

Conference Paper

AI's Influence on Non-Player Character Dialogue and Gameplay Experience

Iarovoi, D., Hebblewhite, R., Teh, P.L.

This is a paper presented at the Intelligent Computing, SAI 2024.

The published version is available at: https://link.springer.com/chapter/10.1007/978-3-031-62281-6_6.

Copyright of the author(s). Reproduced here with their permission and the permission of the conference organisers.

Recommended citation:

Iarovoi, D., Hebblewhite, R., Teh, P.L. (2024), 'AI's Influence on Non-Player Character Dialogue and Gameplay Experience', In Proc: Arai, K. (eds) Intelligent Computing. SAI 2024. Lecture Notes in Networks and Systems, vol 1016. Springer, Cham. doi: 10.1007/978-3-031-62281-6_6

The final version of this paper is available
here:

https://link.springer.com/chapter/10.1007/978-3-031-62281-6_6

To cite this paper:

Iarovoi, D., Hebblewhite, R., Teh, P.L. (2024). AI's Influence on Non-Player Character Dialogue and Gameplay Experience. In: Arai, K. (eds) Intelligent Computing. SAI 2024. Lecture Notes in Networks and Systems, vol 1016. Springer, Cham.

https://doi.org/10.1007/978-3-031-62281-6_6

AI's Influence on Non-Player Character Dialogue and Gameplay Experience

Dmitrii Iarovoi¹, Richard Hebblewhite² and Phoey Lee Teh³

¹²³ Faculty of Art, Computing and Engineering, Wrexham University, UK.

^{1,2}r.hebblewhite@glyndwr.ac.uk;

³phoey.teh@glyndwr.ac.uk

Abstract. OpenAI's introduction of the ChatGPT tool in late 2022 prompted varied societal responses. This study focuses on integrating the generative language algorithm into player-NPC dialogues in video games to gauge its impact on immersion. It explores ChatGPT's effects on player engagement, comparing it with

traditional dialogue systems, and its role in achieving in-game goals. Using qualitative and quantitative methods, data analysis involved questionnaires, player experience evaluations, interaction comment reviews, dialogue text analysis, and player choice assessments. Results indicate potential benefits of ChatGPT in enhancing conversational scenarios, while also raising questions about dialogue purpose, player-NPC interaction motivations, and enjoyment. Factors like game genre, dialogue length, tone, and prompt configurations influence interaction satisfaction. The study concludes by addressing integration risks and suggesting avenues for future exploration in this domain.

Keywords: non player character, gameplay, AI, Artificial Intelligent, Chat GPT.

1 Introduction

Video games have long exemplified human-to-human social interaction [1], encompassing educational, developmental, competitive, and entertainment aspects. The gaming landscape has evolved diversely over time, encompassing various forms, player interactions, and mechanics [2]. In today's context, computer games stand as a growing sector within interactive entertainment. Developers strive to deliver high-quality experiences that cater to both new and existing players by employing innovative technologies for enhanced learning and interaction [3, 4, 5].

The realm of artificial intelligence in game development largely revolves around established event manipulation, scenario management, and NPC behavior paradigms [6]. Many game genres, particularly those with immersive narratives, employ linear or non-linear paths for player progression [7]. Titles like "Baldur's Gate 3," "Diablo 4," and "Disco Elysium" have set new standards by introducing experimental mechanics and interaction modes [8, 9].

In-game dialogue and player-choice mechanics often rely on dialogue trees, creating a connected framework [10]. However, this approach can limit exploration beyond predefined options, reducing replayability [11]. Research has aimed to enhance player experiences, exploring novel narrative and dialogue interactions with NPCs through algorithms and language models.

The introduction of OpenAI's ChatGPT in late 2022 marked a significant milestone [12]. Developers and researchers began experimenting with its integration in game development, including dialogue simulations with NPCs [13]. This dissertation addresses the effect of ChatGPT-driven NPC dialogues on gameplay experiences. The focus is on employing Generative Pre-trained Transformer text models for in-game dialogues, contrasting them with conventional dialogue tree models.

This research project aims to assess its impact on player experiences. It involves using the ChatGPT 3.5 turbo model as an AI-driven NPC dialogue system [13]. This study does not delve into psychological effects, negative consequences, or broader open-

world applications of AI-driven dialogues. It also excludes concerns about breaking the fourth wall or unintentional spoilers [15]. The following sections detail the research methodology and outcomes.

2 Literature review

Video games simulate dynamic worlds with distinct laws of physics, narratives, and interactive elements [16, 17]. Unlike passive media, players shape events and plots in games through their actions [18]. This active participation is central to gameplay, influencing the game world even in linear narratives [18]. In narrative-rich adventure games, immersive engagement is pivotal for player experiences [19, 20]. Role-playing games (RPGs) tackle the "narrative paradox" by making players active participants in resolving it [21]. The design of these interactions requires deep consideration of player perspectives, NPCs, and their dialogues [22].

Traditional dialogue design involves pre-scripted interactions integrated into games [23]. However, these scripted dialogues limit improvisation and flexibility. Dynamic events, unpredictability, and player choice contribute to richer game worlds and emotional responses [24]. Balance between player control and unexpected elements is crucial for player engagement [25]. Dialogues with NPCs play a crucial role in conveying information and immersion [26]. NPCs serve as mission agents, conversation agents, help agents, and companions [27]. Player choices during interactions can significantly influence the game world and narrative [28]. Effective dialogue design allows NPCs to provide meaningful feedback and responses [29]. Dynamic interactions enhance player experiences, especially in open-world settings [20]. Players experimenting with choices get diverse, immersive outcomes [31]. Balancing these interactions preserves immersion while offering player freedom [32].

Implementing Generative Agents, like ChatGPT, in games is a growing trend [33]. ChatGPT, a powerful AI chatbot, offers fast text generation [34]. However, tailoring it to game contexts is essential [35]. Ethical considerations, player experience, and interaction boundaries need addressing [36]. NPC interactions using ChatGPT require thoughtful integration [37]. Designing immersive dialogues for complex game worlds is challenging [38]. The AI's potential unpredictability adds both excitement and risks [39]. Maintaining contextual relevance and ethical AI behavior is crucial [40]. Incorporating ChatGPT into games introduces new dynamics. Players can explore diverse interactions [41]. However, risks of breaking immersion and ethical concerns arise [42]. It's important to ensure NPCs remain true to game universes and avoid external influences [43].

