International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Gaming with Etiquette: Exploring Courtesy as a Game Mechanic in Speech-Based Games

Nima Zargham, Leon Tristan Dratzidis, Dmitry Alexandrovsky, Maximilian A. Friehs & Rainer Malaka

To cite this article: Nima Zargham, Leon Tristan Dratzidis, Dmitry Alexandrovsky, Maximilian A. Friehs & Rainer Malaka (14 Aug 2024): Gaming with Etiquette: Exploring Courtesy as a Game Mechanic in Speech-Based Games, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2024.2387901

To link to this article: https://doi.org/10.1080/10447318.2024.2387901

Published online: 14 Aug 2024.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2024.2387901

Gaming with Etiquette: Exploring Courtesy as a Game Mechanic in Speech-Based Games

Nima Zargham[a], Leon Tristan Dratzidis[a], Dmitry Alexandrovsky[b], Maximilian A. Friehs[c], and Rainer Malaka[a]

[a]Digital Media Lab, University of Bremen, Bremen, Germany; [b]HCI and Accessibility, KIT, Karlsruhe, Germany; [c]Psychology of Conflict Risk and Safety, University of Twente, Enschede, Netherlands

ABSTRACT
As the gaming industry advances, it seeks increased immersion. Concurrently, recent advancements in speech recognition substantially enhanced its accuracy, creating opportunities for speech-based interactions in gameplay. Integrating speech into single-player games with non-player characters (NPCs) introduces a new social dimension. Yet, like toxicity in social media and multiplayer games, anti-social behavior can infiltrate NPC interactions, generating a toxic gaming atmosphere. We designed a speech-based game called "A Day at the Office" and conducted a within-subject study with 26 participants to explore player interactions with NPCs possessing different human qualities. Furthermore, we introduced courtesy-based game mechanics, where NPCs react to players' choice of words and loudness, investigating player experiences. Results revealed that participants spoke differently with NPCs based on their human qualities, such as gender and hierarchical roles. Moreover, the courtesy-based interactions fostered greater cognitive engagement, suggesting that tactful interactions contribute to enhanced immersion and presence in the game.

CCS CONCEPTS
• Human-entered computing → Natural language interfaces;
• Applied computing → Computer games.

KEYWORDS
Voice-controlled video games; game design; speech-based systems; voice interaction; NPCs

1. Introduction

Voice interaction has been advancing rapidly and has gained considerable attention in recent years. This intuitive form of interaction finds application in a wide range of areas, including education, customer service, translation, healthcare, and smart homes, among others (Pyae & Scifleet, 2018; Zargham et al., 2022), making speaking to computer systems a part of everyday life. Voice input is now a feature in various devices, including mobile phones, cars, and home assistants (Zargham et al., 2022). With the advances in artificial intelligence and language processing, voice user interfaces (VUIs) are expected to grow even more prominent in the coming years (Murad & Munteanu, 2020; Zargham et al., 2023).

More recently, this technology has also been attracting a lot of attention in the entertainment industry due to its inclusive, intuitive, and natural form to communicate (Allison, 2020; Allison et al., 2019; Zargham et al., 2022, 2024). Over the past years, a significant number of video games have incorporated voice-controlled functionalities. An increasing number of gaming devices are equipped with built-in microphones, which presents a vast potential for incorporating speech interaction into video games (Allison et al., 2017). This interaction modality has demonstrated the capacity to foster social presence, enhance immersion (Lee et al., 2006; Zargham et al., 2022; Zhao et al., 2018), and encourage social engagement among players (Hicks et al., 2018).

With technological advancements, the concerns associated with speech recognition issues are rapidly being addressed (Zargham et al., 2022, 2024). As technology progresses, speech recognition algorithms are becoming more sophisticated, accurate, and capable of understanding diverse accents, languages, and speech patterns. Moreover, integrating machine learning and artificial intelligence techniques allows these algorithms to continuously learn and improve over time, adapting to individual users' speech characteristics. The emergence of large language models (LLMs) and generative AI technologies like ChatGPT has ushered in notable improvements in this regard (Bubeck et al., 2023). These technological advancements not only enhance the reliability of speech recognition but also open up new opportunities for developing innovative and immersive speech-based games.

In a recent study, Zargham et al. (2024) conducted an in-depth assessment of speech interaction within the context of single-player video games. Their findings showed several

CONTACT Maximilian A. Friehs m.a.friehs@utwente.nl Psychology of Conflict Risk and Safety, University of Twente, Enschede, Netherlands

potential advantages associated with this modality, including heightened immersion, more natural interactions, increased engagement, and improved accessibility. Speech was also identified as a tool for fostering social connections and teamwork, even within the confines of single-player games.

Verbal communication has shown to be a vital component of multiplayer gaming, fostering a social atmosphere and a sense of camaraderie among players. However, this social aspect is often missing in single-player games (Zargham et al., 2020). With advancements in natural language processing, game developers can integrate speech interaction in video games regardless of multiplayer functionality, enabling players to communicate with non-player characters (NPCs) using their voices. This integration of speech interaction in single-player games with NPCs brings a new dimension of social interaction within the gaming realm. Furthermore, utilizing the player's voice not only as a control input but also as an interaction mechanic may pave the way to a more inclusive gaming environment (Pradhan et al., 2018).

One common problem in multiplayer gaming in online contexts is toxic behavior (Adinolf & Turkay, 2018; Beres et al., 2021). Toxicity in games refers to harmful behaviors of players, including abusive communications toward other players and intentionally disruptive gameplay, violating the game's rules, such as spamming, and cheating (Adinolf & Turkay, 2018; Beres et al., 2021; Foo & Koivisto, 2004; Shen et al., 2020). However, toxic behavior does not necessarily have to be limited to player communications. It can also extend to interactions with NPCs, and this behavior may subsequently spill over into player-player interactions (Zargham et al., 2023). Although the verbal abuse of NPCs in single-player games may be perceived as a victimless act, it can be problematic if there are no consequences for such behavior within the game. The absence of repercussions for abusive actions may create an environment where players perceive abusive behavior as permissible, leading to adopting similar interaction styles with other players. Furthermore, the impact goes beyond social dynamics. The failure of in-game characters to react to profanities from the player can diminish the game's realism. The believability and immersion in the game world can suffer when characters do not respond to inappropriate language. Here, we see an opportunity for game designers to incorporate game elements that effectively prevent such negative behavior. To guide our research, in this work, we investigate the following research questions:

RQ1: How do players speak to characters with different human qualities?

RQ2: What are the effects of a courtesy-based game mechanic on player experience in a speech-based game?

To examine the players' interaction with NPCs in speech-based games and how the speech modality affects the player experience, we developed a speech-based adventure video game incorporating interactions between the player and NPCs possessing various human qualities and hierarchical roles (Figure 1). In this game, we implemented a courtesy-based game mechanic, where the NPCs' responses were influenced by the players' politeness, considering both the content of their commands and the volume of their voices. In a within-subjects user study with 26 participants, we sought to investigate players' speech interaction with different NPCs. Furthermore, we compared two versions of the game, one that incorporated the courtesy-based game mechanic (referred to as "sensitive") and another that did not include it (referred to as "non-sensitive").

This research aims to extend our understanding of speech interaction in video games. It provides insights for creating inclusive and pro-social gaming experiences, offering guidance for designing immersive and socially engaging video game interactions with NPCs. The findings can have potentially practical implications for game developers, informing the design of speech-based interactions to enhance player engagement and immersion. Additionally, the study can shed light on the effective use of speech in single-player

Figure 1. A screenshot of our game called "a day at the office," displaying the main menu.

games, highlighting the potential benefits of incorporating speech interactions beyond multiplayer contexts.

2. Related work

The expansion of voice user interfaces has extended its influence to the entertainment sector, including the gaming industry. The intuitive characteristics of voice interfaces have prompted game designers to increasingly explore the integration of voice interaction as a novel game mechanic in video games (Zargham et al., 2022). In this section, we discuss previous research on voice interaction in games, social interaction in video games, and toxicity in video games.

2.1. Speech interaction in games

The early voice-based video games emerged in the 1970s (Allison et al., 2017; Reddy et al., 1973), where a small number of experimental games incorporated voice interaction as a novel feature (Zargham et al., 2024). Over time, with the advent of gaming devices, this form of interaction gained more prominence.

Even though the number of studies concerning voice user interfaces has increased extensively in recent years, research specifically focusing on games where voice control plays a central role remains limited (Carter et al., 2015). A survey study by Allison et al. (2018) indicates that academic research has mainly focused on a narrow subset of design patterns, especially pronunciation. As a result, various aspects and questions regarding voice interaction in games and its appropriate integration into gameplay remain unexplored (Allison, 2020; Zargham et al., 2022).

Due to technological limitations, particularly challenges with voice recognition (Di Petta & Woloshyn, 2001; Zargham et al., 2022, 2024), speech input remains an optional component rather than a fundamental aspect of the game (Carter et al., 2015). However, several attempts have been made to address these challenges.

To enhance speech recognition, Zargham et al. (2024) introduced a context-aware speech recognition technique. This approach utilizes contextual information from the game environment and player actions to supplement speech recognition in a game. Their findings demonstrated that this method improved speech recognition accuracy, leading to an enhanced player experience and improved usability of the speech system. In a different study, Zargham et al. (2022) explored anticipatory error handling to maintain the game's flow and minimize player frustration. In their approach, the game would perform a locally optimized action when players' intents were not recognized, taking into account goal completion and obstacle avoidance. Their results indicated that while this method improved the usability of the speech system, it did not necessarily result in a better player experience if the system made decisions contrary to the players' anticipated actions, even if those decisions moved game progress forward.

Further, speech interaction in games can be particularly important for individuals with disabilities where conventional controls may not be suitable (Harada et al., 2011; Mohammad Mustaquim, 2013; Wilcox et al., 2008). Such games have also demonstrated potential for speech therapy and enabled remote treatment methods (Ahmed et al., 2018; Lopes et al., 2016). For example, Navarro-Newball et al. (2014) developed a voice-controlled video game specifically designed for rehabilitating children with early-diagnosed hearing disabilities. The study revealed that the inclusion of narrative and entertainment elements in the game created an engaging experience, thereby supporting the repetitive approach required for effective speech mechanization sessions.

Very few studies have explored the impact of speech interactions with NPCs on the overall player experience. For instance, in a study by Zargham et al. (2020), researchers examined speech interaction with NPCs in a VR game. They compared a version of the game where players could engage in natural language conversations with multiple characters to a version where they only interacted verbally with a single character. The findings indicated that participants preferred conversing with a group of characters and found this version more entertaining. Another study explored expert opinions on the potential integration of speech interaction within video games (Zargham et al., 2024). Their evaluation revealed that experts recognize a great potential for speech interaction in games, facilitating greater immersion, engagement, and entertainment. Additionally, the experts point to pertinent concerns, including privacy considerations and play environment limitations. All in all, HCI research has yet to examine this area extensively.

