Check for updates

Interviewing a Virtual Suspect: Conversational Game Characters Using Alexa

Gonçalo Baptista(✉), Diogo Rato, and Rui Prada

INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
{goncalo.baptista,diogo.rato,rui.prada}@tecnico.ulisboa.pt

Abstract. The video game industry is constantly innovating, with new mediums and ways for players to interact with the game environment. Voice interaction in games is an ever evolving field, especially with advances in Natural Language Processing. In that vein, there has been a increasing number of conversational agents with natural language interaction capabilities deployed into video games. In this paper, we improve the Virtual Suspect game with a natural language interaction using the tools provided by Amazon Alexa. We followed an iterative, user-centered approach when designing the new interaction, collecting feedback and data from three User Studies in order to improve the interaction with the Virtual Suspect. Our findings suggest that the usage of natural language to support the interaction with game characters can improve the player experience.

Keywords: Conversational agents · Voice games · Interactive narrative

1 Introduction

Videogames offer the player an opportunity to be immersed in virtual worlds enriched with fantastic elements and interesting characters. Interacting with these characters allows players to uncover narrative beats while exploring and progressing through the game's story. These interactions are usually supported by dialogue systems that provide the player with a familiar mechanism to engage characters and extract information: Natural Language. However, such systems often display restrictions that conflict with players' expectations and limit their ability to express themselves. Researchers and developers have applied different paradigms to endow Non-Player Characters (NPCs) with conversational capabilities: from dialogue supported by social exchanges [8,9], to parser-based approaches [7,12] and vast authored branching dialogues [11], all are designed on top of a limited set of utterances. Although these interactions might resemble

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UIDB/50021/2020.

I. Barbedo et al. (Eds.): VJ 2020, CCIS 1531, pp. 98–112, 2022.
https://doi.org/10.1007/978-3-030-95305-8_8

conversations between social beings, they are restrictive and do not allow players to engage NPCs in a natural and free fashion.

Video games are constantly evolving and, as an interactive medium, they welcome the application of new technologies to support more believable and interesting player experiences. As such, there has been a significant amount of video games relying on Natural Language Understanding to support the interaction between the players and NPCs [6]. The hardware to support such interaction is becoming more accessible and the increasing number of developer-friendly APIs has boosted the number of applications, in particular games, that have verbal behaviour as the core of the interaction.

Among other solutions commercially available, Alexa is a virtual assistant developed by Amazon and released with the Echo smart speaker that allows developers to add new third-party functionalities, called Skills, which include voice-enabled games. Since this technology allows for the easy creation of natural language applications, and has already been proven to work as a platform for voice interactive games, we decided to use it to create a new natural language interaction with the lying agent created by Rato et al. [14].

Rato et al. [14,15] designed a model for a Virtual Suspect - an autonomous agent that was capable of lying, within an interaction in the context of a police interrogation. This model has the potential to be integrated into a video game as an NPC that players have to interrogate. The focus of that work was designing the architecture of the agent and devising the lying algorithm. When it came to testing the agent, the interaction between users and agent was very limited, only allowing for a small pre-defined number of questions, and not fully showcasing the capabilities of the lying model. In their conclusion, the authors posited that a natural language interaction could improve the quality of the interaction with the agent.

The limited interaction with the original Virtual Suspect game is the problem that we are trying to solve. We believe the core of the original work, the Virtual Suspect framework and lying capabilities, has potential to make for a compelling character in a video game environment, but the original work did not go far enough to showcase that potential, and the interaction they created in order to test their work was insufficient.

We will test the hypothesis that a natural language interaction will improve the interaction with the agent, and create an interaction that showcases the capabilities of the original model and shows its potential as a lying NPC in a video game. From investigative games centered around interrogating a suspect to just a side character in a larger narrative, a good lying NPC with a good interaction could be an interesting component to add to a video game, providing players with a rich and unique experience.

This work will focus on the creation of a new natural language interaction with the Virtual Suspect, using Alexa, as a means of showcasing the lying agent's capabilities and potential. We looked at other works with certain similarities to our own, analysed how the Virtual Suspect [14] was designed and developed, and how we can create our new interaction with the Alexa Skills Kit (ASK). We then combined that knowledge to create a Virtual Suspect Skill and followed a user-centered approach to improve the quality of the interaction. Our goal is a natural language interaction that is fluid and provides good User Experience (UX), and showcases the potential of the lying agent model as a video game component.