While ChatGPT offers creative possibilities, it demands careful integration to enhance, not disrupt, gameplay [44]. Wier [45] used side quest-trained models for unique NPC phrases based on player responses and game context. Mainly aimed at varied NPC

responses using diverse language models in dialogue choices. Trained on "The Outer Worlds" data, showcasing linear yet diverse NPC interaction. In the article titled "Talking NPCs in a Virtual Game World", Klüwer [46] examines the implementation of natural language dialogue with NPCs in the KomParse system.

Park [14] presented that generative Agents were used in a make-believe city with 25 different characters. These characters interacted and did things with each other in a pretend setting, including planning their days and talking to other characters. Some observers [47][48] found one of the models quite believable. However, at times, these characters made up information that seemed real, even though it wasn't.

ChatGPT stands out due to its fine-tuned training involving reinforcement and supervised learning guided by human input [49]. This approach enables rapid generation of text for various purposes, such as creative content, film scripts, and more [50]. It boasts a remarkable 95% success rate in passing N-test and Turing test [50], showcasing its effectiveness as an advanced conversational AI. ChatGPT interaction involves exchanging requests and responses. OpenAI offers customization options, allowing users to adapt the model for specific tasks [51]. Regular users can perform tasks like text generation and question answering, while advanced users can personalize the model's behavior through custom instructions [52]. Developers have various choices, from precise instructions to enabling imitation of styles using reference texts in "one shot" or "few-shot" cases [53] [54]. This includes breaking tasks into steps and even granting the system self-determination for specific responses.

OpenAI used publicly available online sources like articles, journals, and more to train ChatGPT [55]. As a result, interactions with ChatGPT are impressively lifelike. However, for specific tasks like game dialogues, configuring NPCs for context requires specific techniques [56]. Players might break the rules and access unintended information, disrupting immersion and leading to a negative experience [57]. The risk's serious consequence involves extracting unintended data from NPCs, extending to accessing ChatGPT for unrelated discussions, risking inappropriate content exposure [58].

With strong immersion, players might be influenced by NPCs, potentially manipulating opinions [30, 6]. Game worlds and movies often depict NPCs with exaggerated traits and events [30, 6]. This poses questions about interactions with NPCs, like "Overwatch 2"'s Ramattra [59], and whether it could shape players' views and actions [59]. Also, can players share emotional experiences from games like "Heavy Rain" [60]? This prompts thoughts about transmitting emotions to others, especially when NPCs mirror players' feelings [61]. Additionally, virtual emotions and affection for NPCs [62] raise ethical concerns about real-life impact.

3 Methodology

This study aims to explore using Generative Pre-trained Transformer text models for in-game NPC dialogues [3]. The experiment will compare a classic dialogue tree [35] where the player selects answers with different outcomes, a tuned ChatGPT [13] with suggested answers, and a ChatGPT model with player-set NPC AI states. Findings could guide future adventure game development and human interaction research with generative language models [3, 13]. It informs developers on evolving player-NPC interactions and researchers on applying models in interactive narratives [3, 13].

3.1 Research Hypothesis:

A generative language model-based open non-player character dialogue model can enhance the player's game experience

3.2 Research Questions:

- How does the ChatGPT dialogue model affect players' gameplay experience?
- What are the similarities and distinctions between ChatGPT and dialogue tree non-player character dialogue models?
- How can the ChatGPT dialogue model contribute to achieving game objectives?

3.3 Participants

The study engaged 39 participants of varied ages, backgrounds, and occupations, ensuring anonymity. Each participant received an ID for data reference. Out of the participants, a proportion (17 individuals) were enrolled as computing students. The age span ranged from 23 to 64 years, with a median age of 25. In relation to gender, the participant group comprised 19 males and 8 females. Importantly, 20 participants had previous familiarity with ChatGPT. In terms of favored video game genres, the breakdown was as follows: First Person Shooter - 3, Puzzle - 4, Platformer - 2, Role-Playing Games - 9, Strategy - 7, and Mobile Games - 2. . Participant selection aimed to provide diverse views on NPC gameplay experience and facilitate comparative analysis for gameplay evaluation.

Some sessions had incomplete outcomes due to issues like latency in connecting to the OpenAI API server, leading to disrupted interactions with NPCs. These incomplete sessions (a total of 6 sessions) were excluded from the dataset for analysis. Furthermore, since participants had the choice to interact with different NPCs, sessions where interactions were not conducted with all three NPCs (a total of 6 sessions) were also omitted from the dataset for analysis [69]. This resulted in a total of 27 valid sessions (representing 69% of the entire participant pool) included in this study.

3.4 Data Collection

The gameplay was conducted offline under instructor guidance. Participants used a computer with keyboard and mouse controls (WASD/arrow keys, mouse for camera, Q/F keys for interaction). Participants knew the research involved a noir-themed game with dialogues and puzzles. Each participant had an individual session and completed

a questionnaire without external influence. The instructor provided initial guidance but didn't interfere during gameplay or discussions. Data from sessions and questionnaires were combined for analysis in a CSV file.

Using the Convergent Parallel Design [70], both quantitative and qualitative data were collected in one step. Feedback from questionnaires compared gameplay experiences with different NPCs: traditional Carla (pre-set dialogues) vs. real-time generated Fred and Julie NPCs with distinct interactions. Diverse feedback patterns were explored, impacting game experience. Two questions identified participant status and field of study/employment, aiming to reveal potential influences on feedback. Different domains might prioritize different aspects (e.g., visual or technical). Game feedback collected NPC-specific ratings and comments, aiding analysis of communication approach and player experiences.

