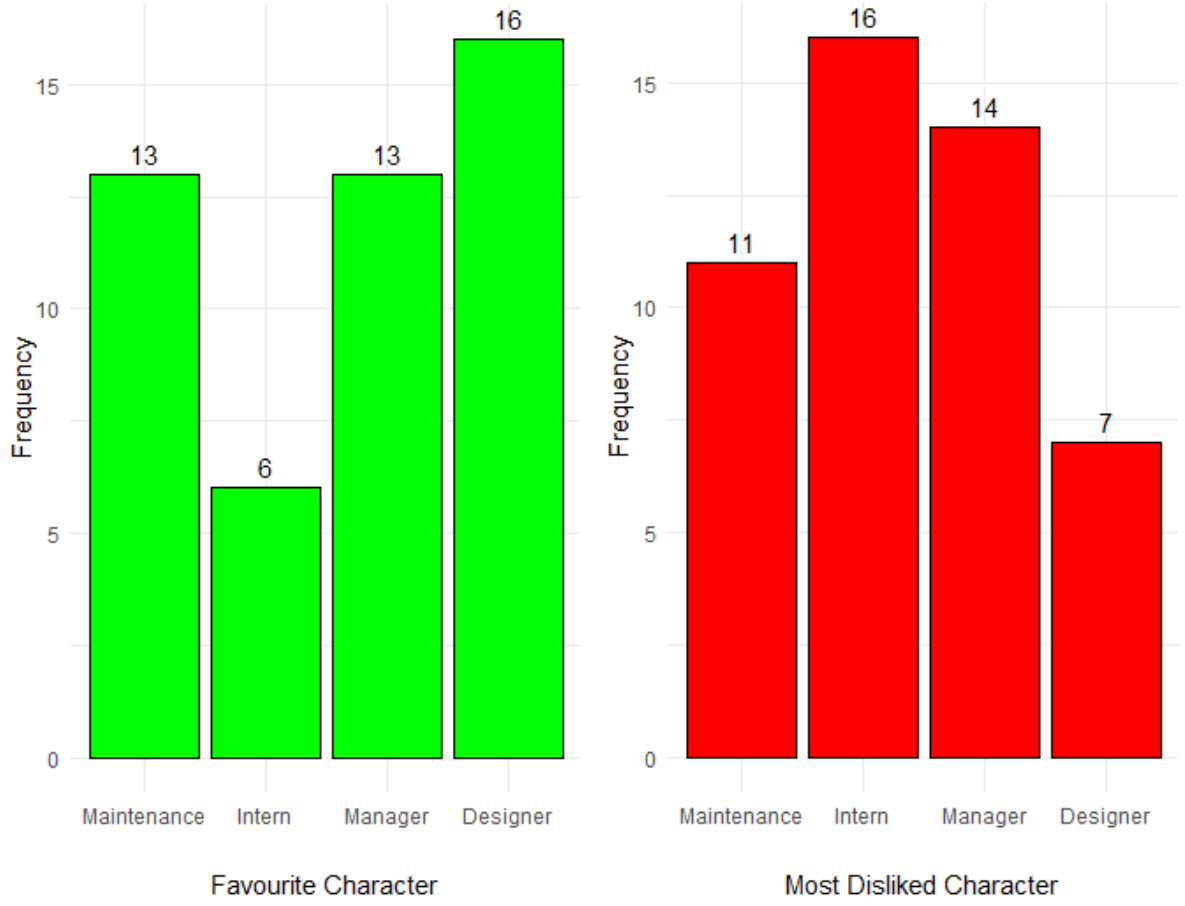
In addition to the Player Experience Inventory items, additive items listed in the survey were analysed. In terms of performance, player self-assessments average around a mean of 4.27 (*SD* = 1.27), implying that most players rated themselves around the middle mark – neither performing very well nor very bad at the game according to their own assessment. Continuing to the overall game experience, player ratings average around an overall positive assessment (*M* = 5.75, *SD* = 0.93), mirroring the generally more positive results of the PXI. Lastly, inquired willingness to play similar games as to the one featured in this study averages at 5.73 (*SD* = 1.33), indicating that participants overall felt more positively inclined towards playing games of a similar type.

Character Interaction

Moving on to character interaction variables, frequencies of most liked and disliked character, as well as means and standard deviations of attractiveness and interaction enjoyment ratings were investigated. The highest frequency character denoted as the favourite character of the selection was the character of the Designer (Parker) (*N* = 16*)*, while the highest frequency character named as the most disliked character of the line-up was the Intern (Aria) (*N* = 16). Mirroring these results, both characters also share the lowest frequency in the opposite category, viz., the Intern as the least favourite character (*N* = 6) and the Designer as the least disliked character (*N* = 7). Consequently, participants tended to like the Designer the most of the cast of characters, while the Intern tended to be the least liked. An overview of the frequencies can be found in Figure 6.

Figure 6

Frequency Charts of Favourite and Most Disliked Characters

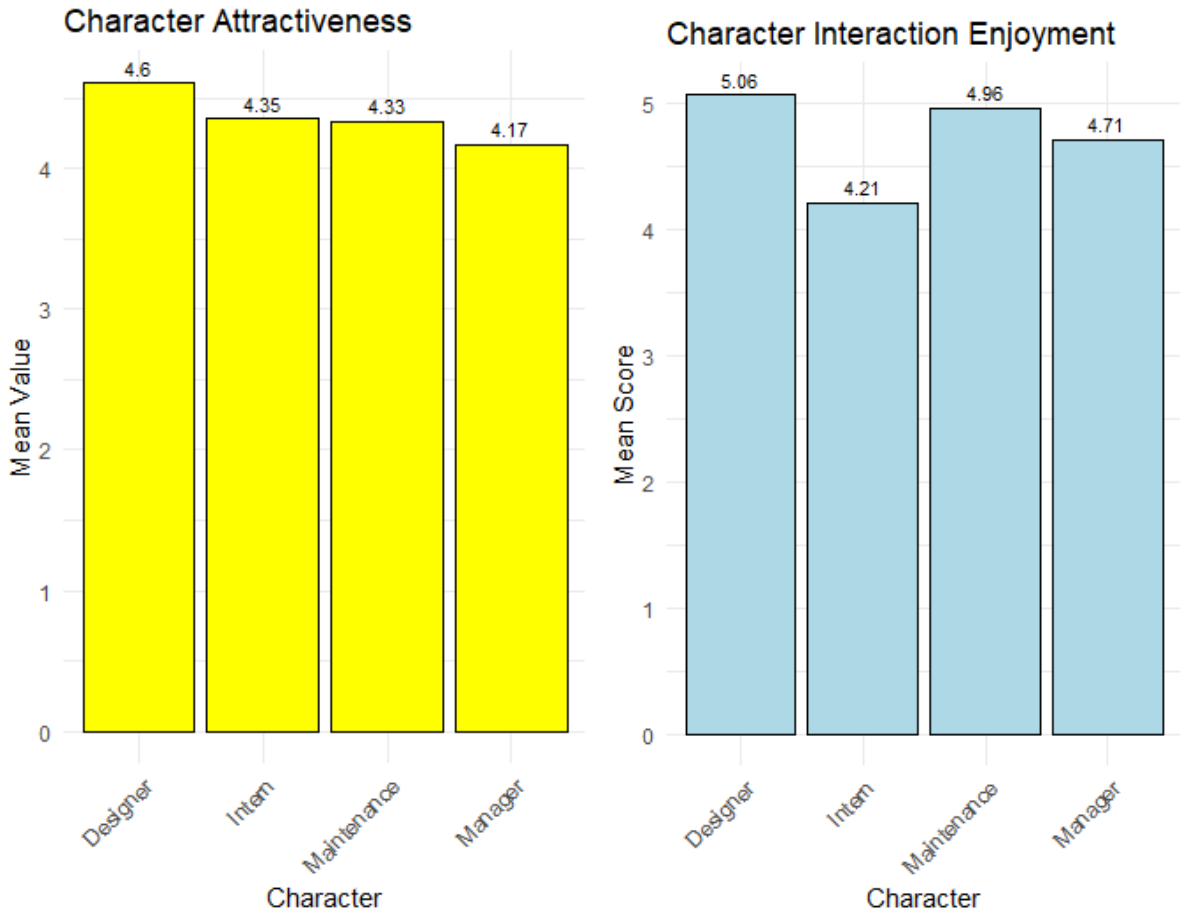


Adding on to favourability ratings, investigation of character attractiveness and interaction enjoyment ratings yields reflective results, as illustrated in Figure 7. Starting with attractiveness ratings, the highest mean score is found in the attractiveness ratings of the Designer (M = 4.6, SD = 1.43), while the lowest attractiveness rating is visible in the attractiveness rating of the Manager (M = 4.17, SD = 1.42). In terms of interaction enjoyment, interactions with the Designer yielded the highest mean scores (M = 5.06, SD = 1.41), while participants tended to enjoy the interaction with the Intern the least, as indicated by the lowest mean score (M = 4.21, SD = 1.62). In summary, participants overall tended to like the character of the Designer the most, whether in interaction or attractiveness. On the other side,

the character of the Intern tended to be the least liked, as is especially visible in the average interaction enjoyment.

Figure 7

Mean Scores of Character Attractiveness and Interaction Enjoyment



Attractiveness and Interaction Ratings Correlations. Participant ratings in connection to the perceived attractiveness and the enjoyment of interactions were further explored and correlated to foster insights into non-player character perception and interaction of players in the game. Results of the correlational analysis between attractiveness and interaction enjoyment ratings are shown in Table 2. As illustrated, statistical significance can be found in correlational tests between attractiveness and interaction enjoyment for each

character with moderate to large effect size. Largest effect sizes are visible in correlation of variables connected to the non-player characters of the Designer and Maintenance. Overall, the results indicate that based on a moderate to large effect, players assessing a character as more attractive also tended to rate the interaction enjoyment with said character higher and vice versa.

Table 2

Correlation Results of Attractiveness and Interaction Enjoyment Measures

NPC	Correlation Coefficient	p-value
Designer (Parker)	.57	< .001
Intern (Aria)	.38	.007
Maintenance (Ali)	.59	< .001
Manager (Kai)	.46	.001

Gameplay Sentiment Analysis

In order to further investigate how players interacted with the game's featured non-player characters, verbal input of participants employed for in-game voice communication with NPCS was analysed through the sentiment analysis tool LIWC and the AFINN and NRC lexica using the tidyverse package in RStudio. The results of the analysis are illustrated in the following.

AFINN and NRC Sentiment Analysis

AFINN. Sentiment analysis employing the AFINN lexicon resulted in an overall sentiment score of 782, with an average word sentiment score of 0.77 (*SD* = 1.75), indicating a stronger presence of overall positive sentiments. A frequency distribution chart of sentiment

scores can be found in Figure 8. As illustrated, words assigned a positive sentiment score of 2 show the highest occurrence overall, with comparatively smaller frequencies across other sentiment scores. Table 3 shows an overview of the highest scoring words in positive and negative sentiment scores, respectively. Listed words and scores underline the stronger presence of positive sentiments in participant input, especially highlighting the present usage of polite or positive language directed at NPCs through words such as “Thanks” or “Wonderful”. Taken at face value, sentiment analysis results based on the AFINN lexicon propagate participant input to contain more positive sentiments than negative ones and illustrate a more present use of positive and polite words when conversing with the game’s NPCs.

Figure 8

Frequency Distribution Chart of AFINN Sentiment Scores

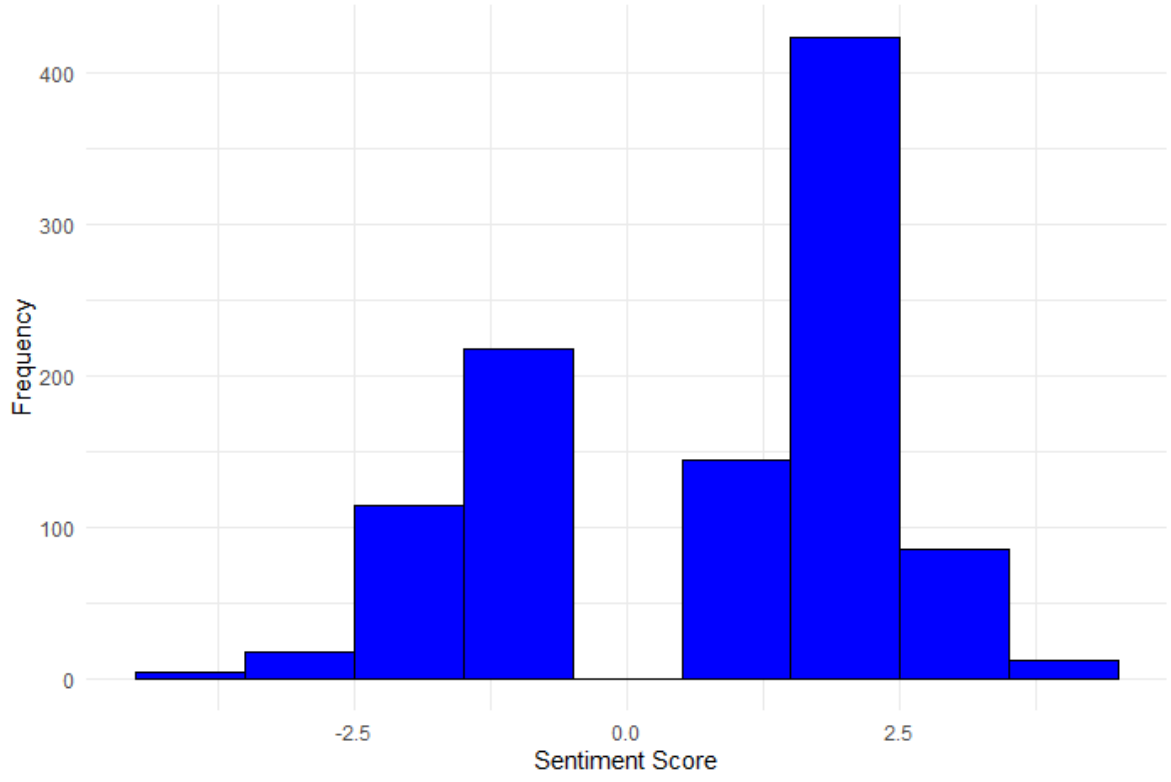


Table 3

Highest 20 Positive and Negative Score Words of AFINN Sentiment Analysis Output

Word	Sentiment Score	Word	Sentiment Score
Gift	330	Annoyed	-122
Thank	202	Broke	-113
Like	72	No	-58
Help	66	Problem	-20
Yeah	62	Sorry	-17
Good	60	Bad	-15
Thanks	54	Damn	-12
Happy	51	Excuse	-9
Great	48	Annoying	-8
Nice	48	Wrong	-8
Wonderful	40	Broken	-7
Perfect	27	Hate	-6
Yes	23	Liar	-6
Want	16	Lost	-6
Fine	14	Mad	-6
Cool	13	Useless	-6
Excellent	12	Accident	-4
Hope	12	Confused	-4
Kind	12	F*ck	-4
Please	11	Leave	-4

NRC. Sentiment Analysis utilising the NRC lexicon yielded a count of overall 577 positive and 237 negative words. Figure 9 displays a frequency distribution chart, showing the highest count of emotionally connotated sentiments were observable in words connected to Joy (*f* = 454), Anticipation (*f* = 448), and Surprise (*f* = 412). Additionally, Table 4 shows an overview of the highest sentiment count words. As listed, the highest count words were gift, birthday, and broke by a substantial margin – notably words that are connected verbatim to one or more objectives presented to the players. Present results imply an overall higher usage of positive words by participants during their playthrough and a higher presence of words connected to emotions such as joy, surprise or anticipation in their utterances directed at the game's featured NPCs.

