Euphonia: Reflecting on the Design of an AI-Powered Voice-Controlled Narrative Game

Lynda Clark† & Jung in Jung
InGAME
University of Dundee
Dundee, Scotland lcclark001@dundee.ac.uk; j.z.jung@dundee.ac.uk

Stuart Anderson, Paul Blackham & Divij Sood
InGAME
Abertay University,
Dundee, Scotland s.anderson@abertay.ac.uk; p.blackham@abertay.ac.uk

ABSTRACT
This paper reflects on the design process for a work-in-progress AI-powered voice-controlled narrative game created by Innovation for Games and Media Enterprise (InGAME). This paper describes the steps which led to the final design decisions, and how the background research, research questions and initial prototyping may be traced through to the work-in-progress game. The design process is then reviewed for its suitability as a practice-based research and development workflow, before finally suggesting next steps the project will take.

CCS CONCEPTS
• Applied Computing • Arts and Humanities • Human-centered computing • Interactive systems and tools
• Interaction Paradigms • Natural Language Interfaces

KEYWORDS
Interactive Narrative; AI-Powered Storytelling; Videogames; Authoring Tools

ACM Reference format:

1 Introduction
There seem to be some common assumptions within the field of conversational design for AI in general and AI-powered storytelling in particular. Namely, that when designing AI-powered conversational toys, stories and games, two qualities are universally desirable: general intelligence (the ability to respond to a wide variety of inputs) [11, 21] and natural language (the ability to return human-like responses to input) [12]. However, many ‘natural’ human-like conversational mechanics such as interruption, digression and hesitation are considered aberrant when found in AI conversational systems or other conventional narratives [8, 21]. Therefore, much of the current literature around conversational UX design is geared towards completing commercial transactions with minimal friction [12] rather than crafting an interesting narrative experience.

†Corresponding author
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI PLAY ’20 EA, November 2–4, 2020, Virtual Event, Canada
© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISRN 978-1-4503-7587-0/20/11...$15.00
https://doi.org/10.1145/3383668.3419913

We sought to explore alternative paradigms outside of ‘natural’ conversation through the creation of our work-in-progress AI-powered voice-controlled narrative game, Euphonia. This paper is intended as a reflective exercise considering the design choices made during the early stages of Euphonia’s development and situates these choices within a context of ‘unnatural’ conversation. It is not intended to be a full evaluation of the research and its goals.

2 Conceptual Framework
If we take the position that ‘natural’ conversations with AI systems are not the desirable outcome, what are the alternatives? One possibility is to move away from anthropocentric design. This could mean allowing AIs to interact with one another, to have their own goals, interests and behaviors that do not necessarily involve serving the human participant. [1, 5, 7] And/or a move away from ‘natural’ conversational design might mean accepting that AI behaviors may be strange and alien and this should be embraced rather than mitigated against [2, 7]. It is also worth noting that existing ‘unnatural’ narratives and storytelling techniques [15] may have elements which are relevant to ‘unnatural’ AI-powered stories. The final perspective considered is that the entanglement of humanity and AI is inevitable, and therefore separating ideas of ‘human’ and ‘AI’ is an impossibility. The two are co-dependent and the development and advancement of one affects the development and advancement of the other. Therefore the only option is to accept AI as an extension of humanity and vice versa. [14, 20]

Since our research focusses on extending humanity’s capabilities via emerging technologies, we may take this final position as a given, and therefore are left with decentering the human player and concepts of ‘unnaturalness’ as design starting points.

3 Design Reflection
3.1 Exemplar Review
The first phase of the project involved a review of exemplar games and artefacts split into three categories – historic examples, AI-powered stories, and voice-controlled games. In the interests of brevity, only the most relevant examples are discussed below. Key takeaways that were later incorporated into the design are italicized for emphasis.
The two most significant historic examples were ELIZA [23] and the original Euphonia [3]. ELIZA was a text-based natural language system, created in the 1960s at Massachusetts Institute of Technology by Joseph Weizenbaum. Framed as a computerized psychiatrist, participants typed in their queries and the system responded in the same manner [22]. Euphonia was far older, an early example of text-to-speech (TTS) technology created by Joseph Faber in the mid-nineteenth century. Faber typed phrases into a piano keyboard, and the machine spoke them aloud via an automaton [3]. Despite being very different systems with very different end goals, one phenomenon occurred in relation to both. Participants assumed the two systems not only engaged in conversation, but also understood the content of said conversations, despite the fact that their creators never claimed such a high level of technical capability, and even attempted to dissuade participants of these notions [3, 22]. In the case of Euphonia, this was in spite of misgivings about the uncanniness of the automaton and its voice [3].

