

NLP AND VOICE TECHNOLOGIES FOR COMPUTER GAMES

APRIL 2021



Rafal Rzepka
Hokkaido University



Dusan Radisavljevic
Hokkaido University



Eugene Joseph
North Side



Mattia Ferrari
Indomitus Games



Joe Montgomery
Gamedev freelancer



Bartosz Ziółko
Techmo



Piotr Żygadło
All in! Games

Dear Readers,

We present to you the report on the voice and language processing in game industry. We hope you will find it interesting reading, which will allow you to look at the issue of those technologies from a broad perspective. Our goal was to create a forum for presenting views and plans, integrating the environment and perhaps establishing new relationships.

The report is far from being complete, and the choice of participants was a bit random. Please accept our apologies for that. We did not moderate the content of the interviews with games authors. However, we did a technical review of other texts and for some of them it resulted in minor modifications. As you can see from the report, applying natural language processing (NLP) and speech processing is still a niche, however, its usage is considered in larger and larger circles. Two main game winners for making games ;-) are still good story and valuable license for the title/world. Even though, the technological attractiveness of games is more and more important. Good example of this trend is a recently announced long term strategy of CD Projekt where the redefining and empowering the CTO role was declared as one of the key point of the company transformation.

Thank you very much to everyone who wanted to take part in the NLP and voice technologies in games report. We will make sure that it reaches wide audience of people who want to change the world. At Techmo, we have the ambition to inspire and be both the engine of change and its active participant. See you in the next report. Let's try to find out what the future of game industry will be like...

We hope the reading will be a valuable source of good ideas. We invite you to read!

Techmo Team

[Vision of applications of NLP and voice technologies in entertainment in 3 years from now](#)

Rafal Rzepka

page 3

[General view on NLP tools for games](#)

Dusan Radisavljevic

page 5

[Bot Colony](#)

Eugene Joseph

page 6

[Verbis Virtus](#)

MattiaFerrari

page 7

[Chant Savant](#)

Joe Montgomery

page 7

[What is ParlAI and Who Should Use it?](#)

Dusan Radisavljevic

page 8

[How NPC sentiment system in Dragon Age dialogues works?](#)

Dusan Radisavljevic

page 9

[ASR in Games](#)

Bartosz Ziółko

page 10

[Natural Language Processing](#)

Piotr Żygadło

page 11

Vision of applications of NLP and voice technologies in entertainment in 3 years from now



Rafal Rzepka

Language Media Lab
Hokkaido University, Japan



My vision of applications of NLP and voice technologies in entertainment in 3 years from now?

My first encounter with voice-based games was “Seaman” made for SEGA Dreamcast console. Heavily scripted chatbot-pet represented as a creature with a human face was supposed entertain players by allowing them to talk to a mysterious grumpy 3D Tamagotchi. It was my first encounter with a talking dialog agent other than simple IRC bots just when I started my research on artificial minds in 1999, and I must say it made me think about the future of entertainment. Now, more than twenty years later, I am fully aware that a natural dialog “about anything” is still beyond the reach of AI, but when it comes to games, it is more like creating closed “task-oriented” conversation flows. Of course there is a tremendous amount of possibilities to generate natural utterances when there is an unexpected input from a user, but even the latest models like GPT-3 suffer from keeping consistency in longer dialogs. My experiences with transformers in predicting usual flow of actions (what is usually done by whom and where), although showing some possibilities for games, tell me that automatic planning of sophisticated NPC’s behaviors is still far away.

Today’s AI heavily depends on machine learning, which in turn depends on vast amounts of data. There are not many attempts to generate dialog agents that consequently maintain one specific personality with a “history” because we do not have enough data on single entities, and the data in language models intermingles millions of human experiences. In the case of games, this is crucial to interact with entities which do not mirror the whole Internet-based corpora, we need personas “grown” in a given world, let it be a fantasy or sci-fi. An astronaut from 2099 could know about what happened in 1314, but a mage asked about Internet should be puzzled or mistake the word for a spell. For example when I asked GPT where Geralt of Rivia learned his sword skills, the model gave several answers: Dragon

Ball Z, YouTube, catfights and World of Warcraft. People constantly come up with ideas how to deal with this problem. Dialog systems build user models (personas), models are fine-tuned with specific data, etc., but it is worth to remember that the latest achievements are not magic wands working out of the box, and a lot of additional work is still needed.