To bridge this gap and to improve our understanding of the effects of player speech interactions in gaming contexts, our study aims to investigate players' speech interactions with three different NPCs, each embodying distinct human qualities.

2.2. Social interaction in video games

Online multiplayer games have drawn in millions of players around the world (Quandt et al., 2013). These games have added a social component to video games, engaging players in interactive and collaborative gameplay with other individuals in virtual environments (Lina, 2015). Previous research has shown that the social aspects of playing online games are the most important factor for many players (Griffiths et al., 2004). The presence of other individuals, or even the perception of their presence, can exert an influence on the enjoyment and performance of individuals in gaming contexts (Bowman et al., 2013). Ravaja (2009) argues that engaging in gameplay with others is associated with heightened physiological arousal and increased engagement compared to solitary gameplay experiences (Ravaja, 2009). Social interaction in multiplayer gaming plays a crucial role in the game experience, providing players with opportunities to engage, collaborate, and compete with one another (Trepte et al., 2012). Through various forms of communication, such as in-game chat or voice chat, players communicate with one another and build social bonds, form communities, and develop relationships (Zargham et al., 2020). Previous research has demonstrated that factors, such as the relationship between the players (Lina, 2015; Ravaja, 2009) and the nature of interactions among them (Velez et al., 2014) influence the game experiences.

McCoy et al. (2011) introduced the social AI system Comme il Faut (CiF) which uses a rule-based approach to model social norms, relationships, and character behavior within the game. It uses a collection of social rules, cultural knowledge, and character-specific traits to guide the interactions and reactions of characters in various social situations (McCoy et al., 2010, 2011, 2014). By incorporating these rules and understanding, the CiF system allows players to engage in more complex, dynamic, and believable interactions with the NPCs (McCoy et al., 2011). Utilizing the CiF system, McCoy et al. (2012) designed the game called "Prom Week," set during the week leading up to a high school prom, where players take the roles of high school students. Players can change the social dynamics between characters by interacting with other characters. The game incorporates over 5, 000 social considerations and rules, encompassing factors like preferences, emotions, relationships, social norms, and cultural knowledge, resulting in complex and lifelike social simulations.

Verbal exchange is a fundamental component of successful cooperation and cohesion in social groups (McGrath, 1984; Shaw et al., 1981; Zargham et al., 2020). In multiplayer games, players engage in strategic discussions, coordinate actions, and share knowledge and experiences. However, in single-player games, this social aspect is often absent. Single-player games can also incorporate cooperative elements, typically involving interactions between the player and in-game characters (Zargham et al., 2020). Though this aspect is often facilitated through dialog boxes or scripted sequences. These interactions lack the dynamic and responsive nature of real-time interactions with other humans. The usage of natural language is the most common form of communication among humans (Hockett & Hockett, 1960). However, only a very limited number of single-player games allow for natural language communication between the players and the game characters, therefore, little is known about player's experience where they are able to communicate with NPCs using speech. In our study, we explore the dynamics of player-NPC interaction where players can speak with the game characters using natural language.

2.3. Toxicity in video games

While online multiplayer games offer players a platform for social interaction, community-building, and relationship development (Trepte et al., 2012), they also face toxic behavior and harassment (Adinolf & Turkay, 2018; Beres et al., 2021) as a prevalent issue. Toxic behavior in games encompasses a range of negative behaviors exhibited by players, including abusive communication toward other players, intentionally disruptive gameplay, and violations of the game's rules, such as spamming and cheating (Adinolf & Turkay, 2018; Beres et al., 2021; Foo & Koivisto, 2004; Shen et al., 2020; Zargham et al., 2023). Encountering toxic behavior from other players can harm the overall player experience, leading to negative consequences, such as players' decision to disengage from the game altogether and ultimately threatening the victim players' well-being (Beres et al., 2021). The presence of toxic communities within games can also pose a financial risk to companies, as their reputation for fostering such environments can adversely affect revenue generation (Kordyaka et al., 2020). Research by Depping et al. (2018) suggests a strong positive relationship between social capital in games and players' psychological well-being. The study identified that toxicity in games has a negative association with both bridging and bonding ties.

All of this underscores the critical significance of proactive measures to deter and mitigate such forms of antisocial behavior. However, it is important to recognize that toxic behavior is not necessarily confined solely to interactions between players but can also extend to interactions with NPCs, which may later carry over to player-player interactions (Zargham et al., 2023). Although toxicity is commonly associated with multiplayer games, it can also manifest in single-player games. A study by Neely (2019) explored moral decision-making in single-player games and found evidence of toxic behaviors exhibited by certain players, even without direct interactions with other individuals.

To combat toxicity in online multiplayer games, developers have employed several techniques, including reporting and banning anti-social players (Reid et al., 2022). For instance, players in some multiplayer games, such as League of Legends (Riot Games, 2009), could receive warnings, penalties, or account bans for negative behaviors. On the other hand, there are also measures to encourage users to engage in positive behavior. Players could receive praise or points from other players after a match. In League of Legends (Riot Games, 2009), for instance, players could receive" Honor Ribbons" for being decent and friendly, earning community badges, and obtaining in-game cosmetics, such as champion skins.

2.4. Morality in single-player games

In single-player games, acts of kindness and positive behavior are frequently rewarded through various game mechanics. One of the earliest examples of such a system that was not only well-implemented but also well-received was Fallout: New Vegas's reputation system. In this game, the player can earn reputation with specific factions by helping them achieve their objectives. This in turn influences how the different factions interact with the player (Obsidian Entertainment, 2010). But also other Triple-A games, such as Red Dead Redemption 2 contain an Honor system that tracks the player's moral choices and actions throughout the game (Rockstar Studios, 2018). This measures whether the player's character behaves as honorable or leans toward a more dishonorable path. Players can influence their Honor level by making decisions during various in-game activities, such as interactions with NPCs, completing missions, and random encounters. Acts of kindness, such as helping strangers, sparing enemies, and adhering to the law, typically increase Honor. Conversely, committing crimes, harming innocent civilians, or antagonizing NPCs decreases Honor. The Honor system impacts various aspects of gameplay, including how NPCs react to Arthur, the outcomes of certain storylines and missions, and the general atmosphere of the game world. Players with high Honor may receive discounts at stores, access unique items or missions, and experience a more positive reception from NPCs. Conversely, low Honor may result in higher store prices, increased hostility from NPCs, and potentially different story outcomes. A game that only allows for positive player interactions is Death Stranding with its social strand system (Kojima Productions, 2019). This system enables players to leave structures, equipment, and messages behind that appear in the singleplayer world of others through the game's asynchronous multiplayer. Helpful contributions can be rewarded with likes from other players, strengthening the" strand" between the players and increasing the likelihood of more structures appearing in the world from the contributing player. Warpefelt and Verhagen (2017) explored incorporating kindness as a design parameter to foster a more reflective player experience. By examining a range of kindness-oriented games, the authors explored the feasibility of imbuing a sense of intrinsic kindness into a game through a careful selection of game mechanics and the overall atmosphere of the game's environment. Previous literature suggests that video games with an explicit focus on kindness-orientated mechanics (often referred to as "persuasive" or "prosocial" games) may even inspire a more open-minded willingness toward acts of compassion within their players and nourish a deeper sense of humanity within oneself (Greitemeyer, 2013). A study by Whitaker and Bushman (2012) suggests that helpful and kind behavior while playing video games increases acts of prosocial behavior after the play session has finished.

In this article, we tackle the areas above within the context of a single-player game, where players engage in speech-based interactions with NPCs. We have designed a novel game mechanic centered around courtesy, where players' selection of words and volume of speech directly influence their interactions with the NPCs. To the best of our knowledge, this work is the first exploration of such a game mechanic. We aim to examine the effects of such mechanics on the overall player experience.

3. Methods

3.1. Game design

To address our research questions, we developed "A Day at the Office," a speech-based adventure puzzle game where players have to communicate with the game's characters using natural language. In this game, the players take on the role of an office employee who interacts with different colleagues to solve the game's puzzles and finish the levels. The story features four non-player characters: a male boss, a female colleague, a male intern, who are used for the main level (see Figure 2), and a male vendor for the tutorial. The selection of the NPCs' gender and their hierarchical role in the office was determined randomly, without any specific criteria or deliberate choice. The game design follows general recommendations and guidelines from the literature (cf. Adams & Dormans, 2012; Schell, 2008), aiming to facilitate enjoyable game experiences with a usable interface and an overall engaging narrative structure. To complete a level, players were required to engage in a minimum of three interactions with each character. These interactions may not occur consecutively within the game's state, as players might need to engage with other NPCs before resuming their conversation with the initial NPC they were interacting with.

3.1.1. Levels

The game consists of two levels and a tutorial level. The tutorial is set in a gas station shop where players are tasked with obtaining a cup of coffee by talking to the vendor. The tutorial level aims to teach players how the in-game speech interaction works. On top of that, the functionality of the courtesy-based game mechanic is explained and demonstrated. The tutorial would guide players on how to approach in-game characters and demonstrate the characters' reactions to specific phrases and potential repercussions. This level unfolds through a guided tour, with players receiving instructions from an invisible narrator. The main levels are both set in an office space. The office space includes open working space, three separate offices, an entrance area, a cleaning closet, and a server room. In both

Figure 2. In the main levels of the game, players engage with three characters: a male intern (left), a female colleague (middle), and a male boss (right).

Figure 3. Screenshot of the game where the player interacts with the "boss" character in his office. The objective box (top-left) and the volume visualizer are visible on the left side of the screen. The subtitles at the bottom of the screen present the dialogue spoken by the NPC.

levels, the player must talk with all game characters to solve the puzzle and finish the workday.

Both main levels are set in the same office environment. The tasks and puzzles in the two levels are similar but differ in solution as players had to complete both levels and would have been familiar with the solution during the second playthrough. Both levels start with a situation where the player should gather information by talking to the characters around the office. In one level, players are tasked with locating an important office file, while in the other level, their objective is to discover a hidden trophy. In both levels, the player has to solve a cipher in their office and present the solution to the colleague who gave them the task. Afterward, the player needs to collect certain items around the office and find an object for the Boss. When all the tasks are completed, the player can collect the level reward, which ends the level and takes the player back to the main screen.

3.1.2. Implementation

The game environment and logic were created with Unity 3D.[1] The speech recognition was implemented using Picovoice,[2] which offers a free API that can be easily integrated with Unity. We created builds for Windows and MacOS.

For the game, we used a low-poly visual style. For the environment, we utilized the 3D models by Synty.[3] Additionally, for assets that explicitly required custom designs, we modeled 3D Objects with Blender.[4]

To promote better accessibility and understandability of the NPCs' statements, we added subtitles displayed in the HUD (see Figure 3). To design the NPC dialogs, we used the text-to-speech program Tortoise[5] to generate the character's voices. Tortoise allows for creating highly realistic and customizable voices (Betker, 2023). An instruction page was also implemented to inform users about the game's procedure, controls, and goals, accessible from the game's main menu.

