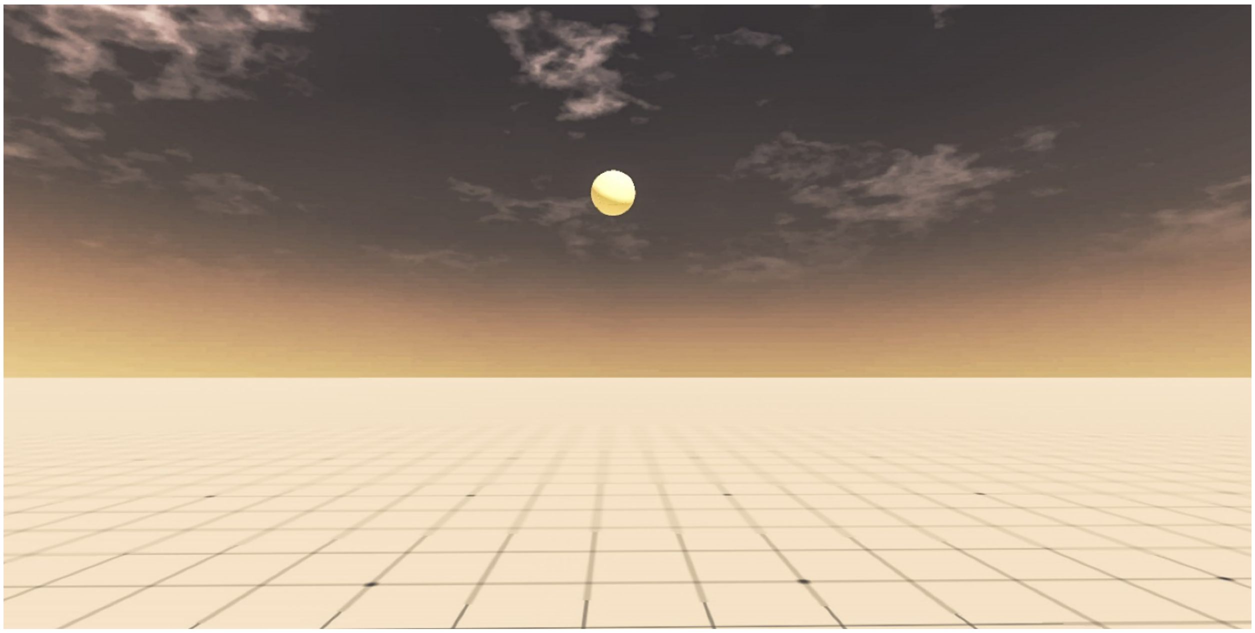
MACHINE LEARNING AND THERAPEUTIC STRATEGIES IN VR

Ray LC
Parsons School of Design
Cornell Medical College

Yuka Fukuoka
School of Visual Arts
Design for Social Innovation



ABSTRACT

Cognitive behavioral therapy is one of the few treatment options for individuals suffering from various psychological and physiological abnormalities like phobias, post-traumatic stress, stuttering, and anxiety. They are expensive in terms of both cost and time, and are often difficult to evaluate for effectiveness. Virtual Reality (VR) is a technology that can potentially supplant ineffective and costly therapies, because it is an experience immersive enough to be effective in reproducing stressful environments, and mobile enough to be used at home at any time. However, psychological therapies are supervised by highly trained psychiatrists whose years of experience help them diagnose problems of individual patients in a personalized manner. Can we use Machine Learning to glean information from patients that allow for such tailored-made therapy training for VR systems? In particular, Machine Learning of subject training history and selection of situations allows the system to adapt therapies to individual levels and needs. Meanwhile, Machine Learning of sensor data that provide details of patient stress levels in time allows us to interpret moment-to-moment fluctuations in symptomatic responses, and provides us ways to customize their training in terms of difficulty and responsiveness. The use of Artificial Intelligence (AI) in VR technology serves to enable customization and adaptive training while maintaining the ease-of-use and comparatively low cost over traditional therapeutic strategies.

INTRODUCTION

Machine Learning (ML) has been applied in areas as diverse as image recognition, medical diagnosis, and customer service. It has been envisioned that in the near future, we will be surveillanced by sophisticated image algorithms, analyzed by AI doctors, and talk to virtual agents with domain-specific knowledge.

Virtual Reality (VR) is currently revolutionizing the gaming industry, but its use in installation art, professional services, and training paradigms are sure to disrupt the industry, for immersive experiences can bypass traditionally expensive and ineffective strategies like physically impossible to construct buildings, psychotherapies, and mission critical tasks that can only be executed once in real life.