The quality of voice recognition and synthesis has been remarkably improved, so we have basic tools available. Why then it is rare to use voice in games? What makes it difficult to design a naturally talking agent is the slow progress in natural language understanding (NLU) field. Today’s AI heavily depends on machine learning, which in turn depends on vast amounts of data. There are not many attempts to generate dialog agents with one specific personality and a “history” because we do not have enough data on single entities, data intermingles millions of human experiences. In the case of games, this is crucial to interact with entities which do not mirror the whole Internet-based corpora, we need personas “grown” in a given world, let it be a fantasy or sci-fi. An astronaut from 2099 could know about what happened in 1314, but a mage asked about Internet should be puzzled or mistake the word for a spell. This is not an easy thing to program, so I am interested in AIs that can “confabulate” within the realms of a given world or area of expertise. This can lead to new ideas in science but also in making believable game characters with whom everybody would like to have conversation with. AI researchers have been recently obsessed with “tabula rasa” approaches, where an algorithm is a blank slate learning from scratch, ignoring that we humans have innate mechanisms built in. Game industry could prove them wrong and help building new approaches in which basic instincts and knowledge about world mechanics are provided before learning process. I see the potential of game developers’ imagination combined with sophisticated machine learning – the creators provide a story line and personality traits, and the AI builds a reasoner which allows a given character to say more than “Looking to make some coin

Vision of applications of NLP and voice technologies in entertainment in 3 years from now

– Rafał Rzepka

too?” (beggar in “Witcher”). The immediate problem for a longer conversation with an NPC is to assure it will not reveal too much about the storyline or confuse the player with lies, but imagine that AI can generate consistent “life stories” for thousands of characters which are now just an element of the background. Such generation does not have to be limited to language (including specific voices, accents, mannerisms) – knowledge and reasoning capability about the world could lead to novel actions, looks or items (including an automatic sanity check deciding if it is normal for a dog to use a sword). I guess that AI-assisted “controlled generation” will be one of the trends in game developing. “AI-assistance” could also simplify developer’s work by becoming a voice-based tools which “translate” natural language into narrative branches aggregation, personality and story additions or graphic detail changes. If just saying “make this soldier’s voice angrier” or “add a tragic family accident to this beggar’s life story” could automatically adjust the code, it would save time and money allowing non-programmers to experiment with the game during development.

It is always hard to predict future in tech, even in a span of few years. From the AI perspective, it might be safe to guess that new games and application will borrow also from the latest NLP solutions. Replika is a good example of a virtual “friend” which uses not only augmented reality to make the dialog agent more “human-like”, but it also utilizes GPT-3, one of the latest language models, to deal with input which was not predicted by the software creators. Replika is also an example of a chatbot which blurs the boundary between task-oriented “psychological help” represented by WoeBot and virtual “pets” like Seaman. It seems important “who” takes care of your mental health. You will feel different when a robot ignores your out-of-question utterances, and when a human-like avatar more or less properly reacts to whatever you say. Another blurring of boundaries can happen between “video games” and “education software”. In Final Fantasy X there is a tribe of technologists called AI Bhed. They use their own language, and although it is not important to the game itself, learning a few words can become handy. There is a chance that this approach can be used to learn foreign (existing) languages as well. Learning by immersion is said to be the fastest way to acquire a language, especially when understanding more gives the player advantages in the game itself. Also, it might be easier to sell a game to a parent who knows that a “standard-looking” shooting game also teaches their kid Spanish or Chinese faster than a school teacher. Of course, it does not have to be limited to languages. There are plenty of professions requiring talking

