Semantic Chat: Enabling Greater Believability through Voice Avatars in Multiplayer and Story-Driven Games

Sherol Chen, Anna Kipnis, Kory W. Mathewson[2], Scott Ysebert,
Ben Pietrzak, Dan Cary, Erin Hoffman-John
[1]Google, [2]DeepMind
sherol,annakipnis,korymath,sysebert,bpietrzak,dcary,ehj@google.com

Abstract

A majority of games keep to discrete inputs and have not easily realized the expressivity of spoken language interfaces. Furthermore, natural language processing systems had limitations understanding language intent. For this paper, we define a type of language interface, **Semantic Chat**, and the challenges of achieving this functionality for interactive fiction and multiplayer games. In the past, games accepted text chat, through a keyboard, or voice chat, through a microphone; however, the inputs were often read verbatim and, at most, pattern matched to a desired intent. With recent advancements in deep learning, language models are able to more effectively derive the semantic meaning behind the textual input, and machine learning models have become increasingly better at transcribing voice. Even so, Semantic Chat is still rarely found in games. In practice, the application of these neural language models is an open problem, with non-trivial challenges in deployment. Using techniques like transfer learning, we discuss the obstacles in realizing believable voice avatars.

1 Introduction

Semantic Chat is a game feature whereby a player's spoken language is transcribed and then semantically matched with intent. It contrasts voice and text chat which use the exact words of the input from the player, and outputs sentences back to the player. *Facade* (2007) is an interactive drama and, along with its alternate-reality installation (See Fig. 1 and Fig. 2), represented an early example of Semantic Chat. For this implementation, developers used a human scribe (i.e. a wizard behind the curtain) to transcribe spoken text from the in-person player. Once the human-spoken input was entered, the language was matched to an intent, which would activate a dramatic segment via a reactive planner [Mateas and Stern, 2006]. This created a more immersive story experience, combining speech-to-text (STT) with natural language understanding (NLU) [Dow et al., 2007]. In recent work, STT has enabled voice assistants and accessible computing (without the aide of a human wizard), and enabled interactive dramatic performances through live transcription for translation [Mirowski et al., 2020] and interaction with chatbots [Mathewson and Mirowski, 2018]. However, is still rarely used in games, with the exception of interactive radio dramas.[1] We describe both the opportunities and the limitations of what STT today provides for games.

[1] http://codenamecygnus.com/

Wordplay: When Language Meets Games Workshop @ NeurIPS 2020,
https://wordplay-workshop.github.io/

2 Speech to Text

Speech interactions in games require audio input and speed-to-text (STT) functionality to process the audio (Fig. 3). Hands-free interactions within games, and particularly with virtual agents, is underutilized and underexplored. With the increase of machine learning (ML)-enabled tools, pre-trained STT functionality is readily available and can be used in place of typing natural language responses or further extending language capabilities of game engines. Today, enterprise application programming interfaces (APIs) are available off the shelf for application needs[2], specifically, and still uncommonly, in games. With this functionality, there is a class of opportunities that explore new capabilities which are unlocked by enabling speech interactions within game.

3 Multiplayer

Aside from experimental installations, the most common use for a microphone in games is for voice chat in multiplayer games. Often, this is challenging to moderate without human supervision. In Figure 4, *Super Smash Bros Ultimate* (2018)[3] gives players a menu of commonly expressed sentiments with pre-approved dialog. For usability, this list must be quick to use and often relatively sparse, like in *Hearthstone* (Fig. 5).

As you take on a avatar, you can look, move, and sound appropriately for the fictional world that is embodied by your character. In the case of *Hearthstone*, you are given the expressivity of 6 sentiments that each hero character can express, consistent within their narrative context. The pre-scripted nature of these interfaces, while capturing intent, is limited, and much closer to Choose Your Own Adventure Books[4] than what is capable today with NLP.

4 Believable NLP

By design, voice assistants (like Siri, Alexa, Google Assistant, and Cortana) use STT and NLP to build the narrative, where you are the main character in need of an assistance, much like Cortana's namesake from *Halo* [Cuddy and Chen, 2011]. Such API's exist whether as the Alexa Skills Kit[5] or, a more platform agnostic API, like DialogFlow[6]. As the primary input method, STT enables a number of story games in the form of interactive fiction [Bizzaco, 2020].

