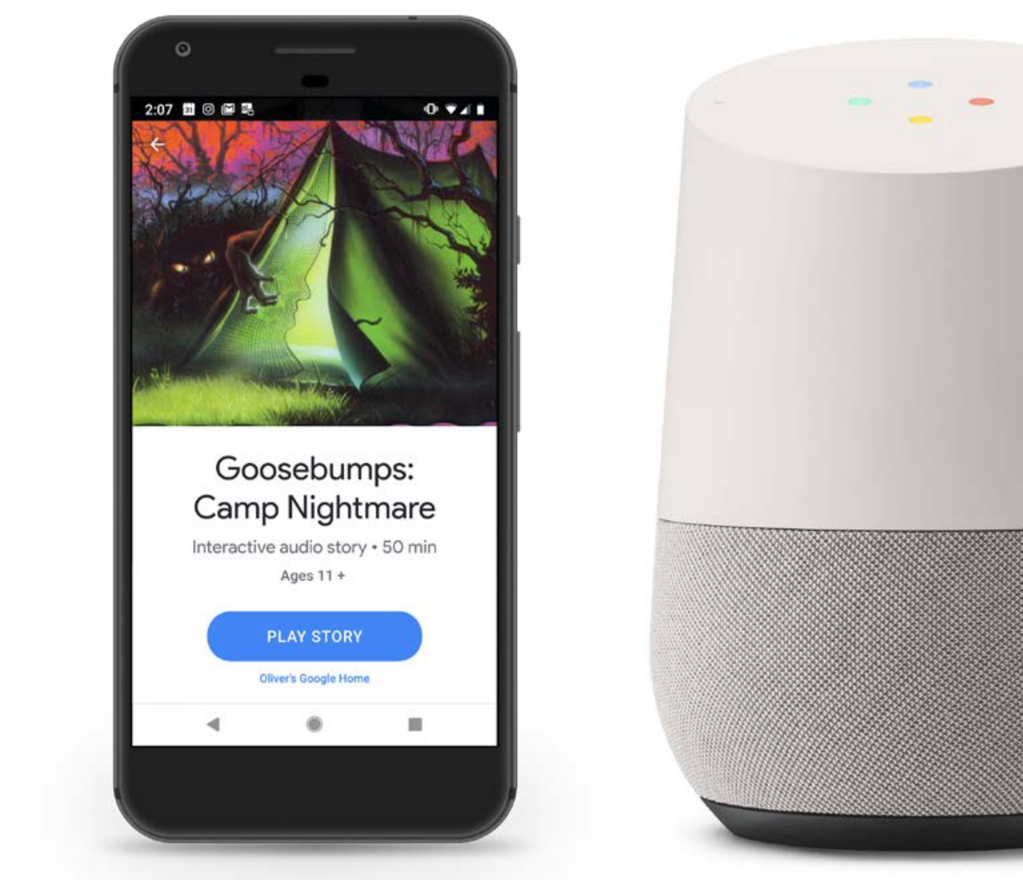
Protagonist: an Interactive Storytelling Experience

Oliver Engel, HCID 520A

Protagonist



I. Concept

Protagonist is an immersive choose-your-own-adventure narrative experience where the user becomes the protaganist in an audio-only story. When prompted, they use their voice to respond to other characters in the story and will subsequently affect the narrative arc, depending on how they respond.

Users can give open-ended responses, which are analyzed for content and emotional qualities to choose which narrative path should be taken. This differs from existing choose-your-own-adventure media in that the user is not restricted to selecting from a predefined list of options, but rather can be expressive in their response and let the algorithm sort out the rest.

II. Overall User Experience

Vocal entry point
A user can enter the experience by saying “Hey Google, open *Protagonist*”. A minimal orchestral soundtrack signifies that the experience is beginning... ”Welcome to *Protagonist*”, a hushed voice says. “Would you like to continue your last story or search for something new?” If the user wants to begin a new story, they vocally navigate a menu that sorts stories into genres and eventually pick one by saying the name of the story (see “Menu navigation” for specifics).