to have the job done. Imagine a court game, where you become a lawyer. To be prepared for defending a client, naturally you need to know the case and the law, but then you must convince the jury, adapt to new witnesses’ statements and express weaknesses in what the prosecutor just said. Algorithms for measuring confidence or eloquence are already here, but as mentioned above, we still work on the reasoning part. If the “making no sense” problem is solved, a new depth in voice-based simulations will open. I assume that still the gamers want to simulate a lawyer, an interpreter or a negotiator are not the majority (maybe only when we add these tasks as episodes to standard games), but we cannot be sure. I myself decided to buy Nintendo DS for my son only after knowing there is a vast range of software that will allow me to learn and test myself. And I was not much into games before that. Finding new potential players seems to be an important step to predict what opportunities may arise for voice-based gaming. It is hard to talk to a console on a train, that is for sure. But on the other hand, for people driving long distances, hands-free, voice only MUD-like games not requiring immediate response (see Amazon Alexa-powered Vortex by Doppio) could open new doors to people who have never been good at moving their thumbs and virtual reality makes them dizzy.

My last prediction contains a little bit of wishful thinking. As I wrote above, there is not enough data for what I dub a “tailored learning”. Every one of us is a specific individual who lives a single storyline which determines what we talk about, to whom and how. Due to privacy constraints, we might never have access to the enormous amount of multimodal data we gather during our lifetimes, but games can provide quite rich snapshots of what we do and what we do not do. About our preferences and choices, behaviors and misbehaviors. Using games to support knowledge acquisition and AI testing is not at all new, but not many researchers have resources as big as DeepMind, for example (they employed one of my students just for working on games). Storing detailed storylines of all individual game sessions is probably a standard for internal analysis and debugging. But sharing them with researchers is still rare. I believe that more data-driven co-research with academia, although being often a bumpy road, could bring the game industry more benefits than it is currently anticipated. Without the deeper understanding of the game’s world, conversational characters will always be limited. More and more students are interested in AI for games and working together with game developers could also become a good investment for the industry. ■

General view on NLP tools for games – there's not much on the table apart from tools supporting translation



Dusan Radisavljevic
Language Media Lab
Hokkaido University, Japan



Natural Language Processing, or NLP for short, is a subfield of artificial intelligence, linguistics and computer science that deals with processing and analyzing spoken and written language data. End goal often tends to be a computer capable of understanding documents and commands that are given using a natural language, allowing for interaction with it through it as a medium. As such, it has been applied to many areas of research with great success. However, work with NLP in the gaming industry has been quite scant so far.

When one thinks of video games, a first image that comes to mind might be that of First Person Shooter games or sport simulations that used to dominate the market a couple of decades ago. So the logical question might be: how and why would we use NLP for these games? Well, although these genres of video games are still very popular, recent years saw the rise in popularity of multiplayer online gaming that puts a strong accent on interaction between players. Additionally, role playing video games that have players interact with broad, often fantasy based setting, and utilize large amounts of textual data in order to increase a sense of immersion into the game's world, have also saw a rise in popularity, with The Elder Scrolls V: Skyrim being one of the all time best selling video games, and a recently released Cyberpunk 2077 selling over 13 million copies in just 10 days after release.

Online games that are centered around teamplay tend to have a strong accent on communication between players, while role playing games oftentimes communicate large amounts of information to players in order to increase a sense of immersion. This makes the communication aspect of these games quite important, and considering that the majority of our communicative capabilities are passed through either written or spoken language, use of NLP seems like a logical thing in the area. However, apart from some translation tools, use of NLP in developing and improving commercial games hasn't yet been thoroughly explored.

In contrast, NLP found great success in use for Serious games, a term that describes video games that are designed with a purpose other than pure entertainment. Considering that a large amount of these games tend to include the communication aspect in order to better interact with the end user, it comes as no surprise that NLP found its use here. Application of these games in areas of education or therapy might have led to NLP being utilized here, as these areas tend to be more popular in research fields than entertainment is. However, with some research suggesting that commercial video games can also be beneficial, either by increasing hand-eye coordination or by improving mental skills such as abstract thinking or logistics, it can be argued that research in entertainment can also be benefactory in the long run.