Figure 9

Frequency Distribution Chart of NRC Sentiment Categories

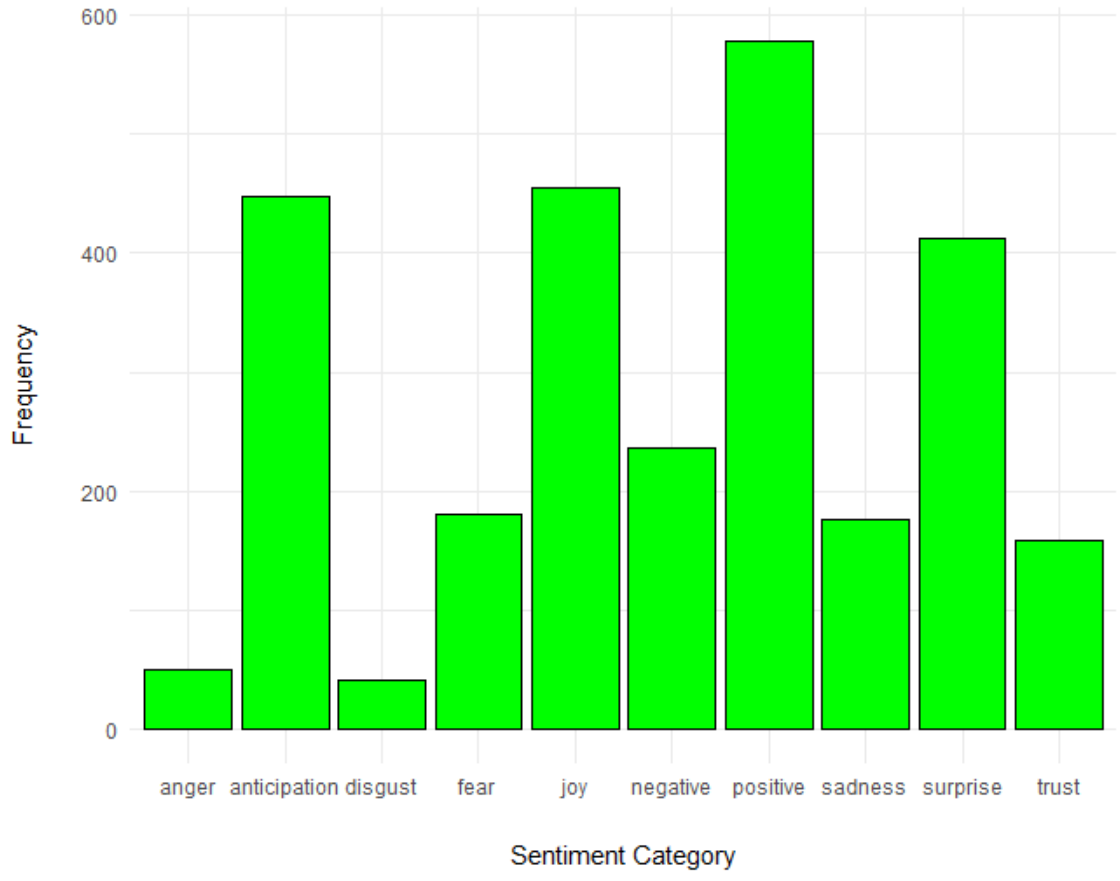


Table 4

NRC Sentiment Analysis Output – Highest 25 Counted Words

Word	Count	Sentiment (Emotion)	Sentiment
Gift	165	Anticipation, Joy, Surprise	Positive
Birthday	149	Anticipation, Joy, Surprise	Positive
Broke	113	Fear, Sadness	Negative
Good	20	Anticipation, Joy, Surprise, Trust	Positive
Happy	17	Anticipation, Joy, Trust	Positive
Talk	15	-	Positive
Cool	13	-	Positive
Present	13	Anticipation, Joy, Surprise, Trust	Positive
Break	10	Surprise	-
Problem	10	Fear, Sadness	Negative
Wonderful	10	Joy, Surprise, Trust	Positive
Excuse	9	-	Negative
God	9	Anticipation, Fear, Joy, Trust	Positive
Perfect	9	Anticipation, Joy, Trust	Positive
Information	8	-	Positive
Maintenance	8	-	-
Broken	7	Anger, Fear, Sadness	Negative
Found	7	Joy, Trust	Positive
Happen	7	Anticipation	-
Surprise	7	Fear, Joy, Surprise	Positive
Hope	6	Anticipation, Joy, Surprise, Trust	Positive
Kind	6	Joy, Trust	Positive
Lying	6	Anger, Disgust	Negative
Proper	6	-	Positive
Time	6	Anticipation	-

LIWC Sentiment Analysis

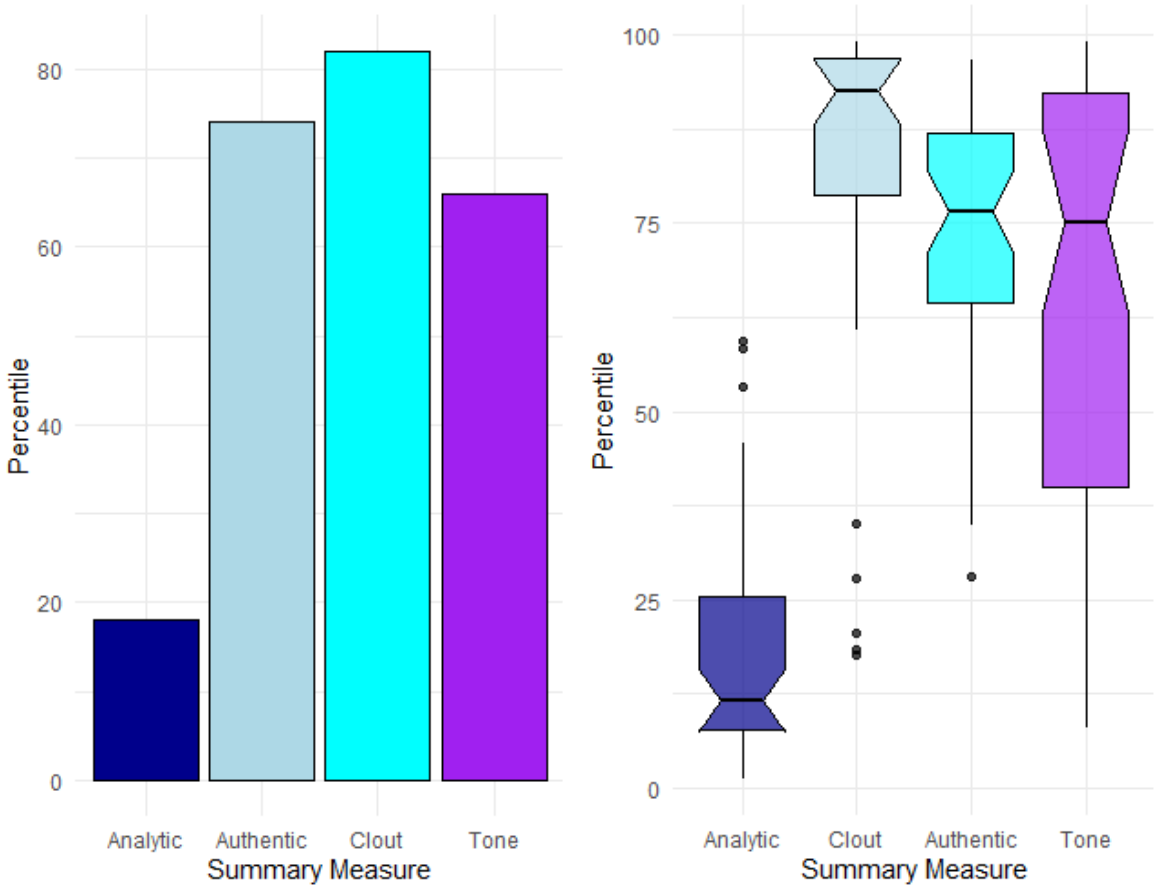
Sentiment Analysis employing internal dictionaries of LIWC-22 resulted in an extensive sentiment analysis data frame containing multiple measures. For purposes of data exploration based on the research questions presented in this paper, only key summary measures are examined. LIWC sentiment analysis output entails four summary measures: Analytical Thinking, Authenticity, Clout, and Emotional Tone. To briefly elaborate, Analytical Thinking signifies the degree to which the language is shaped by formal and logical thinking patterns, while Authenticity is positioned to reflect the degree analysed speakers appear to self-monitor i.e., how inhibited or, conversely, how "frank" they appear in their choice of words (*LIWC Analysis*, 2015). Furthermore, Clout as a key measure represents the degree to which the analysed language features terminology of social hierarchy. Lastly, Emotional Tone analysis is operationalised as a summary variable interlacing two tone dimensions, positive and negative, into one, with higher numbers signifying a more positive tone and lower numbers denoting a more negative tone – positioning 50 as the midway point. As detailed by Tausczik and Pennebaker (2009), each of these summary measures represent algorithms derived from LIWC variables based on previous empirical research. Resulting numbers signify standardized scores converted into percentiles ranging from 1 to 99.

Analysed participant input extracted from the gameplay resulted in a key measure mean percentile distribution illustrated by Figure 10. As presented, the key summary measures Authenticity (M = 74.02, SD = 16.76) and Emotional Tone (M = 65.94, SD = 29.74) scored above the 65th percentile, while Clout scoring the highest above the 80th percentile (M = 81.99, SD = 23.46). On the other hand, the key measure for Analytical Thinking (M = 17.93, SD = 15.64) scored the lowest, around the 20th percentile. Results indicate that participants generally tended to employ a more positive and authentic or spontaneous speech, while also being less analytic and more friendly and personable.

Additionally, respondents also tended to use more vocabulary connected to social status and hierarchy, while talking to the in-game characters. However, sizable standard deviations for the summary measures Clout and Emotional Tone should be noted, underlining a large variance in the observed data around the mean – consequently implying especially the overall emotional tone of participants tended to vary to a greater extent when interacting with the game's NPCs.

Figure 10

LIWC Mean and Median Summary Measure Percentiles



Interview Results

Coding Scheme

The coding process of open survey questions derived a coding scheme entailing 5 themes that partially correspond to the questions asked in the study interview. The themes extend to: Control & User Interface, Game Design, NPCs, Virtual Reality and Voice Interaction. An overview of the entire codebook can be found in the Appendix D.

As the first theme, Control & User Interface extends to codes that relate to the intuitiveness and/or ease of use of the controls of the game, as well as codes signifying sentiments relating to the UI featured in the game, similarly connected to factors of ease of use and intuitiveness. The second theme, Game Design, encompasses coded sentiments pertaining to the game's visual and internal environmental design, its narrative structure and incorporated objectives and general sentiments concerning the overall game. Included in this theme are participant statements regarding features they felt were missing from the game, as specifically inquired in one of the interview's questions.

Featured in the third theme, NPCs, are codes that envelop every notion regarding the showcased non-player characters in the game. Central topics that make up the theme extend to codes connected to the believability of the NPCs, their personality or character, their motion and expression, as well as the interaction with said NPCs. Following this theme, Virtual Reality is the fourth theme, which entails coded sentiments relating to the technological implementation of the virtual reality environment and notions specifically related to VR, such as technical difficulties or admiration of VR as a novel technology. Finally, the last theme, Voice Interaction, encircles codes signifying sentiments related to the impact and functionality of voice interaction as the primary tool of interaction in the game.

Controls & User Interface

The theme Controls & User Interface encompasses 2 code groups with 7 subcodes. The two code groups featured in this code block extend to codes connected to controls and codes related to the user interface of the game, denoted as Controls and User Interface respectively.

Controls. As part of the overall theme of Controls & User Interface, the Controls code group features codes that entail sentiments participants expressed regarding the method of control or control input in the game, i.e. the intuitiveness of how the VR environment controllers were mapped with in-game actions and the general "feel" of how users could execute actions within the given control scheme. Here, responses were coded as a specific code when respondents shared a sentiment either directly or indirectly alluding to features or function of the control scheme and how the game utilizes its controllers. The most frequent code in this code group is found in Cumbersome Controls, signifying that most participants felt that the game felt hard to control or master and was often subject of negative attention. As one participant puts it: "Sometimes I had difficulty grabbing some objects. And sometimes the teleportation wouldn't work, so I had to click the buttons several times."

User Interface. Following Controls, User Interface builds up the second code group of the theme, entailing 3 codes representing respondent sentiments connected to (elements of) the user interface employed in the game. Responses were coded in cases participants shared their thoughts and feelings towards certain elements of the user interface or the implementation of the overall UI. Here, the most frequent code crystallises in Facilitative UI, highlighting that most sentiments expressed by participants towards the UI contained the notion that the UI served a facilitative purpose and positively contributed to the gameplay experience. Elements of the UI that were highlighted the most included the extensive objective list featured in the game, as well as the "speech bubble" chat log system displaying

a live transcript of the conversation to the player when conversing with an NPC. According to one respondent, “it is really useful that the text of what you say and what they say appears at the top of their heads, because … it was easier to understand and have the information there:”

Game Design

The second theme, Game Design, entails 6 code groups with 37 subcodes. The 6 code groups contain codes related to the topics of Game Context, Gameplay Structure, Game Objectives, Missing Features, Virtual Environment, and Visual Style. Here, Missing Features as a code group was constructed to parallel question 8 of the post-study interview (see Appendix B).