A user can also enter the experience by avoiding the menu, and selecting a specific known story. They might say “Hey Google, go to *Camp Nightmare*”. If the user chooses a specific story, it begins immediately playing the introduction to the story, bypassing the normal product introduction.

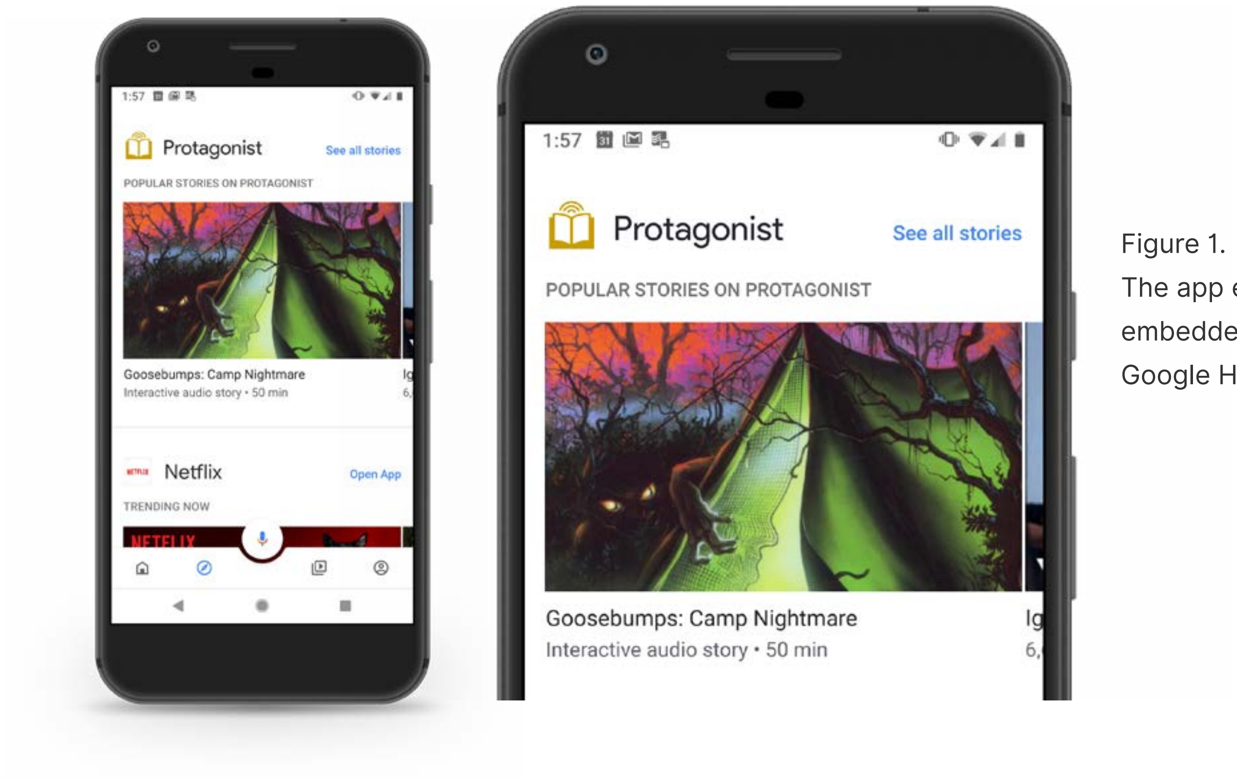

1:57
Protagonist
See all stories
POPULAR STORIES ON PROTAGONIST
Goosebumps: Camp Nightmare
Interactive audio story • 50 min
Netflix
Open App
TRENDING NOW
1:57
Protagonist
See all stories
POPULAR STORIES ON PROTAGONIST
Goosebumps: Camp Nightmare
Interactive audio story • 50 min

Figure 1.
The app entrypoint, embedded in the current Google Home android app.

App entry point
A user can also trigger a story through their Google Home app (Figure 1), by navigating to the browse tab and tapping on the story.