2 Related Work

We analysed two examples of Virtual Suspects, one developed by Bitan et al. [1] and another by Bruijnes et al. [2]. While the Virtual Suspect developed by Rato et al. [14] was mostly focused on modelling the lying behaviour of the agent, these two works were more focused on modelling the agent's mental and emotional state [1], and different personality characteristics of the agent [2]. Their work was centered around how the agent's answers would differ according to those parameters, with all the different possibilities being pre-programmed, while the Virtual Suspect that we will base ourselves off of allows the agent to come up with its own answers dynamically. Given that their answers were already pre-defined, the Virtual Suspects' behaviour could eventually become predictable with enough interaction, if incorporated into a video game, while our agent could introduce more variance to the experience.

We studied three examples of Conversational Agents in video game-like environments, to see how other authors have tackled these issues. The work by Falk et al. [3] centers around guiding players through an interactive narrative, how to model knowledge representation in an agent's memory, and how to model a player's perception of that knowledge. They developed an agent that players can talk to during an interactive narrative, they can ask questions about the story, and the agent can guide users to important parts of the narrative they might have missed. Kenny and Huyck [5] developed an agent that can talk with the player and understand what the player is talking about through context, using Referring Expressions. Last but not least, we looked at the work of Morris [10] who proposed a model for an agent with shallow models of emotion and personality that would be capable of engaging players in conversation while playing a Cluedo-style game. These three works highlight the potential of conversational agents, such as the Virtual Suspect developed by Rato et al. [14], to be added to video game environments, and to use Natural Language as the interactive medium. Things like modelling context and player's perception of knowledge will be extremely useful to us when creating the Natural Language interaction with the Virtual Suspect.

We also examined some examples in the industry, with companies like Doppio Games developing and releasing voice-first games for platforms like Amazon Alexa and Google Assistant. These games focus their interaction and gameplay loops on voice interaction and Natural Language Processing, to create an interesting and innovative experience. Examples include The Vortex by Doppio Games, Starfinder by Paizo - Alexa Games, and Escape the Room by Stoked Skills LLC. Not only are companies and developers taking advantage of this new paradigm, but even Amazon itself is fostering the development of new applications using the Alexa Prize [13], a competition for creating open-domain social bots. These games and works serve as an example that this sort of interaction in games is viable.

While these works have aspects in common with our own, none is too similar to what we are trying to achieve, so we could not base ourselves too much on any of them and had to come up with a model for our new Virtual Suspect interaction on our own.

3 Virtual Suspect

In order for us to create a new interaction with the Virtual Suspect, we have to obtain a deep understanding of how it was designed and how it works. In "Virtual Suspect - A Lying Virtual Agent" Rato et al. [14] laid out the architecture and functionality of their lying Virtual Suspect.

First, the agent has a memory, its **Knowledge Base**, which contains its story. The agent's story is composed of *events* and *entities*. Entities are the most basic memory fragment and can represent people, locations, objects, time spans. Events represent distinct episodes in the agent's story and they are composed of an Action and several entities, in different roles. For example, *"John stole a chocolate from the store on September 5th at 4:30 pm"* can be an event where *"Steal"* is the Action, and *"John"*, *"chocolate"*, *"store"* and *"September 5th at 4:30 pm"* are all represented by entities. Entities can have different roles in events. In the previous example, those roles were Agent, Theme, Location, and Time, respectively, but you can also have Manner and Reason. These roles indicate the relation between those entities and the Action in that event, *who* was involved, *what* was the target of the action, *where* it happened, *when* it took place, *how* it happened, and *why* it happened. Events and entities exist separately in the **Knowledge Base**, and events reference the entities that were involved in them, this way the same entity can be referenced by (and thus have participated in) several events. Events can also be true or false, where true events are what really happened in the agent's story and false events are the events the agent uses to lie. These can have an incriminatory value from 0 to 100, depending on how incriminatory each event is, in relation to the crime our Virtual Suspect committed.