If future technologies are so intricately tied to ideas of using and dreaming of new uses for ML and VR, are there use cases for the use of both, which combines the ability to customize and predict user experiences of ML with the immersive and portable expressiveness of VR? We explore a use case tailor-made to utilize the capabilities of ML and VR: the case of therapy-based training for dealing with anxiety and stress, in particular, in public speaking and stuttering speech in social interactions.

BACKGROUND

1 in 68 children in the US have been diagnosed with Autism Spectrum Disorder, and the majority of them suffer from communication difficulties. At the same time, 1% of the total population suffer from stuttering, which is speech disorder. These people with communication disorders often find social scenarios stressful. Current treatments often include practicing “pretend-play” (Applied Behavioral Analysis) with therapists to prepare them for real world scenarios. In that case study, slow repeated reading and emotion expression exercises were practiced to improve the mental condition of a 10 year old (Kenjo, 2003).

However, these techniques do not awaken similar emotions in the patient as would the real situation due to the fact that it is not considered a sufficiently realistic simulation of the scenario. Moreover, therapists do not always understand the gravity of the specific situation because they don’t have the full sensory tracking data of their patients, who frequently suffer from anxiety disorders (Iverach et al., 2009). To address both of these issues, we propose a flexible, portable VR system that uses ML to interpret user responses and physical states to create individualized training paradigms to complement or reduce the need for expensive, time consuming behavioral therapies. Previous work using ML for therapy has mainly been using chatbots that check in daily with the patient, but its effectiveness has been questioned (Fitzpatrick et al., 2017). VR environments for therapy like Mimerse, CleVR, and SimSensei all lack the supervision and customization that comes with ML strategies, even though studies have shown that adaptive strategies benefit training for children with autism (Lahiri, 2011).

Machine Learning as a discipline emerged originally from the musings of Canadian psychologist Donald Hebb, who postulated that when two neurons fire together, their strength of connection is strengthened (Hebb, 1949). This simple insight postulated how associative learning, a model for how we learn everything from behaviors-rewards to names-people, can arise in the neural network. Subsequent development of the idea took a computational turn with the work of Frank Rosenblatt, who codified the idea computationally, however later machine learning research was hampered by the discovery that simple computational model of a neuron cannot code for certain mathematical functions such as the XOR (Minsky & Papert, 1969). ML research activity was partly revived with the insight that computational neurons connected together can implement complex functions (Figure 1), and importantly, an efficient algorithm devised for calculating weight changes that result from learning (Rumelhart, et al., 1986). More recently, more complex algorithms that uses multiple layers of nonlinear neural networks adapted from earlier models (Russell & Norvig, 2009) have been used to win Chinese character classification contests, chess, and famously the game of Go, using Google's AlphaGo.

Virtual Reality emerged from the experience theatre concept in the 1950s, but its digital manifestation came from work on input devices as diverse as head-mounted displays and the Aspen Movie Map, created by MIT as a virtual tour of Aspen, Colorado. In that program, textures were mapped onto 3D models, providing an early immersive experience from first person points of view. Gaming is the platform that transformed VR. Consumer headsets like Sega VR and Quick Time VR began to make the technology available to everyone, while today, applications like Google Street View are making VR on the web possible. 360 degree cameras have make video production for VR possible, while companies like Oculus and HTC are optimized enough for tracking head directions to make popular use possible.

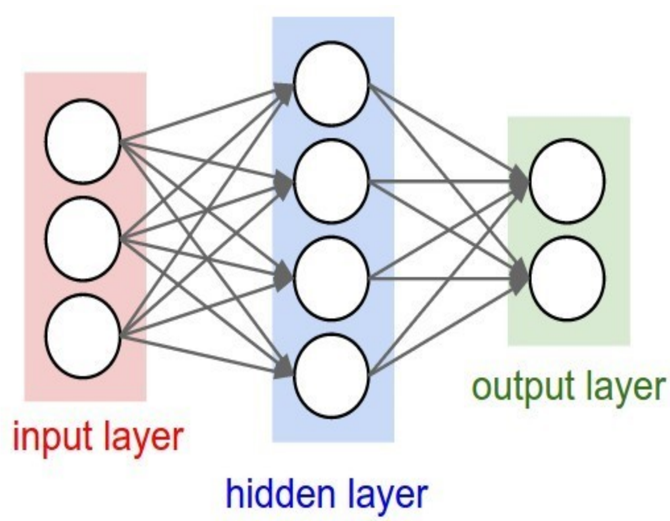


Figure 1. ML: A multilayer neural network. Input layers relay info to hidden and output nodes, where the desired output "trains" the weights in the network to make subsequent inputs fall more in line. Source: Glosser.ca.