A microphone calibration functionality was implemented where players would adjust the volumes for the detection of different types of speaking before starting the game. The calibration was done in a process that would guide players through the necessary steps. During calibration, players are instructed to speak loudly or scream for three seconds and then remain silent for another three seconds. The program would then calculate the microphone's volume range by subtracting the minimum value from the maximum and dividing it into ten equal parts. This calculation is later used to visualize the volume with the Audio Bar.

3.1.3. General mechanics

The player's movements are controlled using the mouse and keyboard. Players could move using the WASD keys and look around the room with the mouse movement. The game included several interactable objects that the player could pick or modify, some of which were needed to solve the game's puzzles. Examples of such items include documents, files, or a mobile phone.

A volume visualizer, a bar separated into ten sections, was designed to represent the player's current volume level, ranging from green to red (see Figure 3). This is only visualized when the player is near an NPC and can interact with the character using speech. An objective box was displayed at the top left corner of the screen, highlighting the player's current task in hand (see Figure 3). The text would change to the next objective as soon as the objective was fulfilled. In the top right corner of the screen, a hint box was placed, which would only be displayed when players pressed the H-key. The hint would display a phrase the player could use to interact with a specific character to proceed with the story. The game also included a pause menu toggle by pressing the Escape key. In this menu, the main game is paused.

The player can resume the game, restart from the last checkpoint, or go to the main menu.

3.1.4. Speech interaction

When the player gets close to an NPC, the character turns their head and looks at the player, indicating they can engage in a conversation. All NPCs had three primary states: working, listening, and talking. For instance, when the player approaches a game character, the NPC state transitions the character from working to listening. If the system recognizes the speech input, the character would transition to the talking state and communicate with the player based on their inquiry. To give the speech input, players do not need to press any buttons. They can communicate with the NPCs anytime during the game using speech. The recognition system would automatically await commands as soon as the NPCs were in the listening state. The system was able to handle several phrases per action. For instance, if players wanted to say "Have you seen my files?" they could also use phrases, such as "Do you know where my files are?" "Where are my files?" or "Where them files at?" If the voice recognition system recognized a command, the characters would respond with a corresponding reply. If no matching command is found, the system will consider that to be a failed attempt. The NPCs would respond with a message indicating they did not understand the player, such as: "I did not quite catch that." The NPCs were also capable of holding small talk with the players. Examples of such interaction include "How are you?" or "How is your day so far?"

3.1.5. Courtesy-based mechanics

We designed a courtesy-based game mechanic where the NPCs react to the player's choice of words and speaking volume. This could result in characters helping players solve the game puzzles faster when they use polite wordings and are not shouting or hindering their progress in cases where they speak too loud or use impolite words in their interactions. We categorized player speech interaction with NPCs into three groups based on intent and tone: (1) friendly, (2) neutral, and (3) unfriendly. Polite and appreciative language, such as apologies and expressions of gratitude, characterized friendly responses. Examples of this would be starting a sentence with a greeting similar to "I am sorry to bother you," using "please" in the inquiries, or thanking the characters for their support and efforts. Unfriendly interactions are those which contain impolite or offensive language. Examples include profanity, shouting at characters, or accusing characters of wrongdoings, such as them entering your office without permission. The volume visualizer feature could indicate if players' speech is too loud. When the visualizer displays colors, such as orange or red, it warns that the players' volume is too loud, potentially causing characters to perceive the interaction as impolite. Those responses that do not fit in either one of the two categories of friendly or unfriendly are marked as neutral.

If the response is marked as unfriendly, a 30-second timer starts, indicating that the NPC is unhappy with the form of communication and will only respond to the player once the timer runs out. During this period, the player could apologize to the NPC to end the timer.

3.2. Study design

To evaluate our two conditions, we conducted a within-subjects design user study with ($N = 26$) participants. Each participant had to play two levels of the game, one utilizing the courtesy-based mechanic (CBM) and one without it. We chose a within-subjects design as it allows for a direct comparison within the same group of participants, reducing the influence of individual differences on the results. Moreover, variations in performance or experience can be attributed to the specific condition rather than participant variability. The game environment and the NPCs were the same in the two levels. The levels were structurally equivalent and differed only in the puzzles and objectives, although they had comparable difficulty levels. The puzzle solutions were different to avoid repetition, as the players had to play both conditions one after another. We counterbalanced the order of the conditions to avoid sequential biases. Both levels were capable of running with or without the CBM.

Participants were instructed to complete the tutorial and play both levels of the game "A Day at the Office" from beginning to end. They had the choice to skip to the next level or quit the game early if desired. Players accessed the game on their PC or laptop. An executable version of the game was provided to each participant before the session. The experimenter verified that participants had a functioning microphone and that the game ran smoothly before starting the play session. To ensure the reliability of the results, external factors that could potentially affect the game experience, such as the type of computer, monitor size, and background noise level in the room, were also controlled for. Additionally, participants were informed about the differences between the two conditions before the study and learned how the CBM functions during the tutorial level.

3.2.1. Procedure

The sessions were held remotely *via* Discord[6] to ensure the convenience of the participants and accommodate a broader range. Upon joining the Discord server, participants were greeted and briefed about the study procedure. They were further briefed about the game and its controls while reminding them to read the in-game instructions. The participants were then directed to give informed consent and fill out the demographics questionnaire. They were then instructed to download the game, launch it, and read the instructions in the game. The experimenter ensured the functionality of the setup. Participants were then instructed to calibrate their microphones. After this, the tutorial level would start. After completing the tutorial, participants played one of the two levels, either with CBM (sensitive level) or without it (non-sensitive). Once they finished the first level, participants were asked to complete the post-exposure questionnaires related to the condition they had just played. Subsequently, participants proceeded to the second level featuring the alternative condition, followed by another round of post-exposure questionnaires. After the two levels, the participants responded to a custom-designed questionnaire to assess the overall game experience followed by a brief semi-structured interview that concluded the session. The players were instructed to send the log files containing the game analytics to the experimenter. Finally, participants were thanked for their time and contribution.

During the experiments, the participants would share their screens with the experimenter while playing the game. During gameplay, the study conductor would typically mute their microphone unless the participant requested assistance. Participants could ask questions or express their needs anytime during the session. The experimenter noted verbal statements and in-game observations while assisting when participants encountered difficulties. Participants were also allowed to take short breaks between levels. The sessions were audio-recorded and later transcribed for further analysis. The interviews took an average of 3.21 min ($SD = 0.54$). Each session lasted ~30–45 min, with the average gameplay time being around 15 min.

3.2.2. Measures

To find answers to our research questions, we used a combination of standardized questionnaires, customized questions, interviews, and game logs. Before the session, the participants filled out a demographics questionnaire. The post-exposure questionnaires included the Player Experience of Need Satisfaction (PENS) (Ryan et al., 2006) questionnaire with the subscales of *Competence*, *Autonomy*, *Presence/Immersion*, and *Intuitive controls* while excluding *Relatedness* as it was not relevant to the scope of the study, as well as the Immersive Experience Questionnaire (IEQ) (Jennett et al., 2008) throughout five factors of *Cognitive Involvement*, *Emotional Involvement*, *Real World Dissociation*, *Control*, and *Challenge*. Both questionnaires are validated and established measurement instruments in games user research and the HCI community. PENS is commonly used for determining the player experience within multiple sub-scales. IEQ is also a widely used tool to measure the subjective experience of immersion in interactive media, including video games. Additionally, we recorded a series of customized questions regarding players' experience with the game. These were implemented via seven-point Likert scales, including questions about their enjoyment of the game, perceived performance, overall game experience, and willingness to play similar speech-based games in the future. We also asked participants to indicate which of the two conditions they preferred to play: sensitive or non-sensitive. We also conducted a short semi-structured interview with each participant to evaluate further the qualitative factors of the player experience and individual preferences (Wilson, 2013). The interview included questions about likes and dislikes concerning the game, the most and least interesting aspects, and the players' thoughts on talking to NPCs and the CBM. Gameplay and audio recordings were collected to allow for additional observations at a later time.

3.3. Data analysis

3.3.1. Quantitative data

Concerning the statistical analysis, we performed the Shapiro–Wilk normality assumption test to assess the distributional assumption of our data. Based on the results of the normality tests, we conducted unpaired t-tests and the Mann–Whitney U Test (when data was not normally distributed) to identify the differences between the conditions with regard to the standardized questionnaires. Additionally, we performed repeated-measures ANOVAs to identify differences in conversation with different NPCs and between the two game conditions. We applied an alpha level of .05 for all our statistical tests.

3.3.2. Qualitative data

The audio recordings of the sessions were transcribed verbatim. The interview data underwent analysis and coding using domain summaries (Braun et al., 2019; Connelly & Peltzer, 2016). The themes are organized around a common topic rather than shared meaning, aiming to encompass the diversity of interpretations related to a specific subject or focus area (Morgan, 2022). The analysis began with data familiarization and categorization (Braun & Clarke, 2019), where two researchers read through the responses to get a sense of the content and context to understand the patterns, ideas, and concepts present in the responses. To develop a coding system, the transcripts of a random selection of 15 interviews were independently coded by two researchers using inductive coding (Chandra & Shang, 2019; Thomas, 2006), where a single quote could be assigned to multiple codes. The researchers then agreed upon a coding system after a thorough discussion. In cases of disagreements, an additional author was consulted to reach a consensus. A coding manual was established through an iterative discussion between the two authors. One author coded The remaining transcripts individually, following the established coding manual. Noteworthy and unique player statements were also collected during this process.

3.3.3. Game logs

After each game-play session, a log file was generated containing information on the recognized commands, the total number of given commands, playtime, and the recognition rates.

3.4. Pre-study

Before running the main study, we conducted a preliminary study with two participants. This pre-study aimed to identify issues related to game mechanics, speech recognition, puzzles, and the overall study procedure. Several gameplay issues were identified during these sessions, and appropriate measures were taken to address them. Additionally, we

Table 1. Descriptive statistics and results of the Wilcoxon signed-rank tests for PENS and IEQ.

Questionnaire	Subscale	Non-sensitive mean (*SD*)	Sensitive mean (*SD*)	*p*-Value	*Z*-value	*W*	Effect size
PENS	Competence	3.80 (0.73)	3.71 (0.71)	0.38	−0.86	73	−0.19
	Autonomy	3.49 (0.68)	3.61 (0.80)	0.43	0.78	59.5	0.19
	Presence/immersion	3.17 (0.77)	3.19 (0.68)	0.92	0.10	146	0.02
	Intuitive controls	4.37 (0.65)	4.25 (0.83)	0.52	−0.635	70.5	−0.14
IEQ	Challenge	4.03 (0.74)	4.17 (0.68)	0.35	0.92	88.5	0.20
	Control	5.30 (0.87)	5.28 (0.65)	0.51	−0.65	106	−0.13
	Real world involvement	4.14 (1.00)	4.20 (1.00)	0.90	0.11	122.5	0.02
	Cognitive involvement	4.75 (0.82)	5.34 (0.72)	0.007**	2.67	56	0.54
	Emotional involvement	4.79 (0.83)	4.67 (1.06)	0.66	−0.43	158	−0.08

Notes: Statistically significant results are marked with asterisks.

considered the feedback from participants regarding the length of the post-exposure questionnaires, leading us to reduce the number of questions.