While the use generative models like we see with AI Dungeon[7], makes the most of language models, there is still a wide space to explore between a Choose Your Own Adventure and GPT-3 [Brown et al., 2020] to be explored. We believe that this open space can be steered towards specific believable experiences that only STT allows.

For example, semantic ranking via transfer learning enables natural language inputs to be matched with a significantly larger set of dialog options [Cer et al., 2018], preventing out of character behavior without limiting the expression space to a small set of sentiment options. Such controls create believability by providing a more natural human interface through spoken language. As such, experimental applications have demonstrated examples of advanced NLP towards believablity [Wiggers, 2020, Wallace, 2020]. We conclude that STT can accelerate and inspire solutions for Semantic Chat.

5 Conclusion

The ability to build voice avatars into game characters is vastly extended by Semantic Chat. While voice chat is useful for socializing and strategizing, it often breaks the immersion of a believable world. To handle this, games often restrict the expression space to a few sentiments or employ pattern matching, like *Inform7* [Reed, 2015] in order to force the player to stay in character. With

[2] `https://cloud.google.com/speech-to-text`

[3] `https://smashbros.com/`

[4] `https://cyoa.com/`

[5] `https://developer.amazon.com/en-US/alexa/alexa-skills-kit/start`

[6] `http://dialogflow.com/`

[7] `https://play.aidungeon.io/`

advancements in STT, there is even more utility to be gained from a hands-free Semantic Chat. We present an open challenge of making speech controlled voice avatars a reality for enabling greater believability in story-driven and multiplayer games.

References

M. Bizzaco. The best games to play with alexa, Sep 2020. URL `https://www.digitaltrends.com/home/best-games-to-play-with-alexa/`.

T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

D. Cer, Y. Yang, S. yi Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil. Universal sentence encoder, 2018.

L. Cuddy and S. Chen. *Halo and Philosophy: Would Cortana Pass the Turing Test*. Popular Culture and Philosophy. Open Court, 2011. ISBN 9780812697285. URL `https://books.google.com/books?id=IWBA0eAt7SMC`.

S. Dow, M. Mehta, B. MacIntyre, and M. Mateas. AR façade: an augmented reality interactve drama. In A. Majumder, L. F. Hodges, D. Cohen-Or, and S. N. Spencer, editors, *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST 2007, Newport Beach, California, USA, November 5-7, 2007*, pages 215–216. ACM, 2007. doi: 10.1145/1315184.1315227. URL `https://doi.org/10.1145/1315184.1315227`.

M. Mateas and A. Stern. Façade: Architecture and authorial idioms for believable agents in interactive drama. In J. Gratch, R. M. Young, R. Aylett, D. Ballin, and P. Olivier, editors, *Intelligent Virtual Agents, 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21-23, 2006, Proceedings*, volume 4133 of *Lecture Notes in Computer Science*, pages 446–448. Springer, 2006. doi: 10.1007/11821830_37. URL `https://doi.org/10.1007/11821830_37`.

K. W. Mathewson and P. Mirowski. Improbotics: Exploring the imitation game using machine intelligence in improvised theatre. *CoRR*, abs/1809.01807, 2018. URL `http://arxiv.org/abs/1809.01807`.

P. Mirowski, K. W. Mathewson, B. Branch, T. Winters, B. Verhoeven, and J. Elfving. Rosetta code: Improv in any language. In *International Conference on Computational Creativity*, 2020.

A. A. Reed. Telling stories with maps and rules: using the interactive fiction language “inform 7” in a creative writing workshop. *Creative writing in the digital age: theory, practice, and pedagogy*, pages 141–152, 2015.

C. Wallace. Can machine learning revolutionise game development? *MCV/DEVELOP UK Media Outlet*, Mar 2020. URL `https://www.mcvuk.com/business-news/can-machine-learning-revolutionise-game-development-stadia-thinks-it-can/`.