The interaction between the agent and the user is done through questions and answers. The user asks the agent a question about its story and the agent responds, with either the truth or a lie (as we will see later). These questions are internally represented as *queries* in the agent's system, and they can either be Validation questions (yes or no), or Information Gathering questions (who, where, when, why, etc.). Each query contains a series of conditions which it seeks to validate in order to find an answer. For example, the question *"When did John steal the chocolate?"* is an Information Gathering question that seeks to retrieve the Time entity from an event that matches the conditions *"Agent Equals John"*, *"Action Equals Steal"* and *"Theme Equals Chocolate"*. The **Query Engine** receives this query, tries to find all the events that match those conditions, and returns a *query response*, which in this case contains the value

"September 5th at 4:30 pm". After this step, the Virtual Suspect also contains a **Natural Language Generator** that transforms the query result into a proper English sentence to be returned to the user as its answer.

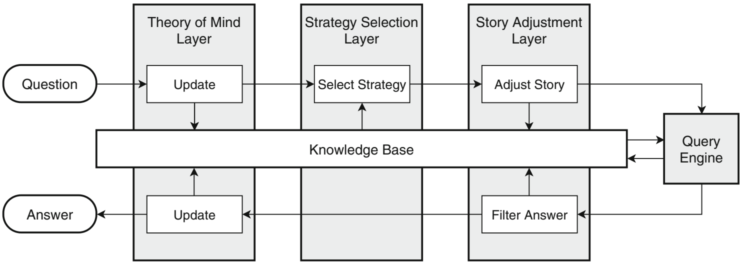

Fig. 1. Virtual Suspect architecture

What enables the agent to lie is a three layered two-pass control system, as illustrated in Fig. 1. When the agent receives a question, it passes through each one of the three layers before being processed by the Query Engine, and the answer passes through the layers again before being returned to the user. The **Theory of Mind Layer** keeps track of what the user already knows about the story, by analysing the information contained inside the query. If the user already knew about John stealing the chocolate, for example, it would not be productive to try to lie about that. The **Strategy Selection Layer** selects an appropriate lying strategy based on the current context, and the **Story Adjustment Layer** creates the fake events that the agent uses in its lies. When the agent encounters a question that would lead it to reveal incriminating information, it instead creates a new fake event with less incriminating information to take the place of the incriminatory event in the version of the story the agent is presenting the user. The agent always keeps track of the true version of events, but is capable of having alternate versions of those events in its memory in order to hide information from the user. After the question has passed through all the layers, it is processed by the Query Engine, and thus the information about the fake events is retrieved instead of the real information, and the result then passes back through the layers again, before being returned to the user.

Figure 2 shows how this was implemented in the original work [14], with the **Response Model** representing the conjunction of the Query Engine and the three layers. The prototype that was originally used to test the Virtual Suspect was a visual interface that contained information about the suspect and the case, and a set number of pre-defined questions that users could select from. All this information, along with the events and entities of the agent's story were all defined in a separate Story file. When the user selected one of the pre-defined questions from the visual interface, it automatically sent the corresponding query to the Response Model, which was then processed as previously described and the answer was displayed back to the user in the interface.

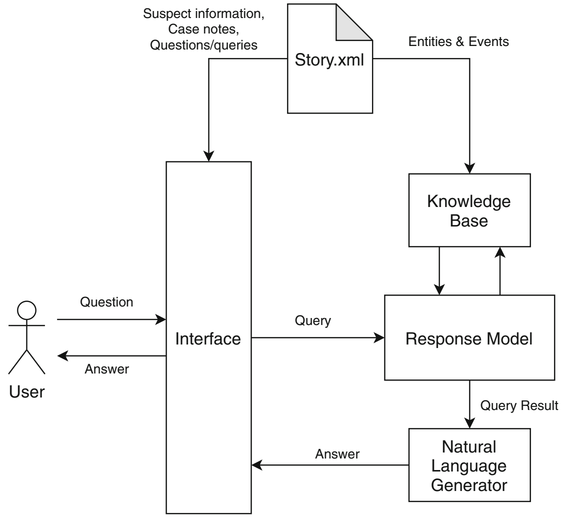

Fig. 2. Virtual suspect prototype implementation

4 Alexa

The Alexa is a virtual assistant developed by Amazon and released with the Echo smart speaker, that is capable of a wide range of features, but the one that is of interest to us is the ability to create third-party applications using the Alexa technology, called Skills. These Skills are made through the Alexa Skills Kit (ASK) and they have two components: the **Interaction Model**, and the **Skill Service**. Figure 3 shows the typical workflow of an Alexa Skill. The user asks a question or gives a command to Alexa, which sends it to the Skill Interaction Model. The Interaction Model disambiguates the meaning of the user's message and sends that information to the Skill Service, which computes the appropriate response and sends it back through the system until it reaches the user.