Some of the areas where NLP can really help when coupled with commercial games is analyzing how soft skills are used in an online environment or how character and personality traits impact our interpretation of virtual agents. Additionally, NLP in this area can also help us better analyze and understand cognitive or emotional motivations of players within the games setting.

It should be noted however, that even though there might be prospect of NLP for commercial video games, it is not without its limitations. For example, the automated analysis of text data is based on computer science and linguistics, so there is no guarantee of it being valid outside of these fields for tasks that analyze semantics. However, we believe that the only way something can be improved upon is by conducting experiments in the field and striving for better results. That being said, even though the work so far has been limited, we believe that it is only a matter of time before we are amazed by improvements NLP makes in the area of video games as it provides both benefit as well as entertainment. ■

Comapnies

Bot Colony / Verbis Virtus / Chant Savant

The first application of our NLU technology was in video games, with our Bot Colony title (www.botcolony.com). We are working now on a new, yet to be announced title, as well as a reboot of Bot Colony with better graphics and NLU. Conversation with intelligent characters (embodied or chatbots) in video games and VR is a game-changer, taking gameplay to a new level of immersion. Games relying on English conversation are an engaging way to practice speaking and writing English and have considerable educational potential, in addition to being entertainment products.

M What stimulates and what slows down introduction of voice and NLP technologies to video games? Did you experience problems with game developers willing to use new more automatic tools? How does it look regarding costs? What publishers think about applying new technologies?

E.J.: With the exception of Bot Colony where language is the main gameplay mechanic, language as a means of interface was used very little in games. Support for voice in games (again, with the exception of the new Bot Colony Redux that will rely on Microsoft Cognitive speech-to-text) is a rarity. Most voice applications in games are for simple commands.

This is probably due to the fact that most publishers don't dare risk going outside known games genres, and prefer building game experiences that have been commercially validated (most high-budget, financially successful videogames are shooters).

M What do you expect from 2021 in the context of your company? What have you learned in the last year? How do you rate 2020 in the context of your business and new technologies in game development? What are your plans for 2021?

E.J.: We've learned that it is extremely difficult to build a robust game based on NLU (and voice), and unless the game is robust, it won't succeed.



Eugene Joseph

CEO and Founder of North Side

North Side of Montreal, Canada develops software to support conversation in unrestricted English between people and machines. We support natural, unrestricted English interaction with embodied characters, aware of their 3D environments and past perceptions. The company is engaged in Natural Language Understanding (NLU) research since 2002.



In the last couple of years, Microsoft Cognitive (and the Google speech-to-text offering) offer precision which enables integrating voice in games without degrading the performance, provided this is accompanied by a strong NLU pipeline for disambiguation. However, a strong NLU pipeline is like the Holy Grail - extremely difficult to build.

M How will NLP and voice based games look like in 3 years? How common will these solutions be? What will they not be? What will their functions and application be?

E.J.: It is hard to predict. We believe that there's a lot of pull from the VR side for good NLU, as good NLU has the potential to redefine the VR experience (keep in mind that you still need to move a mouse to direct a character in VR - VR with autonomous, conversation capable characters doesn't exist - North Side is working in that direction). Publishers have not jumped on the voice/NLU bandwagon for the reason evoked above. At this time, there's still very little understanding among publishers about what kind of VR experiences and games can be built with NLU and voice.

M How different types (age/gender/background/countries etc.) of players react to NLP and voice tech in games? Who feels it naturally, who doesn't and how to change it? How do you adapt speech and language technologies to various types of players and their expectations?

E.J.: You need to type or speak English to our game, and people who are not native English speakers make mistakes. Hence, the need for even more robust NLU and voice technology. As mentioned, NLU is not yet mature enough, and without it speech-to-text, even if it works well, has few applications.

Adapting these technologies to different types of players is something that should be envisaged only after NLU and voice work really well.

M What is a computing and memory budget for dialogues in your game? How many processors and memory can you spare for voice and language calculations? Does it depend on how good new NLP and voice technologies are? Does it vary in various projects?