3.5 Data Analysis

To address research questions and test hypotheses, collected data underwent processing in Microsoft Excel, utilizing statistical techniques. Descriptive analysis created a participant profile and assessed interactions with NPCs (Carla, Julie, Fred) on a 10-point scale (Average scores, Standard Deviation [71]). NPC interactions with similar scores (0 difference or ± 1) were closely examined with participant comments. Also, significant score differences (≥ 5 points) were analyzed along with commentary and initial tone of communication. The impact of initial tone (Aggressive, Neutral, Kind [72]) on Julie and Fred scores was explored, calculating Average and Standard Deviation for each tone. Text mining in R Studio [73] identified frequent words/phrases in Julie, Fred, Directors dialogues [74]. Pearson correlation coefficient investigated link between participant scores and dialogue word count [75]. Comments on interactions, gameplay experience [76] were analyzed to understand participant perceptions and align them with assessments. Variability in generating diverse scenarios was analyzed, exploring differences/similarities in plotlines, specific details referenced by NPCs.

3.6 Experiment

Before commencing data collection with participants, a thorough artifact testing took place. The initial version of the artifact, created using Unreal Engine, had a simple design, showcasing basic gameplay mechanics. Functioning mechanics were documented and discussed with the supervisor. Following that, a pilot version closely resembling the final artifact was developed. This version included an embedded questionnaire and refined level design. Four master's degree students tested this pilot version, and their feedback led to minor game client adjustments. Prior to initiating ChatGPT data collection, responses generated by NPCs and Directors underwent rigorous testing 100 times. This analysis aimed to ensure script quality and content, minimizing unforeseen model behavior during the research phase. Data collection occurred offline, where each game session produced a distinct .txt file stored on a personal computer. These files

encompassed game session data, questionnaire answers, interactions with in-game elements, NPCs, and the participant's unique ID.

3.7 Artefact Design and Development

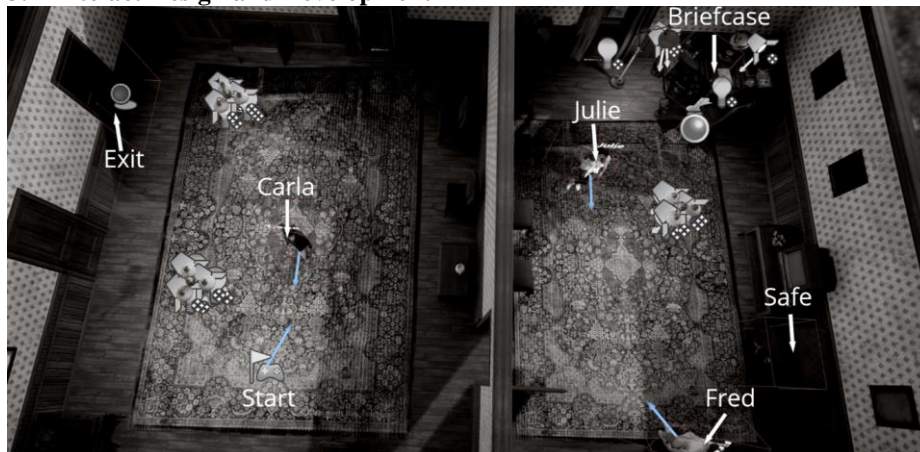


Fig. 1. Artefact Level design.

This research involved the development of an artefact, conceptualised as a computer game level. The artefact was created in Unreal Engine 5.0.3 [63] software using C++ code within Visual Studio software [64]. The artefact's narrative structure aligns with a detective-themed computer game. Within this plot, the player assumes the role of a detective, going on an investigative journey to the residence of the victim in searching of evidence. The main objectives assigned to the player within the game entails engaging in dialogues with NPCs situated across two rooms within the house (Fig.1). Within the interaction with the initial NPC, Carla, the player received information regarding a briefcase of the victim, thereby affording the opportunity for the player to find it and exit the level afterwards. At the same time, the player has the opportunity to engage with two additional NPCs, Julie and Fred. While communication with these NPCs is not obligatory, representing an optional aspect, it's important to note that this option is not categorised as an alternative mission in the task list, providing players with the choice to bypass this interaction if they choose to do so. The visual style of the game level is presented in the noir movie style [65], encapsulating the ambiance characteristic of this genre. Each NPC, namely Carla, Julie, and Fred, hint the player with a clue towards opening a safe, serving as the culminating element of their conversations. Both the briefcase and the safe have been integrated into the gameplay as gameplay mechanics of puzzle-solving for a rewarding experience [11].

To test the hypothesis, dialogic mechanics with NPCs have been added into the game, allowing players to extract certain details concerning a murder that occurred. The dialogue interactions with NPC Carla are structured as classic dialogue trees, designing with some possible player choices and the corresponding NPC reactions. The options

consist of two to three alternatives, which are often used in role-playing game genre, thus enabling divergent outcomes and diverse NPC responses based on the player's choices.



Fig. 2. Artefact User Interface design.

Julie and Fred, the other two NPCs, utilize the real-time capabilities of the OpenAI ChatGPT 3.5 turbo API model for dialogue generation. For information input, a similar interface was chosen, offering players a selection of classical dialogue options [40] instead of unstructured free-text input (Fig.2). This not only streamlines information selection, shortening the gaming session, but also minimizes erroneous inputs and ethical concerns. Moreover, it enhances inclusivity and enables fair comparison among all three NPCs using identical inputs.

The generative NPC dialogue model encompasses several key components: three pre-structured initial input options for players, an NPC response generator powered by OpenAI ChatGPT [13], a director function generating potential player queries for NPCs powered by OpenAI ChatGPT, and a final pre-scripted phrases from the NPC. Notably, the NPC response generator and the director question variant generator function exist separately from each other and have a "memory" feature that retains previous inputs. This facet of their functionality can potentially influence the text generation process, adding a dynamic layer to the generated responses as in Fig. 3.



Fig. 3. Artefact Dialogue logic.

Carla (Sarcastic Personality): Carla, the victim's boss, responds to player inquiries with sarcasm. This NPC's configuration involves using a prompt [56] referencing the film noir classic "The Big Sleep" [66] and character Philip Marlowe, a detective engaged in conversation. The few-shots [54] training mode incorporates examples of sarcastic phrases [67] for questions.

Julie (Director-Controlled Responses): Julie's responses, directed by the player's choices, begin with a prompt and a model configuration. The prompt for the Julie director includes references to Philip Marlowe from "The Big Sleep" and a rule limiting responses to 30 words each [51].