3.5. Participants

We recruited $N = 26$ individuals using a convenience sampling approach. 21 participants self-identified as themselves as male, and five as female. The age ranged between 15 to 57 years ($M = 26.84$, $SD = 10.62$). Participants below the age of 18 engaged in the study with explicit parental consent and authorization. The recruitment was conducted on social media platforms, word of mouth, gaming communities, and university mailing lists. Participation in the study was voluntary and uncompensated. The participants were divided into two groups. Players in the first group started with the non-sensitive condition, and the second group started with the version including CBM. The sample consisted of one native English speaker, and the rest were fluent non-native English speakers. The experiment sessions were conducted in English. However, the interviews were also conducted in German when participants requested it if they felt more comfortable speaking in their native language. All interviews were later transcribed and translated into English by the interview conductor. Most participants (80.76%) played video games frequently (Six daily and 15 several times a week). 19.23% of participants reported not playing video games often (one once per week, three once monthly, and one never). Only a few participants (11.53%) reported experiencing a game with voice interaction. However, 73.08% of the participants had previous experience with voice-controlled applications.

4. Results

This section presents the results of our study. We present the quantitative results first, followed by the qualitative findings from the interviews and experimenter's observations. Post-hoc tests conducted for the evaluation are included in the Appendix.

4.1. Quantitative findings

4.1.1. Player experience and immersion

Concerning the PENS questionnaire, we employed the Wilcoxon signed-rank test to examine the responses due to the violation of normal distribution, as indicated by the Shapiro–Wilk test. No significant differences were witnessed in the analysis ($p > 0.05$), indicating that participants generally had similar experiences regarding *Competence*, *Autonomy*, *Presence/Immersion*, and *Intuitive Controls* in both game conditions, suggesting comparable experiences between the two conditions (see Table 1).

The responses from the IEQ were also subject to the Wilcoxon signed-rank test, as the Shapiro-Wilk test highlighted a deviation from normality. We observed a significant difference in the subscale of *Cognitive Involvement* (see Table 1). Participants in the sensitive level rated a significantly higher *Cognitive Involvement* ($M = 4.75$, $SD = 0.82$) in comparison to the non-sensitive level ($M = 5.34$, $SD = 0.72$; $Z = 2.67$, $W = 56$, $p < .007$). We did not observe any significant differences with regard to *Emotional Involvement*, *Real World Dissociation*, *Control*, and *Challenge* ($p > 0.05$).

4.1.2. Game logs

In the following, we present the results obtained from the game logs. Due to the deviation from normality observed in the data, as indicated by the Shapiro–Wilk tests, we employed the Wilcoxon Signed-rank test to examine significant differences in the log data between the groups.

On average, participants took $M = 6.54$ min ($SD = 2.271$) to complete a level without the courtesy-based game mechanic. When utilizing the courtesy-based game mechanic, participants took an average of $M = 7.75$ min ($SD = 3.073$) to complete a level. As the Shapiro-Wilk test indicated a significant departure from normality, we conducted the Wilcoxon Signed-rank test to identify significant differences. The results indicate that players took significantly longer to play the levels with courtesy-based game mechanics ($Z = 2.155$, $W = 90$, $p = .03$), with a medium effect size ($r = 0.42$). We observed significant differences in the total number of voice commands given at a level. Participants in the sensitive condition had a significantly higher number of intents ($M = 31.65$, $SD = 13.0$) in comparison to the control group ($M = 24.30$, $SD = 9.42$; $Z = 3.46$, $W = 38.5$, $p < .001$), indicating a large effect size ($r = 0.68$). The recognition rate in the control condition was 58.4% and 59.4% for the sensitive condition, showing no significant difference ($p > .05$). It is important to note that the unrecognized commands include those not appropriately recognized by our speech system, as well as out-of-scope commands that were not included in the game's vocabulary. Participants in the sensitive condition ($M = 11.86,\ SD = 6.40$) had a similar number of unrecognized commands compared to the control condition ($M = 9.30,\ SD = 4.37$).

4.1.3. Conversation intents

We quantified the participants' conversation intents during the gameplay (Table 4). For each intent, we performed a sentiment analysis and assigned each request into one of the three categories of *friendly*, *neutral*, or *unfriendly*.

To examine if the two game versions and the conversation with the characters yielded different behaviors, we conducted an RM-ANOVA with Sensitivity (indicating the presence or absence of CBM), Hierarchy (representing character roles: boss, colleague, intern), and Sentiment (reflecting language tone: friendly, unfriendly, neutral) as independent variables and intent occurrences as the dependent variable. The analysis showed significant differences for all three variables Sensitivity, Hierarchy, and Sentiment. The *post-hoc* test, concerning *Sensitivity*, indicates that participants employed a significantly higher number of intents in the sensitive condition (see Table A1). In terms of *Hierarchy*, the *post-hoc* test indicates that participants engaged in significantly more conversation with the colleague than with the boss and intern. Additionally, they spoke significantly more to the boss compared to the intern (see Table A2). The *post-hoc* test regarding *Sentiment* revealed that participants used significantly more neutral intents than both friendly and unfriendly intents. Furthermore, they also used significantly more friendly intents compared to unfriendly intents (see Table A3).

Further, the RM-ANOVA yielded interaction effects for Hierarchy × Sentiment and Sensitivity × Hierarchy. However, no significant interactions were found between Sensitivity × Sentiment, nor Sensitivity × Hierarchy × Sentiment. The corresponding results are depicted in Table 2 and Figure 4. The *post-hoc* tests for Hierarchy × Sentiment indicate that participants employed significantly more friendly intents toward the boss compared to the colleague and the intern. Additionally, they used significantly more friendly intents toward the colleague compared to the intern (see Table A7). On the other hand, the analysis also revealed that players employed significantly more unfriendly intents toward the colleague compared to both the boss and the intern.

Furthermore, we investigated if interacting with the characters yielded different types of special intents (Table 5). Therefore, we coded the participants' Intent into three categories *Chat*, *Thank*, and *Excuse*. Intents containing apologetic terms like "sorry," "excuse me," or "forgive me" were categorized as *Excuse*. Intents featuring off-task comments, such as casual chit-chat like "How are you?" or "How is it going?" were classified as *Chat*. Finally, comments expressing gratitude with words like "thank you" or "thanks for the help" were assigned the code *Thank*. We conducted a Sensitivity × Hierarchy × Intent RM-ANOVA (see Table 3 and Figure 5). The RM-ANOVA showed significant differences for Sensitivity, and Hierarchy as well as significant interactions Sensitivity × Intent, and Sensitivity × Hierarchy.

Table 2. RM-ANOVA table of intent occurrences.

Cases	*df*	*F*	*p*	η_p^2	ω^2
Sensitivity	1, 25	11.63	< 0.0001	0.32	0.13
Hierarchy	2, 50	30.90	< 0.0001	0.55	0.30
Sentiment	2, 50	91.56	< 0.0001	0.79	0.59
Sensitivity * Hierarchy	2, 50	4.27	0.02	0.15	0.05
Sensitivity * Sentiment	2, 50	2.17	0.12	0.08	0.01
Hierarchy * Sentiment	4, 100	25.89	< 0.0001	0.51	0.34
Sensitivity * Hierarchy * Sentiment	4, 100	0.09	0.99	< 0.0001	0.00

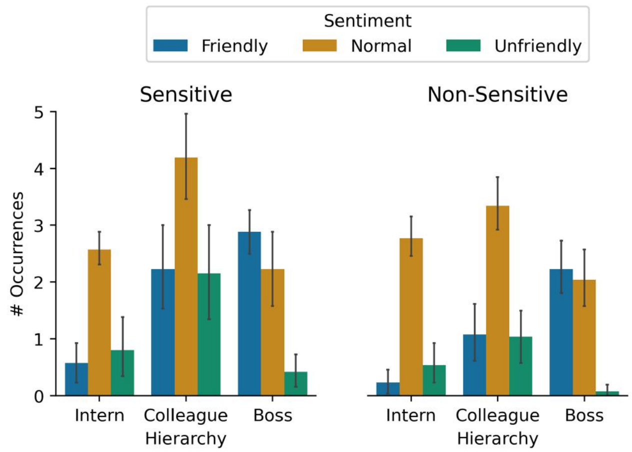

Figure 4. Means and standard deviations of intent occurrences split by game versions (sensitivity), different characters (hierarchy), and intent sentiment.

Table 3. RM-ANOVA results of the special intents.

Cases	*df*	*F*	*p*	η_p^2	ω^2
Sensitivity	1	9.35	<0.01	0.27	0.10
Intent	2	0.19	0.83	<0.01	0.00
Hierarchy	2	18.08	<0.01	0.42	0.18
Sensitivity * Intent	2	7.53	<0.01	0.23	0.07
Sensitivity * Hierarchy	2	6.73	<0.01	0.21	0.06
Intent * Hierarchy	4	0.76	0.56	0.03	0.00
Sensitivity * Intent * Hierarchy	4	1.03	0.40	0.04	<0.01

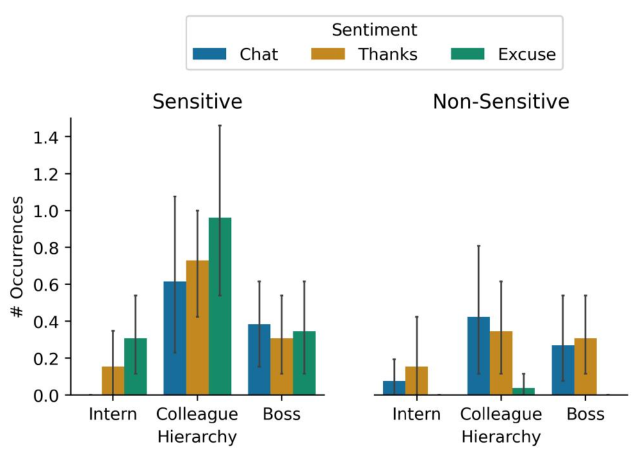

Figure 5. Means and standard deviations of special intent occurrences split by game versions (sensitivity), different characters (hierarchy), and intent.

Table 4. Descriptive statistics of the intent distributions among the characters.

	Type of intent	Non-sensitive mean (*SD*)	Sensitive mean (*SD*)
Intern	Friendly	0.23 (0.58)	0.57 (0.98)
	Neutral	2.76 (0.90)	2.57 (0.80)
	Unfriendly	0.53 (0.90)	0.80 (1.32)
Colleague	Friendly	1.07 (1.38)	2.23 (1.96)
	Neutral	3.34 (1.29)	4.19 (2.03)
	Unfriendly	1.03 (1.21)	2.15 (2.14)
Boss	Friendly	2.23 (1.24)	2.88 (1.10)
	Neutral	2.03 (1.34)	2.23 (1.83)
	Unfriendly	0.07 (0.27)	0.42 (0.75)

Table 5. The descriptive statistics of the occurrence and distribution of special intents, including chitchats, thankful intents, and apologetic intents.

