

Geoffrey Shmigelsky

GeorgiaTech, Educational Technology, Fall 2016

AIVRE

Artificial Intelligence in Virtual Reality for Education

Abstract

The purpose of this project is to create AIVRE, an Artificial Intelligence for Virtual Reality in Education. The topic is the Serengeti; AIVRE will act both as an instructor and a tutor, content sourced from Wikipedia. A student will wear an HTC Vive Virtual Reality headset, communicate to the AIVRE using only voice, and experience an interactive lecture. The Virtual World built with Unity3D; includes animals, terrain, and vegetation. AIVRE is powered by a Google's chatbot engine API.ai while her voice is by IBM Watson. AIVRE is unusual in that she can express emotions visually and verbally.

Keywords: AIVRE, artificial intelligence, virtual reality, intelligent tutor, Wikipedia, chatbot, Unity 3D, HTC Vive, api.ai, Google, IBM Watson, emotional intelligence, Serengeti.

Artificial Intelligence in Virtual Reality for Education

Motivation

According to Edgar Dale the Cone of Experience, reading and viewing an article will result in a recall rate of only 10-30%. Youtube and Khan's Academy have potentially increased



the learning retention rate into the 50% range with the use of online and on-demand video. In 2011, Clive Thompson's article [How Khan Academy is Changing the Rules of Education](#); he highlights the positive impact on-demand Internet-based video education. Video training provides better retention than reading articles.

In Davis's and Summer's 2014 paper titled "[Applying Dale's Cone of Experience to Increase Learning and Retention](#)," the role of experiential education in learning outcomes is clear; both traditional and non-traditional students retain far more skills. More so those skills are successfully transferred to the workplace if the students experience the lesson, elevating students in the Cone of Experience and providing better retention than video education.

Wikipedia is arguably the world's most generic comprehensive database with over 5 million articles. It has become the world's largest encyclopedia; unfortunately, it is almost entirely text and image based. Resulting in a lower retention rate overall.

My motivation is to develop a process to take Wikipedia content and elevate it into the 90% retention range, as defined by the Cone of Experience. Allowing students to reach a higher level of recoverable knowledge for the same amount of hours invested in learning.

How does one elevate Education into a 90% retention rate?

My Plan

My plan is to leverage new technologies to improve the learning experience. Advances in Computer Science have enabled tools that just did not exist a few years ago. My approach will utilize two key advances:

1. Artificial Intelligence to create a next generation Intelligent Tutor.
2. Virtual Reality to create an immersive interactive experience.

In 2016, Bill Gates in an article in The Verge, explains how education could be customized to meet the needs of each student. Gates sees a future where each student has an AI mentor/tutor proactively helping every step of the way.

“The idea that you could talk to a [virtual] advisor that would understand different misconceptions and arbitrary linguistics around it, that’ll certainly come in the next decade.”

I propose to train a chatbot to act as an intelligent cognitive agent for AIVRE. Using Wikipedia as a knowledge source, the chatbot can be both an instructor and a tutor. It can give lectures as a professor would, while gesturing and moving like a human - all in Virtual Reality. Likewise, it can answer questions about a topic in a conversational manner. The student will interact with the agent using both Voice to Text/Text to Speech and the HTC hand controllers.

I propose a learning environment where the student will experience the benefits of a one-on-one instructor/tutor and experiential education in Virtual Reality.

Wikipedia will be the source of domain expertise; the topic is the Serengeti - History, Geography, and Wildlife. Virtual Reality is to be created and managed by Unity3D. The student will put on the HTC Vive headset and appear to be standing on the Serengeti Plains of Africa. The student will be able to walk around the scene, peering over bushes, under rocks or in the water.

AIVRE will also have emotional intelligence. She will express feelings in her face and her voice. Likewise, she will analyze student input to determine the student’s emotional state.