The AI-powered narratives were selected on the basis of the AI tool used to create them in order to explore as many different potential platforms as possible. Restless [17,18] and Sherlock Holmes Mysteries [13] were both created with tools which were taken forward for further experimentation.

Restless is a Unity game powered by Spirit AI’s Character Engine (CE). The player takes on the role of a ghost haunting a young woman’s apartment. The play experience is free-form and non-linear. CE’s Dynamic Menu system means that the menu of player dialogue options is generated on the fly by CE and can be altered by changing the player-character’s mood. Multiple moods can be selected at once, generating different menu options. Keywords can also be discovered and pursued as topics of conversation. The game makes use of ‘unnaturalness’ through casting the AI-powered player character as a playful, disembodied ghost. Therefore, when it generates disjointed or strange menu text, the player is more likely to interpret this as intentional characterization rather than a failure of the system.

Sherlock Holmes Mysteries takes an entirely different approach. Separating Google’s Assistant from the narrative, the assistant instead acts as the narrator for a ‘Choose Your Own Adventure’-style story, asking players what decisions they would like to make. While this simplifies the conversational options in the manner described by Moore [12] it also means that any frustrations encountered are more likely to be directed at the system, rather than explained away as character quirks as in Restless. Sherlock Holmes Mysteries is reliant on a keyword system similar to Restless, but here the lack of a robust set of synonyms sometimes led to dead ends, or attempting to guess the exact phrase required.

Finally, The 3% Challenge [6] based on the Netflix television show of the same name is a puzzle-based voice-controlled game. Players face various challenges including memory tests and hearing tests. Characters frequently ask players to repeat words and phrases after them under the guise of trying to fit in with the society presented in-game. This trains players in key word usage while also being in keeping with the narrative frame.

Having analyzed these (and other) exemplars the next steps arising from this phase of the design research were determined as follows: 1) to further study the different affordances offered by Google Assistant and Spirit AI’s CE; 2) to explore the challenges and opportunities associated with single versus multi character AI-powered games; 3) to consider how ‘unnaturalness’ might be incorporated as a design feature; 4) to further investigate the use of keywords in AI-powered narratives.

### 3.2 Initial Experiments

Next, short demo pieces were made with Google Assistant and CE, using both the Dynamic Menu and Natural Language input formats, and these were then presented to the wider team for feedback. Initially, each used the same basic narrative input formats, and these were then presented to the wider team for feedback. Initially, each used the same basic narrative input formats, and these were then presented to the wider team for feedback.}

Having analyzed these (and other) exemplars the next steps arising from this phase of the design research were determined as follows: 1) to further study the different affordances offered by Google Assistant and Spirit AI’s CE; 2) to explore the challenges and opportunities associated with single versus multi character AI-powered games; 3) to consider how ‘unnaturalness’ might be incorporated as a design feature; 4) to further investigate the use of keywords in AI-powered narratives.

Next, short demo pieces were made with Google Assistant and CE, using both the Dynamic Menu and Natural Language input formats, and these were then presented to the wider team for feedback. Initially, each used the same basic narrative input formats, and these were then presented to the wider team for feedback. Initially, each used the same basic narrative input formats, and these were then presented to the wider team for feedback. Initially, each used the same basic narrative input formats, and these were then presented to the wider team for feedback. Initially, each used the same basic narrative input formats, and these were then presented to the wider team for feedback.
The next round of experiments involved the creation of a multi-character AI game where characters argued amongst themselves if the player did not intervene (and sometimes even if the player did). It was found that CE was unable to offer this functionality satisfactorily – while it is possible to create multiple AI agents in CE, control over multiple characters is limited. Characters can hold knowledge about one another as well as the player, but only one character can present in a scene at a time, and the player must speak each time a character has spoken – a line of character dialogue cannot be followed by another line of character dialogue. This meant that it was not possible to implement the kind of interrupting behavior required for an automated argument using CE in this instance. Therefore, another tool was sought to test this particular scenario.