E.J.: Microsoft, Google, Apple, IBM speech-to-text and text-to-speech solutions, all work in the cloud, so this is not an issue.



Mattia Ferrari
Lead Programmer

Indomitus Games is an Italian video game development and consulting company. It released *In Verbis Virtus*, a voice-controlled fantasy adventure, and *5 Minutes Rage*, a cross between a party game and a shooter, both built with Unreal Engine.



M What stimulates and what slows down introduction of voice and NLP technologies to video games?

M.F.: Voice has always been used in videogames, not for interacting with the videogame itself but with other people. Nowadays services that allow players around the world to talk and share each other gaming experience have grown a lot. Voice not only communicates information but also emotions. We think that speech recognition in games can greatly increase immersion, but it must be kept in consideration in every part of the game design process. If it is inserted merely as a side feature, it easily becomes irrelevant or harmful.

From our point of view, most publishers don't have any specific interest in games with speech recognition. They just consider it valuable if it can create buzz, just like any other unusual feature.

M What do you expect from 2021 in the context of your company? What have you learned in the last year?

M.F.: In 2020 we started a new project featuring speech recognition, that is still in pre-production. As for new technologies, we are pretty excited about Unreal Engine 5 and we are keeping an eye on developments in speech emotion recognition.

M How will NLP and voice based games look like in 3 years?

M.F.: We have no idea. We hope our game will be out there, but we don't know how many other projects will make use of NLP.

M How different types (age/gender/background/countries etc.) of players react to NLP and voice tech in games?

M.F.: We had positive feedback from players. In *Verbis Virtus* makes you play as a wizard in a very engaging way: you pronounce a magic formula and the spell is cast. Players liked this concept and liked it even more when they realized that the recognition system worked well. To optimize recognition accuracy, we developed a fictional language that avoids phonemes that we found problematic, for instance 'r' and 's'. Moreover we designed magic formulas so that they can be easily distinguished by the system. Doing so we managed to make it work well out of the box for the vast majority of players, regardless of their native language. As a side note, we found that the system tended to be less precise with the voices of women and children.

M What is a computing and memory budget for dialogues in your game?

M.F.: The algorithms we use have negligible impact on the desktop systems we target.

M What stimulates and what slows down introduction of voice and NLP technologies to video games?

J.M.: Ease of use and difficulty of debugging were the main issues for me. If I remember correctly I used a MS library for speech recognition (the built-in .Net speech recognition library, e.g. System.Speech.Recognition.), and by and large it was easy to work with, however it had to run as a separate process which then had to communicate with the Unity game, and that was a bit of a hassle. The other main hassle was just the inaccuracy of the library, e.g. correctly and reliably identifying the words spoken by the player. The easier to incorporate the library and the more reliable it is, the more people will incorporate the technology.

M What have you learned in the last year?

J.M.: In 2020 and this year I ,m mostly working on dead-simple, 2D, board-game type video games.

M How will NLP and voice based games look like in 3 years?

J.M.: No idea! :D At least with the technology from a few years ago, I feel like voice-based games are a trap. E.g. they limit the player base that will be interested in the game, and they impose technical limitations and pitfalls that prevent people from playing the game

that might otherwise enjoy it. As an example, take Zoom, which is a great and top of the line voice app. Even with Zoom though, my meetings run into sound issues depending on various mics, background sounds, and Macs vs PCs vs Phones, the browser version vs the installed app, and on and on. So by adding voice to your game, you are adding a giant failure point that will catch out many of your users and add an extra layer of debugging and play testing difficulty.

M How different types (age/gender/background/countries etc.) of players react to NLP and voice tech in games?

J.M.: No idea! I will say though that the library I was using was not very reliable, and so you had to use careful and precise intonation for it to work correctly. I could definitely see it breaking when exposed to different voice ranges.

M What is a computing and memory budget for dialogues in your game?