	Type of intent	Non-sensitive mean (*SD*)	Sensitive mean (*SD*)
Intern	Chat	0.07 (0.27)	0 (0)
	Thank	0.15 (0.61)	0.15 (0.46)
	Excuse	0 (0)	0.30 (0.61)
Colleague	Chat	0.42 (0.94)	0.61 (1.13)
	Thank	0.34 (0.62)	0.73 (0.77)
	Excuse	0.03 (0.19)	0.96 (1.21)
Boss	Chat	0.26 (0.66)	0.38 (0.63)
	Thank	0.30 (0.61)	0.30 (0.55)
	Excuse	0 (0)	0.34 (0.62)

4.1.4. Customized questions

The customized questionnaire was only administered once after participants had completed both game levels. Therefore, other than the specific question regarding the preferred level, here, the participant responses are generally about the speech-based game "A Day at the Office."

Regarding the overall game experience, players provided a relatively high rating ($M = 5.46$, $SD = 1.04$). Similarly, participants expressed high levels of enjoyment while playing the game ($M = 5.34$, $SD = 0.99$). They also indicated a strong interest in playing similar video games in the future ($M = 5.53$, $SD = 1.66$). Participants rated their own performance with $M = 4.96$ ($SD = 1.12$). Moreover, the majority of our participants (76.92%) preferred the sensitive condition over the non-sensitive condition.

4.2. Qualitative responses

Qualitative data was collected through brief semi-structured interviews conducted at the end of the session. Overall, participants expressed enjoyment and found the game novel and entertaining.

4.2.1. Positive impressions

Regarding specific aspects that participants liked about the game, 88.46% highlighted the ability to use natural language to speak to characters as a novel and appealing feature. Additionally, 30.77% mentioned their appreciation for the game's aesthetics. A notable finding was that 26.92% of players reported feeling more immersed in the game when interacting with the characters through speech. One participant mentioned: "I felt like I was a part of the office and could express myself freely" (P9). Three players (11.53%) noted that playing the game using speech felt "more engaging than selecting dialogue options in decision-making games" (P11). Eight participants (30.77%) specifically found the CBM interesting: "I liked the possibility to talk to the characters and having to pay attention to what I should say and how to say it" (P6). 42.31% of the participants expressed their desire to play similar games in the future. Seven players (26.92%) expressed interest in a larger version of the game with an expanded world, more interaction possibilities, riddles, and additional characters. Two other players raised interest in having the possibility to play this game in a different language.

Regarding the game's most interesting aspect, a significant majority of participants (73.07%) found the ability to talk directly to the characters to be the most interesting. Seven (26.92%) highlighted the courtesy-based game mechanic and the fact that they had to be especially mindful of what to say and in what tone they say as the most interesting aspect of the game: "I never experienced this in a single-player game where you have to be careful of your wording and loudness. This gave a similar feeling as multiplayer games" (P1). Another player mentioned: "I liked that I had the opportunity to make them feel if I was happy or angry with them" (P15).

4.2.2. Negative impressions

Turning to aspects that participants disliked about the game, 50% of participants mentioned issues with voice recognition, particularly citing problems related to the limited vocabulary and restrictive dialogue options. One participant mentioned: "I think the dialogue options were restricted, which was frustrating" (P25). Additionally, four participants (15.38%) did not enjoy the quests and puzzles, with one explicitly noting difficulties in grabbing required objects. One player disliked the game's graphics style: "I like more realistic looking games" (P16). One participant specifically mentioned that the timer in the sensitive version was "annoying" (P11). Additionally, four participants reported experiencing microphone calibration issues, where the audio bar displayed higher volume than they actually spoke, resulting in the characters interpreting their intents as impolite.

With regards to the least interesting aspect of the game, 73.07% of participants considered the quests and puzzles to be too simple and, consequently, the least interesting part of the game. Other aspects mentioned as least interesting included the game characters (mentioned by two participants), the game's graphic style (mentioned by one participant), the no-response timer (mentioned by one participant), and the background music (mentioned by one participant).

4.2.3. CBM impressions

When discussing the sensitive level, eight participants (30.77%) (seven males and one female) mentioned that the female colleague character was more sensitive to their intents, triggering the no-response timer more frequently. It is worth noting that the technical implementation of triggering the timer was identical for all characters. Two participants (7.69%) reported similar concerns regarding the intern character. Furthermore, six participants (23.07%) stated that the game felt more realistic at the sensitive level, for example: "It resembled real-life interactions where I must be mindful of what I say" (P2). Another player emphasized: "This aspect made the characters feel lifelike and more than just a bunch of pixels" (P5). Eight participants mentioned that the CBM added a new layer of challenge to the game: "It was more challenging when I had to be mindful of what to say" (P7). Among those, five specifically mentioned that they enjoyed this added challenge.

4.3. Researchers' observations

Throughout the sessions, we noticed certain patterns of behavior among the participants. Specifically, when the game did not recognize their voice input, many participants tended to move closer to the in-game characters physically. Conversely, some participants instinctively backed away from the characters if they felt their voices were perceived as too loud.

Another observation was that several participants engaged in sarcastic and humorous interactions with the characters. However, it is important to note that our speech system was not designed to understand such intents, resulting in frequent instances of non-recognition.

We also witnessed a phenomenon that we referred to as "Sorry Spam" among a few participants (11.53%). When these players realized they had upset a character, they would repeatedly apologize in an attempt to stop the no-response timer. This behavior often involved interrupting their previous intents.

5. Discussion

The study examined players' experiences conversing with different NPCs and their impressions of the courtesy-based game mechanics. Overall, players expressed enjoyment and found "A Day at the Office" to be a novel and entertaining game. Despite encountering frustrations with recognition issues, a majority of players appreciated the ability to communicate with NPCs using natural language. The customized questionnaire results confirmed the positive feedback, with participants providing high ratings for overall game experience and enjoyment. Many participants expressed interest in playing similar games or even an expanded version of the same game in the future. The qualitative data from the interviews further supported these findings, with participants highlighting the game's novelty, entertainment value, and immersive nature.

During the post-gameplay interviews, participants reported feeling immersed in the game when conversing with NPCs. They felt a sense of being part of the office environment, aligning with previous research indicating that voice interaction in games can enhance immersion (Lee et al., 2006; Osking & Doucette, 2019; Zargham et al., 2022, 2023; Zhao et al., 2018). This sense of immersion was further supported by our observations during gameplay sessions, where participants adjusted their character's proximity to the NPCs based on their understanding of their intents, moving closer when intent recognition was challenging and further away when intents were perceived as loud.

We analyzed the results of this experiment to address the following research questions:

RQ1: How do players speak to characters with different human qualities?

RQ2: What are the effects of a courtesy-based game mechanic on player experience in a speech-based game?

5.1. NPC interactions

With regards to **RQ1**, the analysis of conversation intents revealed that players exhibited different speaking behaviors toward the NPCs.

We witnessed that our participants initiated more conversations with the colleague character compared to the boss and intern. Since the required minimum interaction with each NPC for puzzle-solving in the main levels was deliberately balanced during the game's design, this finding could suggest that the observed disparity in conversation engagement is not an outcome of the game's structure but rather implies that participants autonomously chose to involve themselves more in conversations with their peer, the colleague. This finding highlights the players' tendency toward social interactions with characters perceived as equals within the in-game office hierarchy. The preference for engaging with a colleague over superiors or subordinates could reflect a desire for relatable or less authoritative interactions. Game designers should consider these player preferences in character development, emphasizing the significance of relatable and socially engaging NPCs to enhance player involvement and satisfaction in speech-based video games. Furthermore, considering the NPC's gender, with the colleague being female and the intern and boss being male, and considering that most of our participants were male, this finding could also be indicative of gender-related dynamics in player interactions with virtual characters. Further examining the gender dynamics within our game, a noteworthy finding emerged, revealing that participants directed significantly more unfriendly intents toward the colleague than toward the boss and intern. Moreover, the qualitative responses added depth to this observation, indicating that some participants perceived the female colleague as more prone to sensitivity. This perception contrasts with the technical implementation of all the NPCs being uniform and programmed in the same way to respond to users' conversational sentiments. This finding raises questions about potential gender stereotyping within the game, as players exhibited distinct speech behaviors influenced by the NPCs' gender. This insight emphasizes the importance of considering gender portrayals in video game design to prevent unintended reinforcement of stereotypes. The observed tendency to use unfriendly intents toward the female colleague could stem from existing biases or societal norms that impact player behavior. Game developers should critically assess character representations and narratives to ensure inclusivity and avoid perpetuating harmful stereotypes.

Notably, participants exhibited significantly more politeness toward the boss in comparison to the colleague and the intern. Additionally, their interactions were notably more polite toward the colleague than the intern. This hierarchical influence on conversational dynamics within the game suggests that players were attuned to the NPCs' roles and adjusted their speech accordingly, mirroring real-world workplace social norms.

We observed that our participants predominantly maintained conversations with a neutral tone while favoring friendliness over unfriendliness. Considering the office environment, this conversational pattern may be attributed to the context of the game. We can assume that players adapt their speech interactions in the game to align with the social dynamics typically observed in real-world settings.

5.2. CBM impact

Concerning **RQ2**, we witnessed that a significant majority of participants (76.92%) preferred the sensitive levels of the game that utilized the courtesy-based game mechanic. Additionally, participants found the sensitive levels of the game to be more realistic. This finding also aligns with the previous work that suggests players' in-game voice commands can be associated with a feeling of taking on a character in the game's world (Allison et al., 2019).

The analysis of the PENS questionnaire showed the game provided a sufficient level of player experience with similar ratings for competence, autonomy, presence/immersion, and intuitive controls in both the sensitive and non-sensitive conditions. This suggests comparable experiences between the two conditions and that the CBM did not significantly affect these aspects of player experience. The results of the IEQ indicated that the participants mainly experienced a similar immersion level in both conditions. However, a significant difference was observed in the subscale of *Cognitive Involvement*, where participants in the sensitive condition rated higher cognitive involvement compared to the non-sensitive condition. This suggests that the CBM enhanced cognitive engagement, suggesting that players were more actively involved in the game when their speech interactions had consequences based on their choice of words and volume of speech. This result was expected, as the participants had to be more cautious about how they conversed, adding an additional level of cognitive load during gameplay. Furthermore, our qualitative data highlighted that our participants tended to sense higher levels of realism at the CBM level. This increased realism could contribute to higher cognitive involvement by immersing players in a gaming environment that closely mirrors real-world consequences, aligning with real-life decision-making processes. The analysis of game logs revealed that participants took significantly more time to complete the sensitive levels. This can be attributed to the increased cognitive involvement required by the courtesy-based mechanic. The courtesy-based mechanic likely required players to carefully consider their speech and choose appropriate responses, leading to more thoughtful and deliberate gameplay. The longer gameplay sessions may also be influenced by the reactions of NPCs to impolite intents, triggering the no-response timer and causing delays in the game. Participants exhibited significantly more intent in the sensitive condition than the non-sensitive one. One key factor contributing to this difference could be the implementation of the no-response timer in the sensitive condition. In this condition, participants had to apologize to the NPC to resume interactions after the NPCs detected unfriendly interactions. Consequently, this may have prompted participants to speak more in the sensitive condition. Moreover, the novelty of the CBM likely fostered a more engaging gameplay experience, encouraging them to explore the possibilities and experiment with different conversation intents. Participants also showed a higher frequency of interactions with the colleague. This pattern could also be associated with increased occurrences of *Excuse* cases directed toward this specific NPC.