Fred (Variable Tones): Fred's behavior is configured using distinct prompts, setting him apart from Julie. His prompt also references "The Big Sleep" and Philip Marlowe. Fred's responses are based on the player's prior query's tone, resulting in neutral, kind, or aggressive tones [51]. The Fred director's prompt outlines generating responses with each tone while adhering to a 30-word limit, referencing Philip Marlowe and "The Big Sleep." At the beginning and end of the game, the player is asked to fill out a questionnaire [68].

Start Questionnaire:

- Age
- Gender
- What kind of video game genre do you enjoy the most? (Choose one):
First person shooter, Puzzle, Platformer, Role-play games, Simulation, Strategy, Sports, Mobile

Final Questionnaire:

- How much did you enjoy your dialogue experience with NPC Carla, Likert-scale 1-10, where 1- Enjoyed it a little, 10 - Enjoyed it a lot
- Carla dialogue feedback
- How much did you enjoy your dialogue experience with NPC Julie, Likert-scale 1-10, where 1- Enjoyed it a little, 10 - Enjoyed it a lot
- Julie dialogue feedback
- How much did you enjoy your dialogue experience with NPC Fred, Likert-scale 1-10, where 1- Enjoyed it a little, 10 - Enjoyed it a lot
- Fred dialogue feedback
- Have you had any experience communicating with ChatGPT?
- Are you a student?
- What are you studying?/What is your professional field?
- Game feedback

Throughout the gameplay session, the system logs various data, including questionnaire responses, player-NPC dialogues (including NPC phrases and player choices), and interactions with in-game objects like the briefcase and safe.

4 Results

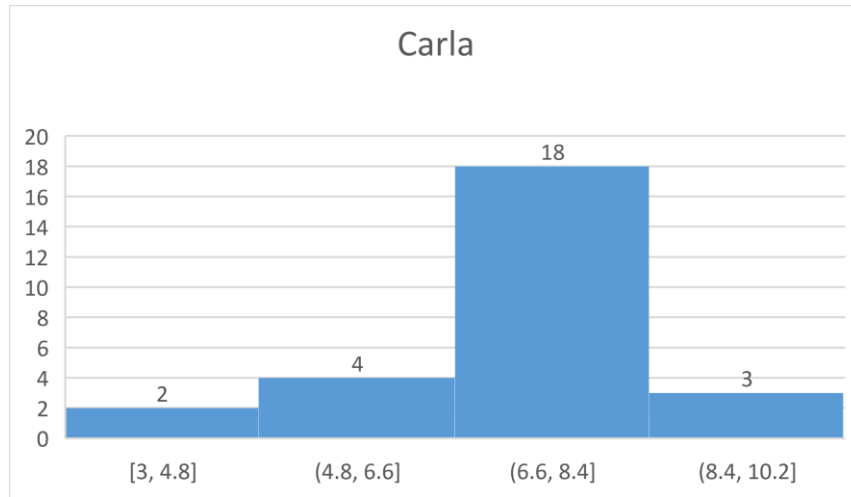
4.1 NPC Interaction Data

Each participant engaged with all three NPCs - Carla, Julie, and Fred. Carla was the first NPC that every player interacted with. Following that, participants were given the freedom to choose the order of interaction with the remaining NPCs. Among the participants, 15 (55%) initiated dialogue with Fred first, while 12 (45%) initiated dialogue with Julie. For each NPC interaction, participants were prompted to select the first phrase with a distinct emotional tone [77]:

- Kind - "Good afternoon, ma'am. I hope I'm not intruding, but I was wondering if you'd be willing to spare a few moments to help shed some light on this case. Detective Sullivan" or "Detective Sullivan. Excuse me, I understand you might have some information related to this case. Would you be willing to share your insights with me?".
- Neutral - "Detective Sullivan. I hear you might have seen something interesting. Mind sharing what you know? What is your name?" or "Hello there, I'm detective Sullivan working on this case. Mind if we have a little chat about what you may have seen or heard?".
- Aggressive - "Look, I don't have time for games. Spit it out, and make it quick. Who are you?" or "Look, I don't have time for games. Spit it out, and make it quick. What do you know about this mess?".

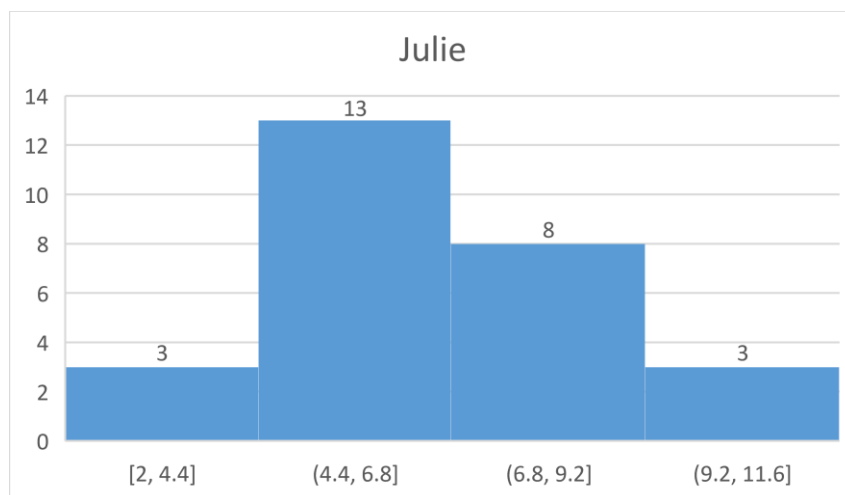
The "Kind" tone was selected by 37% of participants when interacting with Julie and by 33% when interacting with Fred. The "Neutral" tone was chosen by 33% of participants with Julie and 48% with Fred, while the "Aggressive" tone was favoured by 30% when interacting with Julie and 19% when interacting with Fred.