J.M.: The MS library that I used was at least efficient, I never remember it taking more than a fraction of my (old, second tier) PC's resources. It was basically a non-factor in making the game run well. I would have been happy to devote a large fraction of the game's resource budget to voice recognition if the voice recognition had been reliable.



Joe Montgomery
Gamedev freelancer

Austin area programmer and gentleman of leisure.

What is ParlAI and Who Should Use it?

– Dusan Radisavljevic

Whether it is digital assistants like Siri or Alexa or use of conversational agents used to provide quick answers and help on corporate websites, dialogue agents have found many purposes in today's world. Improving on their dialogue capabilities has been the main goal for the longest time as it helps make them more believable. This comes as no surprise, since we often use dialogue to rate other media with how realistic they are, from movies to comic books. Dialogue can make the agent seem almost human-like, making the end user feel like they are actually conversing with another person rather than with a machine. However, a unified framework that would allow users to train and share language models for dialogue has not existed until recently.

ParlAI represents a python-based platform created by Facebook AI Research team in 2017, as a unified framework for testing, training and sharing dialogue models with a goal of enabling easier AI dialogue research. Additionally, ParlAI makes many popular datasets available at one place and further allows integration with Facebook Messenger, enabling users to connect their agents with chatting services, as well as integration with Amazon Mechanical Turk. Users of the platform can submit new tasks and codes and share them with the community, contributing to creating a general dialogue AI that can perform well at every task at once.

ParlAI has several core concepts, and uses its own specific terminology for them, with the most basic concept being agent. Agents have a bit of broader term than that of a dialogue agent. They can be human, a dataset being read out or a perfectly tuned neural network. Basically, anything that can interact with the environment. An agent has two primary methods that help

it interact with its environment: observe and act. The first method, observe, consumes a dictionary that is usually a result of some other agent, and is used to update the calling agent's state accordingly. The second method is act, which produces an action that interacts with the environment. For example, the act method of a dataset reading agent would return a next batch of data, while a neural network agent's act method calls a training or evaluation step.

Other core concepts are worlds, which basically represent an environment in which agents interact with each other, teachers which are agents that are used to train other agents in a world and messages, which are basically objects that are produced and consumed (or rather observed) by agents. Worlds need to implement a parley method in which a scenario of how agents, which interact through messages, should behave. A teacher agent, for example, asks a question to other agents and then depending on answers updates the internal report accordingly, allowing the users to get insight by calling the teachers report method.

Similarly with how agents act in ParlAI worlds, where they can observe and act, with some of them teaching and others learning, ParlAI allows the users to do the same thing - learning about other users tasks and models as well as teaching them about new tasks and models and thus working towards the goal of creating a general dialogue AI. All the basics are detailed in the paper titled "ParlAI: A Dialog Research Software Platform" by Miller et al. (2017) and the work is available at <https://parl.ai/> with the hope of the platform enabling systematic development and evaluation of dialog agents as well as pushing the state of the art in the field further. ■

How NPC sentiment system in Dragon Age dialogues works?

– Dusan Radisavljevic

Dragon Age: Origins is an action role playing video game developed by Canadian developers Bioware studios in 2009, and is the first installment in the Dragon Age series. Like many role playing video games, Dragon Age series is inspired by tabletop role playing games, such as Dungeons and Dragons. As such, the game allows players to create their character by selecting from a set of predefined backstories, put available attribute points into attributes, such as strength and stamina, and to select class and specialization (for example berserker warrior) as well as race (dwarf, human or elf) for their characters. The game also allows players to influence the world by choosing different actions and dialogue options when interacting with games NPC characters and the world itself, through which players ultimately shape the ending of the game.

Most action role playing video games tend to focus on character development, and are usually set in mythical or fantastical worlds. Due to that, characters that players create in these games tend to be projections of themselves in that universe, and for that reason some of them provide different dialogue options and actions as response to certain situations in game, allowing players to choose their own identity for their in-game character. Dragon Age: Origins is one of such games where dialogue plays the central role in how players interact with the world. Through dialogue, players can further the story as well as unlock additional events that can enrich the plot or acquire bonus items that can help them progress through the game's combat system more easily. However, not all dialogue options are visible to the player, as some only appear if their in-game character is of a certain race, class or has a certain origin story as well as enough attribute points spent on a certain attribute.