We witnessed no significant differences in the use of *Chat* and *Thank* intents between the two conditions. However, significantly more *Excuse* intents were observed in the sensitive condition, primarily attributed to the no-response timer. These results indicate that the courtesy-based mechanic did not profoundly alter participants' communication approach.

The aim of designing courtesy-based game mechanics is to encourage civil and respectful conduct among players. However, it is important to acknowledge that implementing such a game mechanic comes with certain challenges. One significant challenge arises from the fact that different cultures have varying standards for what is considered appropriate behavior. Actions perceived as courteous in one culture may be seen as rude or inappropriate in another. Additionally, the context of the game and its environment can influence the appropriateness of certain actions. For example, in a game set in a war zone with loud explosions and chaos, shouting may be necessary for effective communication. However, in a peaceful or domestic setting, such as an apartment or home, shouting directly at someone may be viewed as impolite or threatening. While a courtesy-based game mechanic can be a valuable tool for promoting positive social behavior in video games, it is crucial to recognize the variations in social norms and expectations due to cultural, competitive, and environmental factors. Game designers and developers must be aware of these nuances and strive to create game environments that are respectful and inclusive for all players, regardless of their background or the context in which the game is played. It is also worth mentioning that the intention behind implementing courtesy-based game mechanics is not to act as a form of policing, but rather to discourage toxic behavior. The goal is to foster a gaming community that values positive interactions and encourages players to engage in a manner that enhances the overall gaming experience for everyone involved. Further, even beyond gaming applications, interactions with adaptive NPCs may make it possible to practice difficult social interactions and enhance perspective-taking (e.g., when considering cultural sensitivity training or victimoffender mediation) (Ganschow et al., 2021; Howard & Gutworth, 2020).

5.3. Limitations and future work

The results of this study offer valuable insights into player interactions with NPCs, as well as the courtesy-based game mechanics in speech-based games. However, our work has certain limitations that require acknowledgment.

We recruited ($N = 26$) participants for our study. While this sample size is adequate for an initial exploration of players' interactions with NPCs with different human qualities and the implementation of CBM, future research should validate our results by investigating a broader population. Moreover, in our experiment, the majority of our sample consisted of male participants (80.7%). While the current findings remain valid and provide valuable insights, it is essential to acknowledge that the gender bias present in the sample raises the need for caution in generalizing findings to a more diverse population. Future studies should aim for a more balanced and representative sample to ensure a comprehensive understanding of the topic.

Even though the game "A Day at The Office" allowed players to use various intents to interact with the NPCs, the set of accepted voice commands was still limited. This enabled us to have a structured procedure with high comparability (Porzel & Baudis, 2004). Nevertheless, we suggest that future studies broaden the range of potential actions and expand the command vocabulary to assess the scalability of our findings across diverse application domains.

One of the biggest challenges encountered by our participants during gameplay was regarding the accuracy of speech recognition. The average rate of correctly recognized commands for the session hovered around 60%, a suboptimal figure indicating room for improvement. Notably, the challenges were not solely attributed to intent misrecognition. Participants frequently issued commands that fell outside our vocabulary's scope and could not be accurately recognized. For example, sarcastic or humorous interactions, not accounted for in the game's vocabulary, presented notable obstacles. Moreover, participants' unfamiliarity with the mechanics of speech interactions contributed to interruptions in character dialogues. Instances where players interrupted ongoing character speech or overlapped their commands also led to recognition issues. While recognizing these issues, we acknowledge them as inherent limitations in our game design. In light of these challenges, future studies should place a heightened emphasis on refining speech recognition accuracy.

In our designed game, we incorporated three distinct characters: a male intern, a female colleague, and a male boss. While this selection of characters served our initial purpose of examining player interactions with NPCs possessing varying human qualities, it may not fully represent the broader population. Future research should encompass a wider range of characters, incorporating diversity and inclusivity to ensure a comprehensive exploration of speech interaction between players and NPCs in speech-based games.

This study specifically examined a puzzle game to explore CBM, and as such, the findings might not be universally applicable to other genres of video games. To demonstrate the broader relevance of our CBM, additional research is needed to explore its effectiveness across diverse types of video games.

Overall, while our study has certain limitations, they do not negate the implications of our findings. We witnessed that most participants expressed significant enthusiasm for the game, describing it as interesting and exciting. Notably, only 11.5% of our participants had prior experience with voice-controlled games, indicating a novelty factor for most. This unfamiliarity likely contributed to the heightened interest in our game, given its unconventional nature in speech-based video games. The results underscore that incorporating courtesy-based mechanics in speech-based games can lead to heightened engagement. In light of these positive outcomes, we strongly encourage researchers and designers to delve further into this specific category of video games. The potential for high engagement suggests promising avenues for exploration and development within this gaming niche. There are still several technical challenges when it comes to utilizing speech in video games, particularly with recognition (Zargham et al., 2024), specifically pronounced for non-native speakers and individuals with specific accents or dialects (Pyae & Scifleet, 2018). Nonetheless, the ongoing advancements in AI and natural language processing offer promising solutions, facilitating a smoother and more reliable integration of speech into gaming experiences while addressing these technical hurdles (Bubeck et al., 2023; Zargham et al., 2024). However, thoroughly examining players' perceptions regarding this interaction form in games is essential to fully grasp the intricacies of using speech as a modality for interaction in video games. Further research is needed to explore player experiences, highlighting the challenges and benefits associated with this form of interaction.

6. Conclusion

In this work, we investigated a speech-based video game where players interacted with three NPCs, each possessing unique human qualities. We designed a courtesy-based game mechanic where the player's choice of words and volume of speech could influence their interactions with the game characters. We conducted a within-subjects design study with 26 people to explore player speech interactions with the NPCs and their impressions of the courtesy-based mechanic. Our findings unveiled that participants adjusted their speech patterns based on NPC attributes, such as gender and hierarchical roles. Additionally, the courtesy-based interactions not only showcased diverse player behaviors but also facilitated increased cognitive engagement. The findings of this study can provide valuable insights for researchers and developers in the gaming industry, offering a nuanced understanding of player interactions with speech-based NPCs possessing diverse human qualities. The incorporation of courtesy-based game mechanics, which demonstrated the impact of tactful interactions on cognitive engagement, suggests a promising avenue for enhancing immersion and realism in single-player games. These results support guiding future endeavors to create more immersive and socially dynamic gaming experiences while addressing potential challenges related to anti-social behavior and toxicity in NPC interactions.

Notes

1. https://unity3d.com/unity.
2. https://picovoice.ai.
3. https://syntystore.com/collections/polygon-series.
4. https://www.blender.org/.
5. https://github.com/neonbjb/tortoise-tts.
6. https://discord.com.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partially funded by the Klaus Tschira Foundation, by the FET-Open Project 951846 "MUHAI—Meaning and Understanding for Human-centric AI" funded by the EU program Horizon 2020, as well as the German Research Foundation DFG as part of Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE—Everyday Activity Science and Engineering," University of Bremen (http://www.ease-crc.org/).

ORCID

Nima Zargham http://orcid.org/0000-0003-4116-0601
Leon Tristan Dratzidis http://orcid.org/0009-0005-3504-0621
Dmitry Alexandrovsky http://orcid.org/0000-0001-9551-719X
Maximilian A. Friehs http://orcid.org/0000-0002-9362-4140
Rainer Malaka http://orcid.org/0000-0001-6463-4828

References

Adams, E., & Dormans, J. (2012). *Game mechanics: Advanced game design*. New Riders.

Adinolf, S., & Turkay, S. (2018). Toxic behaviors in esports games: Player perceptions and coping strategies. Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY '18 Extended Abstracts), Melbourne, VIC, Australia (pp. 365–372). Association for Computing Machinery. https://doi.org/10.1145/3270316.3271545

Ahmed, B., Monroe, P., Hair, A., Tan, C. T., Gutierrez-Osuna, R., & Ballard, K. J. (2018). Speech-driven mobile games for speech therapy: User experiences and feasibility. *International Journal of Speech-Language Pathology*, *20*(6), 644–658. https://doi.org/10.1080/17549507.2018.1513562

Allison, F. J. (2020). *Voice interaction game design and gameplay* [PhD dissertation]. University of Melbourne.

Allison, F., Carter, M., & Gibbs, M. (2017). Word play: A history of voice interaction in digital games. *Games and Culture*, *15*(2), 91–113. https://doi.org/10.1177/1555412017746305

Allison, F., Carter, M., Gibbs, M., & Smith, W. (2018). Design patterns for voice interaction in games. Proceedings of the 2018 Annual Symposium on Computer–Human Interaction in Play (CHI PLAY '18), Melbourne, VIC, Australia (pp. 5–17). Association for Computing Machinery. https://doi.org/10.1145/3242671.3242712

Allison, F., Newn, J., Smith, W., Carter, M., & Gibbs, M. (2019). Frame analysis of voice interaction gameplay. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Glasgow, Scotland, UK (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300623

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. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21), Yokohama, Japan (p. 15). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445157

Betker, J. (2023). Better speech synthesis through scaling. arXiv:bib1 [cs.SD].

Bowman, N. D., Weber, R., Tamborini, R., & Sherry, J. (2013). Facilitating game play: How others affect performance at and enjoyment of video games. *Media Psychology*, *16*(1), 39–64. https://doi.org/10.1080/15213269.2012.742360

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). *Thematic analysis* (pp. 843–860). Springer Singapore. https://doi.org/10.1007/978-981-10-5251-4_103

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv:bib2 [cs.CL].