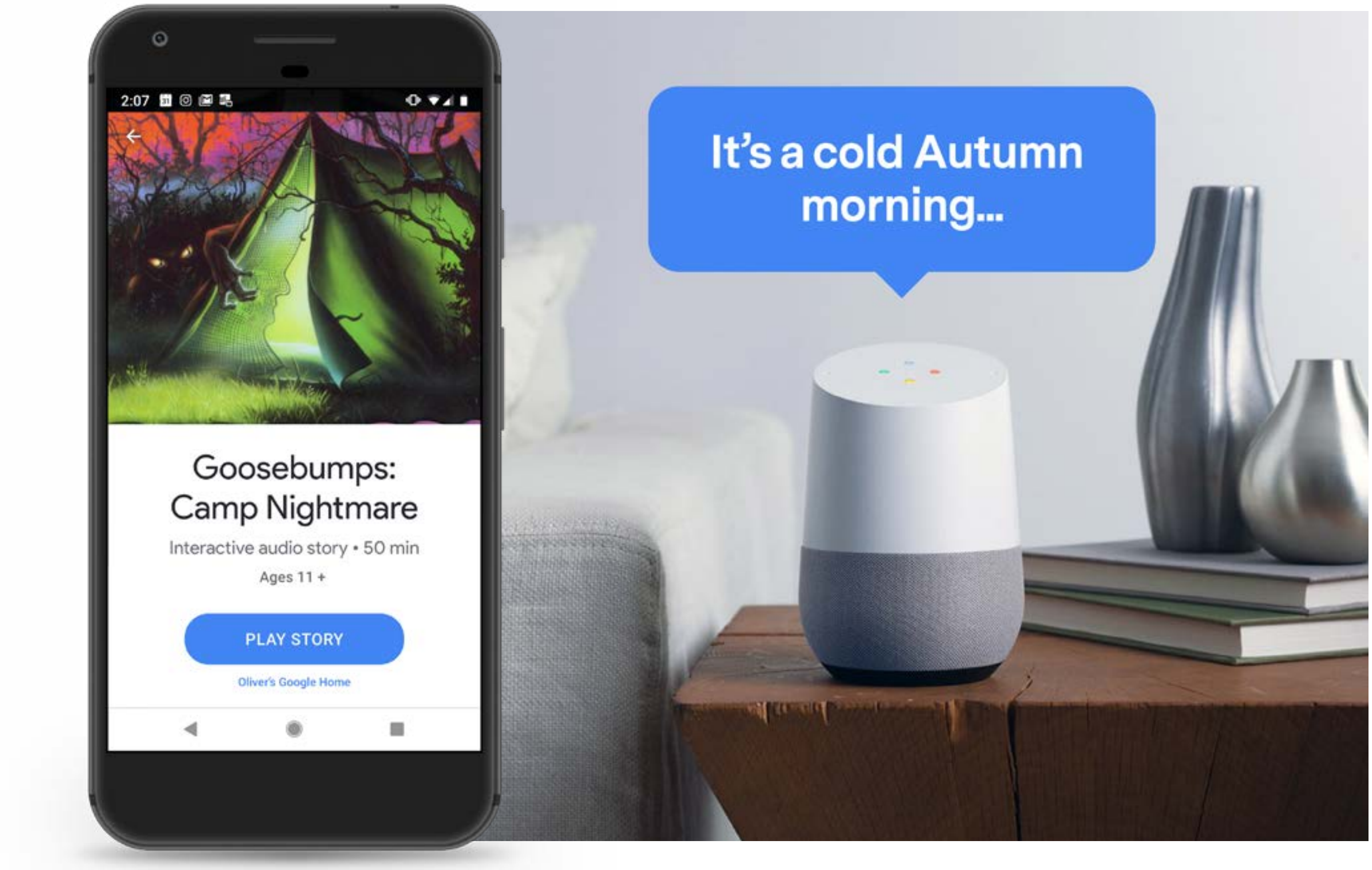

2:07
Goosebumps: Camp Nightmare
Interactive audio story • 50 min
Ages 11 +
PLAY STORY
Oliver's Google Home
It's a cold Autumn morning...

Figure 2. Triggering a story via the Google Home app entry point.

Bulk of the experience
The bulk of the experience is the user listening to the story and responding to prompts. If the user is playing *Goosebumps: Camp Nightmare*, a reimagination of a *Goosebumps* classic, it might begin like...