Game Context. The code group Game Context is comprised of 4 different codes. The codes pertain to different topics that may embed or contextualise the game featured in this study and sentiments expressed from a more general or objective perspective. For example, codes in this group encoded sentiments about possible future application of games similar in type as the one featured in this study or allude to suitable genres for such a game. The most frequent code of this group is found in the code denoting participant sentiments applauding the overall game experience as a novel and positive experience. As an example, one participant shares that “it was unlike anything I've ever experienced before. This was novel. It was so fluid and … fun.” Here, many participants highlighted that specifically the combination of virtual reality and voice interaction resulted in their conception of the game as something novel or positive. As one respondent shared:

> “As I get older now, I find games more of … a time waste and more of a ‘it's too much, it takes too much of my time’ and sometimes I get very bored of it. [The game] was very fun because I'm talking to an AI. And I'm doing

tasks in the world, and I'm being active and I'm actually up on my feet. So, I think it's very, very interesting."

Gameplay and Structure. Following Game Context, Gameplay and Structure constitutes the second code group of the theme. The group encompasses 5 codes that relate to the inherent gameplay of the VR game and its connected overall narrative or mechanical structure. Responses were coded when respondents explicitly or implicitly shared thoughts or notions regarding features of the gameplay, story or their expression or role as a player of the game. To illustrate, when a participant explicitly shared that they felt a high sense of freedom in their role and possible action as a player or implicitly alluded to being able to engage in any activity or interaction they wanted, as opposed to other video games, the response would be coded as Player Freedom. The highest frequency code in this group takes form in Immersive Gameplay, wherein participants illustrated the gameplay of the VR game as absorbing, very present or very immersive, in a sense that they felt as part of the world. As one respondent describes it: "I felt like I was into this, I was really in this office. So later on, I didn't feel as weird anymore because I actually thought I was talking to somebody."

Missing Features. As the third code group, Missing Features is comprised of 12 subcodes associated with sentiments underlining features that are missing from the game from the participant's point of view. Respondent statements were encoded into one such code when participants gave feedback in response to question IQ8 "What did you miss with regard to the conversation with the characters?" and detailed their conception of features potentially missing. As an example, a respondent stating a lack of verbal backchanneling by the NPCs as conversation partners was consequently coded as Backchanneling. In this code group, the highest frequency code is found in UI prompts, highlighting that most respondents elaborating on potentially missing features underlined or alluded to a lacking guidance through interface prompts, specifically during the voice interaction between NPCs.

Game Objectives. Continuing, Game Objectives entails 5 codes as a code group, encoding sentiments connected to the VR game's employed list of objectives and any and all comments on the quality of implementation or design. To illustrate, respondent statements characterising the objectives featured as achievable, possible to accomplish, or simple to solve were coded as Achievable Objectives. The highest frequency code of this code group is found in the code Interesting Objectives, which is closely followed by Uninteresting Objectives. This contrast signifies that participants often found parts of the objectives interesting, but criticised other parts. A notable sum of respondents shared that they felt the objectives connected to interacting with the NPCs were more interesting than tasks that asked players to navigate the virtual space or search for certain objects. According to one participant, when asked what they found the most interesting aspect of the game, the tasks were perceived as "kind of challenging, like to get to the goals and everything. And because I already thought: 'Did he lie? Did he not lie?' I was always questioning everything. That was interesting.", but later also stating that the least interesting aspect of the game was "having to walk around to find stuff.", further adding: "I think the talking part is the interesting one. And then going around finding stuff, maybe not."

Virtual Environment. Following Game Objectives, Virtual Environment builds the fifth code group of the Game Design Theme. It consists of 5 subcodes signifying participant sentiments regarding the virtual environment employed in the game in perspective of its realism, interactivity, intuitiveness, and ability to spark interest in players. Accordingly, sentiments were encoded that assess or communicate perceived qualities of the environment (or objects as part of it) either explicitly or implicitly. As an example, sentiments shared by respondents exclaiming that they felt the virtual environment represented a realistic environment were coded as Realistic Environment. The highest frequency code in this code group is Uninteresting Environment, closely followed by Interactive Environment and

Intuitive/explorable Environment. One participant commented that “the environment, like the office, could have been more interesting”, while another added that “The least interesting was probably the office setting. I'm just not very interested in offices. It's just a little bit of a boring environment.” In terms of the code Interactive Environment, one participant had the following to say: “I think grabbing the stuff and giving it to people is quite fun, and also looking for the different codes and then being able to open door and see what's behind that and see if there's more things to do there”, while another respondent felt that “the fact that I could open the doors and go to the next room, to the other rooms and see the other rooms, it was also nice, interesting for me. That was attractive”, as encoded in Intuitive/Explorable Environment.

Visual Style. As the last code group comprising the theme, Visual Style entails 4 codes featuring encoded sentiments related to visual elements in the game and design choices of the game’s elements. To elaborate, if a participant alluded to finding the visual stylisation of the NPCs implemented in the game unattractive, estranging, or otherwise unappealing, the sentiment was coded as Unappealing NPC Visuals. In this code group, the highest frequency code is Attractive Visual Style, wherein sentiments pertaining to the overall visual style of the game and characterising it as attractive or appealing are encoded. As one participant puts it in response to what they found attractive in the game: “Um, I don't know, maybe just style of the game in general. Like, it was something that you are used to when you play with your games and, um, it's, uh, attractive and maybe a pleasant thing for your eyes.”

Non-player Characters

Non-player Characters, forming as the third theme, consists of 5 code groups with a sum of 35 codes. The code groups featured in this group extend to Believability Rating, Believability Aspects, Interaction, Motion & Expression, and Personality & Character.

Believability Rating. The first code group of non-player characters takes shape in Believability Rating. Here, codes concerning the overall assessment of the believability of the conversation with the game's NPCs and said NPCs as a realistic conversation partner by participants are contained. The coding of this group was informed by question 9 of the post-study interview questions (see Appendix B). The group entails 5 ordinal subcodes, ranging from Unbelievable Conversations to Believable Conversations. The highest frequency code in this code group can be found in Somewhat Believable Conversations, wherein participants were generally positive in their perception of the in-game conversations as believable, but did note various aspects negatively impacting said perception, therefore resulting in a somewhat positive attitude. Notably, many participant sentiments coded as Somewhat Believable Conversations contained an explicit rating of "7 or 8 out of 10", with the highest number signifying human and conversely the lowest number signifying machine. A common denominator for less believability that was listed as part of a respondent's explanation of how they arrived at their assessment was a seeming disillusionment in the form of realising one or more traits or patterns in the NPC's answering or behaviour that made them appear less humanlike. One participant puts their thoughts in terms of this effect this way:

> "I think especially the first time you talked to them, they were rather believable, I would say it's really good – a nice believability. And then at some point, some of the prompts started to repeat themselves and then it starts to become a little less realistic in a way. But I think it was super nice in terms of [them having] enough answer possibilities to give something that was fitting all the time."

Believability Aspects. Thematically tying into the previous code group, Believability Aspects builds the second code group of the theme. The group entails 8 subcodes encoding sentiments related to the different aspects participants viewed to be important or noteworthy

for the overall believability of NPCs in conversations. For example, if a respondent conveyed a sentiment detailing the importance of gestures and facial expression for a believable NPC conversation partner or the lack thereof as a negative factor, the sentiment would be coded as Gestures & Facial Expression. The highest frequency code of this group is observable in the code Personalities, which entails sentiments underlining the noteworthy role an apparent personality of an NPC appears to play for a perception as believable in conversations, meaning that participants either mentioned such a notion or criticised the lack thereof. Here, one participant states that the NPCs were "all pretty interesting and they all had their own characteristics", furthermore stating that they "noticed that Parker is more of an expressive guy, and he seemed to try to be in tune with [the conversation content]. Those little things help bring more life to the character."

Interaction. Following Believability Aspects, Interaction forms the third code group of the theme. The group encompasses 18 subcodes containing sentiments connected to the user interaction with the game's featured NPCs, verbal or otherwise. To expatiate, if participants reported the conversation with an NPC feeling "artificial" or criticised the overall "flow", meaning the general structure or smoothness of a conversation, as unnatural or unlike a real conversation, the sentiment would consequently be coded as Artificial Conversation Flow. As another example, if a participant highlighted an apparent deceptive nature of an NPC, possibly because of erroneous statements or factual lies in the context of the game's narrative or objectives on the NPCs part, a sentiment would be coded as Deceptive NPCs. The highest frequency code in this code group crystallises in Realistic/Natural Conversation, highlighting the high occurrence of participant sentiments expressing the perception of the interaction with NPCs as rather realistic or natural to them, conversing as they would do similarly in real life. One participant states that "It felt like a human interaction mostly.… I got really the responses that I expected and that surprised me. It didn't feel like a robot or

something.” Additionally, many respondents point out that using verbal communication as their main mode of conversation is likely the cause for the realistic impression, making the interaction more “personal”, or to put it in a participant’s words:

> “I thought [voice interaction] could be used for a lot of things because it's a totally new way of interacting with NPCs. It makes it feel more personal in a way, because.… Yeah, like I said before, it's like them interacting to exactly what you said. Yeah. So that's the part I liked about it. I think very much deeper connections can be formed to the characters, honestly.”

Motion & Expressions. Continuing, Motion & Expressions shapes the fourth code group of Non-player Characters. The code group entails 4 codes pertaining to sentiments addressing or alluding to the movement and facial expressions of the NPCs in the game. The highest frequency code in this group is Mismatching Facial Expression, incorporating participant sentiments underlining perceived problems with the facial expressions of the NPCs, such as wearing an expression unfitting for their supposed emotional state or continually showing the same facial expression regardless of what was said, tied with No NPC Movement, entailing encoded sentiments highlighting the stationary nature of the NPCs in the game.

Personality & Character. Lastly, Personality & Character forms the fifth code group of the theme, encircling 12 codes. Said codes consolidate respondent sentiments connected to the personality and characteristics of the NPCs, including the impression of the NPCs on participants and recounts of like and dislike by participants of NPCs. Accordingly, sentiments were coded when participants explicitly or implicitly expressed notions connected to NPC characteristics or personality traits or matters related to said topics. For example, if a participant illustrated that an NPC was relatable, friendly, or easy to talk to, the sentiment would be encoded as Approachable NPCs. In terms of the highest frequency code, Bland

NPC Personality condenses as such, containing respondent sentiments related to a perceived bland personality of the NPCs in expression, voice, or character traits. One participant comments: “I think the other thing is that they're all just kind of like the same person. When you get down to it, they don't have any interesting facets of them as people [would have].”

Virtual Reality

The theme of virtual reality is comprised of 6 codes each pertaining to different facets of the VR technology employed by the study, for instance the game’s programming or the effects of a virtual reality environment. Sentiments were encoded when participants thus expressed or otherwise alluded to opinions or notions connected to relevant topics. To elaborate, in cases where participants underlined experiencing problems with the head-tracking of the NPCs meant to follow the player, sentiments would be coded as NPC Gaze Problems. The highest frequency code in this theme is visible in Technical Difficulties, wherein sentiments concerning bugs and game disruption due to technical failures such as NPC conversational loops or API connection problems are encoded. As an example of one such case, one participant highlights an interaction with an unresponsive NPC in the following: “The fact that [the NPC] didn't talk to me. It was just sitting and was looking at me, so I didn't know how to talk to him. And I tried so many times, but he just said nothing.”

Voice Interaction

As the last theme, Voice Interaction entails 3 code groups comprised of 22 codes. The code groups featured in this theme extend to codes grouped based on the interaction utilising the participant’s voice and the recognition of the voice input central to these interactions, thus named Interaction, Recognition, and, directly tying into the prior, Recognition Problem Causes.

Interaction. The code group of Interaction encompasses 8 codes connected to the verbal interaction central to the game. Sentiments were encoded in cases when participants either explicitly or implicitly brought forth notions and thoughts related to the voice interaction itself. As an example, if respondents reported or described an initial or ongoing unease about using their voice as the main mode of interaction, sentiments would be encoded as (Onset) Uncomfortableness. The highest frequency code in this code group is Voice Interaction Boundaries. If a respondent illustrated or explicitly addressed problems with the boundary of voice interaction, i.e. the distance threshold between the player and an NPC until the latter will start to recognise the voice input of the participant as an utterance targeted at them and start a conversation, a sentiment would be coded as Voice Interaction Boundaries. One participant expressed the problems in the following manner:

> "Thing is I, unlike in real life, I couldn't tell the distance between me and the NPC were – how… how far should I be? Or how close should I be to start the conversation? I approached him, I tried to talk to him, and I look up and the text is not showing. Okay, so I had to move around a little bit, and once I almost hit him on the face [sic], his face was right here."

Recognition. Lastly, Recognition entails 5 codes associated with the recognition of the verbal input given by the player in the game. Sentiments were coded when participants made statements or expressed opinions related to the voice recognition employed in the game, either implicitly or explicitly. To illustrate, in case a participant reported or alluded to a perception of a well implemented or well-functioning voice recognition, the sentiment would be encoded in Good Recognition. Contrasting this example, Recognition Problems crystallises as the highest frequency code, containing encoded participant sentiments related to descriptions or reports of problems in the recognition of verbal user input. As one respondent reports "sometimes [the game] didn't recognize some words. Maybe it's due to my

accent or, uh, like how I say words, but that can always be improved, I guess", while another states that "sometimes my voice wouldn't be recognized correctly. So, they would mistake server room for a different room, or they would recognize the name of Kai for Cardi B or anything else."

Recognition Problem Causes. Lastly, the final code group of the theme takes shape in encoded causes of recognition as illustrated by participants in the form of 4 subcodes. Sentiments would be encoded if a respondent alluded to or detailed a possible cause for the event of misrecognition during a speech interaction with an NPC. The highest frequency code in this code group is found in the code Pronunciation/Accent, denoting encoded sentiments by participants that convey the idea that mispronunciation or a stronger accent was the suspected cause for a misrecognition event. As one participant puts it, "sometimes it just didn't pick up what I said correctly, but when I repeated myself and spoke more clearly that wasn't really an issue anymore." Notably, many participants illustrating such instances tended to also shift the blame to themselves, rather than towards their digital conversation partner or technical failures. One respondent states here that "I guess sometimes they don't pick up what you said, but it probably is fault of my part. Wasn't clear enough", while another conveyed that "I would attribute it to my accent and just like, say that it's hard to hear." Here, one participant even explicitly takes the blame, stating that "sometimes, my pronunciation was not very good, so they understood me wrong, but I clarified them…. It was my fault. And not really theirs."