Execution

The execution plan is broken into three distinct phases, each with its challenges, risks, and milestones:

1. Training AIVRE's chatbot engine
2. Creating AIVRE's Serengeti classroom in Virtual Reality
3. Building AIVRE in Virtual Reality

Chatbot Engine

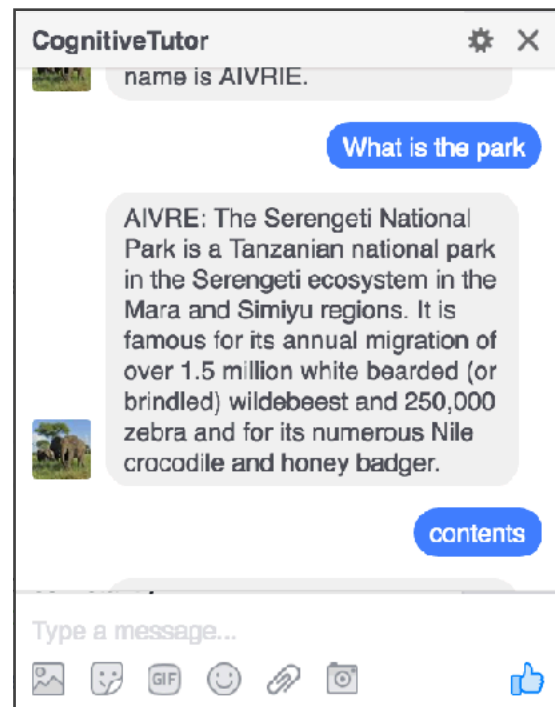
Chatbots have become a big topic in 2016. With Facebook announcing it as a new medium to replace mobile apps as the next generation of computer to human interface software. Given Facebook's dominant market position in texting, it is easy to see why. There are at least three main chatbot platforms, Facebook's Wit.ai, IBM's Watson and API.ai. During this project, Google acquired API.ai and announced significant improvements. I deployed API.ai to act as AIVRE's intelligence engine.

A Chatbot engine, in itself, is not usable by an end user. It is simply a middleware product. A third party product is required to interact with a chatbot engine. Facebook's Messenger platform with over 1 billion users, provides one avenue of interaction. API.ai offers an interface for many texting clients.

As a result, I implemented AIVRE inside Facebook Messenger, Slack, and web page. In all cases the results were identical, text messages sent to AIVRE where responded in kind. Typical response time was sub-second.

Trying out the Chatbot is best done Facebook Messenger. Each Messenger app requires a corresponding Facebook page, found here:

[Facebook Page](#)



Facebook Messenger

A chatbot is defined with its Entities and Intents. Entities serve to act as nouns or items in the conversation. Intents are samples of conversation used to train the chatbot. To keep track of a specific topic in a conversation, a chatbot will retain context.

A sample sentence would from a student be “I would like to learn about the Serengeti.” AIVRE would recognize the Serengeti as an Entity. A user could have easily just said “The Park,” and AIVRE would have understood, mapping the incoming text to an “About Park” Intent. Her Intent’s response could be “I can teach you about the Wildlife and History.” She will also set the context to the Serengeti, so that if the student asks, “How big is it,” she understands the topic is the Serengeti Park.

Intents are much more than a simple set of rules. They represent a many to many multiplexer of logic. Using Machine Learning, the AIVRE chatbot will be trained to recognize many versions of an English phrase to mean the same thing. Likewise, an Intent can have many responses to the same input. The consequence is that an Intent can encapsulate a significant portion of a conversation. For example, if a chatbot recognizes ten versions of an input sentence and has five different responses, that is fifty different human to computer interactions about a single topic.

Entities represent a many to one relationship. The term “King of the Jungle,” “Maasai Lion,” and “Panthera Leo” all mean Lion. Internally a chatbot will tracking the Entity as a Lion but can refer to with it using any of the aliases.

Domain expertise from Wikipedia can be transferred directly into the chatbot with copy and paste. It is possible to automate the task using JSON and Python.