Sound of a schoolbus shaking and churning as it barrels down a highway, the excited chatter and laughter of children.

[Narrator]: *"It's a cold Autumn morning, and you and your friends are on your way to a school camping trip..."*

[Dori, one of the characters]: *"My cousin told me not to wander off into the woods while we're out camping. She said there's ghosts out there. Do you believe in ghosts?*

[User responding with voice]: *Hmm, I don't really believe in them.*

[Dori]: *What! My cousin says she's seen them out there. If you look for long enough you can see their red eyes between the trees.*

In this scenario, the program is listening for an affirmative or negative response. The response doesn't have to be as binary as "yes" or "no"–natural language processing picks out the general sentiment and keywords to classify the response given by the user, and then select the appropriate next piece of the story.

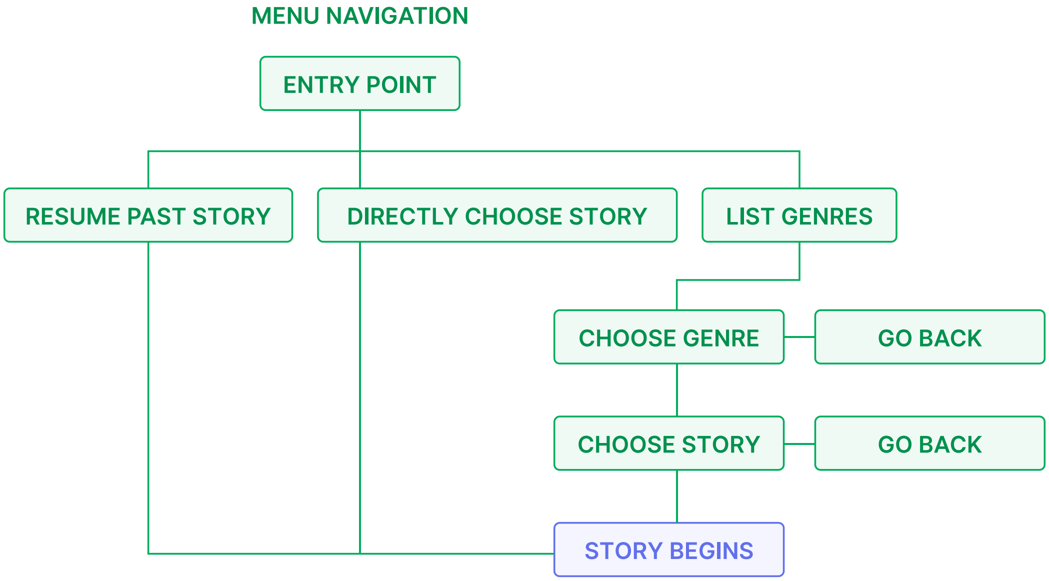


Figure 3. Navigating the audio-only menu in *Protagonist*. In the genre and story selector, the user hears 5 options at a time, and can either choose from those 5 or ask to hear the next set of 5 genre or story options.

Menu navigation

In order to select a story in the audio-only interface, the user must be able to browse a potentially large database of content with only their voice. Upon starting, *Protagonist* asks if the user would like to resume a story or browse story genres. If the user says browse, then *Protagonist* lists off the first five available genres, like "Adventure" or "Mystery", and asks if the user would like to hear more genres, if they are available. The user can select a genre by saying its name. This minimizes cognitive load and follows best-practice guidelines on menu navigation in audio interfaces.[1]

Upon selection of a genre, *Protagonist* lists off the first five stories available and asks if the user wants to select one, hear more about one, or list off the next 5 available stories.

The user can select a story by saying its name, or can request to "hear more about ____" for a brief audio description of what the story is about.

Exiting the experience

A user can ask Google to pause, play, or quit a story. This could be as simple as saying "Google, play / pause / quit", but we can also extend the natural language processing to map to these functions–a user might say "stop", "end", or "finish", all of which are semantically similar to "quit". When the user returns, the story will resume from where it left off.

Context of use

Protagonist is designed to work in the home setting, with a Google smart speaker device located in the same room as the user. The experience can be similarly extended to a mobile-only experience, but we will focus first on the in-home experience since it is more socially acceptable to vocally engage with a story in a private setting.