Figure 2. VR: Sensorama, an early VR progenitor. An immersive display with stereoscopic images and stereo sound. Source: scriptanime.wordpress.com

METHODS

To solve issues of time-intensive and costly therapies for speech and anxiety in social interaction, I worked with a team of designers, computer scientist, speech researcher, and education technologist to propose an alternative solution utilizing Virtual Reality to train subjects, fortified by Machine Learning algorithms to 1. Process understanding of speech commands and patterns, 2. Incorporate sensor input that tell us about the current state of the trainee, and 3. Provide customized step-by-step training tailored to, and predictive of the difficulty level and progress of individual users.

We proposed to use galvanic skin response to keep track of user stress levels as training progresses, recognizing points of difficulty. In addition, computer vision would be utilized to help subjects interpret the expressions they observe, providing an alternative way to understand emotions of others through visual cues that make up an expressive language. Machine Learning enables a virtual environment that adapts to the user's needs as training progresses. Unlike weekly therapies, ML tracks user feedback and her focus of attention so that they can learn in a customized way whenever they are ready to train. These technologies provide a socially safe and optimized learning experience that can build communication skills. We aim to replace the currently outdated practices of Applied Behavioral Analysis, and produce an example of the level of high precision and realism that a customizable VR experience could have in the future when fortified with ML.

We wish to focus on an individual user, who has a communication disorder, seeing how she interacts with our customizable VR experience. We built a specific scenario in VR that our target audience typically practices in a role-playing setting (i.e. walking through a public park and interacting with strangers, speaking in an auditorium, a one-on-one meeting, etc.). The objective of the prototypes are to illustrate with precision how our concept would work on a specific use-case. By integrating human-centered design into our prototyping process and developing our minimum viable product over multiple iterations, we integrated user testing results to refine a VR demo of the specific scenario involving customization through ML.

The scope of this project goes beyond behavioral therapy and special education. VR fortified by ML can also lead to innovations in entertainment, advertising, and digital media. The implementation of a smart Augmented or Virtual Reality environment could not only have the potential to create more realistic new media experiences, but also further the idea of customization in these newly emerging spaces.

RESULTS

We first interviewed speech language pathologists to see what the current treatment strategies are like. The consensus appear to be that exposure therapies, in which the patient is exposed to the circumstance which usually causes stuttering, is required. It appears that often cognitive behavioral therapies proceed using systematic desensitization, which is the substitution of a none-threatening relaxation response in place of the fear response that came with speaking. In this counter conditioning procedure (Davidson, 1968), subjects learn that the lack of fear response that comes with relaxation can be paired with fearful situations, so that subsequent exposure leads them to conjure up the feeling of relaxation every time speech opportunities are triggered. However, the amount of fear associated with desensitization is often not set ideally, because only appropriate levels of relaxation should accompany appropriate levels of fear. Inspired by this strategy, we attempt to implement systematic desensitization in a graded, customized manner using ML that utilizes sensor data to understand user mental states.

When we talked to stutterers, another piece of the puzzle came into focus. According to them, how they perceive the audience they are talking to causes them either stutter or speak fluently. For example, Daniel M. says that if he knew the person he's speaking to is aware of his stuttering, he'll actually stop stuttering, because of decrease in social pressure to engage this knowledge. He also suggests that what he learns from therapy in the therapist's office is not generally applicable to the outside environment of loud noises and distractions because of the drastic differences in the two environments. He wishes there's a way to prepare for social interactions like job interviews in a progressive and immersive way. In summary, the perception of other's state of mind appear to influence how stutterers perform in stressful situations.

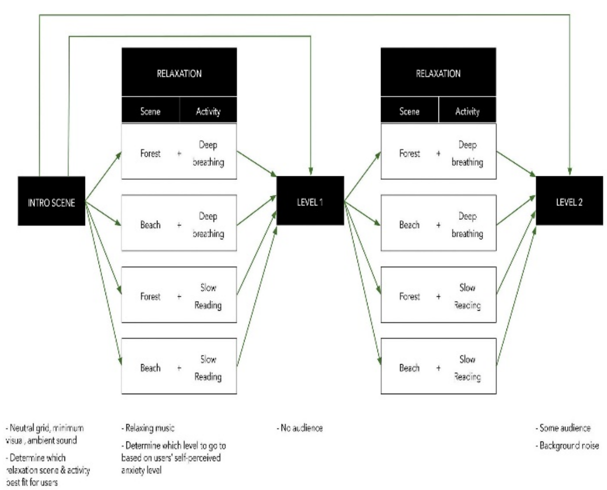


Figure 3. User flow of the training and relaxation sessions. Sensor data from GSR and other user movements are used by the ML to determine the difficulty of the next training.