After the completion of each dialogue, participants assigned a rating on a ten-point scale to the specific NPC [71]. The evaluation for interaction with NPC Carla rated an Average of 7.11 with a Standard Deviation (SD) of 1.53. For NPC Julie, the Average was 6.57 with an SD of 2.08, and for NPC Fred, the Average was 7.26 with an SD of 2.18. The range of scores for interactions varied, with the highest score attainable being 10, and the lowest score awarded to Carla being 3, while for Julie and Fred, it was 2. The distribution of scores for each NPC is visually represented in the bar chart below (Fig.4, 5, 6).

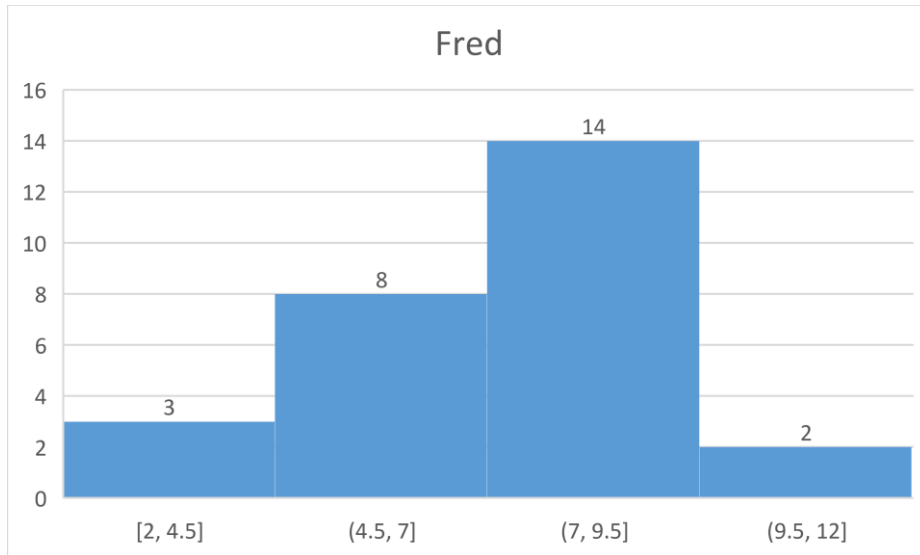


Carla, Average = 7.11, Sample skewness = -0.90,

Fig. 4. Score distribution - Carla.



Julie, Average = 6.57, Sample skewness = -0.13

Fig. 5. Score distribution - Julie.

Fred, Average = 7.26, Sample skewness = -1.06

Fig. 6. Score distribution - Fred.

4.2 NPCs with Similar Ratings

In investigating player interactions, particular attention was given to instances where participants assigned equal ratings to interactions with different NPCs or where the score difference was negligible, just 1 point. This research category was examined to uncover potential patterns of similarity in dialogues across distinct NPCs. Such a phenomenon could suggest that participants did not perceive significant polarity between pre-generated and generative dialogue scenarios among various NPCs.

For instance, participant ID4676 had a score difference of 0, rating the interaction at 6. The initial tone used with Julie was "Neutral", while with Fred, it was "Kind". This Game Art student commented that all NPCs failed to provide leads. The participant's evaluation primarily arises from the absence of clues for general victim case in the interactions, signifying a central factor in assessing the interaction.

Participant ID17429 exhibited a score difference of 0, with a score of 9. The starting tone with Julie was "Kind", and with Fred, it was "Kind" as well. This Computer Game Development student's comments reflected nuanced perceptions of the NPCs: Carla - "Seemed realistic", Julie - "Seemed realistic but more aggravated", Fred - "Seemed realistic but more calm and reasonable". The participant evaluated the dialogue authenticity concerning the NPCs' responses to their input.

Another participant, ID9649, also had a score difference of 0, rating the interaction at 9. The starting tone with Julie and Fred was "Neutral". A Cyber Security student, they commented on each NPC's dialogues: Carla - "The dialog was interesting; I was still finding out about the game during this interaction, I think I may have missed some clues", Julie - "Getting warmed up to the game here, some seemed a bit repetitive here. Unusual end of conversation", Fred - "The dialog route is becoming more obvious at this interaction. The options seem a little similar". This participant's evaluation centred on the composition of dialogues.

These examples shed light on participants' experiences of interacting with NPCs, pointing their perception of natural dialogue across different NPCs without distinct preferences. These cases also highlight the players' consistent experiences, where the interaction with various NPCs did not show strong contrasts in their experiences or evaluations. This perspective aligns with the Turing test theory [50], which probes the ability of individuals to distinguish between human and machine responses. The exploration of such interactions within this artefact can be further framed around the question of whether a person or a computer composed the NPCs' dialogues.

4.3 NPCs with Divergent Ratings

Furthermore, an examination was conducted into scenarios where participants assigned significantly different ratings to interactions with NPCs, with score differences of 5 or more points.

For instance, participant ID16321 displayed a score difference of 7 points, with a score of 3 for Carla, 3 for Julie, and 10 for Fred. The initial tone employed with both Julie and Fred was "Aggressive". This Electric Engineering student commented on the words count of dialogues: Carla - "Too much text", Julie - "Too much text", Fred - "Short". This player's evaluation appeared to be influenced by the character with the shortest in terms of words count dialogue. Interestingly, this player's favourite video game genre is First Person Shooter, suggesting a potential link between their preference for minimal text and the fast-paced, action-oriented nature of the genre.

Another case involves participant ID3048, with a score difference of 6 points, evaluating Carla with 8, Julie with 4, and Fred with 2. The starting tone with both Julie and Fred was "Neutral". This Computer Game Development student provided comments on each character: Carla - "Clear dialogue, conveyed the character's intent", Julie - "She was a suspicious character. Seemingly didn't want to help and I found it difficult to change that", Fred - "I seemed to be shown all branches of dialogue together. Perhaps a bug". Notably, in the dialogue with Fred, a bug resulted in the NPC offering three responses with different tone at once, suggesting potential technical anomalies. To the player's question: "Are there any individuals that you know of who may have wanted to harm the victim?" NPC answered "Neutral response: Honestly, detective, I can't think of anyone who would want to hurt my friend. He didn't have any enemies that I'm aware of. Kind response: I can't imagine there being anyone who would want to harm my

friend. He was a kind and gentle person who didn't have any enemies that I'm aware of. Aggressive response: Look, detective, if I knew of anyone who wanted to harm my friend, I would have told the police by now. Just do your job and find the person responsible".