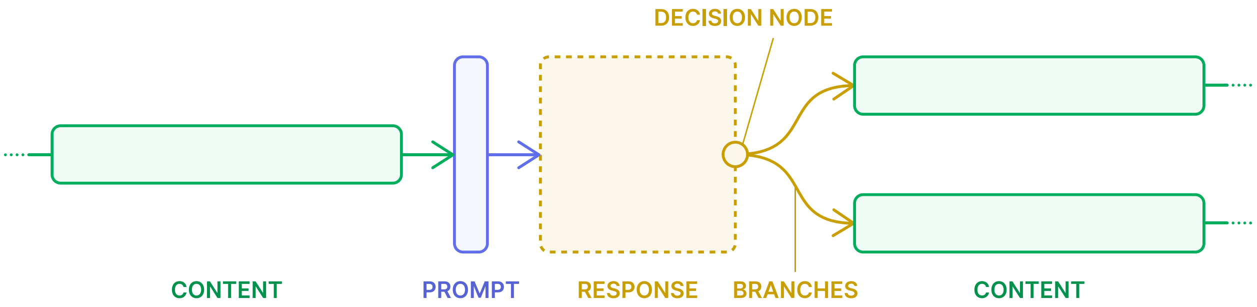


Figure 4. Individual components of the story experience. These pieces are modular and are largely arranged by the content creator in relation to a specific narrative arc.

III. Architecture of a story

Story

A story is the top-level wrapper for a complete narrative structure, which the user can navigate end-to-end. It is comprised primarily of content, prompts & responses, decision nodes, and branches. Figure 4 outlines the architecture of a story.

Content

Content is the bulk of the story, taking the form of .wav files that are supplied by the content creator. Content is static and non-interactive, and is simply played back to the user through the Google Home's built-in speaker. Content is up to the content creator, but is likely to contain character dialogue, ambient sounds, and musical scores that help to create an immersive story.

Prompts

Once the story reaches a point at which the user can interact with the narrative, they are given a prompt. Prompts are questions that are posed to the user, typically by the story narrator or by another character in the story, and are immediately followed by an audio cue that informs the user that they can speak to enter an input (see sound design section). Prompts are supplied by content creators, as they are specific to the narrative in which they exist.

Responses

Responses are user-provided voice inputs that occur directly after a prompt is given. These are the inputs that are processed via the text-to-speech → NLU pipeline. They can take any form that the user chooses, which introduces potential problems in recognition and content analysis issues. For specific directions in error recovery and response analysis, see Section VI.

Decision nodes

Decision nodes are the points at which a story diverges into two or more different directions, or "branches". Which branch is taken depends on the output of the response analyzing algorithms. See Section V for an in-depth explanation of the types of nodes and their algorithms.

Branches

Branches are the different paths taken throughout a story. One decision node can lead to one or more separate branches.

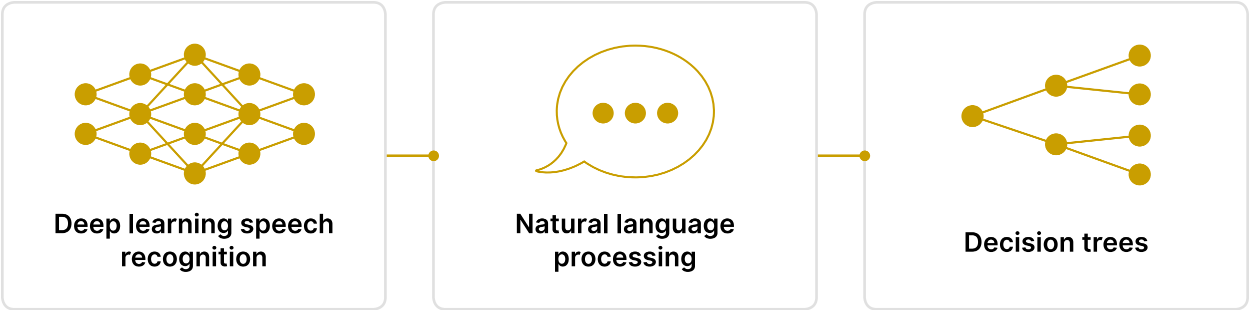


Figure 5. A simplified data pipeline flow. The first two stages record and process data, while the third stage uses that data to make a decision.