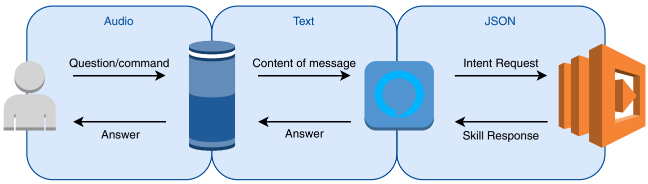

Fig. 3. Workflow of an Alexa Skill

The **Interaction Model** is the front-end of the Skill, and it is composed of *intents*. An intent contains a selection of sample phrases that could be uttered to invoke that intent. For example, a *HelloWorldIntent* could contain the utterances

"Hello", *"Hi World"*, and *"Hey"*, so that when the user says one of these phrases or something similar, Alexa can correctly identify the *HelloWorldIntent* and provide the proper response. The more sample phrases an intent has, the more accurate Alexa can be when detecting it, as Alexa trains a model with the information in our Interaction Model to be able to detect meaning from a wide variety of phrases, although the exact method is not publicly known.

Besides having a set of sample utterances, an intent can also have *slots*, which are essentially variables that can be fulfilled by certain values. For example, we can have the sample phrase *"My name is*{name}*"*, where {name} represents a slot that accepts English first names as values. This way, both the sentences *"My name is John"* and *"My name is Mary"*, would equally fulfill that intent. A slot type can be one of many provided by Amazon (like the First Name slot type), or can be a custom list of possible slot values according to the skill's domain. These slot values can also contain synonyms. Slots cannot be iteratively defined, so a slot cannot contain another slot.

This information (intent and slot values), once processed by the Interaction Model, is sent to the **Skill Service**, which is the back-end of the Skill, through a JSON file. The Skill Service takes the information sent by the Interaction Model and computes the appropriate response (for example, *"Hello John"*), and sends it back to the Interaction Model through another JSON file.

5 Solution

In order to create the new Natural Language interaction with the Virtual Suspect, we combined what we studied in the previous sections to create a Virtual Suspect Skill. Our Interaction Model has different intents for the different question types, and we use slots to create the query conditions. Each of our intents needs lots of sample utterances so our model can cover a wide range of questions, and our slot values include the possible entity values for each type. Our Skill Service was created in the same environment as the original Virtual Suspect was developed, so we can use the original Virtual Suspect Response Model as a sort of code black box. The Skill Service takes the intent and slot information from the Interaction Model and uses it to create a query object that can then be sent to the Virtual Suspect Response Model. We also use the Virtual Suspect Natural Language Generator to transform the query result returned by the Virtual Suspect Response Model into a proper answer, before returning it to the Interaction Model.

Figure 4 shows a representation of our Virtual Suspect Skill, showing the Skill Service interacting with the Virtual Suspect modules, and combining what we had already seen in Figs. 2 and 3. The connection between the Story file and Interaction Model is merely symbolic, as we cannot directly connect those two entities, but it represents the entity values that populate the slot values.

We used an iterative, user-centered approach when designing the Virtual Suspect Skill. We started by recreating the functionality of the original prototype, ensuring basic coverage for all the different types of questions, and then we did

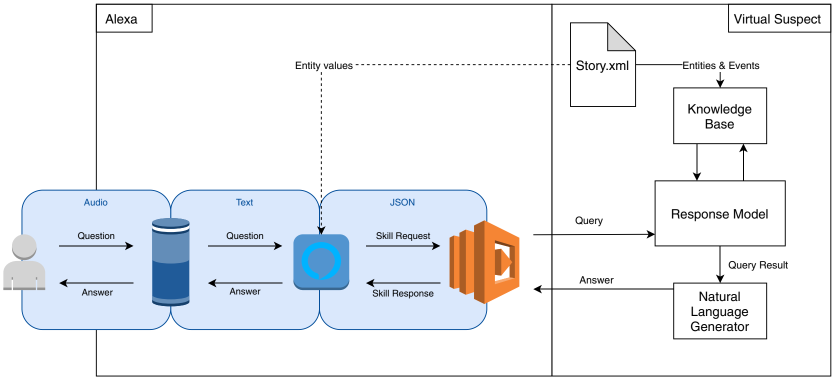

Fig. 4. Solution architecture

a User Study to collect data on how users interacted with the Virtual Suspect, what kinds of questions they wanted to ask and how they asked them. We also measured the performance of the agent, collecting data on the problems of the interaction, so we could have a baseline performance to compare to later.