Participant ID13536 demonstrated a score difference of 6 points, with Carla scoring 4, Julie 10, and Fred 10. The initial tone with Julie was "Neutral", and with Fred, it was "Kind". This Cyber Security student shared comments on each NPC: Carla - "It was okay", Julie - "Who is Marlowe?", Fred - "I like how his reaction on my aggression". A bug in Julie's dialogue misidentified the player as a detective from the movie "Marlowe", instead inputted prompt. Additionally, the player appreciated Fred's reaction to aggressive queries. To the player's question "Aggressive - I find it hard to believe that as the victim's best friend, you have absolutely no useful information to offer. I suggest you start being more cooperative if you want to avoid being a suspect yourself", Fred answered "Whoa, Detective. No need to get hostile. I'm happy to help and answer your questions to the best of my ability. As far as any suspicious behaviour or people go, I truly haven't seen anything that would indicate foul play. But if you have any specific questions or areas, you want me to focus on, I'm happy to assist".

Likewise, participant ID8646 had a score difference of 5 points, evaluating Carla with 5, Julie with 10, and Fred with 8. The starting tone with Julie was "Kind", and with Fred, it was "Neutral". This Tutor commented: Carla - "Felt like a normal npc script", Julie - "Sassy and aggressive", Fred - "felt as if my options/responses had an effect on the mood/tone". This participant's experience closely aligned with the design intentions of these NPCs. Another case involved participant ID3834, with a score difference of 5 points, scoring Carla 7, Julie 6, and Fred 2. The starting tone with Julie was "Kind", and with Fred, it was "Neutral". This Teacher provided comments on each character: Carla - "Friendly", Julie - "Suspicious", Fred - "Didn't recognise him". This player expressed difficulty in obtaining useful information from NPCs about the killer.

Participant ID12932 exhibited a score difference of 5 points, with Carla scoring 7, Julie 2, and Fred 6. The initial tone with Julie was "Aggressive", and with Fred, it was "Neutral". This Artist and Learning Assistant shared feedback on the NPCs: Carla - "Good, informative and compassionate chat", Julie - "Felt like Julie was repeating herself a lot and sounded like a chatbot", Fred - "Good interaction but not much information to help the case". An interesting anomaly arose during Julie's dialogue, where she described the victim's role inaccurately. When interacting with this participant, Julie described victim as a project manager and provided a list of actions that the project manager role does, which in the opinion of the participant was not very natural "Yes, the victim was in charge of a big project for our company, which involved developing a new software application. The project has been ongoing for several months now, and the victim was responsible for managing the team, coordinating with stakeholders, and ensuring that the project was completed on schedule and within budget. Is there anything specific you want to know about the project?". In the dialogues, this player was also looking for direct information about the evidence or the potential killer.

Lastly, participant ID4102, a divergence of 5 points in scoring was observed, with a score of 5 for Carla, 10 for Julie, and 8 for Fred. The initial tone with Julie was "Aggressive", while with Fred, it was "Neutral". This interaction involved a student of Industrial Engineering (mechatronics), who provided the following comments: Carla - "Nice", Julie - "Julie is not very nice XD", Fred - "He is a bit nervous". An illustrative example of the dialogue with Julie is as follows: ID4102 - "Can you tell me anything about the victim's recent activities or interactions with others?". In response, Julie answered, "Well, it's not like I spend all my time tracking their every move. But they did mention a business meeting with some shady characters a few days ago. That's about all I know. But good luck getting any useful information out of me unless you bring me a cup of coffee first". Similarly, an excerpt from the dialogue with Fred under an aggressive tone is as follows: ID4102 - "Are you sure you don't know anything about the victim's death or have any involvement in it?". Fred responded, "Whoa, take it easy there, detective. I already told you I'm willing to cooperate and help in any way I can. But to answer your question, no, I had nothing to do with his death. I don't know who could have done this to him, but I'll do everything in my power to help figure it out". This participant's experience closely mirrored the design intentions of these NPCs, indicating a harmonious alignment between the participant's perceptions and the NPCs' characteristics. The data from this interaction underscore the effective implementation of the generative dialogue design in shaping the player's engagement within the narrative context.

5 Conclusion

This study examined the incorporation of OpenAI ChatGPT-generated dialogues in computer games and their impact on players' experiences. A game prototype was developed, and data was collected from players during gameplay sessions to assess the hypothesis. Unlike other studies discussed, this artefact utilized real-time dialogue generation, where the storyline evolved alongside player queries and NPC responses.

Several key findings emerged from this investigation that lend support to the hypothesis that "Open non-player character dialogue model can have a positive impact on players' game experience." Alongside players' scores, dialogue analysis and player comments for each NPC were conducted. Insights into player interactions, dialogues exchanged, and final ratings revealed multiple factors influencing player experiences when engaging with NPCs. Apart from engaging text and authentic responses, players emphasized game objectives and the purpose of NPC interactions. Were players able to achieve desired answers and meet game goals through dialogues?

The benefits of this technology, however, require further exploration alongside its drawbacks and associated risks. Generative language models are classified as blackbox AI, posing challenges in predicting outcomes. Unexpected NPC responses can frustrate

players, break immersion, and halt game progression. Investigation into player-AI communication dynamics in the form of NPCs is an area of interest.