API.ai offers a review and correction tool. Allowing for human verification to improve training based on human to student interactions, in real-time. Combined with Facebook Messenger, it allows for a quick build, deploy and test process with users.

API.ai includes additional libraries, such as news, mathematics, and personalization. Intelligence components easily added to AIVRE’s core functionality.

Serengeti in Virtual Reality

Using Unity3D, I re-created a scene from the Serengeti. Including a terrain, vegetation, wildlife, weather, and an environment.

A satellite elevation map of a river basin served as the foundation for the ground. This approach is remarkable as it allows the accurate recreation of almost any mapped point on Earth. With elevation maps, it is simple to replicate a physical location in Virtual Reality - from mountain ranges to city streets.

With procedural texture placement, the terrain is painted with dead grass, rock, and sand. The Serengeti is largely filled with dry grasslands, photographs from Megascans are used to fill the terrain based on elevation and slope. In total eight textures were used.

Vegetation is created using SpeedTree, including trees native to Africa such as the Acacia and Baobab. Logs, rocks, and Elephant Grass populate the ground. With procedural placement, items are mapped to the terrain, creating impressive results as shown in this [YouTube Video](#).



Wildlife consists of a high-resolution mesh and material, combined with a modest artificial intelligence. Allowing for the animals to roam, sleep, hunt, flee and interact with the environment. In total, almost 20 species are included; Antelopes, Monarch Butterfly, Lions, Elephants, Giraffes, and Zebras. Due to efficient use of Level of Detail, a herd of 100 photorealistic baby elephants, created procedurally, were deployed without any noticeable performance degradation.

AIVRE AI in Virtual Reality

AIVRE in VR is a sophisticated collection of software and services that bring her to life. Her core functionality includes text to speech, speech to text, emotional analysis, emotional expression, movement, morphing mesh, and interfacing to the chatbot.

AIVRE communicates with API.ai, the chatbot engine, using a non-blocking asynchronous interface. When a user speaks, AIVRE uploads the audio to API.ai, where is converted to text. The chatbot then matches it to an Intent, sending back a text response. Typically this process is under a second and will include instructions.

AIVRE sends the text response to IBM Watson for text to speech conversion. Using Speech Synthesis Markup Language (SSML), AIVRE can add emotional tones to the audio response. For example, AIVRE can make her voice sound happy, excited, worried or sad. She can even make herself sound younger or older - all in real time.

AIVRE will also use IBM Watson Bluemix for emotional analysis of the student's input.



In the image above, the VR Chat Log illustrates all the components together. Note the extensive use of human and AI emotion. Related Topics is meta-data instructions sent back from the API.ai chatbot. The chatbot can direct AIVRE with an extensive list of commands. AIVRE is in T-pose lower right.

AIVRE Avatar in Virtual Reality

AIVRE's female form is represented as a Morph3d humanoid and includes 150+ morph settings. Allowing for her to change shape and clothing as required, while using a sophisticated mesh on bone structure. On the previous page, it is mentioned AIVRE can sound younger. Correspondingly her visual form can match the transformation and convert to a 10-year-old child - all at runtime in real-time, with a single setting.

The animator system consists of 170 different bone mapped animations. She can climb, crawl, run and pick up items. Using Unity's Mecanim system Forward Kinetics, it is possible to add even more movements to her plethora of options. BlendShapes make its possible to mix and match in countless variations. With Inverse Kinetics, her hand and foot movement interact with the environment correctly and precisely.



AIVRE receives emotion commands from the Chatbot engine and can express at least ten feelings. Not shown here, her facial muscles randomly move to appear more human, such as raising an eyebrow or curling a lip.

To the image to the right, AIVRE is doing something remarkable; she is speaking - in real time. By analyzing the incoming audio feed from IBM, she is shaping her mouth to match the words. Here she is making the "Errr" sound. In total, she has fifteen phonemes.

Finally, AIVRE's eye's will track the student eyes, in virtual reality, making eye contact accordingly.