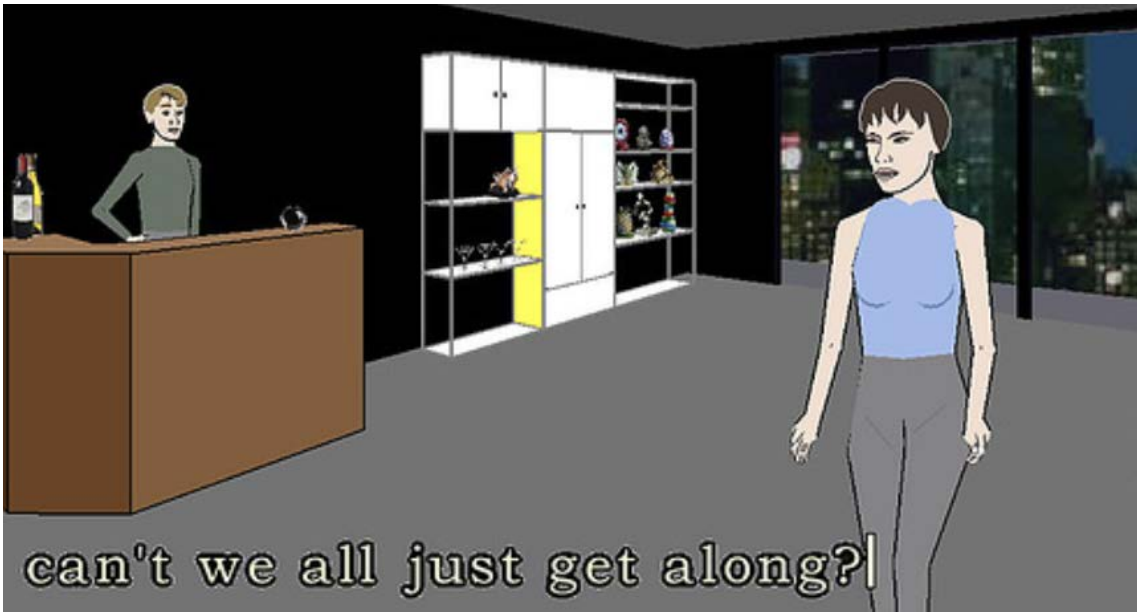
K. Wiggers. Google releases semantic reactor for natural language understanding experimentation, Mar 2020. URL `https://j.mp/2Iqa1Gx`.

6 Appendix

6.1 Sample of games that utilize voice controls

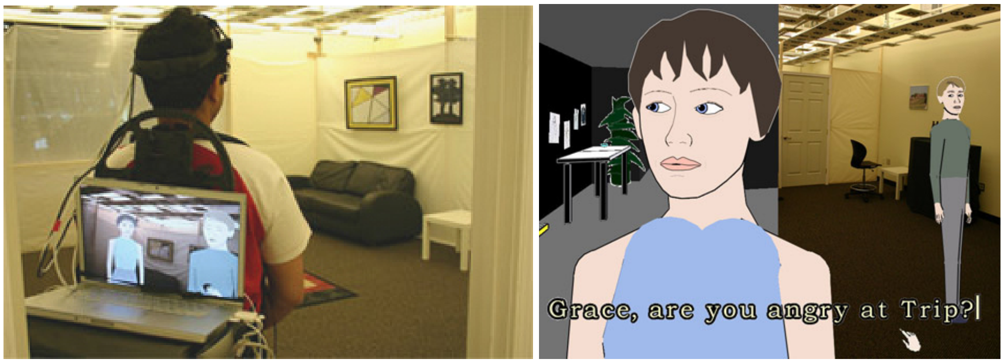
- Bot Colony: `https://en.wikipedia.org/wiki/Bot_Colony`
- Seaman: `https://en.wikipedia.org/wiki/Seaman_(video_game)`
- Mass Effect 3: `https://masseffect.fandom.com/wiki/Mass_Effect_3`
- Nevermind: `https://nevermindgame.com/about`
- There Came an Echo: `https://en.wikipedia.org/wiki/There_Came_an_Echo`

6.2 Figures



can't we all just get along?

Figure 1: Natural Language driven interactive drama, Facade.



Grace, are you angry at Trip?

Figure 2: Augmented reality version of Facade.