Discussion

Summary of Results

The purpose of this study was to investigate player sentiments and attitudes concerning virtual reality games that feature verbal communication systems as their central

mode of communication, and how potential players will engage with implemented non-player characters in such games. Therefore, this study set out to answer the concrete research questions of how players experience a VR game featuring AI NPCs using their voice as the main tool of interaction and how players would perceive and interact with said NPCs. The results of the present study offer a myriad of insights and multiple points of interest for both questions. Overall, the player experience of the game was a positive one, with players valuing its goal clarity and immersive qualities resulting in higher levels of enjoyment. Despite some shortcomings in, for example, the flow of conversation, NPCs were outlined as flexible and organic conversation partners that offered an immersive conversational experience that was, however, often interrupted by problems in the recognition of player voice input. The more embedded insights in their relation to the specific research question are examined under the premise of both research questions in the following.

RQ1: How do players experience a virtual reality game using their voice as the main tool of interaction?

Informed by the data of the PXI and additive player experience questions, the player experience of the VR game was generally positive. Player assessments of their overall game experience averaged around a clear positive sentiment, where the average player generally felt positive about their experience of the gameplay and the game's overall features, as implied by the above average PXI mean scores. This sentiment is also further corroborated by a general positive attitude towards playing games of similar type, as shown in the results. Players saw the strong suits of the virtual reality game in its goal clarity, its immersive nature, and, as a result, its ability to provide fun and entertaining moments. However, players may have experienced a relatively lower sense of their own skill while playing, in addition to finding it harder to see personal meaning in playing the game, as indicated by the results of

corresponding items. While the latter may be explained by considering the premise of the study exposing participants to the game regardless of prior personal interest in video games of this kind, the prior may have occurred as a consequence of either low experience of playing VR environments or the inability to complete objectives in the game and feel a sense of skill through accomplishment. The generally averaging performance self-assessments could be an indicator of this conundrum.

The results of the correlational tests shed further light onto potential underlying factors that influence player experience. First, based on negative correlation between age and autonomy and control items, it is likely that older players had more difficulties navigating the game and achieving a sense of control than younger players. Secondly, correlations imply that higher player enjoyment was most likely to occur in instances where players found the game to be personally meaningful, intriguing, and/or appealing. Lastly, players also tended to feel a higher sense of autonomy and mastery of the game, when they felt they were able to exert a high degree of control in their playthrough.

Sentiments extracted from post-study interviews further contextualise these findings. Multiple participants illustrated their experience with the game as a positive and novel one, on the one hand paralleling positive indications of overall gameplay experience in the survey data, and on the other hand underlining the possible role of novelty found in the type of game presented, which in turn may elicit curiosity about the game. Furthermore, reports of fun experiences and interactions with the game's character crystallise with high frequency among the participants' responses, with many participants explicitly outlining their experience of the game as an immersive one. Additionally, players mentioned feeling a sense of freedom and agency through the open-ended structure of the dialogue and praised helpful features such as objective lists and live voice transcription, while some players also illustrated being intrigued by the game's overall structure.

Nevertheless, participants also outlined some overall weaknesses they perceived in the game featured in the study by expressing predominantly negative sentiments regarding the intuitiveness of controls, being somewhat split on the content of objectives, and sharing very present concerns with the capabilities of the voice recognition featured in the game, together with the oftentimes apparent unhelpfulness of responses given by the NPC during conversations.

RQ2: How do players perceive and interact with artificial intelligence non-player characters in a virtual reality game using their voice as the main tool of interaction?

In terms of NPC perception, players indicated a general preference for the character of the Designer, while a less favourable view of the character of the Intern. Regarding attractiveness, while a slightly more positive tendency towards the Designer is visible, no extreme difference in attractiveness rating of the NPCs can be discerned in overall participant ratings. Examining the interaction ratings, clearly lower values for the Intern arguably provides a potential reason for the less favourable indicators for the character. Additionally, correlational tests outlined a potential relationship between attractiveness rating and interaction enjoyment, indicating that higher personal attraction tends to co-occur with likewise higher interaction enjoyment and vice versa with lower values.

Moving on, the results of analysing player to NPC conversations during gameplay and sentiments surrounding NPCs conveyed by participants in the post-study interviews illustrate a complex picture of perception and mode of interaction players carried throughout their time interacting with the game's NPCs. In general, players tended to use more direct, social, and positive language, as evidenced by sentiment analyses. Apart from inquiring about the game's objectives when talking to the game's NPCs, participant input analysis showed a

strong presence of polite and supportive words in the language of the players. Conversely, distanced, and analytical language was less present in player speech interaction input.

A more nuanced picture of player to NPC perception and interaction is presented by the analysis of the post-study interviews. Most participants indicated the game's NPCs as somewhat believable conversation partners, oftentimes described as more human than robotic, but with sometimes obvious flaws. Praise was often given when discussing the flexibility of answers and very organic manner of response the NPCs employed in conversation, which often compared to real conversations and highlighted in its natural feel. A considerable number of participants also felt the need to highlight their surprise or need to mention the occasional deceptive nature of the game's featured NPCs, more often than not underlining the ability of NPCs to lie to them with mixed feelings of surprise, unease, or irritation.

Points of criticism, however, were brought forward in connection to the general interaction with the NPC in terms of structure or flow, with several participants pointing out perceiving a somewhat artificial conversation flow that often fell into a prompt-answer pattern when conversing with the NPCs, wherein conversation would only happen upon the players verbally approaching the NPCs and initiated by the other side. Consequently, NPCs were often perceived as lacking any real personality and some players even reported increasingly skipping verbal pleasantries and courtesies. Some technical limitations were further underlined by the participants as well, extending to concerns with the technical boundary of when and where conversation with an NPCs could be engaged as well as questionable head-tracking. Lastly, participants reported an initial unease in engaging with the NPCs via their voice, but likewise often also reported gradually easing into this form of interaction.

Examining Findings

Player Experience

Immersive and Engaging. Player assessments connected to increased immersion and entertainment, as well as player sentiments outlining their positive experience playing the game, highlight the potential that VR games incorporating VI hold for providing a novel, engaging, and immersive experience for players. These findings resonate with the results of the study by Zargham et al. (2020) in their exploration of voice interaction with NPCs within a single-player VR game, wherein participants were found to share a similar sentiment in their overall player experience. The findings also further strengthen the characterisation of voice interaction and virtual reality as technologies facilitating immersion (Allison et al., 2018; Pan & Hamilton, 2018; Slater, 2018; Zargham et al., 2022; Zargham et al., 2024), showing that a combination of the two in a VR environment does indeed retain this immersive quality and possibly even enhance it, as seen in the unique opportunity to afford players the liberty to freely ask any question and give personalised answers through a combination of AI backed NPCs and VI.

Due to their engaging and immersive features, VR games featuring AI-backed NPCs and VI may therefore be suitable for novel or enhanced VR training programmes, language learning environments, and interventions. Previous studies have already found positive results by implementing combinations of these elements in interventions and games (Li et al., 2020; Zargham et al., 2022). Furthermore, given the positive sentiment regarding goal clarity, future games should continue to pay attention to the clarity and ease of access of given objectives in such environments by implementing drop-down in-world displays, as featured in the present study VR game. Games featuring voice interaction are similarly recommended to include a live transcription of voice in- and output, considering the generally positive feedback players have expressed regarding this feature.

Higher Age, Lower Control. Furthermore, lower reported sense of control with higher player age may indicate that older players may have a harder time accessing and a different way of experiencing the game in comparison to younger players. The pattern parallels the findings of Birk et al. (2017) that outlined a decline in experienced intuitive control with higher player age, highlighting that while technological innovation in the digital entertainment industry can provide novel and positive experiences for many players and consumers, the effect may be inhibited by the increasing disconnect or absence of required technological expertise in higher age groups. This paradigm underlines the importance of inclusive design choices in future similar games and products to the one featured in the study, so that older generations do not become "sidelined" in the spurt of innovative technologies (cf. Köttl et al., 2021).

NPC Perception and Interaction

Worse Interaction, Worse Favourability. Moving on to NPC perception and interaction, the results of favourability ratings, wherein a less favourable assessment of an NPC co-occurred with lower interaction ratings, imply that one of the major drivers for likability of an NPC for participants is found in the quality of the overall interaction they had, more specifically how well they responded to the player and how informative their responses were to inquiries, as outlined by players in the interviews. Given the fact that players were required to interact with the NPCs to complete the given objectives and many participants reported experiences of unhelpful responses in addition to reported recognition problems, this suspected causal relation appears sensible. In their study on game character attachment, Bopp et al. (2019) highlight the definitive role gameplay competency of an NPC plays for a potential player attachment throughout their definition of character types. Players disliking some characters more than others as reflected by the results of this study may thus have been

a consequence of an inverse effect of attachment based on perceived competency occurring – or in other words, lower player attachment due to lower gameplay i.e. conversational competency. Future endeavours to build a similar game should thus ensure that the interaction with AI integrated NPCs are optimised in terms of the smoothness of interaction, as well as the helpfulness of responses in terms of their content if said content is central to the gameplay experience. The latter is still an especially apparent problem for generative AI (cf. Cox & Ooi, 2024), which is further discussed later in this section.

Players Converse Friendly and Authentic. The findings of sentiment analyses imply that players generally would tend to approach AI integrated NPCs in VR using voice interaction in an overall friendly, authentic, and polite manner. A possible reason for this behaviour may be a high perceived attractiveness that may have in turn led to a more enjoyable and thereby friendly conversation between the player based on found correlations. Another explanation might be the perception of these NPCs as more aware or socially apt conversational partners in comprehension and response. Consequently, players may have perceived their digital conversation partners as agents that can (a) realise impolite or socially inadequate speech patterns and (b) verbalise appropriate responses, possibly even chiding or reprimanding players if they were to use offensive or inappropriate language. This awareness could thus have resulted in the usage of an overall more positive language during conversation, as players verbally approached the NPCs similar to how they would have human actors. Players outlining skipping verbal courtesies and politeness when NPCs were more perceived as response machines, may further substantiate this proposed relationship. This possibility highlights the need for further research diving specifically into how the player perception of NPCs as realistic social agents and their capabilities as conversation partners can influence the actual content of the conversation and whether this relationship could be manipulated, or even capitalised on for future interventions, e.g. mitigating player

toxicity as proposed by Zargham et al. (2023) in their paper on combating player toxicity through the implementation of a courtesy mechanic.

AI NPC Conversation Partners are Realistic. Tying into the previous point, players did indeed outline the NPCs featured in the game as flexible and natural conversation partners that were able to organically answer questions posed by players and fluidly switch between and discuss different topics without being forced into predefined, scripted answer or question choices. These features appear to have moulded the NPCs to a substantial degree of believability as conversation partners – often described as more human than not. This illustration implies that AI integrated NPCs for conversation hold the potential to drastically alter and innovate the traditional position of NPCs in video games, promising a likely unprecedented depth of realism and immersion for games centred around such interactions and intelligent NPCs.

The perception as actors with a rather high believability promises AI-backed NPC to be an effective implementation and drivers of immersion and engagement according to the earlier outlined criteria of preceding papers investigating NPCs (Emmerich et al., 2018; Lim et al., 2012; Warpfelt, 2016). Thus, coupled with VI, implementation appears attractive in suitable genres of games such as RPGs, detective games, or immersive tactical simulations (Zargham et al, 2024), while other possibilities can be realised in, for example, serious games or language learning environments, as seen for example in studies featuring NPCs for emergency simulations (Li et al., 2020), perspective taking exercises (Ho & Ng, 2020), and simulations of foreign language environments (Cheng et al., 2017). Interventions in the realm of psychology utilising VR environments may also find use in such intelligent conversational agents, such as current explorations of the role of AI in VR-based offender mediation environments as conducted by Gerritsen & Zebel (Vrije Universiteit Amsterdam & University of Twente, 2024) or for simulating intelligent NPCs as deterrent guardians in VR

studies exploring the behaviour of burglars in digital neighbourhoods (cf. van Sintemaartensdijk, 2022).

As shown earlier, implementation of this new iteration of intelligent NPCs is already well on its way in the digital entertainment industry (Inworld AI, 2023; O'Brian, 2024); however, little research is currently diving into the promise and limits of generative AI NPCs, not just for games, but specifically for educational and/or interventive purposes as well. The promise the implementation of these NPCs yields for these areas forms a direct imperative for future research to explore and investigate how and where generative AI NPCs can be integrated, while additionally contributing to a clearer understanding of the associated risks and possible problems endeavours to integrate LLM NPCs may face.

AI NPC Conversation Partners are not Flawless. On the topic of associated risks and problems when integrating generative AI NPCs, the results of the present study also produced a condensed list of problems that occur when utilising AI integration in NPCs. These problems may extend to an artificial conversation flow and rigid conversation pattern, a lack of personality in NPCs, partially very apparent lying by NPCs, and a general unawareness of the context and the environment the conversation occurs in – all oftentimes culminating in unhelpful responses. The result is a conversational experience with NPCs that is occasionally disrupted and contaminated by either "hard" technical issues such as the current "turn-based" and artificial structure of conversations with NPCs and more "soft" issues such as unhelpful responses through mishaps in response generation by the AI or a perceivable flatness in character expression through a unilateral control of all NPCs by the same AI model that is, in a nutshell, acting out an attempt to impersonate a certain character. These issues in turn may inhibit the entertaining and immersive experience that is afforded by the implementation of AI-backed NPCs.

Overall, the found problems strongly resonate with the challenges outlined by Cox and Ooi (2024), Huang and Huang (2023), and Cabe et al. (2024). Specifically focusing on the problems associated with AI generated responses, while no instance of generation of directly harmful content as outlined by Huang and Huang (2023) and observed during conversations with other AI conversational agents such as Replika (Depounti, 2022) have occurred in conversations analysed in this study, motives of manipulation can be recognised in some of the answers given by the NPCs. These answers frequently were, next to the occurrence of "hallucinations" (Cox & Ooi, 2024), factual lies presented as truth even though the correct information was available to the NPC, while said NPCs oftentimes also insisted they were telling the truth when questioned. Players thus, justifiably, often perceived the non-player characters as deceptive actors. As current AI integration is not capable of recognising the actual virtual environment without prompts or player input and is therefore conversely vulnerable to player manipulation, the overall resulting conundrum is, essentially, a two-way avenue of possible "gaslighting" between the player and the NPC; a conversation dynamic that could platform misinformation or inappropriate sentiments from both sides in other games.