Results

AIVRE works as good or better than expected, I consider her to be an incredible success. She can interact with a human in a revolutionary way to teach and educate. The image summarizes much of the functionality:



Successful

Training a chatbot to act as a teacher worked remarkably well, using what was initially a manual process. It is possible to transcribe a dozen pages of Wikipedia into a chatbot engine to create the impression of intelligence and expertise rather quickly. With an automated Python to JSON process, it would be possible to scale a chatbot to thousands of Wikipedia pages.

Virtual Reality proved to be up to the task of creating almost photo-realistic worlds. Thanks to high-speed GPU's, today's hardware enables Physically Based Renderers (PBR) with high-quality 4K textures. In comparison, state of the art 4K residential TV is a single image. In VR, up to ten 4K images can be used for just a rock; ground, plants, and animals appear almost real.

AIVRE seems to be near human in VR. She walks, talks and acts much like a human would. With facial expressions, movement, and interaction.

Challenges

A few challenges remain which stifled development. The Unity platform running with 50+ packages is a little unstable and crashes, significantly slowing down development.

API.ai's Unity wrapper has a critical limitation; its client to server text communication is only synchronous. Causing the main game thread to block, forcing the VR to stop momentarily. It can lead to motion sickness, which is why human to AIVRE communication is limited to asynchronous voice. Thus preventing some of the more advanced HTC Vive controller commands from being activated.

Many of the animal models are having issues with the root motion, requiring a workaround which is sub-par. The animals motion does not always match its footing.

API.ai's processing of context is not efficient, it requires duplicate rules for most Intents. I would recommend a detailed study of Watson and Wit.ai, they may be more effective.

Next Steps

Ultimately, I intend to release AIVRE on the Steam VR app marketplace. The outstanding issues include:

1. Activate the missing animal animations, currently less than 20% utilized. Animations are in need of tuning and timing.
2. Make AIVRE more human, enabling her with many more commands that the chatbot can issue; such as laughing or dancing. Add facial expressions including iris dilation, blushing and sweat. Given AIVRE human skills like playing catch or tag.
3. Create more realistic animal herds with herbivores and carnivore behaviour.
4. Activate weather and a night/day cycle.

To turn the Serengeti VR into a tool usable by human teachers, it is necessary to tie it into a class management system, such as Google's Course Builder - Open Online Education.

Conclusion

AIVRE offers an insight into what education can be one day. Intelligent Cognitive Tutors have the potential to revolutionize the educational experience. Combined with Virtual Reality, a student can relive almost any experience. Together, these two technologies can significantly increasing a student's retention as defined in the Cone of Experience.

References

Benjamin Bloom. (1984) The Two Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One tutoring., *American Educational Research Association.*

Davis B, Summers M. (2014). Applying Dale's Cone of Experience to increase learning and retention. *QScience Proceedings (Engineering Leaders Conference)*

Gates, Bill. (Apr 25, 2016) Educational Software with Artificial Intelligence.
The Verve

Thompson, Clive. (2011). How Khan Academy is Changing the Rules of Education. *Wired Digital.*

Software Utilized

| ToolSet | Source |
|-----------------------------|-----------------------------|
| Chatbot | API.ai |
| 3D Game Engine | Unity 3D |
| PC Hardware | Windows 10, Nvidia GTX 1080 |
| Virtual Reality Toolset | HTC Vive by Valve |
| Speech to Text | Google's API.ai |
| Text to Speech | IBM Watson |
| Wikipedia Parser | Python Library: Wikipedia |
| World Creator, StateMachine | GAIA, Playmaker, Gena |
| Water, Weather, Sky | Aquas, Tonkoku |
| Animal Models | African Big Animal Pack, |
| Animal AI | Ice Creature Control |
| Plant Models | SpeedTree |
| AIVRE Model | Moprh3D |
| AIVRE Animations | Female Movement AnimsetPro |
| Eye Tracking, Audio Facial | Salsa 3D, Random Eyes 3D |