After we collected the data from the First User Study, we used it to improve upon the interaction, making changes and improvements to fix those issues. We then conducted a Second User Study after those changes, to validate if those changes had improved the quality of the interaction, and to measure the User Experience (UX).

During the development and first two User Studies, we kept the lying component of the agent deactivated, so we could better measure its responses to the various questions without the lies obscuring that information. After we did the Second User Study, we turned the lying component of the agent back on and did a final User Study to measure the UX of that interaction, to see how well the original lying algorithm fit into the new interaction.

The next sections describe the development process of the Virtual Suspect, and the three User Studies, respectively.

6 Development

6.1 First Steps

In order to be able to do our fist User Test and collect data on how people interact with the Virtual Suspect, we needed a functional Virtual Suspect Skill prototype. We started off by replicating the functionality of the original Virtual Suspect Prototype [14] whose visual interface only included 13 pre-defined questions. Since those questions were already predefined and the corresponding queries would directly be sent to the Virtual Suspect Response Model (as seen in Fig. 2), there was no concern about being able to interpret those questions using Natural Language Processing, and as such they did not conform to a consistent style, often having sentences before the question and information that was not relevant for the query. In order to recreate the functionality of being able to ask those original questions (or their corresponding queries at least) and obtaining the same answers, we had to restructure the questions into a more consistent style that we could then expand to the rest of the question types in our Interaction Model.

We ended up with a style where a question like *"Where did you meet John Frey?"* was modeled as *"Where* {question_verb} {subject} {filler_verb} {agent}*"*. In this example, *"Where"* indicates the type of question, {subject} and {agent} are slots that contain information relevant for the query conditions, while {question_verb} and {filler_verb} are slots that allow for a wider range of questions with the same meaning to be identified. This way, questions like *"Did you meet John Frey?"* and *"Have you met John Frey?"* can both be represented by the same utterance, as they both have the same meaning.

After we established a consistent style of question, and managed to recreate the functionality of the original 13 questions, we expanded our Interaction Model to include more questions of each type, by looking at the events of our story and figuring what types of questions could be asked, with which conditions. As we mentioned before, this was done with the lying component turned off, so we could better understand how the agent was processing the information. With a lying agent, it would be more difficult to tell if the agent answered *"No"* because he understood the question and decided to lie, or if he did not understand the question at all.

With this functioning prototype, we realized our First User Study to collect data on how people interacted with the Virtual Suspect, so we could expand our Interaction Model with more possible questions, and to measure the performance of the agent, so we could note the problems with the interaction and work to improve it.

6.2 Improving the Interaction

With the data collected from the First User Study, we were able to make a lot of changes and improvements to the Virtual Suspect Skill, to address problems such as:

- **Missing intents:** questions that the users wanted to ask but the agent was not capable of answering.
- **Pronouns:** both direct pronouns (it, him, there) and indirect pronouns (something, anyone).
- **Context:** a knowledge of what was previously asked.
- **Synonyms:** adding more synonyms to the Interaction Model.
- **Missing information:** information about the story that users wanted to know about but it was not represented in the story.
- **Answer generation:** improving the Virtual Suspect Natural Language Generator.
- **Time conditions:** add more cases for different possible time conditions in questions.

- **More utterances:** add more variety of questions to the Interaction Model.
- **Filters in the Skill Service:** to make sure that things are being processed correctly.
- **Feedback:** providing better feedback to the user when the agent cannot answer a question for some reason.

By addressing these and other problems and making all the necessary changes to the Virtual Suspect Skill, we were able to improve the interaction with the Virtual Suspect. We realized the Second User Study in order to verify that improvement and measure the quality of the interaction.

6.3 Last Adjustments

After we validated the improvements we made with the Second User Study, we made a few final minor adjustments before turning on the lying component again and making sure it was still working as intended with all of our changes. After that, we moved on the Third and final User Study, to test how the lying algorithm was working with the new interaction.