IV. Technology specifications

Platform
This product will be developed for the Google Home and Google Hub devices, with entry points also available via a linked Android device. The mobile device entry point will serve as a visual navigation aid to help users learn about available content, while the interactive audio experience will take place within the Google Home.

Technology components / APIs
The design necessitates three main technology components: first is a deep learning speech recognition algorithm that will be used to capture and transcribe user voice input into machine-readable text. We will use the Google Cloud Speech-to-Text REST API[2] for speech recognition, with audio input from the Google Home built-in microphone.

The second component is natural language processing, for which we will use the IBM Cloud Natural Language Understanding (NLU) API[3]. Text inputs generated from the Speech-to-Text API will be passed to NLU for processing, particularly to analyze for keywords and sentiment of the user input.

The third component is decision trees, which will form the basis for the branching nature of the choose-your-own-adventure narrative. While "playing" the audio experience, users will be navigating the branching structure, determined by the outputs of the NLU API for keywords, and sentiment, which are mapped to different branches of the narrative. You can read more about the specifics of the decisions in Section V.

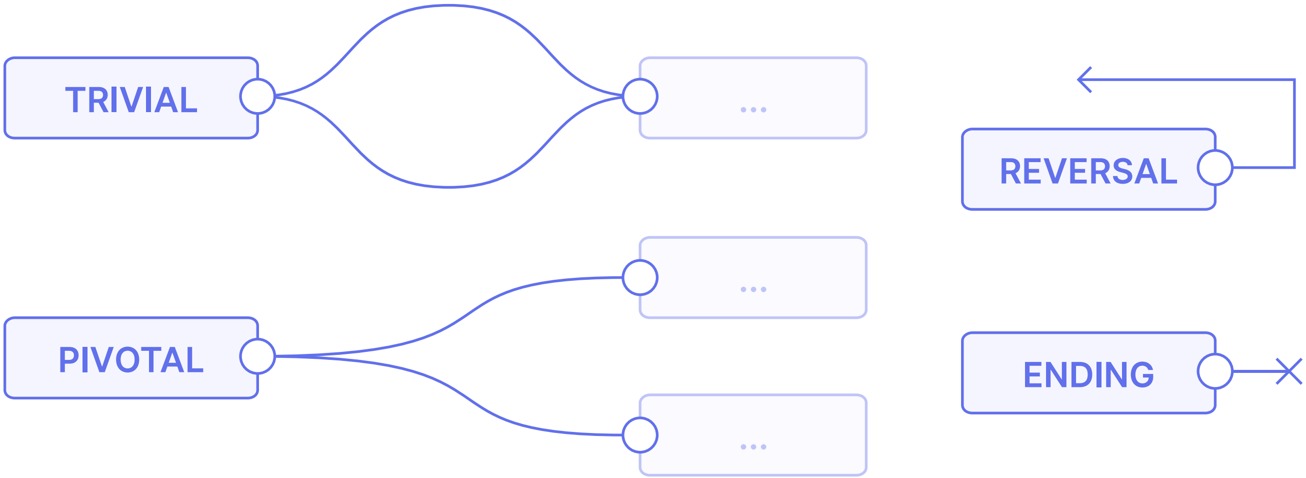


Figure 6. The four different types of decision nodes, explained below.

V. Decision Nodes

Concept
Decision nodes are the most important components of the story: they respond to user voice input and help to move the story forward. There are multiple types of decision nodes: trivial nodes, pivotal nodes, and reversal nodes (see Figure 6).

A *trivial node* is a section of the story where branches converge on the same path. As the name indicates, the user's selection is trivial; it could be a question like "what color shirt do you want to put on?" These types of questions give the user a more personalized experience; it's not necessary for every node to lead to a major plot pivot.

A *pivotal node* occurs when there *is* a major plot pivot. In this case, the user is prompted with character dialogue, and based on their response will significantly affect the story.