When it comes to dialogue, impact of players choices is most noticeable when they interact with one of the 10 available companion characters that can follow the player on his quest. Companions, much like the players character, have a certain background, class, race and personality traits that influence how they look on players actions. A more righteous companion character, for example, will disapprove of players actions if they are considered to be morally gray. How much the companion approves or disapproves of players overall actions is visible in game as the said characters approval rating, which is on a scale from -100 to +100. This can be viewed as a sentiment system, where each action makes a certain companion happy or unhappy leading to them being more or less satisfied with players in-game actions. This approval rating can, if low enough, lead to companion characters leaving the players party and in some cases even confronting the player in combat. On the other hand, a high approval rating can lead to companion characters asking for players' help with a certain quest that is related to their backstory, or can even lead to a possibility of in-game romance with that companion.

If we consider only the interaction through dialogue with these non-playable characters, we can view them as dialogue agents in the area of NLP. From here we can conclude that having a certain sentiment system, or an approval rating, can help the agents feel more human-like and believable, as it gives them a certain persona and motivation that they don't abandon if the player chooses conflicting actions. This approach can help increase the user engagement with the agents and make the overall experience feel more purposeful as well as memorable, as believable agents and characters help with better storytelling. ■

ASR in Games

Using voice technologies in games - tips from Techmo



Bartosz Ziółko

CEO

TECHMO



There are a few important questions one should ask before applying automatic speech recognition or synthesis in video games. First, what is the task? Is it to make playing smoother (easier to control various elements of the game) or is the voice element going to be the part of the game itself bringing fun or challenges to the player. I can imagine doing both at once but I would rather suggest to start with only one of those approaches. In the first case, the game developers should consider ergonomics of graphical and voice interfaces, and the balance between them. In the second case, voice has to be part of the game mechanics and hopefully of the scenario. An interesting study in this topic was made by Allison et al. They employed a group of people who played 3 different games with voice interfaces and asked them several questions about their experiences.

“The enjoyment of EndWar was attributed to its greater sense of control. Participants found its voice commands easy to understand and remember due to their familiar terminology and the on-screen menu of command phrases. ATCV lacked a visible command menu and used unfamiliar phrases such as “descend maintain two thousand”, which increased the difficulty of its voice commands. The Howler’s lack of constraints on voice input was cited by both those who enjoyed it most and those who enjoyed it least. Those who preferred The Howler praised its non-verbal voice interaction as offering a low barrier to operation and enabling an unusual style of

control. Those who disliked The Howler explained that the non-verbal voice interaction was uncomfortable in a way that they did not enjoy, and which distracted from their focus on the game.

When asked “What were you thinking about or concentrating on during the game you just played?”, three participants (3/22) said that they were more focused

on what they sounded like while playing The Howler than the actual gameplay. Several others reported being somewhat distracted by self-consciousness about their voice, although their main focus remained on the game itself. All participants reported being more focused on the gameplay than how they sounded for both EndWar and ATCV. Among those who reported concentrating on the gameplay, there was a split between those who focused on strategy (such as where to send units or how

to approach a certain level) and those who focused on function (such as what intonation or volume to use to improve their control). For EndWar, the majority said that they concentrated on strategy (16/20) over function (4/20). For ATCV, there was an even split between strategy (8/17) and function (9/17). For The Howler, half of the participants concentrated on strategy (11/22), with the remainder focused on function (8/22) or how they sounded (3/22).

Perhaps corresponding with this, most participants (11/18) said that The Howler was best suited to being played in a social context. Only a minority said the same about EndWar (5/15) and ATCV (1/13), which were generally agreed to be best played “locked up in a room by myself” (P24). While The Howler provoked the greatest sense of concern about the perceptions of other people, it also had the potential to create an enjoyable shared social feeling, as described in the thematic analysis.”