In sum, while an integration of AI-backed NPCs may offer a substantial number of benefits as elaborated previously, such implementation currently may introduce various corrosive factors and risks into a game – from characters appearing bland to shattered player immersion caused by nonsensical conversations with NPCs. While leading developers looking to implement this technology are aware of these potential drawbacks (cf. O'Brien, 2024), future plans to incorporate similar NPCs should pay special attention to these factors in order to not offset any benefits gained through the implementation in the first place, especially in areas of sensitive topics, such as commonly occurring in matters of psychological interventions. More research is required to ascertain whether an

implementation of this new iteration of intelligent digital conversational partners is a viable option and attractive opportunity or still too volatile in risks and disruptive elements to be a suitable implementation. Additionally, further research is also needed to investigate how and if these potentially disruptive patterns take shape in different iterations of video games and how they could be mitigated not only through further refinement on a technical level, but also on the user and developer side. Possible solutions might lie in more accurate model conditioning or manipulation of environmental factors, such as specifically curated conversational context, e.g. interrogation scenarios in games, or a more conservative approach in its implementation through a more mixed-methods approach in games, i.e. featuring partially scripted and partially AI controlled NPC conversations. Another approach might lie in narratively framing AI characters as possessing machine-like qualities in the game (e.g. humanoid robots) to internally justify some of the disruptive experiences caused by LLM integration.

Voice Interaction is Awkward. Examining the findings surrounding voice interaction, participants reporting initial unease but gradually becoming more comfortable engaging with NPCs via voice interaction reflects common findings of earlier literature investigating voice interaction, wherein a certain awkwardness or uncomfortableness was often described by players when talking to a computer (Allison et al., 2018; Allison et al., 2019; Carter et al., 2015; Dow et al., 2007, as cited in Allison et al., 2018). The occurrence of this phenomenon in this study shows that even in a more technologically sophisticated environment, previously commonly encountered difficulties may still resurface – and VI is no exception to that trend. Notably, while research peripherally outlining this inherent awkwardness in VI exist (cf. Reicherts et al., 2022), the phenomenon is still relatively unexplored, as the occurrence is oftentimes chalked up to more abstract terms such as the artificial feel of the conversation or the awareness that users are not talking to a real human.

Future studies pursuing research of voice interaction should thus consider investigating the causes of this inherent awkwardness and whether it is mitigated by conversational agents and simulated social settings becoming incrementally realistic or may continue to pose a “barrier of awkwardness” for players and negatively affect their experience even in future iterations of the technology.

Recognition Remains a Problem. Building on the trend of commonly encountered problems for voice interaction, the findings of the present study spotlight that inaccurate voice recognition remains a major disruptive factor for the player-to-NPC interaction in games featuring VI as the main mode for this interaction. As outlined by previous studies (Zargham et al., 2022; Zargham et al., 2024), misrecognition is a well-known problem for the application of voice interaction and often cited as one of the reasons why users usually prefer traditional, mechanical input methods over VI. It appears that even with integration into a VR environment, any gains in immersion and engagement through voice interaction implementation can be hindered by these inherent problems. This underscores the importance of further efforts to develop more reliable and efficient voice recognition technology for future devices to fully capitalise on the benefits that voice interaction can offer. The nature of existing VR systems makes them inherently less conducive for traditional, mechanical input methods, which commonly entails a restriction of player freedoms in, for example, answer choices in games or narrow conversation topics. VI has the potential to become the preferred method of communication for these systems if it can overcome its persistent challenges and move beyond its current status as a mere fun "gimmick" in video games.

Limitations and Strengths

After discussing the study's findings in relation to existing literature, it's essential to also acknowledge the study’s potential limitations and strengths. First, in terms of limitations,

rudimentary concerns arise regarding the sample size and composition, as the sampling methods extended to snowball and convenience sampling and may therefore introduce potential bias in the data and limiting extrapolation of insights to the general public. Specifically, a large portion of the sample resided in western cultures, while a great majority of the sample was also collected from a predominantly academic environment. Consequently, participant selection may have introduced certain underlying biases in the answers of participants such as, for example, social desirability biases due to occupational homogeneity or even personal acquaintance with the researchers, given their field of work.

Secondly, additive survey questions apart from the items of the PXI, as well as the questions utilised in the interview have not been piloted in a previous study or otherwise tested in terms of statistical reliability or validity, as well as overall clarity and exhaustiveness. This may introduce a degree of uncertainty regarding the robustness of the data obtained and may leave room for inconsistencies in participant responses. Therefore, while these questions were drafted and designed by experts to provide extensive and additional insight into the overall player experience, their limitations should be acknowledged and considered in the interpretation of the study results.

Thirdly, while the game featured in the study was subject to extensive QA-testing by the researchers and IT personnel, the final version of the game was still plagued by bugs, errors, and some other technical difficulties. Said problems may have compromised part of the player experience in some parts, i.e. players not being able to complete objectives due to quest-essential items disappearing through the floor or NPCs being unresponsive because of an unexpected loading loop. A potentially disruptive effect is further substantiated, given the number of participants that explicitly remarked problems with bugs, unresponsive NPCs, and other technical difficulties. Needless to say, these disruptions of gameplay may have very

well had further effects on the perceived immersion of participants, their entertainment and sense of control or mastery.

Lastly, results of the sentiment analyses should be met with greater scrutiny than other results, as said analysis tools tend to be a rather crude method of converting essentially qualitative data into quantitative ones. Additionally, as often visible in the highest scoring and overview of the analysis output, participant utterances toward NPCs were contaminated by terminology as given by the game's listed objectives by necessity, as participants had to engage NPCs verbally in order to complete them in the first place. Consequently, usage of phrases and words extracted from given objectives such as "gift", "birthday", or "annoyed" may be overrepresented and therefore distort results of said analysis tools.

Despite its shortcomings, however, the study also boasts several strengths. First and foremost, to the best of the author's knowledge, it is the first of its kind to examine player experiences, perception of and interactions with AI-backed NPCs in a virtual reality setting with voice interaction as the main tool of interaction, thereby not only boasting potentially new insights into the topics of study, but also setting itself apart from similar studies in the field. In a similar vein, the present study is also very timely. The interest in implementing generative AI and LLMs into video games is incrementally growing and implementation is promising new breakthroughs for digital entertainment sectors such as video games, but also interventions and educational environments – especially when coupled with VR and/or VI. Finally, the study incorporates a comprehensive multimethod approach towards studying areas of interest, suffusing efficient quantitative data from surveys with rich qualitative data gathered from interviews. The resulting data is not only comprehensive, but also rich and contextual.

Conclusion

The present study contributes to the body of knowledge of psychology and human computer interaction, specifically the research on user interaction with generative AI-driven NPCs through voice interaction in a VR environment, as well as contextual factors, benefits, and risks. To the best of the authors knowledge, this study is among the first to investigate the user experience of such a game, as well as how users perceive and interact with said NPCs. Findings indicate that players generally indicate their time with the game as a positive one and found the interactions with the NPCs to be natural and fun. However, problems crystallise within the framework of conversation with the NPCs, such as recognition issues and an artificial structure and flow. Additionally, challenges in large language model (LLM) implementation were observed, such as flattening perceived NPC personalities and difficulties in providing informative answers. The results underline that future endeavours to implement LLMs in VR environments that utilise voice interaction may achieve a substantial increase in immersion and believability of digital non-player entities along with other benefits; however, it's crucial to address the outlined challenges through technological advancements or compensatory game design choices.

Future research is encouraged to further explore the potential and possible risks of LLM implementation into NPCs in an effort to create more realistic game experiences. Current development of the technology promises many new possibilities for emulating convincing non-player entities in the digital realm that may not only open the door to novel and unprecedented immersive experiences, but also introduce entirely new and effective approaches to psychological interventions and applications. Finally, this paper stresses the need for further exploration of voice interaction in video games. Despite its showcased benefits of being able to provide an immersive and natural way to engage with the game, the technology is still rarely implemented in games. Thus, further research and development of

the modality is imperative to reap its potential benefits and address associated drawbacks in video games and interventive methods alike.

References

Abeele, V. V., Spiel, K., Nacke, L., Johnson, D., & Gerling, K. (2020). Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies*, *135*, 102370. doi:10.1016/J.IJHCS.2019.102370

Allison, F., Carter, M., Gibbs, M., & Smith, W. (2018). Design Patterns for Voice Interaction in Games. *CHI PLAY 2018 - Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 5–17. doi:10.1145/3242671.3242712

Allison, F., Newn, J., Smith, W., Carter, M., & Gibbs, M. (2019). Frame Analysis of Voice Interaction Gameplay. *Conference on Human Factors in Computing Systems - Proceedings*, 393. doi:10.1145/3290605.3300623

Allison, F., Carter, M., & Gibbs, M. (2020). Word Play: A History of Voice Interaction in Digital Games: *Games and Culture*, *15*, 91–113. doi:10.1177/1555412017746305

Andrus, B., & Fulda, N. (9 2020). Immersive Gameplay via Improved Natural Language Understanding. *ACM International Conference Proceeding Series*. doi:10.1145/3402942.3403024

Beres, N. A., Frommel, J., Reid, E., Mandryk, R. L., & Klarkowski, M. (2021). Don't you know that you're toxic: Normalization of toxicity in online gaming. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445157

Birk, M. v., Friehs, M. A., & Mandryk, R. L. (2017). Age-based preferences and player experience: A crowdsourced cross-sectional study. *CHI PLAY 2017 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 157–170. https://doi.org/10.1145/3116595.3116608

Bopp, J. A., Müller, L. J., Aeschbach, L. F., Opwis, K., & Mekler, E. D. (2019). Exploring emotional attachment to game characters. *CHI PLAY 2019 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 313–324. https://doi.org/10.1145/3311350.3347169

Boyd, R. L., & Schwartz, H. A. (2021). Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present, and Future States of the Field. *Journal of Language and Social Psychology*, *40*, 21–41. doi:10.1177/0261927X20967028/ASSET/IMAGES/LARGE/10.1177_0261927X20967028-FIG1.JPEG

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. *Human Communication Research*, *48*(3), 404–429. https://doi.org/10.1093/HCR/HQAC008

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., … Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. Retrieved from https://arxiv.org/abs/2303.12712v5

Cabe, L. N., Woolbright, L., & Stephenson, M. (2024). *Character Creation and Digital Storytelling using Generative AI*.

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, *9*, 48–57. doi:10.1109/MCI.2014.2307227

Carter, M., Allison, F., Downs, J., & Gibbs, M. (2015). Player Identity Dissonance and Voice Interaction in Games. *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 265–269. doi:10.1145/2793107.2793144

Cheng, A., Yang, L., & Andersen, E. (2017). Teaching language and culture with a virtual reality game. *Conference on Human Factors in Computing Systems - Proceedings, 2017-May*, 541–549. https://doi.org/10.1145/3025453.3025857

Chowanda, A., Blanchfield, P., Flintham, M., & Valstar, M. (2014). ERiSA: Building emotionally realistic social game-agents companions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8637 LNAI*, 134–143. https://doi.org/10.1007/978-3-319-09767-1_16/COVER

Chowanda, A., Blanchfield, P., Flintham, M., & Valstar, M. (2016). Computational Models of Emotion, Personality, and Social Relationships for Interactions in Games. *Conference: International Conference on Autonomous Agents & Multiagent Systems 2016*. https://doi.org/10.5555/2936924.2937151

Cox, S. R., & Ooi, W. T. (2024). *Conversational Interactions with NPCs in LLM-Driven Gaming: Guidelines from a Content Analysis of Player Feedback* (pp. 167–184). Springer, Cham. https://doi.org/10.1007/978-3-031-54975-5_10

Depounti, I., Saukko, P., & Natale, S. (2022). Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. *Https://Doi.Org/10.1177/01634437221119021*, *45*(4), 720–736. https://doi.org/10.1177/01634437221119021

Emmerich, K., Ring, P., & Masuch, M. (2018). I'm glad you are on my side: How to design compelling game companions. *CHI PLAY 2018 - Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 153–162. https://doi.org/10.1145/3242671.3242709

Eysenbach, G. (3 2023). The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Medical Education*, *9*, e46885. doi:10.2196/46885

Fan, C. W., Jiang, J. R., & Wu, J. W. (2011). Peer-to-peer immersive voice communication for massively multiplayer online games. *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, 759–764. https://doi.org/10.1109/ICPADS.2011.99

Fan, X., Wu, J., & Tian, L. (2020). A Review of Artificial Intelligence for Games. *Lecture Notes in Electrical Engineering*, *572 LNEE*, 298–303. doi:10.1007/978-981-15-0187-6_34

Fischer, J. E. (2023). Generative AI Considered Harmful. *Proceedings of the 5th International Conference on Conversational User Interfaces, CUI 2023*. https://doi.org/10.1145/3571884.3603756

Gutman, J. (1982). A Means-End Chain Model Based on Consumer Categorization Processes. *Https://Doi. Org/10. 1177/002224298204600207*, *46*, 60–72. doi:10.1177/002224298204600207

HTC VIVE. (2023, August 18). *AI Encounter! Getting Social with NPCs in Riff XR* [Video]. YouTube. https://www.youtube.com/watch?v=y9OmqoNyL4Y

Henderson, P., Hashimoto, T., & Lemley, M. (2023). *Where's the Liability in Harmful AI Speech?* http://arxiv.org/abs/2308.04635

Ho, J. C. F., & Ng, R. (2022). Perspective-Taking of Non-Player Characters in Prosocial Virtual Reality Games: Effects on Closeness, Empathy, and Game Immersion. *Behaviour & Information Technology*, *41*(6), 1185–1198. https://doi.org/10.1080/0144929X.2020.1864018

Huang, J., Huang, K. (2023). ChatGPT in Gaming Industry. In: Huang, K., Wang, Y., Zhu, F., Chen, X., Xing, C. (eds) *Beyond AI. Future of Business and Finance.* Springer, Cham. https://doi.org/10.1007/978-3-031-45282-6_9

Inworld AI (2023). *Inworld Origins* [PC]. Inworld AI

Jung, H., So, S., Oh, C., Kim, H. J., & Kim, J. (2019). TurtleTalk: An educational programming game for children with voice user interface. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312773

Kordyaka, B., Jahn, K., & Niehaves, B. (2020). Towards a unified theory of toxic behavior in video games. *Internet Research*, *30*(4), 1081–1102. https://doi.org/10.1108/INTR-08-2019-0343/FULL/XML

Kowalski, J., Skorupska, K., Kopeć, W., Jaskulska, A., Abramczuk, K., Biele, C., & Marasek, K. (2019). Older adults and voice interaction: A pilot study with google home. *Conference on Human Factors in Computing Systems - Proceedings*. doi:10.1145/3290607.3312973