7 User Studies

We conducted three User Studies during the development of our work. In all three studies, users interacted with the Virtual Suspect via a text messaging service, where an account in the name of the Suspect was created to add to user immersion. For all three studies, the conversations between the users and the agent were logged and annotated, in order to identify the problems with the interaction.

Considering that this work is centered around the Alexa, the decision to use text messaging as the means of interacting with the agent might be a strange one, but it was not without reason. Firstly, the global COVID-19 pandemic made it more difficult to organize and conduct user studies, having to do so remotely, which limited our options on how to have users interact with the game. Second, even though Alexa Skills function exactly the same way when interacting via voice or via text, the ability to log the conversations between players and the agent does not, as the Amazon logging services do not save information on the full sentence spoken by the user, only its intent and slots, and as we will see, the ability to thoroughly log the conversations was crucial to our work. Thirdly, and this is not as significant as the other two constraints, our Studies were conducted with non-native English speakers, which could negatively impact the Automatic Speech Recognition capabilities of the Alexa.

For the Second and Third User Studies, a questionnaire was presented to the users after the interaction to measure the User Experience (UX), which used the User Experience Questionnaire (UEQ) developed by Schrepp et al. [16,17].

7.1 First User Study

For the First User Study, we wanted to collect data on how users interacted with the Virtual Suspect (what type of questions they asked, and how they asked those questions) and do an analysis of the problems with the interaction.

Since the interaction was still in an early state and the range of the agent's understanding capabilities was limited, we decided to do two different types of interactions. One where users would interact directly with the Virtual Suspect Skill, and another where users would instead interact with a human pretending to be the Virtual Suspect, answering questions as the Virtual Suspect ideally would without the Skill limitations (following a Wizard-of-Oz technique [4]). This way we could collect data on the problems of the current interaction with the Skill, but also analyse how people would interact without those limitations. When logging the conversations, both types of responses were recorded for either type of interaction, allowing us to do a comparison between what the user actually interacted with and what the other interface would have said instead.

Twelve people participated in this study, none of which had interacted with the Virtual Suspect in any way before. The average number of exchanges per conversation with the Virtual Suspect, across all the participants, was 23.67. The average conversation success rate (which is the percentage of exchanges that were correctly identified by the agent) was 37.29%. We classified problems in two categories: Question Problems (which was anything that caused the question not to be recognized by the agent), and Answer Problems (which was anything that caused the agent to give a bad answer). Only 35.92% of exchanges did not have any Question Problems, while only 22.54% did not have any Answer Problems.

Overall, the results indicate that a more robust Interaction Model was necessary to support a better gameplay experience, as could be expected. Apart from that, we were able to collect valuable insight about how players engage the Virtual Suspect and its capability to reply appropriately, through the use of the aforementioned quantitative annotation of the interactions. This also provided us with a baseline performance with which we could compare future iterations.

7.2 Second User Study

The Second User Study was realized after the changes and improvements made as a result of the data gathered in the First User Study, with the objective of validating those changes, verifying the improvement with the interaction, and measuring the User Experience (UX).

Fourteen people participated in this study, six of which had participated in the First Study and the remaining eight being new participants. This study had an average of 54.07 exchanges per conversation, and an average success rate of 63.39%. This time, 65.13% of exchanges did not have any Question Problems, and 89.83% of exchanges not having any Answer Problems.

These results were a marked improvement over the First Study, validating the changes we made to improve the interaction. Not only that, but the UX results were also very good. Figure 5 shows our results compared to the benchmarks set by the authors of the UEQ [16], and we can see that all of them fall either into the Good or Excellent category.

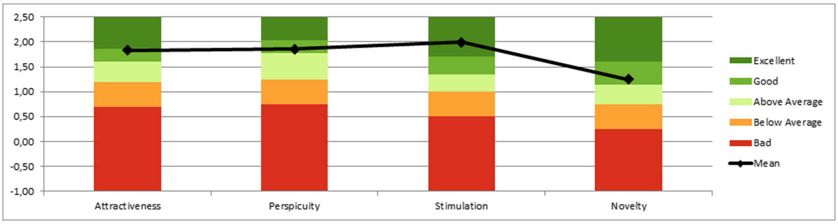

Fig. 5. UEQ benchmarks

This Study was a success, as we were able to definitively show the improvements we made to our Virtual Suspect Skill, validating our previous choices, and we were able to show that our Skill provides a good User Experience.