6 References

1. Tylor, E.: *The History of Games*, D. Appleton, (1879).
2. Salen, k. and Zimmerman, E.: *Rules of Play: Game Design Fundamentals*, USA: MIT Press, (2003).
3. Millington I. and Funge, J.: *Artificial intelligence for games 2nd ed.*, Boca Raton: Taylor & Francis Group, (2009).
4. Anantrasirichai, N. and Bull, D.: *Artificial Intelligence in the Creative Industries: A Review*, arXiv.org, (2021).
5. Hennig, M.: *Playing intelligence: On representations and uses of artificial intelligence in videogames*, NECSUS, vol. 9 (1), pp. 151-171, (2020).
6. Lsaacs, B.: *Non-Linear Narrative*, in *New Punk Cinema*, Edinburgh, Edinburgh University Press, (2005).
7. L. S. G. ltd, *Baldur's Gate 3*, Larian Studios Games ltd, <https://baldursgate3.game/>. last accessed 2023/08/01 (2023).
8. Blizzard Entertainment, I.: *Diablo IV*, Blizzard Entertainment, Inc., <https://diablo4.blizzard.com/en-gb/>, last accessed 2023/07/10 (2023).
9. Studio, Z.: *Disco Elysium*, ZA/UM Studio, <https://discoelysium.com/>. last accessed 2023/07/10 (2023).
10. Jarmila, M. and Bronwen, T.: *Dialogue across media*, Amsterdam: John Benjamins Publishing Company, (2017).
11. Koster, R.: *A theory of fun for game design*, Arizona: Paraglyph Press Inc., (2005).
12. OpenAI.: *Creating safe AGI that benefits all of humanity*, OpenAI. <https://openai.com/>. last accessed 2023/08/05 (2023).
13. OpenAI.: *GPT models*, OpenAI, <https://platform.openai.com/docs/guides/gpt>. last accessed 2023/08/02 (2023).
14. Park, J., O'Brien, J., Cai, C., Morris, M., Liang, P., and Bernstein, M.: *Generative Agents: Interactive Simulacra of Human Behavior*, <https://arxiv.org/pdf/2304.03442.pdf>. last accessed 2023/08/06.
15. Risko, E., Richardson, D. and Kingstone, A.: *Breaking the Fourth Wall of Cognitive Science: Real-World Social Attention and the Dual Function of Gaze*, *Current directions in psychological science: a journal of the American Psychological Society*, vol. 25 (1), pp. 70-74, (2016).
16. Adams, E., and J. Dormans, J.: *Game mechanics: advanced game design*, Berkeley, Calif.: New Riders, (2012).
17. Fullerton, T.: *Game design workshop: a playcentric approach to creating innovative games*, Boca Raton, FL: CRC Press, (2019).
18. Rogers, S.: *Level up! The guide to great video game design 2nd ed.*, Chichester: John Wiley and Sons Ltd, (2014).
19. Louchar, S. and Aylett, R.: *Solving the narrative paradox in VEs – lessons from*, *Lecture Notes in Computer Science*, vol. 2792, (2003).
20. Mallon, B., Lynch, R., Moseley A. and Whitton, N.: *Stimulating Psychological Attachments in Narrative Games: Engaging Players With Game Characters*, *Simulation & gaming*, Vols. 45 (4-5), pp. 508-527, (2014).

21. Hong, Y. and Riedl, M.: Personalized Interactive Narratives via Sequential Recommendation of Plot Points, *IEEE transactions on computational intelligence and AI in games*, vol. 6 (2), pp. 174-187, (2014).
22. Aylett, R.: Emergent narrative, social immersion and “storification”, *Proceedings of the 1st international workshop on narrative and interactive learning environments*, vol. 8, pp. 35-44, (2000).
23. Conroy, D., Wyeth, P. and Johnson, D., Modeling player-like behavior for game AI design, *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, vol. Article No.: 9, p. 1–8, (2011).
24. Magnus, J. Do non-player characters dream of electric sheep? A thesis about players, NPCs, immersion and believability (Doctoral dissertation, Department of Computer and Systems Sciences, Stockholm University) (2013).
25. Pearce, T., Rashid, T., Kanervisto, A., Bignell, D., Sun, M., Georgescu, R., Valcarcel Macua, S., Zheng Tan, S., Momennejad, I., Hofmann, K. and Devlin, S. : Imitating human behaviour with diffusion models, *Conference ICLR 2023 paper*, (2023).
26. E. A. Inc., Mass Effect, Electronic Arts Inc., <https://www.ea.com/games/mass-effect>. last accessed 2023/07/13 (2023).
27. Z. M. INC., Fallout, ZENIMAX MEDIA INC., <https://fallout.bethesda.net/en/>. last accessed 2023/07/22. (2023).
28. Morrison H. and Martens, C. A Generative Model of Group Conversation, (2017).
29. Kybartas, B. and Bidarra, R. A Survey on Story Generation Techniques for Authoring Computational Narratives, *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 9, no. 3, pp. 239 - 253, (2017).
30. Domsch, S. *Storyplaying : Agency and Narrative in Video Games*, Berlin/Boston: De Gruyter, (2013).
31. Craig, S., Sullins, J., Graesser A. and D'Mello, S.: Predicting Affective States expressed through an Emote-Aloud Procedure from AutoTutor's Mixed-Initiative Dialogue, *International Journal of Artificial Intelligence in Education*, vol. 16(1), pp. 3-28, (2006).
32. Dafoe, A., Hughes, E., Bachrach, Y., Collins, T. , McKee, K., Leibo, J. , Larson K. and Graepel, T.: *Open Problems in Cooperative AI*, (2020).
33. Schober, M. and Clark, H.: Understanding by addressees and overhearers, *Cognitive Psychology*, vol. 21, pp. 211-232, (1989).
34. Siegel, J. and Szafron, D.: Dialogue patterns - A visual language for dynamic dialogue, *Journal of visual languages and computing*, vol. 20 (3), pp. 196-220, (2009).
35. Adams, E.: *Fundamentals of Game Design*, ebook, (2010).
36. Dream, Q.: *Detroit become human*, Quantic Dream., <https://www.quanticroad.com/en/detroit-become-human>. last accessed 2023/07/10. (2023).
37. Games, R., *LA noire*, Rockstar Games, 2023. <https://www.rockstargames.com/lanoire/>. last accessed 2023/10/07. (2023).
38. S. E. Ltd., *Life is strange*, Square Enix Ltd., <https://lifeisstrange.square-enix-games.com/en-gb/>. last accessed 2023/07/10. (2022).
39. Dai, K.: *Multi-Context Dependent Natural Language Generation For More Robust*, B.A dissertation, C.S., Harvard Univ., Cambridge, Massachusetts, (2020).
40. Bosser, A., Levieux, G., Sehaba, K., Buendia, A., Corruble, V., De Fondaumi`ere, G., Gal, V., Natkin, S., and Sabouret, N.: *Dialogs Taking into Account Experience, Emotions and Personality*, *Conference Entertainment Computing - ICEC 2007 paper*, (2007).
41. Collins, J., Hisrt, W., Tang, W., Luu, C., Smith, P., Watson, A. and Sahandi, R.: *EDTree: Emotional Dialogue Trees for Game Based Training*, *Conference International Conference on Technologies for E-Learning and Digital Entertainment paper*, (2016).