Charisma AI allowed for unlimited multiple characters present and allowed characters to speak after one another. A more developed (although still experimental) piece, Elevenses With Eddie Murphy and Two Feeding Robots [4] was made to explore this functionality. The user interface of Charisma is very similar to that of the interactive authoring tool Twine [10]. As in Twine, it is possible to zoom in and out of the Charisma interface which helps achieve a clear visualization of how each dialogue response is connected to each potential player input, offering a far more advanced version of Google’s conversational flow diagrams [9]. This led to reflection on the narrative design method to date. In CE it proved difficult to keep track of the overarching narrative once a larger number of conversational options had been authored, even if these were logged in Excel prior to entry into the authoring tool. Perhaps Twine could be utilized as a design, rather than authoring, tool.

At the end of this testing phase, both Charisma and Spirit’s CE were straightforward enough to permit a focus on general writing process rather than the minutiae of technical requirements. Multiple character versus single character pieces were found to be equally engaging, although inexperienced players seemed to find the ‘misbehavior’ of a single character less frustrating than multiple characters. However, this may also have been because the tourist character in the single player tests created more of an expectation of atypical AI behavior than the robot characters of the multiplayer tests. These findings were used to inform the outline planning of three potential projects for the team to choose between as the main game project. It was also decided that regardless of the selected game, Twine would be used for prototyping and planning of the narrative.

3.3 Project Selection

InGAME is a multi-disciplinary team, and therefore any project undertaken needed to foster collaboration between a variety of disciplines. The three potential projects outlined following the previous phases were:

- A voice-controlled VR game made with Spirit AI CE built on the findings of the initial single character test and using Euphonia as the central character
- A multi-character conversational game made with Charisma AI in which the player manages a love triangle between 3 AI characters (this would have built on the findings from Elevenses [4] but was also inspired by Seering et al’s idea for chatbots in love [16])
- A word puzzle game in which the tutorial AI gradually becomes a character in its own right – a wildcard idea which could be made with either CE or a bespoke solution and would build on ideas of ‘unnatural’ AI behavior and the inversion of subservient behaviors developed in the time tourist test

After discussing the benefits and drawbacks of the various options, it was decided that a voice-controlled VR game would allow the most opportunity for exploration of emerging technologies, one of InGAME’s core goals. Spatial sound, text-to-speech (TTS) and speech-to-text (STT) could all be explored as well as further developing ‘unnatural’ characterization and conversational design in an AI-powered narrative.

4 Game Design

Once these basic principles were in place, various others emerged either to ensure particular features identified in the earlier phases were included, or due to the constraints of the technology and timescale. Narrative decisions were then made based on these requirements and constraints. These were recorded in the design document which was edited collaboratively throughout the process. The requirements and constraints and their resulting design decisions were as follows:

Research Requirements:

- Exploration of concepts of ‘unnaturalness’ – Euphoria was selected because not only does it have a fascinating background story (a machine passed down through a family and eventually lost, along with the techniques used to make it work), it also inherently possesses theunnaturality shown to make players more forgiving of technological failings, and a context likely to make players assume a greater level of cunning on the part of the AI [3, 22]
- Experimentation with CE mood settings – Euphoria has two moods which are on sliding scales, Trust & Anger. If Trust, hits a low level, Euphonia becomes Suspicious. Suspicion can increase Anger, but some actions and comments such as being insulting or refusing to help automatically increase Anger without affecting Suspicion. Reaching the highest Anger level ends the game
- Ambient sound as mood feedback - An ‘Anger’ parameter in FMOD Studio is linked to CE’s Anger parameter in Unity to change the ambient sound depending on Euphonia’s mood. This helps the player understand mood changes despite the lack of vocal inflection. e.g. The weather worsens the angrier Euphonia becomes
- Experimentation with TTS – voice-control is the primary mode of interaction
- Twine as a workflow component – draft conversations were created in Twine and feedback taken on the playable Twine prototype before implementation into CE. The visual nature of the interface meant that any gaps in conversational pathways were immediately
obvious. This also aided implementation and testing as the idealized flow of any conversation could be mapped out, along with points for the AI to offer narrative nudges to return to the main plotline. The fact a playable version could be shared to give an approximation of the final narrative also helped give the team insight into how the final game might look and provided further opportunity for feedback and refinement of plotlines and dialogue options.