Carter, M., Allison, F., Downs, J., & Gibbs, M. (2015). Player identity dissonance and voice interaction in games. Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15), London, UK (pp. 265–269). Association for Computing Machinery. https://doi.org/10.1145/2793107.2793144

Chandra, Y., & Shang, L. (2019). *Inductive coding* (pp. 91–106). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-3170-1_8

Connelly, L. M., & Peltzer, J. N. (2016). Underdeveloped themes in qualitative research: Relationship with interviews and analysis. *Clinical Nurse Specialist*, *30*(1), 52–57. https://doi.org/10.1097/NUR.0000000000000173

Depping, A. E., Johanson, C., & Mandryk, R. L. (2018). Designing for friendship: Modeling properties of play, in-game social capital, and psychological well-being. Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18), Melbourne, VIC, Australia (pp. 87–100). Association for Computing Machinery. https://doi.org/10.1145/3242671.3242702

Di Petta, T., & Woloshyn, V. E. (2001). Voice recognition for on-line literacy: Continuous voice recognition technology in adult literacy training. *Education and Information Technologies*, *6*(4), 225–240. https://doi.org/10.1023/A:1012972601388

Foo, C. Y., & Koivisto, E. M. I. (2004). Defining grief play in MMORPGs: Player and developer perceptions. Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '04), Singapore (pp. 245–250). Association for Computing Machinery. https://doi.org/10.1145/1067343.1067375

Ganschow, B., Cornet, L., Zebel, S., & Van Gelder, J.-L. (2021). Looking back from the future: Perspective taking in virtual reality increases future self-continuity. *Frontiers in Psychology*, *12*, 664687. https://doi.org/10.3389/fpsyg.2021.664687

Greitemeyer, T. (2013). Effects of playing video games on perceptions of one's humanity. *The Journal of Social Psychology*, *153*(4), 499–514. https://doi.org/10.1080/00224545.2013.768593

Griffiths, M. D., Davies, M. N. O., & Chappell, D. (2004). Demographic factors and playing variables in online computer gaming. *Cyberpsychology & Behavior*, *7*(4), 479–487. https://doi.org/10.1089/cpb.2004.7.479

Harada, S., Wobbrock, J. O., & Landay, J. A. (2011). Voice games: Investigation into the use of non-speech voice input for making computer games more accessible. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.). *Human–Computer Interaction – INTERACT 2011* (pp. 11–29). Springer Berlin Heidelberg.

Hicks, K., Gerling, K., Dickinson, P., Linehan, C., & Gowen, C. (2018). Leveraging icebreaking tasks to facilitate uptake of voice communication in multiplayer games. In A. D. Cheok, M. Inami, & T. Romão (Eds.). *Advances in computer entertainment technology* (pp. 187–201). Springer International Publishing.

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, *203*(3), 88–96. https://doi.org/10.1038/scientificamerican0960-88

Howard, M. C., & Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Computers & Education*, *144*, 103707. https://doi.org/10.1016/j.compedu.2019.103707

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human–Computer Studies*, *66*(9), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

Kojima Productions (2019). *Death Stranding*. Game [Microsoft Windows, PlayStation 4, PlayStation 5].

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

Lee, K. M., Peng, W., Jin, S.-A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of Communication*, *56*(4), 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x

Lina, E. (2015). Playing video games together with others: Differences in gaming with family, friends and strangers. *Journal of Gaming & Virtual Worlds*, *7*(3), 259–277. https://doi.org/10.1386/jgvw.7.3.259_1

Lopes, M., Magalhães, J., & Cavaco, S. (2016). A voice-controlled serious game for the sustained vowel exercise. Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology (ACE '16), Osaka, Japan (p. 326). Association for Computing Machinery. https://doi.org/10.1145/3001773.3001807

McCoy, J., Treanor, M., Samuel, B., Reed, A. A., Mateas, M., & Wardrip-Fruin, N. (2014). Social story worlds with Comme il Faut. *IEEE Transactions on Computational Intelligence and AI in Games*, *6*(2), 97–112. https://doi.org/10.1109/TCIAIG.2014.2304692

McCoy, J., Treanor, M., Samuel, B., Reed, A. A., Wardrip-Fruin, N., & Mateas, M. (2012). Prom week. Proceedings of the International Conference on the Foundations of Digital Games (FDG '12), Raleigh, NC, USA (pp. 235–237). Association for Computing Machinery. https://doi.org/10.1145/2282338.2282384

McCoy, J., Treanor, M., Samuel, B., Tearse, B., Mateas, M., & Wardrip-Fruin, N. (2010). Authoring game-based interactive narrative using social games and Comme il Faut.

McCoy, J., Treanor, M., Samuel, B., Wardrip-Fruin, N., & Mateas, M. (2011). Comme il Faut: A system for authoring playable social models. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, *7*(1), 158–163. https://doi.org/10.1609/aiide.v7i1.12454

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Prentice-Hall. https://www.jstor.org/stable/41718602

Mohammad Mustaquim, M. (2013). Automatic speech recognition-an approach for designing inclusive games. *Multimedia Tools and Applications*, *66*(1), 131–146. https://doi.org/10.1007/s11042-011-0918-7

Morgan, H. (2022). Understanding thematic analysis and the debates involving its use. *The Qualitative Report*, *27*(10), 2079–2090. https://doi.org/10.46743/2160-3715/2022.5912

Murad, C., & Munteanu, C. (2020). Designing voice interfaces: Back to the (curriculum) basics. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Honolulu, HI, USA (pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376522

Navarro-Newball, A. A., Loaiza, D., Oviedo, C., Castillo, A., Portilla, A., Linares, D., & Álvarez, G. (2014). Talking to Teo: Video game supported speech therapy. *Entertainment Computing*, *5*(4), 401–412. https://doi.org/10.1016/j.entcom.2014.10.005

Neely, E. L. (2019). *The ethics of choice in single-player video games* (p. 341–355). Springer International Publishing. https://doi.org/10.1007/978-3-030-01800-9_19

Obsidian Entertainment (2010). *Fallout: New Vegas*. Game [Microsoft Windows, PlayStation 3, Xbox 360].

Osking, H., & Doucette, J. A. (2019). Enhancing emotional effectiveness of virtual-reality experiences with voice control interfaces (pp. 199–209).

Porzel, R., & Baudis, M. (2004). The Tao of CHI: Towards effective human–computer interaction. Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004, Boston, MA, USA (pp. 209–216). Association for Computational Linguistics. https://www.aclweb.org/anthology/N04-1027

Pradhan, A., Mehta, K., & Findlater, L. (2018). "Accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Montreal, QC, Canada (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174033

Pyae, A., & Scifleet, P. (2018). Investigating differences between native English and non-native English speakers in interacting with a voice user interface: A case of Google Home. Proceedings of the 30th Australian Conference on Computer–Human Interaction (pp. 548–553). ACM.

Quandt, T., Chen, V., Mäyrä, F., & Van Looy, J. (2013). (Multiplayer) gaming around the globe?: A comparison of gamer surveys in four countries. In *Multiplayer* (pp. 23–46). Routledge. https://doi.org/10.4324/9780203627488

Ravaja, N. (2009). The psychophysiology of digital gaming: The effect of a non co-located opponent. *Media Psychology*, *12*(3), 268–294. https://doi.org/10.1080/15213260903052240

Reddy, D., Erman, L., & Neely, R. (1973). A model and a system for machine recognition of speech. *IEEE Transactions on Audio and Electroacoustics*, *21*(3), 229–238. https://doi.org/10.1109/TAU.1973.1162456

Reid, E., Mandryk, R. L., Beres, N. A., Klarkowski, M., & Frommel, J. (2022). Feeling good and in control: In-game tools to support targets of toxicity. *Proceedings of the ACM on Human–Computer Interaction*, *6*(2022), 1–27. https://doi.org/10.1145/3549498

Riot Games (2009). *League of Legends*. Game [Microsoft Windows].

Rockstar Studios (2018). *Red Dead Redemption 2*. Game [Microsoft Windows, PlayStation 4, Xbox One].

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 344–360. https://doi.org/10.1007/s11031-006-9051-8

Schell, J. (2008). *The art of game design: A book of lenses*. CRC Press.

Shaw, M. E., Robbin, R., & Belser, J. R. (1981). *Group dynamics: The psychology of small group behavior*. McGraw-Hill. https://books.google.ca/books?id=4RdHAAAAMAAJ

Shen, C., Sun, Q., Kim, T., Wolff, G., Ratan, R., & Williams, D. (2020). Viral vitriol: Predictors and contagion of online toxicity in World of Tanks. *Computers in Human Behavior*, *108*(2020), 106343. https://doi.org/10.1016/j.chb.2020.106343

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Trepte, S., Reinecke, L., & Juechems, K. (2012). The social side of gaming: How playing online computer games creates online and offline social support. *Computers in Human Behavior*, *28*(3), 832–839. https://doi.org/10.1016/j.chb.2011.12.003

Velez, J. A., Mahood, C., Ewoldsen, D. R., & Moyer-Gusé, E. (2014). Ingroup versus outgroup conflict in the context of violent video game play: The effect of cooperation on increased helping and decreased aggression. *Communication Research*, *41*(5), 607–626. https://doi.org/10.1177/0093650212456202

Warpefelt, H., & Verhagen, H. (2017). A model of non-player character believability. *Journal of Gaming & Virtual Worlds*, *9*(1), 39–53. https://doi.org/10.1386/jgvw.9.1.39_1

Whitaker, J. L., & Bushman, B. J. (2012). "Remain calm. Be kind." Effects of relaxing video games on aggressive and prosocial behavior. *Social Psychological and Personality Science*, *3*(1), 88–92. https://doi.org/10.1177/1948550611409760

Wilcox, T., Evans, M., Pearce, C., Pollard, N., & Sundstedt, V. (2008). Gaze and voice based game interaction: The revenge of the killer penguins. *SIGGRAPH Posters*, *81*(10.1145), 1400885–1400972. https://doi.org/10.1145/1400885.1400972

Wilson, C. (2013). Interview techniques for UX practitioners: A user-centered design method.

Zargham, N., Alexandrovsky, D., Friehs, M. A., Zebel, S., Gerling, K., & Malaka, R. (2023). Speaking up against toxicity: Opportunities and challenges of utilizing courtesy as a game mechanic.

Zargham, N., Avanesi, V., Reicherts, L., Scott, A. E., Rogers, Y., & Malaka, R. (2023). "Funny how?" A serious look at humor in conversational agents. Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23), Eindhoven, Netherlands (p. 277). Association for Computing Machinery. https://doi.org/10.1145/3571884.3603761

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. Extended Abstracts of the 2020 Annual Symposium on Computer–Human Interaction in Play (CHI PLAY EA '20) (pp. 408–412). ACM.

Zargham, N., Fetni, M. L., Spillner, L., Muender, T., & Malaka, R. (2024). "I know what you mean": Context-aware recognition to enhance speech-based games. Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), Honolulu, HI, USA (p. 18). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642426

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*, *0*(0), 1–21. https://doi.org/10.1080/10447318.2024.2338666

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. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), New Orleans, LA, USA (p. 13). Association for Computing Machinery. https://doi.org/10.1145/3491102.3502115

Zargham, N., Reicherts, L., Bonfert, M., Voelkel, S. T., Schoening, J., Malaka, R., & Rogers, Y. (2022). Understanding circumstances for desirable proactive behaviour of voice assistants: The proactivity dilemma. Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22), Glasgow, UK (p. 14). Association for Computing Machinery. https://doi.org/10.1145/3543829.3543834

Zhao, R., Wang, K., Divekar, R., Rouhani, R., Su, H., & Ji, Q. (2018). An immersive system with multi-modal human–computer interaction. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 517–524). IEEE. https://doi.org/10.1109/FG.2018.00083

About the authors

Nima Zargham is a PhD student in the Digital Media Lab at the University of Bremen. His research focuses on conversational user interfaces, game user research, multimodal interaction, virtual reality, and privacy and security.