A *reversal node* occurs when the user is forced back to an earlier branch in the story. This is a common tactic in choose-your-own-adventure novels and can help the user experience more pieces of the story.[4]

Finally, an *ending node* accompanies one of the story's endings; it may or may not be paired with a prompt.

Algorithm

Selecting a branch at a decision node is the key algorithmic piece of the *Protagonist* experience. Here's how it works:

1. Each branch option is assigned a set of sentiment and keyword tags. These could be terms like "sad", "angry", or "selfish".
2. The user is given a prompt from a character in the story, and responds to the prompt with their voice.
3. The Speech-to-Text REST API transcribes the voice input to text.
4. The NLU API analyzes the text for keywords and sentiment, and generates corresponding tags for the user's input.
5. The tags from the user input are compared to the tags from the branch. In the case of a direct match, the matching branch is immediately selected. If there is no direct match, then the closest matching branch is selected.

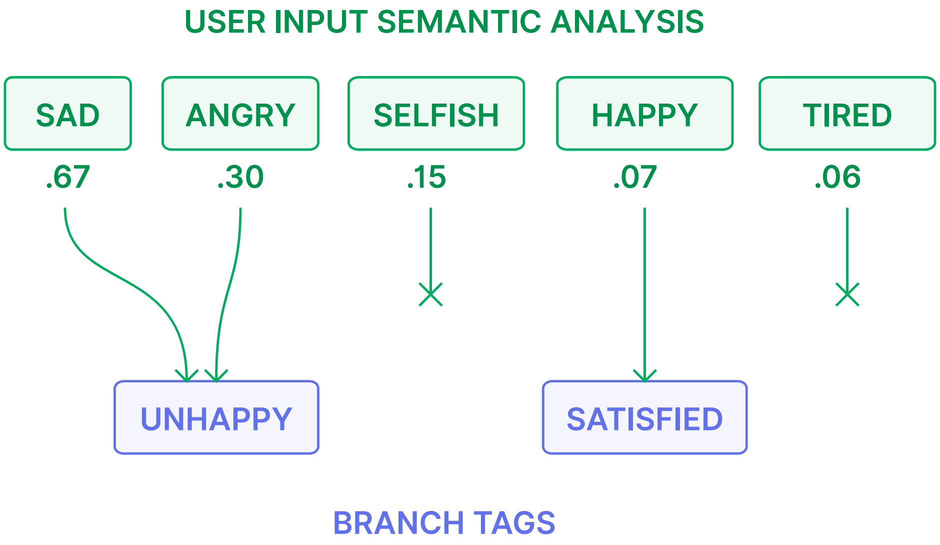


Figure 7. Process of comparing generated tags from user input and then comparing them to the branch tags in order to select which branch to move forward with.

VI. Error Prevention & Recovery

User input semantic analysis

User input analysis has the greatest potential for user frustration, since an erroneous choice of branch could break immersion and make the user feel as if they aren't in control of the story. To minimize errors here, we will use the following method to analyze a user response (also illustrated in figure 7):

1. Using the IBM NLU API, generate at least 5 novel tags based on the voice input.

Tags with similar semantic meaning (e.g. “sad” or “grieving”) should be grouped and counted as a single tag to allow for greater diversity in the list of tags. The API also outputs its confidence in the assigned tag, ranging from 0 to 1.0.

2. Compare the generated tags with the tags assigned to each branch, and calculate the similarity of their outputs. For example, if the API outputs a “sad” sentiment with .67 confidence, and an “angry” sentiment with .29 confidence, then the “sad” sentiment should carry greater statistical weight. So if one of the branches is assigned the tag “unhappy”, and the other is assigned “satisfied”, then the user’s input will trigger the branch with the “unhappy” tag.
3. If the generated tags all have low confidence (< .2), then it could signify an issue with the microphone, an issue with the user’s input, or failure in translating user input into text. The next section addresses these issues.

Recovery from poor vocal input quality or inaccurate voice-to-text

A known limitation[5] of speech recognition algorithms is their relatively high levels of inaccuracy. This may result in unrecognized user input which is unable to be analyzed for sentiment and keywords.