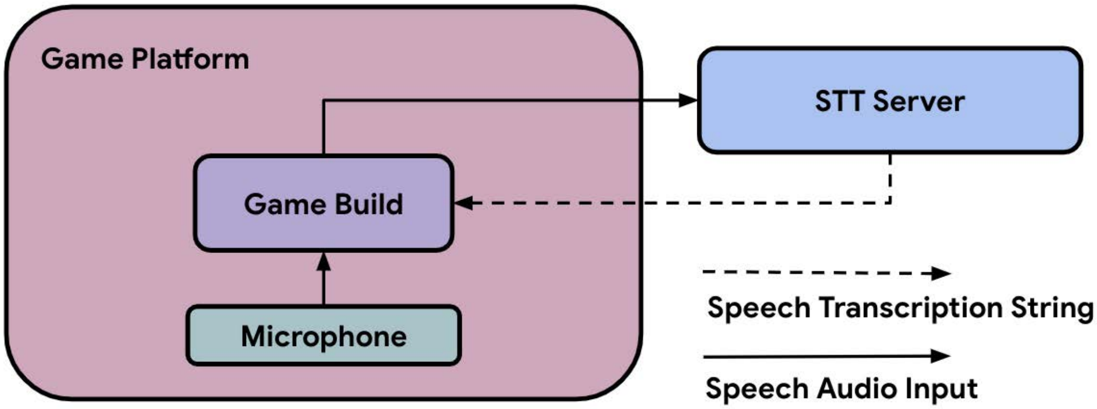
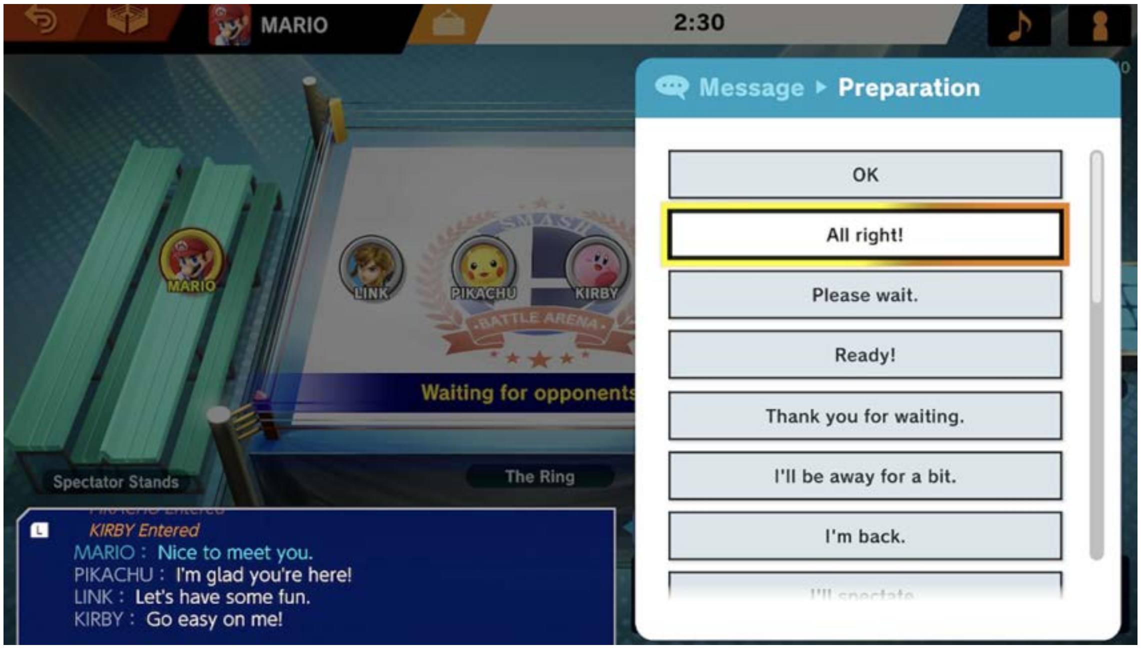


Figure 3: Speech to Text technical design.



PIKACHU Entered
KIRBY Entered
MARIO : Nice to meet you.
PIKACHU : I'm glad you're here!
LINK : Let's have some fun.
KIRBY : Go easy on me!

Figure 4: How Smash Bros handles expressive intent.