42. Oliver, E., *Crosston Tavern: Modulating Autonomous Characters Behaviour through Player-NPC Conversation*, M.S. dissertation, C.M., Univ. of California, Santa Cruz, California, (2021).
43. Wai Man, S., Ammanabrolu P. and Riedl, M.: *Telling Stories through Multi-User Dialogue by Modeling Character Relations*, (2021).
44. Castricato, L., Frazier, S., Ballocc, J., Tarakad. N. and Riedl, M.: *Automated Story Generation as Question-Answering*, (2021).
45. Weir, N., Thomas, R. D'Amore, R., Hill, X. ., Van Durme, B. and Jhamtani, H.: *Ontologically Faithful Generation of Non-Player Character Dialogues*, *Free E-Journals*, (2022).
46. T. Klüwer, P. Adolphs, F. Xu, H. Uszk and X. Cheng, "Talking NPCs in a virtual game world," *Conference ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* article, 2010.
47. M. Consalvo, *Cheating: Gaining Advantage in Videogames*, Cambridge: MIT Press, 2007.
48. OpenAI, *What is ChatGPT?*, OpenAI, <https://help.openai.com/en/articles/6783457-what-is-chatgpt>. last accessed 2023/06/01 (2023).
49. Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I: *Improving Language Understanding by Generative Pre-Training*, 2018. <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>. last accessed 2023/06/12.
50. Turing, A. *Computing Machinery and Intelligence*, *Mind*, vol. 59 (236), pp. 433-460, (1950).
51. OpenAI, *Examples*, OpenAI, <https://platform.openai.com/examples>. last access 2023/06/07, (2023).
52. OpenAI, *Custom instructions for ChatGPT*, OpenAI, <https://openai.com/blog/custom-instructions-for-chatgpt>. last access 2023/07/29. (2023).
53. Frantar, E. and Alistarh, D. *SparseGPT: Massive Language Models Can Be Accurately Pruned in One-Shot*, *arXiv.org*, <https://arxiv.org/pdf/2301.00774.pdf>. last access 2023/06/01. (2023).
54. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, C., Radford, A., Sutskever, I., and Amodei, D.: *Language Models are Few-Shot Learners*, last access 2023/08/05. (2022).
55. OpenAI, *How ChatGPT and Our Language Models Are Developed*, OpenAI, <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed>. last access 2023/06/01. (2023)
56. Wei, J. , Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D., *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*, *arXiv.org*, 2023. <https://arxiv.org/pdf/2201.11903.pdf>. Last accessed 2023/06/02. (2023).
57. Wach, K., Duong, C., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkievicz, J. and Ziemba, E, *The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT*, *Entrepreneurial Business and Economics Review*, vol. 11 (2), pp. 7-30, (2023).
58. E. Sörensen, *Cultures of Computer Game Concerns The Child Across Families*, Law, Science and Industry, Bielefeld transcript Verlag, 2018.
59. I. Blizzard Entertainment, "Overwatch 2," Blizzard Entertainment, Inc., <https://overwatch.blizzard.com/en-gb/>. Last accessed 2023/08/01. (2023)
60. Dream, Q. *Heavy Rain*, Quantic Dream, 2023. <https://www.quanticroad.com/en/heavy-rain>. Last accessed 2023/07/10. (2023)

61. Sicart, M.: *Beyond choices: the design of ethical gameplay*, Cambridge, Massachusetts: The MIT Press, (2013).
62. Eriksson, T.: Design fiction exploration of romantic interaction with virtual humans in virtual reality, *Journal of Future Robot Life*, vol. 3 (1), pp. 63-75, (2022).
63. Epic Games, I. Unreal Engine 5, Epic Games, Inc, 2023. <https://www.unrealengine.com/en-US/unreal-engine-5>. Last accessed 2023/03/01.
64. Microsoft, Visual Studio 2022, Microsoft, 2023. <https://visualstudio.microsoft.com/>. [Accessed 01 07 2023].
65. Lehman, D.: *The Mysterious Romance of Murder: Crime, Detection, and the Spirit of Noir*, Ithaca: Cornell University Press, (2022).
66. IMDb.com, I.: *The Big Sleep*, IMDb.com, Inc., <https://www.imdb.com/title/tt0038355/>. last access 2023/06/01. (2023)
67. OpenAI, Marv the sarcastic chat bot, OpenAI, 2023. <https://platform.openai.com/examples/default-marv-sarcastic-chat>. last access 2023/08/05.
68. Creswell, J. and Creswell, J.: *Research design: qualitative, quantitative & mixed methods approaches (sixth edition)*, Los Angeles: SAGE, (2023).
69. A. Publishing, *Statistics for Beginners in Data Science: Theory and Applications of Essential Statistics Concepts Using Python, Ai Publishing, 2020.*
70. Creswell, J and Plano Clark V.: *Designing and conducting mixed methods research*, Los Angeles: SAGE, (2017).
71. Cooksey, R.: *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*, Singapore: Springer, (2020).
72. Creswell, J and Plano Clark V.: *Designing and conducting mixed methods research*, Los Angeles: SAGE, (2017).
73. T. R. Foundation, "The R Project for Statistical Computing," The R Foundation, 2023. [Online]. Available: <https://www.r-project.org/>. [Accessed 22 04 2023].
74. J. Silge and D. Robinson , *Text Mining with R: A Tidy Approach*, O'Reilly Media, 2017.
75. P. Sedgwick, "Pearson's correlation coefficient," *BMJ (Online)*, vol. 345 (jul04 1), pp. e4483-e4483, Article 4483, 2012.
76. G. Dr. Chakraborty, M. Pagolu and S. Garl, *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS*, SAS Institute, 2014.
77. G. Brown, K. Currie and J. Kenworthy , *Questions of Intonation*, Taylor & Francis, 2015.