7.3 Third User Study

The Third User Study was conducted after we turned on the lying component of the Virtual Suspect and its objective was to measure the effect it had on the interaction, to see how well it was working.

This study had sixteen participants, six who had participated in both previous studies, and six who had participated only in the Second, with four new participants. The average number of exchanges per conversation was 46.13, and the success rate was 65.01%. 66.80% of exchanges did not have any Question Problems, and 88.48% did not have any Answer Problems.

The agent performance was largely the same as the Second Study, with the difference in the average number of exchanges being explained by the fact that the interactions of the Second Study were more free and exploratory, while in the Third Study they were more focused on the crimes of the Suspect.

The UEQ results were more telling, with Fig. 6 showing that the results in the Third Study were noticeably worse than the Second Study. Given that the agent performance remained at about the same level, and given the feedback we received about the agent's lies not being very believable or realistic, it is safe to conclude that it was the introduction of the lying component that caused this drop in UX. Therefore, as it stands, the current lying algorithm is not very suited for this new conversational interaction.

8 Discussion

The ongoing development of the Virtual Suspect aligned with the user-centered studies suggests that the employment of a conversational model based on Natural Language can be beneficial to improve the player experience. The user

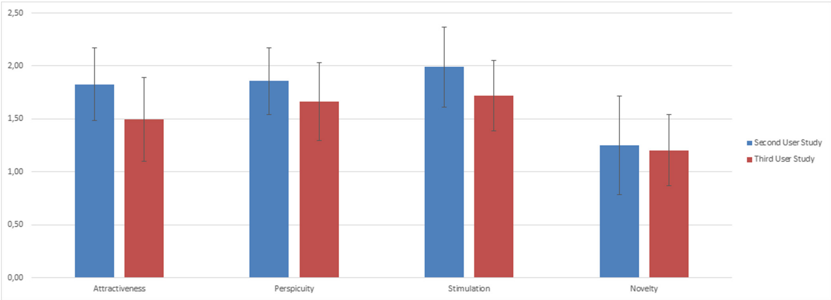

Fig. 6. Comparison of UEQ results

questionnaires further support this claim, since all the evaluated dimensions of the UEQ had positive values. Furthermore, the duration and variety of questions asked during the interaction lead us to believe that introducing such conversational skills affords different interviewing strategies not supported by traditional dialogue systems found in games.

Additionally, regarding the methodology adopted, using a user-centered approach allowed us to better understand the interactive space the players want to and, to an extent, can explore inside the narrative. With each iteration of the Interaction Model, players were able to choose different lines of questioning as well as collect information not accessible before.

However, accommodating conversational intents not widely adopted by the entire sample might introduce a heavy developer and authoring effort, which might not be worthy of consideration. Based on the development of the Virtual Suspect, we realized that the authoring tools provided by Alexa have a restricted set of functionalities that demand a wide variety of sample utterances to generate an adequate interaction model. Constraints like the inability to have slots inside of slots led to an inflation of utterances with the same meaning in the Interaction Model, and kept us from achieving a degree of nuance that would have allowed even more questions to be recognized.

And finally, as we saw in the results of our last User Study, the introduction of the lying mechanism negatively affected the user experience. Although not being the focus of this work, the introduction of new conversational capabilities required some modification to the story representation. The way the lying algorithm was defined in the original work did not adequately accommodate these new story elements and structures. A more nuanced and thorough lying algorithm is needed to better take advantage of and demonstrate the more advanced capabilities and natural interaction of the Virtual Suspect game. Furthermore, during the course of this research, we realized that the development of the Interaction Model should be deeply linked to the querying mechanism that retrieves information. Pursuing a disconnected development of the interaction and the knowledge representation does not benefit the capacity of the agent as a whole and, ultimately, compromising the player experience.

9 Conclusion

The deployment of conversational agents in games should not aim to replace classical dialogue systems but rather explore new gameplay opportunities that are driven by voice interaction as its core mechanic. In this work, we followed a user-centered approach in pursuit of our objective of improving the interaction with the Virtual Suspect, and the deployment of conversational agents in games. We were able to overcome the limitations of the original Virtual Suspect interaction [14], create a Natural Language interaction that showcased the capabilities of the Virtual Suspect, and, most importantly, support conversation with a game character with good UX. On top of achieving our goal of improving the interaction with the Virtual Suspect, we were able to test whether the original lying algorithm [14] was suited to this type of interaction, and concluded that it needs further improvement.