If the classification of user input is impossible, a clarifying question can be used to ask the user for clarification. For example, if Dori asks the user if they believe in ghosts, and the user’s response is unrecognizable, a .wav file can be triggered so that Dori asks “What? I don’t know what you said”, or whatever the content creator chooses. If after two clarifying questions it is still impossible to assign tags to the input, a branch is randomly chosen.

Minimizing noisy inputs

To minimize interference between speaker output and user voice input, the output volume should be reduced to 10% of its current volume while the user is giving their input. Though it could potentially break immersion[6], this is common practice with smart speaker setups so that voice input is not obscured by the Google Home’s sound output.

VII. Sound Design

Sound design is integral to the experience of *Protagonist*. However, a majority of the audio is provided as .wav files by content creators. There is only one sound effect that persists through all stories on the platform: the prompt cue. Whenever the user reaches a prompt, there is a brief audio cue directly following the end of the prompt to let the user know that the story is awaiting vocal input.

SOUND DURING RESPONSE

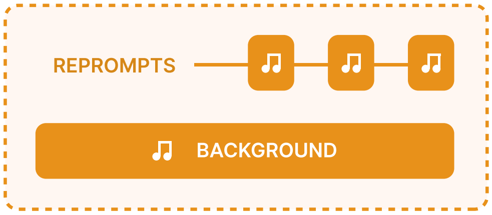


Figure 8.
Layering of sounds during
user response window

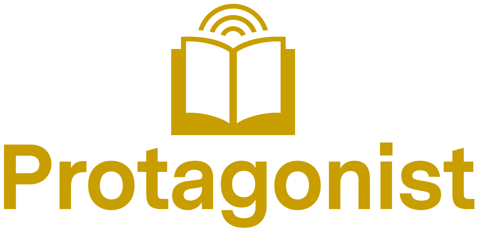
Sound design becomes more complex when the user is expected to give a response (see Figure 8). During the response window, there is either a background musical score or a soundscape, depending on whichever is used by the content creator. This is included to avoid breaking the flow of the story. If the user does not respond for 5-8 seconds, a re-prompt may be triggered: this is when the character that posed the prompt can prompt the user again, in case they don’t remember the question or are unable to answer. This can be either the same prompt, or a rephrasing of the prompt (e.g. “so do you believe in ghosts or what?” These must be supplied as additional .wav files that are triggered on a regular cadence.

After 25 seconds with no input from the user, a branch is randomly chosen.

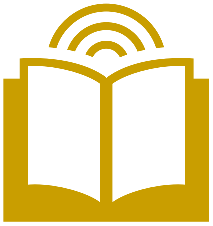
Visual Design

Integrated with the Google Home App
There is no additional visual UI component to *Protagonist*, as it lives in the Google Home App. The only UI customization will be a hero image that is uploaded by a content creator. This hero image appears in Figure 1 and 2.

Brand assets
The only brand asset needed is the *Protagonist* logo, below.



Logo & wordmark



Logo only

VIII. Works Cited

1 “Build for Voice with Amazon.” Ways to Build with Amazon Alexa, 2018, build.amazonalexadev.com/vui-vs-gui-guide-ww.html.

2 “Cloud Speech API | Cloud Speech-to-Text API | Google Cloud.” Google, Google, cloud.google.com/speech-to-text/docs/reference/rest/.

3 “Natural Language Understanding - IBM Cloud API Docs.” Natural Language Understanding - IBM Cloud API Docs, 1 May 2016, cloud.ibm.com/apidocs/natural-language-understanding#emotion.

4 Laskow, Sarah. “These Maps Reveal the Hidden Structures of ‘Choose Your Own Adventure’ Books.” Atlas Obscura, Atlas Obscura, 15 June 2017, www.atlasobscura.com/articles/cyoa-choose-your-own-adventure-maps.

5 Hannun, Awni Y. et al. “Deep Speech: Scaling up end-to-end speech recognition.” CoRR abs/1412.5567 (2014): n. pag.

6 “Lower the Volume of Videos and Music When Using the Google Assistant.” Google Home Help, Google, support.google.com/googlehome/answer/7383040?hl=en.