- Experimentation with keyword usage – keywords are often used to allow Euphonia to switch between conversational topics. Drawing on The 3% Challenge [6], some phrases may be repeated to elicit special responses. However, in Euphonia rather than being a repeated mechanic, uncovering this phrase is part of the central mystery of the narrative, so players may reach the narrative end without discovering it.

- Further development of knowledge model – Euphonia’s experiences prior to the player meeting her were mapped to a timeline to account for questions relating to past and present.

Constraints:

- Limited timescale – the game occurs in a single location (Euphonia’s attic), will have a first person viewpoint, a single Non-Player Character (Euphonia), and uses a default TTS voice to limit asset creation.

- Limited time for animation may lead to uncanny character appearance – Euphonia was selected as primary character due to her embodiment of unnaturalness.

- Default TTS voice lacks emotion – sound and narrative design convey character mood instead. Character mood settings affect the narrative tone of Euphonia’s responses and sound effects. (see Figure 1 & Research Requirements bullet point 3 above)

- VR movement can cause motion sickness for some players – the in-game camera is static, centered on Euphonia and requires no movement to play. Conversation is the game’s primary focus and occurs only in short bursts. Menus and controls are kept to a minimum.

Considering constraints as opportunities to determine some of the required components is by no means a new approach [19] but it helped us balance exploring the features and technologies we were most interested in both collectively and individually alongside the difficulties we would inevitably face. When COVID happened, we were forced to make a further amendment and temporarily remove the VR component, as we felt this would add too much unnecessary complexity to both testing and sharing the work.

5 Conclusion

Ultimately, the design process ran as follows:

- Exemplar Review – current playable examples and historically documented precursors were examined. They were selected in order to cover a wide range of platforms, mechanics and technologies while remaining within scope of the central research interests – AI-powered and voice-controlled narrative.

- Theoretical Background Reading – occurred in advance of and alongside the entire project whenever new concepts arose.

- Experimentation – each made over a matter of days, and intended to get a feel for particular tools and mechanics, with a couple worked up further (although still taking no more than a week) to further analyze key elements. This included experimentation with authoring processes as well as with the tools themselves. These experiments were played and discussed by the team.

- Project Selection – several project possibilities were outlined based on the previous research, and one was selected which best matched the needs of the team.

- Design Creation – a design document was created building out from the fundamental restrictions and needs suggested by the technologies in use and team research goals.

- Prototyping – Twine was used to plot the narrative and gameplay flow, and this prototype was then played and discussed by the team.

- Game Creation – The narrative was implemented into the game engine (CE) and the relevant technologies were integrated (Unity and Microsoft Azure TTS).

This practice-centered workflow provides time to explore the affordances of the relevant technologies, but also allows freedom to discard those which were taking too long to learn or were found to be unsuitable for the desired end product. However, this process would be equally applicable for a wide number of projects and specialisms.

6 Next Steps

Euphonia currently only has bare bones implementation which showcases the TTS functionality and AI-powered storytelling. Further dialogue options will be fleshed out to limit repetition of central storyline phrases. An art pass will add a 3D model of Euphonia to the scene, further increasing the uncanny nature of the experience. Additional ambient sound will be added to create atmosphere. Euphonia’s sounds will be tied to her speech and moods so that it is more apparent to the player when her mood shifts. Accessibility has proved a challenge for one team member who has a vocal disorder, so options such as typing rather than speaking commands may also be added, particularly now that the VR component of the game has temporarily been put on hold.
testing will then be undertaken to further assess and refine the narrative content. A key challenge will be balancing playful 'misbehavior' with a rewarding player experience.

ACKNOWLEDGMENTS

Our thanks to the wider InGAME team; the universities of Dundee, Abertay and St Andrews; the Spirit AI & Charisma AI teams; and the UKRI Creative Clusters program. This project is funded by InGAME: Innovation for Games and Media Enterprise, part of the AHRC Creative Industries Clusters Programme. (Grant Reference AH/S002871/1).

REFERENCES