Regarding the Interaction Model structure, the authoring approach can be improved. A different natural language model, something more non-deterministic and grammar-like, could be beneficial in achieving an even better interaction with a reduced authoring effort.

There were also constraints with the definitions of the Virtual Suspect Architecture, like the way that Time and Reason entities were defined within events, that kept us from being able to achieve a more realistic story that would have matched the more natural interaction we created. A restructuring of the agent's memory, keeping the same basic concepts but improving upon them, could lay the foundation for a richer and more realistic story.

The other objective of this work was to showcase the Virtual Suspect capabilities and demonstrate its potential as an addition to a video game environment as a lying NPC. The fact that we were able to create a Natural Language interaction with good UX does seem to support that hypothesis, and the work that we have done so far can easily be expanded upon in the future. However, further research must be conducted.

The future of this work should be centered on either creating an interactive video game centered around interrogating the Virtual Suspect, or incorporating the Virtual Suspect lying model into an already established video game environment. The improvements to the interaction with the agent itself, like the new Natural Language Model, a restructuring of the agent's memory, or a new lying algorithm, can still be considered depending on the needs and context of the eventual work, as well as writing new stories and characters.

References

1. Bitan, M., Nahari, G., Nisin, Z., Roth, A., Kraus, S.: Psychologically based virtual-suspect for interrogative interview training. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
2. Bruijnes, M., Wapperom, S., op den Akker, R., Heylen, D.: A virtual suspect agent's response model. In: Affective Agents, p. 17 (2014)

3. Falk, J., Poulakos, S., Kapadia, M., Sumner, R.W.: PICA: proactive intelligent conversational agent for interactive narratives. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents, pp. 141–146. ACM (2018)
4. Hajdinjak, M., Mihelic, F.: Conducting the wizard-of-Oz experiment. Informatica (Slovenia) **28**(4), 425–429 (2004)
5. Kenny, I., Huyck, C.: An embodied conversational agent for interactive videogame environments. In: Proceedings of the AISB 2005 Symposium on Conversational Informatics for Supporting Social Intelligence and Interaction, pp. 58–63 (2005)
6. Kiiski, T.: Voice games: the history of voice interaction in digital games (2020)
7. Mateas, M., Stern, A.: Façade: an experiment in building a fully-realized interactive drama. In: Game Developers Conference, vol. 2, pp. 4–8 (2003)
8. McCoy, J., Treanor, M., Samuel, B., Reed, A.A., Wardrip-Fruin, N., Mateas, M.: Prom week. In: Proceedings of the International Conference on the Foundations of Digital Games, pp. 235–237 (2012)
9. McCoy, J., Treanor, M., Samuel, B., Wardrip-Fruin, N., Mateas, M.: Comme il Faut: a system for authoring playable social models. In: Seventh Artificial Intelligence and Interactive Digital Entertainment Conference. Citeseer (2011)
10. Morris, T.W.: Conversational agents for game-like virtual environments. In: AAAI 2002 Spring Symposium on Artificial Intelligence and Interactive Entertainment, pp. 82–86 (2002)
11. Obsidian Entertainment: The outer worlds. [Windows, Playstation 4, Xbox One] (2019)
12. Ocelot Society: Event[0]. [PC Digital] (2016)
13. Ram, A., et al.: Conversational AI: the science behind the Alexa prize. arXiv preprint arXiv:1801.03604 (2018)
14. Rato, D., Prada, R., Paiva, A.: Virtual Suspect. Master's thesis, Instituto Superior Técnico, October 2016
15. Rato, D., Ravenet, B., Prada, R., Paiva, A.: Strategically misleading the user: building a deceptive virtual suspect. In: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, pp. 1711–1713. International Foundation for Autonomous Agents and Multiagent Systems (2017)
16. Schrepp, M.: User Experience Questionnaire Handbook. All You Need to Know to Apply the UEQ Successfully in Your Project (2015)
17. Schrepp, M., Hinderks, A., Thomaschewski, J.: User experience questionnaire. Mensch und Computer 2017-Tagungsband: Spielend einfach interagieren **17**, 355 (2018)