Kwak, H., Blackburn, J., & Han, S. (2015). Exploring cyberbullying and other toxic behavior in team competition online games. *Conference on Human Factors in Computing Systems - Proceedings*, *2015-April*, 3739–3748. https://doi.org/10.1145/2702123.2702529

Köttl, H., Gallistl, V., Rohner, R., & Ayalon, L. (2021). "But at the age of 85? Forget it!": Internalized ageism, a barrier to technology use. *Journal of Aging Studies*, *59*, 100971. https://doi.org/10.1016/J.JAGING.2021.100971

Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., & Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms

from emotional dependence on the social chatbot Replika. *New Media and Society*. https://doi.org/10.1177/14614448221142007/FORMAT/EPUB

Li, M., Sun, Z., Jiang, Z., Tan, Z., & Chen, J. (2020). A Virtual Reality Platform for Safety Training in Coal Mines with AI and Cloud Computing. *Discrete Dynamics in Nature and Society, 2020*. doi:10.1155/2020/6243085

Lim, M. Y., Dias, J., Aylett, R., & Paiva, A. (2012). Creating adaptive affective autonomous NPCs. *Autonomous Agents and Multi-Agent Systems*, *24*(2), 287–311. https://doi.org/10.1007/S10458-010-9161-2

Liu, Y., & Agur, C. (2023). "After All, They Don't Know Me" Exploring the Psychological Mechanisms of Toxic Behavior in Online Games. *Games and Culture*, *18*(5), 598–621. https://doi.org/10.1177/15554120221115397/FORMAT/EPUB

Lv, Z. (2023). Generative artificial intelligence in the metaverse era. *Cognitive Robotics*, *3*, 208–217. doi:10.1016/j.cogr.2023.06.001

Mateas, M., & Stern, A. (2005). Façade [Video game]. Interactivestory.net

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28–37. https://doi.org/10.1016/j.chb.2019.05.009

Ning, Y., He, S., Xing, C., & Zhang, L. J. (2019). The Development Trend of Intelligent Speech Interaction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11518 LNCS*, 169–179. doi:10.1007/978-3-030-23407-2_14

O'Brian, L. (2024, March 19). *How Ubisoft's new generative AI prototype changes the narrative for NPCs*. news.ubisoft.com. Retrieved March 28, 2024, from https://news.ubisoft.com/en-us/article/5qXdxhshJBXoanFZApdG3L/how-ubisofts-new-generative-ai-prototype-changes-the-narrative-for-npcs

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology (London, England : 1953)*, *109*(3), 395–417. https://doi.org/10.1111/BJOP.12290

Kinetic Games. (2020). *Phasmophobia* [PC].

Pradhan, A., Mehta, K., & Findlater, L. (2018). "Accessibility Came by Accident." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, *2018-April*, 1–13. https://doi.org/10.1145/3173574.3174033

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*. doi:10.1145/3173574.3174214

Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services*, *58*. https://doi.org/10.1016/J.JRETCONSER.2020.102283

Reicherts, L., Rogers, Y., Capra, L., Wood, E., Duong, T. D., & Sebire, N. (2022). It's Good to Talk: A Comparison of Using Voice Versus Screen-Based Interactions for Agent-Assisted Tasks. *ACM Transactions on Computer-Human Interaction*, *29*(3), 25. https://doi.org/10.1145/3484221

Sardari, M., Moss, K. A., Zaman, K. A., Aghdaie, N., Kolen, J., Mattar, M. M., & Inc, E. A. (2017). *Interactive voice-controlled companion application for a video game* (Patent No. US10449440B2). USPTO. https://patents.google.com/patent/US10449440B2/en

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, *109*, 431–433. https://doi.org/10.1111/bjop.12305

Sony Interactive Entertainment. (2003). *Lifeline* [PlayStation 2]. Konami, Sony Interactive Entertainment.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology and Behavior*, *7*(3), 321–326. https://doi.org/10.1089/1094931041291295

Türkay, S., Formosa, J., Adinolf, S., Cuthbert, R., & Altizer, R. (2020). See No Evil, Hear No Evil, Speak No Evil: How Collegiate Players Define, Experience and Cope with Toxicity. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376191

Unreal Engine. (2020, May 13). *Unreal Engine 5 Revealed! | Next-Gen Real-Time demo running on PlayStation 5* [Video]. YouTube. https://www.youtube.com/watch?v=qC5KtatMcUw

Unreal Engine. (2024, March 21). *State of Unreal | GDC 2024 | Epic Games* [Video]. Y-ouTube. https://www.youtube.com/watch?v=0H63U0V11HI

van Sintemaartensdijk, I. (2022). *Burglary in Virtual Reality*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam

Vivarium. (2006). *Odama* [GameCube]. Nintendo.

Vrije Universiteit Amsterdam & University of Twente. (n.d.). *VR and AI in mediating Victims and Offenders: Interview Charlotte Gerritsen (VU) and Sven Zebel (UT)*. VU - UT | Research and Impact | Secure Societies. Retrieved April 18, 2024, from https://vu-ut.nl/en/research-and-impact/secure-societies/interviews/

Wadley, G., Gibbs, M. R., & Hew, K. (2005). Factors influencing users' decisions to adopt voice communication in online console games. *International Journal of Advanced Media and Communication*, *1*(1), 41. https://doi.org/10.1504/IJAMC.2005.007722

Waqar, D. M., Gunawan, T. S., Kartiwi, M., & Ahmad, R. (2021). Real-Time Voice-Controlled Game Interaction using Convolutional Neural Networks. *2021 IEEE 7th International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2021*, 76–81. doi:10.1109/ICSIMA50015.2021.9526318

Warpefelt, H. (2016). The Non-Player Character: Exploring the believability of NPC presentation and behavior (PhD dissertation, Department of Computer and Systems Sciences, Stockholm University). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-128079

Williams, D., Caplan, S., & Xiong, L. (2007). Can You Hear Me Now? The Impact of Voice in an Online Gaming Community. *Human Communication Research*, *33*(4), 427–449. https://doi.org/10.1111/j.1468-2958.2007.00306.x

Yan, J. (2021). Text Mining with R: A Tidy Approach, by Julia Silge and David Robinson. Sebastopol, CA: O'Reilly Media, 2017. ISBN 978-1-491-98165-8. XI + 184 pages. *Natural Language Engineering*, *28*(1), 137–139. https://doi.org/10.1017/s1351324920000649

Zargham, N., Alexandrovsky, D., Friehs, M. A., & Zebel, S. (2023). Speaking Up Against Toxicity: Opportunities and Challenges of Utilizing Courtesy as a Game Mechanic Multimodal Algebra Learning View project Usability of access control systems View project. *Proceedings of The CHI 23 Workshop on Combating Toxicity, Harassment, and Abuse in Online Social Spaces*, *1*. https://www.researchgate.net/publication/369542750

Zargham, N., Pfau, J., Schnackenberg, T., & Malaka, R. (2022). "I Didn't Catch That, But I'll Try My Best": Anticipatory Error Handling in a Voice Controlled Game. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3491102.3502115

Zargham, N., Bonfert, M., Volkmar, G., Porzel, R., & Malaka, R. (2020). Smells like team spirit: Investigating the player experience with multiple interlocutors in a VR Game. *CHI PLAY 2020 - Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*, 408–412. https://doi.org/10.1145/3383668.3419884

Zargham, N., Friehs, M. A., Tonini, L., Alexandrovsky, D., Ruthven, E. G., Nacke, L. E., & Malaka, R. (2024). Let's Talk Games: An Expert Exploration of Speech Interaction with NPCs. *International Journal of Human–Computer Interaction*, 1–21. https://doi.org/10.1080/10447318.2024.2338666

Zeng, G. (2023). *A review of AI-based game NPCs research*. doi:10.54254/2755-2721/15/20230827

Appendix A

Consent Form

Study Title: Speech Interaction with Non-Player Characters
Principle Investigator: Lennart Nacke, Associate Professor, University of Waterloo
University of Twente Investigators:
Leandro Tonini, Master Student
Emma Ruthven, Intern
Maximilian A. Friehs, Assistant Professor, University of Twente

The **purpose** of this project is to investigate users' perceptions of voice interaction with non-player characters (NPCs) in a VR game. This research project is being conducted as part of a larger research line on voice interaction in video games. Researchers from the University of Bremen, the University of Twente, the University of Waterloo, and the Karlsruhe Institute of Technology are involved. You are invited to participate in this research project. Your participation involves playing a speech based virtual reality video game where you will interact with game characters using speech.

The **procedure** involves filling in a general demographics questionnaire, playing two levels of the game, filling in the post-exposure questionnaires, and participating in a short interview. The session will take approximately 45 minutes, including a 20-minute VR game session. We ask that you answer the questions truthfully. You are, however, free to quit the session at any point in time by notifying the experimenter.

Your participation is completely **voluntary**, and you can stop without stating reasons and without further consequences. Already collected data can then be deleted. After concluding the study and anonymizing your data, it can no longer be identified or deleted on your request.

As part of the research process, we make **audio recordings** of the session for later, detailed analysis. The recordings will be saved, accessed, and processed by the researcher(s) listed below. Your data will be anonymized before analysis or publication, so it will not be possible to infer your identity. The anonymized transcripts of the recordings from the session may be used in academic presentations or documentation of this study. You can ask to pause the recording at any time during the interview.

All data is stored in a password-protected electronic format. To help protect your privacy and personal data, the survey will not contain information that will personally identify you. Your data will be anonymized and securely stored in an encrypted drive. Data will be stored for a minimum of 8 years. The results of this study will be used for publication in the form of a research report and may be shared with the University of Waterloo representatives as well as the wider scientific community in an anonymized form.

Study Risks: When using virtual reality systems, it is important to be aware of potential side effects that may occur. These side effects can include symptoms of motion sickness, such as nausea, headache, and dizziness. It is worth noting that these symptoms typically subside within a few hours after using the technology. While severe adverse effects are extremely rare (occurring in approximately 1 in 4000 cases), it is essential to be aware of their possibility. These severe effects may include intense dizziness, seizures, eye or muscle

twitching, or even temporary loss of consciousness. It is important to note that the likelihood of experiencing these severe effects increases if you are fatigued, under stress or anxiety, or currently dealing with a cold or other illness.

Who may participate: You may participate in this study if you are 18 or over and have normal or corrected to normal vision. You may not participate in this study if you are pregnant, have any known musculoskeletal (back or bodily injuries) impairments, any vestibular (balance), hearing or visual deficits, neurological disorders (seizures, epilepsy, Parkinson's, etc.), medical implants (e.g. heart pacemaker) or if you are experiencing any anxiety or post-traumatic stress or suffering from cold/flu/headaches.

In appreciation of your participation, you may receive SONA course credit from UT as well as participate in a raffle for a voucher.

Ethics: This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45294) as well as the University of Twente BMS Ethics Board (#230076 & 230787).
If you have any **questions** about the research study, please contact: Leandro Tonini, l.tonini@student.utwente.nl

By signing this form, you acknowledge and **agree** to the following:

1. You have carefully read and understood the information provided above regarding the user study.
2. You are aware that your interview will be audio recorded to ensure an accurate recording of your responses.
3. Your participation in this study is voluntary, and you have chosen to participate willingly.
4. You meet the specified inclusion criteria for participation in this study.

By providing your consent, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

Appendix B

Survey and Interview Questions

Table B1

PXI Constructs & Items

Constructs	**Items**
Meaning	Playing the game was meaningful to me.
	The game felt relevant to me.
	Playing this game was valuable to me.
Curiosity	I wanted to explore how the game evolved.
	I wanted to find out how the game progressed.
	I felt eager to discover how the game continued.
Mastery	I felt I was good at playing this game.
	I felt capable while playing the game.
	I felt a sense of mastery playing this game.
Autonomy	I felt free to play the game in my own way.
	I felt like I had choices regarding how I wanted to play this game.
	I felt a sense of freedom about how I wanted to play this game.
Immersion	I was no longer aware of my surroundings while I was playing.
	I was immersed in the game.
	I was fully focused on the game.
Progress Feedback	The game informed me of my progress in the game.
	I could easily assess how I was performing in the game.
	The game gave clear feedback on my progress towards the goals.

Constructs	**Items**
Audiovisual Appeal	I enjoyed the way the game was styled.
	I liked the look and feel of the game.
	I appreciated the aesthetics of the game.
Challenge	The game was not too easy and not too hard to play.
	The game was challenging but not too challenging.
	The challenges in the game were at the right level of difficulty for me.
Ease of Control	It was easy to know how to perform actions in the game.
	The actions to control the game were clear to me.
	I thought the game was easy to control.
Clarity of Goals	I grasped the overall goal of the game.
	The goals of the game were clear to me.
	I understood the objectives of the game.
Enjoyment*	I liked playing the game
	The game was entertaining
	I had a good time playing this game

** "Enjoyment" is not a construct of the PXI, but represents an instrumental value reflecting resulting user engagement and thereby an integral part of the inventory*

Survey Questionnaire Additive Items

SQ1: “Which character was your favourite?”

SQ2: "Why was this character your favourite?”

SQ3: “Which character did you dislike the most?”

SQ4: "Why did you dislike the character the most?”

SQ5: "Please rate each character in terms of their attractiveness.”

SQ6: “Please rate how much you enjoyed interacting with each character.”

SQ7: “How would you rate your performance in the game?”

SQ8: “Would you be willing to play similar games (being able to talk to game characters using your voice) in the future?”

SQ9: “How would you rate your overall game experience?”

Post-Study Interview Items

IQ1: “What did you like about the game? Why?”

IQ2: “What did you dislike about the game? Why?”

IQ3: “What was the most interesting aspect of the game?”

IQ4: “What was the least interesting aspect of the game?”

IQ5: “What did you think about the game characters?”

IQ6: “How did talking to the game characters make you feel?”

IQ7: “What did you find attractive in the game?”

IQ8: “What did you miss with regard to the conversation with the characters?”

IQ9: “How believable were the conversations with the characters?”

IQ10: “Did you experience any issues with regard to voice interaction?”

IQ11: “What did you think of the speech interaction in the game?”

IQ12: “Do you have any further comments?”