Leon Tristan Dratzidis is a PhD student in the Digital Media Lab at the University of Bremen. His research focuses on haptic user interfaces and virtual reality.

Dmitry Alexandrovsky holds a postdoctoral position at the HCI & Accessibility Lab at Karlsruhe Institute of Technology (KIT), where he works in the area of VR and Accessibility, and develops new research perspectives on disengagement from digital technology.

Maximilian A. Friehs is an assistant professor at the University of Twente (NL) and a visiting Researcher at the Max-Planck Institute for Human Cognitive and Brain Sciences (GER) and the University College Dublin (IRE). His research focuses on behavior change.

Rainer Malaka is a professor of Digital Media at the University of Bremen. He is the head of the Digital Media Lab and the managing director of the Center for Computing Technologies (TZI). His research focuses on multimodal interaction, language understanding, entertainment computing, and artificial intelligence.

Appendix A

The study material, including the customized questionnaire, interview questions, evaluation codebooks, and anonymized transcripts of the interviews, at the following link: https://osf.io/rzwu3/.

Table A1. *Post-hoc* comparisons—sensitivity.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Sensitive	Non-sensitive	0.53	0.15	3.41	0.40	0.0022

Table A2. *Post-hoc t*-test for the number of intents between the different hierarchies.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Intern	Colleague	−1.09	0.14	−7.77	−0.82	1.15×10^{-9}
	Boss	−0.40	0.14	−2.83	−0.30	6.64×10^{-3}
Colleague	Boss	0.69	0.14	4.93	0.52	1.87×10^{-5}

Table A3. *Post-hoc t*-test between the different types of intents (sentiment).

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Friendly	Neutral	−1.32	0.15	−8.71	−1.00	2.68×10^{-11}
	Unfriendly	0.70	0.15	4.61	0.53	2.81×10^{-5}
Neutral	Unfriendly	2.02	0.15	13.32	1.52	1.34×10^{-17}

Table A4. *Post-hoc* comparisons of special intents between different hierarchies.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Intern	Colleague	−0.40	0.07	−5.96	−0.62	7.67×10^{-7}
	Boss	−0.15	0.07	−2.27	−0.24	0.03
Colleague	Boss	0.25	0.07	3.69	0.38	1.12×10^{-3}

Table A5. *Post-hoc* comparisons of special intents between the different game versions.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Sensitive	Non-sensitive	0.24	0.08	3.06	0.37	5.25×10^{-3}

Table A6. *Post-hoc* comparisons of special intents for the interaction between sensitivity × intent.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Sensitive, chat	Non-sensitive, chat	0.08	0.11	0.71	0.12	1.00
	Sensitive, thanks	−0.06	0.12	−0.56	−0.10	1.00
	Non-sensitive, thanks	0.06	0.13	0.49	0.10	1.00
	Sensitive, excuse	−0.21	0.12	−1.78	−0.31	0.63
	Non-sensitive, excuse	0.32	0.13	2.46	0.49	0.21
Non-sensitive, chat	Sensitive, thanks	−0.14	0.13	−1.08	−0.22	1.00
	Non-sensitive, thanks	−0.01	0.12	−0.11	−0.02	1.00
	Sensitive, excuse	−0.28	0.13	−2.16	−0.43	0.37
	Non-sensitive, excuse	0.24	0.12	2.11	0.37	0.38
Sensitive, thanks	Non-sensitive, thanks	0.13	0.11	1.19	0.20	1.00
	Sensitive, excuse	−0.14	0.12	−1.22	−0.22	1.00
	Non-sensitive, excuse	0.38	0.13	2.95	0.59	0.06
Non-sensitive, thanks	Sensitive, excuse	−0.27	0.13	−2.07	−0.41	0.38
	Non-sensitive, excuse	0.26	0.12	2.22	0.39	0.35
Sensitive, excuse	Non-sensitive, excuse	0.53	0.11	4.86	0.81	1.22×10^{-4}

Table A7. *Post-hoc t*-tests for the number of intents for the interaction hierarchy × sentiment.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Intern, friendly	Colleague, friendly	−1.25	0.24	−5.26	−0.94	9.12×10^{-6}
	Boss, friendly	−2.15	0.24	−9.07	−1.62	1.86×10^{-14}
	Intern, neutral	−2.27	0.24	−9.29	−1.71	5.64×10^{-15}
	Colleague, neutral	−3.37	0.25	−13.63	−2.54	1.37×10^{-27}
	Boss, neutral	−1.73	0.25	−7.01	−1.30	1.27×10^{-9}
	Intern, unfriendly	−0.27	0.24	−1.10	−0.20	1.00
	Colleague, unfriendly	−1.19	0.25	−4.83	−0.90	5.17×10^{-5}
	Boss, unfriendly	0.15	0.25	0.62	0.12	1.00
Colleague, friendly	Boss, friendly	−0.90	0.24	−3.81	−0.68	2.05×10^{-3}
	Intern, neutral	−1.02	0.25	−4.13	−0.77	8.02×10^{-4}
	Colleague, neutral	−2.12	0.24	−8.66	−1.59	2.07×10^{-13}
	Boss, neutral	−0.48	0.25	−1.95	−0.36	0.37
	Intern, unfriendly	0.98	0.25	3.97	0.74	1.37×10^{-3}
	Colleague, unfriendly	0.06	0.24	0.24	0.04	1.00
	Boss, unfriendly	1.40	0.25	5.68	1.06	1.16×10^{-6}
Boss, friendly	Intern, neutral	−0.12	0.25	−0.47	−0.09	1.00
	Colleague, neutral	−1.21	0.25	−4.91	−0.91	3.88×10^{-5}
	Boss, neutral	0.42	0.24	1.73	0.32	0.46
	Intern, unfriendly	1.88	0.25	7.63	1.42	4.01×10^{-11}
	Colleague, unfriendly	0.96	0.25	3.89	0.72	1.70×10^{-3}
	Boss, unfriendly	2.31	0.24	9.45	1.74	2.28×10^{-15}
Intern, neutral	Colleague, neutral	−1.10	0.24	−4.62	−0.83	1.33×10^{-4}
	Boss, neutral	0.54	0.24	2.27	0.41	0.22
	Intern, unfriendly	2.00	0.24	8.19	1.51	3.05×10^{-12}
	Colleague, unfriendly	1.08	0.25	4.36	0.81	3.35×10^{-4}
	Boss, unfriendly	2.42	0.25	9.81	1.83	8.46×10^{-17}
Colleague, neutral	Boss, neutral	1.63	0.24	6.88	1.23	3.43×10^{-9}
	Intern, unfriendly	3.10	0.25	12.54	2.33	1.75×10^{-24}
	Colleague, unfriendly	2.17	0.24	8.90	1.64	5.39×10^{-14}
	Boss, unfriendly	3.52	0.25	14.25	2.65	2.36×10^{-29}
Boss, neutral	Intern, unfriendly	1.46	0.25	5.92	1.10	3.81×10^{-7}
	Colleague, unfriendly	0.54	0.25	2.18	0.41	0.24
	Boss, unfriendly	1.88	0.24	7.72	1.42	4.09×10^{-11}
Intern, unfriendly	Colleague, unfriendly	−0.92	0.24	−3.89	−0.70	1.70×10^{-3}
	Boss, unfriendly	0.42	0.24	1.78	0.32	0.46
Colleague, unfriendly	Boss, unfriendly	1.35	0.24	5.67	1.01	1.43×10^{-6}

Table A8. *Post-hoc t*-tests comparisons for the number of intents for the interaction sensitivity × hierarchy.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Sensitive, intern	Non-sensitive, intern	0.14	0.24	0.59	0.11	1.00
	Sensitive, colleague	−1.54	0.21	−7.27	−1.16	1.27×10^{-9}
	Non-sensitive, colleague	−0.50	0.23	−2.20	−0.38	0.22
	Sensitive, boss	−0.53	0.21	−2.49	−0.40	0.12
	Non-sensitive, boss	−0.13	0.23	−0.56	−0.10	1.00
Non-sensitive, intern	Sensitive, colleague	−1.68	0.23	−7.38	−1.27	1.42×10^{-9}
	Non-sensitive, colleague	−0.64	0.21	−3.03	−0.48	0.03
	Sensitive, boss	−0.67	0.23	−2.93	−0.50	0.04
	Non-sensitive, boss	−0.27	0.21	−1.27	−0.20	0.82
Sensitive, colleague	Non-sensitive, colleague	1.04	0.24	4.34	0.78	4.91×10^{-4}
	Sensitive, boss	1.01	0.21	4.79	0.76	7.09×10^{-5}
	Non-sensitive, boss	1.41	0.23	6.20	1.06	2.63×10^{-7}
Non-sensitive, colleague	Sensitive, boss	−0.03	0.23	−0.11	−0.02	1.00
	Non-sensitive, boss	0.37	0.21	1.76	0.28	0.49
Sensitive, boss	Non-sensitive, boss	0.40	0.24	1.66	0.30	0.50

Table A9. *Post-hoc* comparisons of special intents for the interaction between sensitivity × hierarchy.

		Mean diff	*SE*	*t*	Cohen's *d*	p_{holm}
Sensitive, intern	Non-sensitive, intern	0.08	0.11	0.72	0.12	1.00
	Sensitive, colleague	−0.62	0.09	−6.73	−0.94	1.69×10^{-8}
	Non-sensitive, colleague	−0.12	0.11	−1.04	−0.18	1.00
	Sensitive, boss	−0.19	0.09	−2.10	−0.29	0.34
	Non-sensitive, boss	−0.04	0.11	−0.35	−0.06	1.00
Non-sensitive, intern	Sensitive, colleague	−0.69	0.11	−6.27	−1.06	3.34×10^{-7}
	Non-sensitive, colleague	−0.19	0.09	−2.10	−0.29	0.34
	Sensitive, boss	−0.27	0.11	−2.44	−0.41	0.17
	Non-sensitive, boss	−0.12	0.09	−1.26	−0.18	1.00
Sensitive, colleague	Non-sensitive, colleague	0.50	0.11	4.69	0.77	1.75×10^{-4}
	Sensitive, boss	0.42	0.09	4.63	0.65	1.36×10^{-4}
	Non-sensitive, boss	0.58	0.11	5.22	0.88	2.12×10^{-5}
Non-sensitive, colleague	Sensitive, boss	−0.08	0.11	−0.70	−0.12	1.00
	Non-sensitive, boss	0.08	0.09	0.84	0.12	1.00
Sensitive, boss	Non-sensitive, boss	0.15	0.11	1.44	0.24	1.00