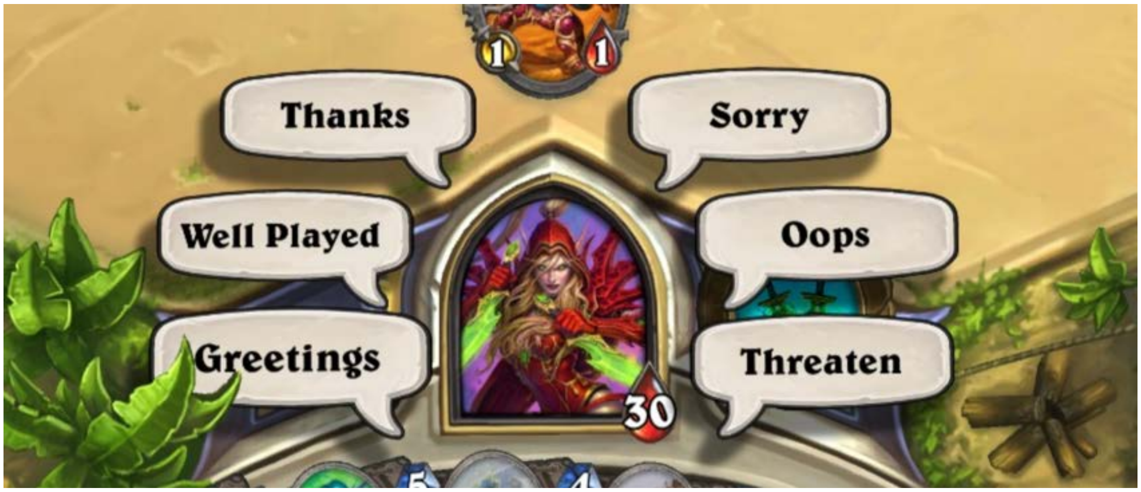


Figure 5: Shortlist of expressions in Hearthstone for Valeera, the Rogue.

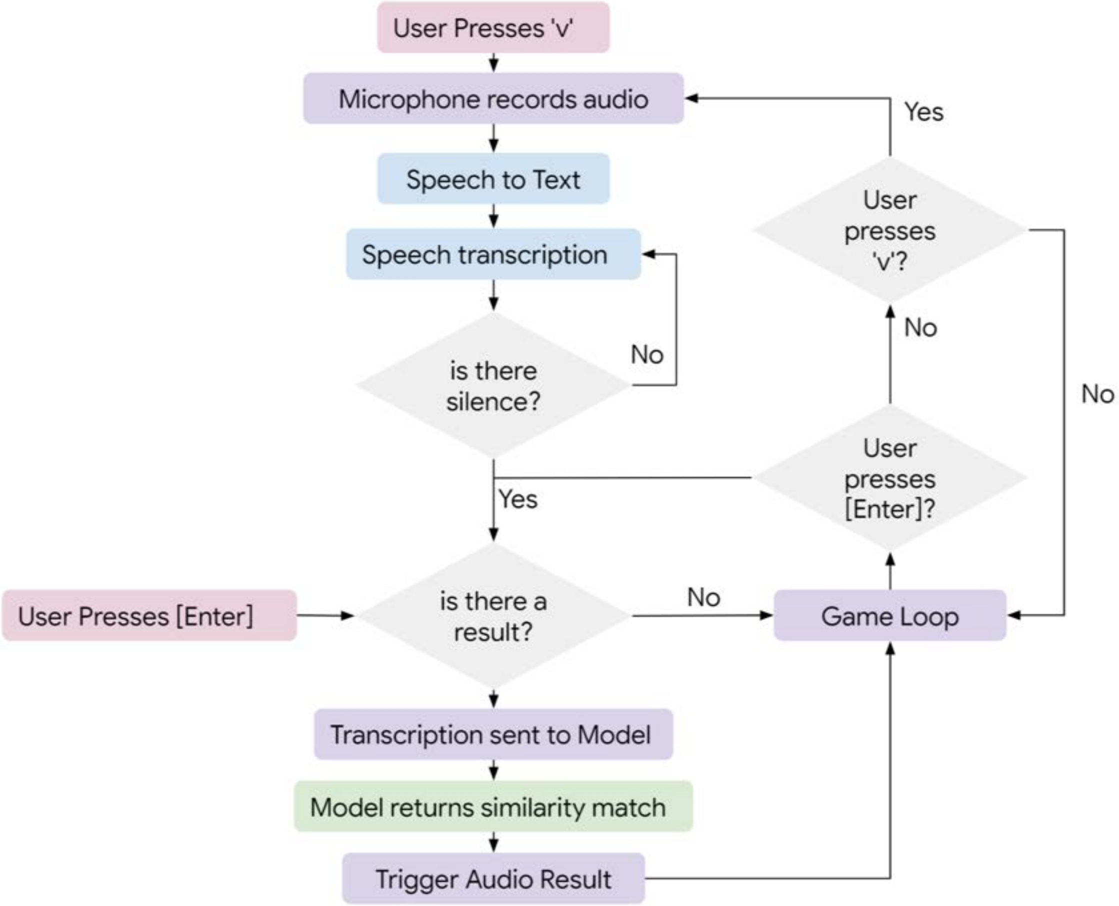


Figure 6: Semantic Chat.