Appendix C

VR Game Non-Player Character Cast

Figure C1

Overview of NPC Cast in Version 1

Figure C2

Overview of NPC Cast in Version 2

Appendix D

Interview Questions Coding Book

Table D1

Finalised Coding Scheme – Controls & User Interface

Code Group	Codes	Description	Example	Frequency
Controls	Accessible Controls	Participant reported the game's controls as intuitive or easy to access	"The controls are very easy to access after a while, playing the game is very intuitive."	6
	Cumbersome Controls	Participant described the controls as hard or unintuitive to access, use or navigate	"Sometimes, I had difficulty grabbing some objects. And sometimes the teleportation wouldn't work, so I had to click the button for several times." "Sometimes, it was really hard to do, like, do something with the controls."	18
	Disorienting Controls	Participant illustrated the controls as disorienting or perceptually confusing	"I disliked, uh, looking around, it was just slightly too choppy. Would have been nice if there were just slightly more frames, because now it was almost like every turn was just slightly too big."	9
	Learning Controls Fun	Participant expressed liking the learning curve the controls posed for players	"I think I like just learning the controls to be honest."	1
User Interface	Facilitative UI	Participant alluded to elements of the user interface being facilitative to player experience or otherwise helpful	"I like the scoreboard where you can see the task and then you had a clear overview of what you had to do, but you also had to figure our and think for yourself a bit." "I think it was really good that you also	13

Code Group	Codes	Description	Example	Frequency
			put [speech bubbles] there. It was really nice to follow."	
	Cumbersome UI	Participant described elements of the user interface as hard to access or unintuitive	"Some of the [UI] objects were particularly small to me and I had to step up closer to read it."	1
	Distracting UI	Participant communicated viewing elements of the user interface as distracting	"I also didn't like the [speech bubbles]. Those were distracting to me."	1

Table D2

Finalised Coding Scheme – Game Design

Code Group	Codes	Description	Example	Frequency
Game Context	Future Application	Participant conveyed a vision or conception of a possible future application of the game concept in a suitable area	"It's a well thought idea [sic] and it works. As I tried it myself, I think it's a good idea to involve such NPCs in video games or interventions." "I think it would be fun in video games." "You can talk to someone; you can practice different languages with people, and you can practice saying out loud those things that you never want to speak in your real life."	13
	Novel Positive Experience	Participant stated they perceived the game experience as a novel and positive experience	"It was unlike anything I've ever experienced before. This was novel. It was so fluid and … fun." "It was good. It was something new. I liked experiencing it."	28
	Suitable Genres	Participant mentioned suitable genres for an application of the voice interaction in other video games	"[You could implement them in] RPGs and stuff because you can pretty much approach anyone and have a conversation with them." "I've never really played a game with speech interaction before, and I could really see, as I said	5

Code Group	Codes	Description	Example	Frequency
			before, that it could be very interesting for also some leisure experiences, like [a] murder mystery, for example.”	
	Too Little Playtime	Participant reported that they felt like they had too little playtime in their playthrough	“It was too short. I didn’t really get to play that much. You know, [experience] the whole story. And what is next.”	2
Gameplay & Structure	Immersive Gameplay	Participant illustrated the gameplay experience as immersive, absorbing, or realistic	“It was kind of immersive. …I did feel sort of [a] physical presence there.” “It seemed so real. It just seemed like you were there, right? That you weren’t sitting and playing the game. You just felt part of the game.”	30
	Interesting Game Structure	Participant described elements or parts of the overall game structure	“I like the most that you had like the different characters that you could interact with and that you had a clear [objective list] on what to do.” “I think this is definitely very intriguing idea. And as I said, we human beings love just talking, socializing with other people. This element of talking. Would definitely help games to become more interactive, more engaging, and of course, more immersive.”	16
	Interesting Story	Participant related parts or the overall narrative of the game as	“I liked the storyline. I liked how it related to our role here”	3

Code Group	Codes	Description	Example	Frequency
		interesting or intriguing		
	No Predefined Answer Options	Participant communicated a sentiment about the absence of predefined answer options	"I think it is very interesting - It does make the game a lot more appealing to me at least, as if you would just have choice ABC or something that you could pick from." "Weird talking to computers that way, I guess. That they're not actual persons. And usually, you can choose which answers, which questions you want to ask, and now you have to formulate them yourself. So weird and a like nuisance." "I thought it was a fun way to interact with the characters instead of picking your dialogue."	16
	Player Freedom	Participant stated feeling or recognizing a sense of freedom or heightened agency as a player in the game	"You could approach the task from whichever point of view you wanted. So, I could have gone to any other character to start the task and made my way around." "It feels more like you control the game instead of the game kind of controls where you are supposed to play it."	18
Missing Features	UI Prompts	Participant expressed missing guidance or nudging by user	"Perhaps a prompt? Like, if you [are] stuck on a certain point for	6

Code Group	Codes	Description	Example	Frequency
		interface prompts, or helpful user interface tools	a bit, [the game] could just prompt off what next to do or a reminder."	
	More Informative NPCs	Participant underlined missing more informative NPC conversations while playing the game	"I would say [the NPCs] could give more information." "It would be nice to get an answer also if they don't understand what you say."	4
	Explicit NPC Names	Participant highlighted a missing feature in insufficient explicitness regarding the NPC's names	"The names with the characters. I couldn't see them on the screen. To know who is who."	2
	Flexible Responses	Participant noted missing flexibility in the content or form of the NPC conversation answers	"Like generating different responses that don't copy what the other person passed beforehand."	2
	In-Game Guidance	Participant illustrated missing in-game guidance, such as explicit help from NPCs or clues/helpful items	"Maybe [the NPCs] could give a bit more hints to who to talk to next, maybe. Or more subtle hints."	5
	NPC to NPC Conversation	Participant described missing conversations occurring between the game's NPCs	"I really like the task to find out what the first birthday gift was for Parker. And then the manager told me that she needed to discuss that with the intern. I really anticipated that she would go over there. And then they [would have] a conversation where you could interject. That [not happening] was unfortunate."	2

Code Group	Codes	Description	Example	Frequency
	Active NPCs	Participant expressed missing NPCs being active, either through interacting with the environment or moving around	"I disliked that I was the only one that could move and do stuff. Maybe they could walk around, too. It would be a little bit more realistic situation."	2
	Background NPCs	Participant communicated missing background NPCs as part of the environment	"The only thing I would change was adding more characters, not even that you interact with [sic], like NPCs."	1
	Backchanneling	Participant mentioned missing natural backchanneling from NPCs during conversation	"They didn't have very natural backchanneling. So, when you are talking to people you expect to have a natural backchanneling. …But not having that proper backchanneling was not very good for me."	2
	Precise Recording	Participant underlined an absence of precise recording/recognition in the verbal input	"Maybe an improved voice recording that I can also talk unclearly, because that makes it feel more natural. So, the more unclear the recording is able to record my language, the more realistic it feels to me."	1
	Verbally Distinct NPCs	Participant stated missing more verbally distinct or unique NPCs in either intonation, pronunciation, or accent	"So, what I miss that is coming from the character interaction… I think it's important to consider the social norm. So, of course this is a professional environment. I would expect it to be professional in how they	2

Code Group	Codes	Description	Example	Frequency
			respond to things. But, also, if you were in a different country or a different place, they might have certain dialects that are important, and which makes it a different experience already. So, tone matters, dialect."	
	Specific VI Boundaries	Participant described missing clear or well-defined boundaries upon which voice interaction with the NPCs would start	"[Being] within a certain distance of them or I have to be facing them or something:"	1
Objectives	Achievable Objects	Participants stated that the objectives given by the game felt achievable or possible to complete	"It was easy to understand the tasks – you could find them easily." "I thought they were doable. I think they were good."	4
	Background Main Objective	Participant criticised the main objective of the game falling into the background or being less present	"I expected [the main objective] to be more important." "Actually, finding the laptop … it kind of got into the background for me personally."	2
	Interesting Objectives	Participant noted interesting aspects or an engaging nature of the game's objectives	"The assignments that we had. Finding the presents of Parker and also finding the numbers for the locks. It was kind of interesting." "It gave me an escape room vibe. It seemed like you had to kind of figure out where the things were. I think the objective was quite fun."	17

Code Group	Codes	Description	Example	Frequency
	Uninteresting Objectives	Participant illustrated the sum or an aspect pf game's objectives as uninteresting or otherwise unengaging	"Having to walk around to find stuff. Well, I think the talking part is the interesting one. And then going around finding stuff, maybe not." "I think it was kind of weird to, just bring stuff from A to B, just – Yeah. Fetching things."	16
	Vague/Unintuitive Objectives	Participant conveyed a perception of vagueness and/or intuitiveness sharing their thoughts about the game's objectives	"I didn't manage to solve or obtain [the objectives] So, not too difficult, but maybe a bit vague."	13
Virtual Environment	Appealing/Realistic Environment	Participant reported the virtual environment as realistic or appealing to them	"I liked how it was like a classical office. I think it's funny for it to just be like an office and just be like a normal day life or something." "I would say the background was really realistic."	10
	Interactive Environment	Participant communicated viewing the world as very interactive	"I like just being able to move around on myself. Being able to just grab random stuff is funny."	11
	Intuitive/Explorable World	Participant underlined the virtual world as very explorable or approachable	"I think I find attractive the opportunity to have an open world. Open things and close things. That was really cool to be able just to go anywhere, even though you do have tasks, to be able to do it."	11
	Uninteresting Environment	Participant reported perceiving	"The least interesting was	13

Code Group	Codes	Description	Example	Frequency
		the virtual environment as uninteresting or unengaging	probably the office setting. I'm just not very interested in offices. It's just a little bit of a boring environment.”	
	Unintuitive Environment	Participant alluded to the virtual environment being hard to navigate or unintuitive in structure or layout	“So, I think I dislike that it took me a really long time to know where everything is. So, I took some time there to find the right rooms.”	2
Visual Style	Attractive Visual Style	Participant noted the overall visual style of the game or elements of it as visually appealing or aesthetically pleasing to them	“Maybe just style of the game in general. It was something that you are used to when you play with your games and it's attractive and maybe a pleasant thing for your eyes.” “I thought the game characters seemed very well designed.”	13
	Unappealing NPC Visuals	Participant mentioned finding the overall NPC visuals or aspects of their visual design unappealing	“I didn't like the way [the NPCs] look. They looked very fake. It just really looked like a scarecrow not like a person or something.”	5
	Unattractive Visual Style	Participant communicated the visual style of the game or elements of it as unattractive	“I am going to be honest; I don’t find it a very visually appealing game.”	4
	Unimmersive Graphics	Participants described the general graphics of the game or elements of it as unimmersive or breaking their suspension of disbelief	“The graphics were not really high and good. So, you know you're in a game.”	2

Table D3

Finalised Coding Scheme – Non-Player Characters

Code Group	Codes	Description	Example	Frequency
Believability Rating	Believable Conversations	Participant assessed the in-game conversations as believable	"They were they were believable. Like they say, 'I don't know' or 'Sorry, I don't know'. That was pretty nice, I think."	7
	Somewhat Believable Conversations	Participant assessed the in-game conversations as somewhat believable	"On a scale of 1 to 10, probably a solid seven. The not having the ten for the minus three that I'm pointing away is probably just for the tone. Um, it's very monotone for me."	20
	Split on Believability	Participant was unsure about the believability of in-game conversations or described NPCs has half human, half machine in conversation	"It's like in the middle because there are kind of believable because when I asked, 'Do you know that?', they gave me an answer. But on the other hand, they always said the same thing. They have like the same sentence structure, every character.	3
	Rather Unbelievable Conversations	Participant assessed the in-game conversations as rather unbelievable	"It was not really that much realistic, I think. I did not get the feeling 'Okay, I'm talking to a real human'. This feeling wasn't there."	8
	Unbelievable Conversations	Participant assessed the in-game conversations as unbelievable	"Not believable…. Like I said, the lack of intelligence. These characters were very limited because all they are trained in is the correspondence of the task within the game. They're not able to branch away from that, which I find very limiting and very frustrating…"	1

Code Group	Codes	Description	Example	Frequency
Believability Aspects	Action	Participant underlined movement and activity as an aspect of believability	"I would appreciate if the character can move and come to me to prove the password for me or something."	3
	Conversation Flow	Participant illustrated the flow of conversation as a denominator for NPCs as a believable conversation partner	"For example, the Manager. I talked to him also about one thing, and then I wanted to finish the conversation, and then he made like a whole paragraph why he was the manager and what he needed to do, the things that he needed to do."	2
	Correct Information	Participant highlighted providing correct or informative answers/conversation as an aspect of believability	"I once asked the girl in the front to where the kitchen is because I got a hint about the kitchen, and she said it was downstairs, but I couldn't go downstairs. So that was kind of confusing. But when I asked, [the Manager], he told me that the code from the storage room is on the kitchen fridge. That was helpful because, um, that actually answered my question."	1
	Emotions	Participant conveyed the expression of emotions as an aspect of believable conversations with NPCs	"I would rather those characters to [have] a bit more real emotions"	4
	Gestures & Facial Expression	Participant communicated conversation motion in the form of gestures and facial expressions as an aspect of believable NPC conversation	"I would like motions in their bodies and hand motions. Body movements and stuff. Gestures."	4
	Inquiry	Participant indicated inquisitive answers or counter-questions by the	"Maybe that they can also ask questions back, maybe that the	3

Code Group	Codes	Description	Example	Frequency
		NPC as an aspect of believable conversation	interaction is a bit more going and that you can get more information from asking and not always responding with the same phrases, maybe."	
	Personalities	Participant illustrated a presence or perception of personality in the NPC as an aspect of believable conversation	"I didn't really feel like they all had a different personality, but that might be because you don't really have the time to really get to know them because you're just focused on the goals, especially."	6
	Politeness	Participant touched on the presence of courtesy as a norm in interactions as an aspect of believable NPC conversation	"I think if it were to be more realistic…. I couldn't just go to you and say, 'What is the password?'. Maybe [the NPC] would be like 'How about you say hello first'?"	3
Interaction	Artificial Conversation Flow	Participant stated perceiving and artificial nature or structure in the mode conversation was held with the NPC	"The conversation wasn't as natural as it would be with a real human because you don't have that ping pong."	16
	Attractive NPC Voices	Participant noted being pleased with the voice tone and volume	"The tones were also nice. It didn't bother me. The tone of the characters were neutral. I like it that way. Yeah. My preference is actually neutral tones. Not very loud noises."	1
	Deceptive NPCs	Participant highlighted the perceived deceptive nature of the NPCs	"I find it interesting that characters don't give you the correct information, but that they lie a little bit…. Had nothing like that in a game before."	23
	Fun NPC Lies/Errors	Participant reported an interesting or fun situation or engagement based on a lie or erroneous statement of an	"I just wanted to add again that it is really funny when the AI just gives you some wrong information. The	6