There is a kind of a trap in using voice interface in games. If it is just interface (the first scenario described earlier) it can bring attention of a player, but often for short time. If it is not connected with game mechanics and scenario, the interest will not last long, and if the game itself will not cause interest, the final attitude towards the game can be negative.

ASR in Games

Using voice technologies in games - tips from Techmo

– Bartosz Ziółko

Another important issue to consider are other people and speaking by a player in front of them. First, you may assume people will not be able to “voice play” in public transport. You need to decide and shape a game either with assumption it is just not a game to play there or give options of playing without voice. Be consistent in this matter – it as a kind of a choice all or nothing. But you need to think not only about playing in public places, but also about playing with friends. People might be more shy in front of other players and friends to speak loud to a game. Using vocalization in this situation is especially risky – for example loudness as part of interface.

And of course, there are general challenges of speech recognition and natural language processing. Anytime someone makes an NLP rule, exceptions will follow fastly. People speak to voice interfaces things different than designers expect. The best way to deal with it, is to test on production with real customers. It is painful and risky, but it is the most efficient way. Of course, limit the group size and make it bigger in future iterations of tests and development. Assume 6 months of checking what players say after launching the games to do voice interface corrections every 1-2 months (of course if privacy regulations allow).

People tend to not like their own voice, especially the recorded version. Consider using voice conversion techniques to allow a player hear their characters with more in-game voice rather than their own voices and playback what they say using this voice. These kind of technologies are available (also in Techmo) and can cause a WOW effect. If one starts to speak like a hero, it is a strong element of immersion.

Finally consider using speech synthesis for rapid prototyping of dialogues. You can have a simple tool to hear possible dialogue lines without any recording tools and losing your time. ■

Future of NLP in entertainment



Piotr Żygadło

CEO

ALL IN!
GAMES

These are fascinating times for Natural Language Processing! This particular use of computers and algorithms to work with natural speech and written language is becoming more and more popular these days.

While I write these words, it is difficult for me to miss the word processing app, this big blue microphone icon in the topmost right corner asking me to just dictate it. And just before I started writing it, I called my mother just by asking my phone to dial the number. I see this speech control icon on TVs (no need to add the word “smart” here – all of them are smart these days) and even in cars. And this is just speech recognition.

While you write a text message on your mobile, you can get contextual suggestions that would be approved by most grammar teachers. And let’s be honest, when was the last time you had a serious issue because of autocorrect? Recently it feels more like a meme-worthy content than a real problem from our everyday life

So it is quite safe to say that Natural Language Processing becomes a household feature. We might not want to use it. We might hate it when a call service uses it and it doesn’t work. But we cannot escape from it.

It is a common knowledge that the computer games industry is the first one to try any novelty. 3D, VR, motion sensors, internet, browsers, user-generated content - you name it. Even social media platforms owe their viral success to free games designed for them.

So how does our industry sees Natural Language Processing? The goal of this paper is to review the current state of tools and solutions available to developers as well as general situation and attitude towards this novelty.

And it looks like the road has been nicely paved. There are tools and toolsets, environments and libraries allowing everybody to work with natural language and speech easily and cheaply. Google is offering its speech recognition for free. Tens of developers are testing new ways of controlling the game and using voice and natural language as additional dimensions of game universes.

Natural Language has become the household tool. Is it ready to get between Muses and start its entertainment career? ■

TECHMO

Voice Technologies

Publisher:

Techmo Sp. z o.o.

ul. Torfowa 1/5

30-384 Kraków

Texts: collective author

Editing and proofreading: Bartosz Ziółko and Iwona Jendernal

Technical review: Tomasz Pędzimąż

Graphic design: Piotr Słota - Pada Deszcz Studio

Kraków-Sapporo, April 2021



The research conducted for this white paper was funded by
National Institute of Information and Communications Technology

More about Techmo:

www
techmo.pl



LinkedIn

/company/techmo-sp.-z-o.o.



Facebook

/TechmoVoice



Twitter

/Techmo_Voice



Youtube

/techmovoice