Code Group	Codes	Description	Example	Frequency
		NPC	3am example, I immediately looked outside - it was sunny…. I think that's really funny."	
	Fun/Interesting Interactions	Participant noted having one or more interesting or fun verbal interactions with the NPCs	"Actually, the interaction with the NPCs, how they would respond, [I] found it quite funny how they responded sometimes, it was pretty enjoyable."	34
	Inconsistent Terminology	Participant criticised an apparent inconsistent terminology between the objectives, written information, or conversation	"Yeah, I think just the terminology depending on who makes the game, because to me, a bin – I was looking for a box bin, not the garbage. Maybe just terminology in that aspect. I felt I wasn't understanding really. And that could have helped me be more successful in the game."	1
	Monotone NPC Voices	Participant underlined a perception of the NPC voices as monotone and not different between the NPC	"Across the characters? No differences. I think most of the tone was monotone. [The NPCs] seem very similar across each character."	5
	NPC Corrective Response	Participant highlighted the NPCs being able to correct or repair participant input in cases of misrecognition or unclear communication	"The people in the game understood what was said and didn't make any mistakes that often or misunderstood what I was saying. I thought they would – maybe I'm not talking to clearly or something like that, and maybe they couldn't understand what I was saying, but they actually corrected something."	2
	Organic/Adaptive NPCs	Participant illustrated the NPCs as being able to adapt to their input	"I think it's very interesting that they have an answer to everything you ask.	28

Code Group	Codes	Description	Example	Frequency
		and/or show flexibility in their answering of question or general communication	So, it's not like you do not feel like it's programmed or something.”	
	Realistic/Natural Conversation	Participant reported in-game conversations with NPCs to feel realistic or close to a natural conversation	“It felt like a human interaction mostly. I got really the responses that I expected, and that surprised me. It didn't feel anything like a robot or something.” “It made me feel like I was somewhat talking to a real person.”	42
	Response Machine	Participant indicated the perception of an NPC as a conversation partner in terms a passive response dispenser or information vehicle	“It feels less like I’m having a conversation with a human being, but just with a response machine.” “They just went ‘I don't know’, or ‘You can do this, or you can do that’. It wasn't really like a conversation. Like, if I would ask a person here like. ‘Hey, do you know this or that?’ I don't think anyone would say, ‘No, I don't.’ and then just go. And then like, stop the conversation. You were in a game. It wasn't like in real life or anything.”	29
	Unaware NPCs	Participant touched on the NPCs being unaware or their actual surroundings and the consequent impact on conversation	“I am able to gaslight certain characters and certain responses; I kept trying to be able to do that. … I think I was successful in some parts, but obviously there might be a limitation on how the AI might respond to a certain way.”	2
	Unclear NPC	Participant noted an element of	“I enjoyed it, but at some point you do	3

Code Group	Codes	Description	Example	Frequency
	Communication	vagueness or unclarity in the communication of NPCs	feel a little frustrated because they weren't answering the question that you were asking … I was trying to figure out how to ask it better so that they would understand."	
	Unhelpful NPC Responses	Participant highlighted receiving unhelpful responses from the NPCs	"I didn't like the fact that, like, the one person was supposed to know the answer to a couple things and just wouldn't give me the answer." "The characters often didn't know the answer to my questions."	27
	Useful NPC Responses	Participant highlighted receiving helpful responses from the NPCs	"I felt like there was quite a wide array of stuff you could ask them, and then they would still produce some kind of useful answer."	5
Motion & Expressions	Good NPC Gestures	Participant underlined the presence of well-defined or visible gesturing by the NPCs during conversation	"I also liked the manners of some of them. Some of them moved a little bit while they talked."	2
	Mismatching Facial Expressions	Participant noted a potential mismatch in conversation content and the displayed facial expression	"Aria always looked very grumpy. I don't know if that was intentional or not. But I was a bit confused because she didn't always look grumpy. She kind of looked grumpy in certain interactions, and I wasn't sure if she actually based on content that I'm saying. Or she's just sometimes like this."	5
	No NPC Movement	Participant criticised the absence of active movement in NPCs	"It could be better if the character, if I could also go on, go around and see the characters in different places."	5
	Disliked Character	Participant noted disliking the	"I guess [I disliked] the style and the	1

Code Group	Codes	Description	Example	Frequency
	Movement	movement of in-game characters	way the characters moved.”	
Personality & Character	Approachable NPCs	Participant alluded to in-game characters being approachable and/or friendly	“I liked how Aria was just there and he was just like a new guy, so it was kind of useless. I just thought it was like, [a] cool guy. I kind of related to him. I'm also like new here.” “[It made me feel] like I was talking to my friends. Like they call me, you know. Kind of relaxing.”	9
	Bland NPC Personalities	Participant conveyed perceiving an absence of character or personality in the game’s NPCs	“Yeah, they just didn't really have any personality.” “Honestly, right now they feel a little [similar]. I could tell that they had some sort of personality. Like they had a bit of personality. One of them even had an accent and such. Um, but currently, right now, they felt very similar. Similar. Okay. Nothing too different.”	12
	Dislike of Most NPCs	Participant stated disliking most or all of the game’s NPCs	“I almost disliked all of them. I don't know why. But of course, if you get answers that don't lead to the solution of a problem, then it can, uh, get very annoying.”	2
	Distant NPCs	Participant illustrated the NPCs as distant and hard to approach or engage	“I think the characters didn't help me that much, honestly. I felt like I was a bit left on my own. They seemed a bit distant…” “I don't know. It still felt quite like a distance from them because they're not that human to me.”	3

Code Group	Codes	Description	Example	Frequency
	Distinct NPC Personalities	Participant mentioned perceiving distinct or different personalities in the game's NPCs	"Some of the characters seem to have their own personalities. They kind of joked around a little bit and had their own interests. I think one of them mentioned like an NFL game at some point. It was nice to see that some of the characters had something other than kind of like 'yes or no' dialogue options gave them real personalities.	6
	Diverse NPCs	Participant underlined a perception of diversity in the game's NPC cast	"I liked the diversity in the characters." "[The characters] were diverse and a sense that they had different roles and therefore could give you a different information."	3
	Favourite NPC: Kai	Participant alluded to the Manager NPC being their favourite character	"For the very limited time I spent talking to them, I did I mean the manager was kind of funny because he kept telling lies."	3
	Favourite NPC: Parker	Participant alluded to the Designer NPC being their favourite character	"Maybe Parker, because he told me who broke into the office or the server room."	5
	Favourite NPC: Ali	Participant alluded to the Mechanic NPC being their favourite character	"I thought Ali was, like, really nice. Sympathetic."	7
	Interesting NPCs	Participant communicated perceiving the game's NPCs as interesting or feeling curious about them	"Maybe getting them a little like – I really like when they said a little bit of the background of all the characters, like when Ali said, 'Oh yeah, Aria. She doesn't know she's new.' Yeah, it's a little background to the questions being asked. That was really cool.	5

Code Group	Codes	Description	Example	Frequency
			Like, getting to know them."	
	Role Representative NPCs	Participant illustrated perceiving role representative functions in the NPCs in the context of their position in the office setting	"I sort of interact with them based on how I would [with] their counterparts in a real office. So, like the manager, I would probably ask just specific questions, whereas the others might be more relaxed towards. I felt that the maintenance person only was very good with keeping their security roles but made it challenging for things."	1

Table D4

Finalised Coding Scheme – Virtual Reality

Codes	Description	Example	Frequency
Bug Fun	Participant mentioned an interesting or fun experience as the consequence of unintended features/bugs	"The fact that the trash disappeared [was interesting]."	1
Motion Sickness	Participant reported feeling motion sick during or after the playthrough	"I get motion sickness, apparently. Yeah. So that's what I dislike." "I don't know if it's anything to do with the glasses or with the game. I still feel really motion sick afterwards."	10
Novel Technology	Participant highlighted the novelty of the overall technology or game experience	"[The speech interaction is] definitely something I've never seen much before. I knew it was possible, but I think that it could be the start of something." "[The most interesting aspect of the game] has to be the voice. Anything with the when talking to the NPCs. I had never done that before."	10
NPC Gaze Problems	Participant stated perceiving unsettling or problematic instances with the headtracking of NPCs looking at the player	"It was kind of creepy that they would all look at me throughout the game…. Even when I was behind the character." "I think it was good, unless they looked at me from [behind] when they turned their head. Yes, a little bit creepy."	4
Technical Difficulties	Participant reported technical difficulties	"The fact that Ali didn't talk to me. It was	12

Codes	Description	Example	Frequency
	while playing the game	like just sitting and was looking at me, so I didn't know how to, talk to him. And I tried so many times, but he just said nothing." "There was some delay between catching my full question. Sometimes I had to wait till I hear full response of the person, wait till they're waiting for my question, and then ask my question."	
VR Fun	Participant conveyed the virtual reality experience as interesting	"I think just having just being able to sort of do things in like a virtual sense. The putting in the passcodes or like just doing stuff in the game [was the most interesting aspect of the game]."	1

Table D5

Finalised Coding Scheme – Voice Interaction

Code Group	Codes	Description	Example	Frequency
Interaction	Negative Story Impact	Participant communicated the negative potential for game storytelling speech interaction may hold	"In terms of we'll having these type of AI NPC characters affects narratively any interests that we can generate for [structured/pre-written] stories…. I think that would defeat the purpose of the story if [I were] able to interact with them versus seeing the intended relationship between the characters as the game as intended."	1
	Solution: Signalling Recording	Participant elaborated on signalling the recording of verbal input as a possible solution for potential issues in voice interaction	"I think they could stop the voice recording between when he answers. Maybe just make dots or something so that you really know 'Oh, he's answering right now, I cannot interrupt him."	2
	VI as Additional Input	Participant poised voice interaction as an additional mode of input for interacting with the game	"Like, I think that quite honestly, if I had some experience in a game, it would actually help with speeding up the tasks of a game to be able to control things with my voice instead of trying to control everything with my hands. I think it makes it feel a	3

Code Group	Codes	Description	Example	Frequency
			little bit more natural."	
	(Onset) Uncomfortableness	Participant described an initial uncomfortableness when utilising voice interaction in the game	"It felt a bit weird because you were sitting in the room it feels great to talk to something that is not there, but you see it. No one else can see it. But I got used to it. So, yeah, it was a bit difficult to start and know where to say something, what to say. So, I felt a bit weird and uncomfortable, but it was gone fast."	14
	Input Commitment	Participant conveyed feeling a sense of commitment when employing voice interaction	"I do like [the speech interaction because] it also helped me speak because I kept stuttering in the beginning and I'm like 'Okay, I am not going to actually think about what I'm going to say and formulate those questions' and instead of typing it out because when you're typing it, you kind of just like read it and nod and close it, whereas this sort of thing, I've already said it. So, I can't really take it back."	1
	Gradual Ease	Participant reported a process of getting used to or easing into the use of voice interaction in the game	"At first it was really weird because you knew that [the NPC] wasn't a real person. So … you needed to get a feeling for [speech interaction]. But I	10

Code Group	Codes	Description	Example	Frequency
			think in the end it was easier to do that. So, if you would do that more often, you would not have any problems to be talking to that."	
	Privacy Concerns	Participant highlighted potential concerns with privacy through the (continuous) recording of verbal input	"And I [think] that could also be a security think, a privacy thing. Because if I am actually playing this game at home, it would be nice if I am talking to an AI, and I am [using push-to-talk] to speak rather than picking up everything that's going on around the house."	1
	Retry Input	Participant underlined instances or the possibility of having to retry verbal commands or input due to error or miscommunication	"When you see a text that you know is badly recognized, sometimes it's beyond salvaging it. You can't. You just have to wait for the response and to start anew."	3
	Turn-based	Participant noted the turn-based nature of voice interaction in the game	"I think that was the ping pong. … I feel like they didn't listen when they were talking themselves. So, they wouldn't stop [talking for me to] try to put words [in] or correct myself or something."	5
	Effective Information Vehicle	Participant illustrated speech interaction as a good way to access information	"[Speech interaction] was a good way to sort of access information, and just look for clues."	1

Code Group	Codes	Description	Example	Frequency
	Voice Interaction Boundaries	Participant stated perceiving and finding trouble with the interaction boundary for voice interaction	"I did notice sometimes when I was kind of talking to myself or when I was walking away. They were hearing my voice, and they were answering and I was like 'Oh, okay'. That's like to myself [sic]. And two times they didn't catch my voice."	16
Recognition	Good Recognition	Participant reported a good voice recognition	"The recognition of the language and translation into text that the part could understand was quite well done."	13
	Misrecognition Fun	Participant communicated one or more fun experiences as a direct consequence of misrecognition	"As I said, like when I walked away, they would sometimes hear me and twice they got a sentence wrong … 'Did you steal the birthday gift?' And it said, 'Did you eat?', which was funny."	2
	Recognition Problems	Participant noted problems in the recognition of verbal input by the game	"Sometimes when I was trying to mention Kai, there was a lot of incongruency about how they would interact with that because they couldn't hear my voice clearly." "Well, they made a mistake in recognizing my voice. For example, when someone would be responding about Cardi B. And there would be no reason to have any information in the	39

Code Group	Codes	Description	Example	Frequency
			game about Cardi B, so why would the bot know anything about it?”	
Recognition Problems Causes	Pronunciation/Accent	Participant denoted their own pronunciation or accent as the cause for recognition errors	“There was [erroneous recognition] a couple of times, but then it was because of my pronunciation with that word I was struggling.” “Sometimes [I experienced recognition problems]. But I would attribute it to my accent and just like, say that it's hard to hear.”	26
	Semantic	Participant stated semantic errors as the cause for recognition problems	“I felt sometimes because of the open-endedness of the questions the characters weren't interpreting the instructions or the questions correctly. So that could get annoying over the course of several tasks.”	4
	Sensitivity	Participant underlined voice input sensitivity as the problem for misrecognition	“I think just because sometimes the threshold for the words that are said sometimes makes up something else. But that's normal, I guess, because of the sensitivity.”	5
	Talking Over NPCs	Participant poised talking over and being talked over by the NPCs as a cause for misrecognition	“They just kept talking when I was talking to them.”	1

Appendix E

Declaration on the Use of Artificial Intelligence

Throughout the writing process of this paper, I have used ChatGPT to spellcheck and improve the grammatical structure of my writing. Furthermore, I used ChatGPT to provide assistance and feedback on error messages in R throughout the coding process of data analysis. Lastly, I have made use of in-built generative AI tools in Atlas.ti during the coding process to assist and check overlapping code structures and identify redundant codes. After using these tools, I thoroughly reviewed and edited the content as needed. I take full responsibility for the final outcome and any and all consequences that may arise because of my usage of these tools.

Enschede, 22.04.2024

Leandro Tonini