Figure 4. A stuttering subject participating in user study to determine the factors that influence how well he is able to proceed during speech training sessions.

Considerations of 1. using relaxation as a method of systematic desensitization and 2. Daniel and other users' assertion that the difficulty comes hand-in-hand with the audience they are speaking to in a social situation, provided us the blueprint for creating a user flow for the patient's interaction with the ML system (Figure 3). The system is based on refinement of the prototype using feedback from 19 user tests that determined the fidelity of the interaction, the speed of the ML engine, blurriness, and other parameters. After a brief intro, we allow users to go into a relaxation scene of their choice based on ML recognition (using IBM Watson SDK) of their voice commands that indicate their preference (either forest or beach). During the scene, they perform a simple task in systematic desensitization that helps prepare them for performance (either a deep breathing exercise or slowly reading text, both suggested by our speech pathologists). Then they are given a task of introducing themselves in a realistic 360 degree video scene to a simple audience. The desired training at the speech scene can be repeated with different amounts of difficulty corresponding to number of people present, noise level, and other cues designed to distract the user.

Throughout the process, a Galvanic Skin Response sensor was attached to the fingers of a user so that we can monitor how much stress he is under (Brundage, et al., 2016). The ML will utilize the averaged stress level over a five second window in the preparatory relaxation scene to determine the level of difficulty level in the speech training scene, just as we do with voice input. Thus ML takes in as input layers the stress level sensor data, voice commands, and previous user training and learns a model which predicts the level of desired difficulty to use for speech training, which is adjusted by having different levels of distractors such as number of audience present, noise loudness, and amount of distracting movements. When fully implemented, the ML agent acts as a guide much like the chatbot therapist (Fitzpatrick, et al., 2017) as she determines user stress levels, previous training, and current speech nervousness patterns, and provides the right scene difficulty level for the systematic desensitization process.

We have shown that in a VR environment for training purposes, Machine Learning using sensor and speech inputs can be used to customize a solution for appropriate types of training that is essential for the gradual process of assimilation critical in, for example, speech therapy.

Figure 5. Example of a VR relaxation scene used to prepare users for subsequent exposure to speech situations. Sensor recordings during this scene is utilized by the ML agent to determine the appropriate level of difficulty in training.

Figure 6. The Galvanic Skin Response (GSR) sensor attached to fingers of a stuttering subject, used to determine stress levels that are used by the ML algorithm.

DISCUSSION

For those suffering from stress and anxiety issues like PTSD, stuttering, and phobias, the possibilities for convenient, low-cost, low-profile, context-specific treatment options have been scarce, but the advent of VR technology, in conjunction with customization by the way of Machine Learning, have made technological solutions possible.

In the clinical test regime, PTSD following September 11 attacks have been treated using exposure therapy in VR with significant reduction in clinical severity scores compared to waitlist controls (Difede, et al, 2007). Although effective in clinical settings, the promise for illness that can be treated with less drastic means in consultation with psychologists may prove to be a more useful use case for VR. For example, anxiety in public speaking have been treated with VR for university students, indicating reduced heart rate and Subjective Units of Stress Scale ratings over the course of therapy (Harris, et al., 2004). Moreover, skin resistance levels are reduced over the course of training in a 20 minute VR session for those who have phobia of flying (Wiederhold, et al., 2009). In both cases, measures of stress and anxiety shows efficacy of VR in treating disorders that do require covert intervention (like PTSD). This suggests that certain disorders that require more frequent treatment sessions without over-exposure may benefit from the convenience and efficacy of VR treatments.

These studies also didn’t take advantage the ability to tailor training during the process of treatment so that patients are getting ideal levels of exposure over time. So far, Machine Learning has only been used to detect stress levels rather than customize the treatment (Cho et al., 2017). ML promises to take therapeutic strategies to a new level of customization that comes with ability to sense bodily states. As proposed in our project, using GSR and other sensors of physiological activity in conjunction with detecting head movements in VR allows us to detect subjective stress levels, which allows customization of treatment strategies over the long haul, so that subtle differences in response to treatment can make a difference in the level of exposure in subsequent trials. Moreover, by changing the number and type of distractors in the treatment environment as well as the location of treatment itself, the exposure can be tailored to conditions like a specific speech at a company or a specific context like at a wedding or at a conference. The ability to change these parameters using ML is a hall mark of VR therapy that has yet to be taken advantage of until now.

At our final demo at an audience of designers and technologists, we found that the responsiveness of the Watson agent is critical to perceived intelligence of the engine. While the ML system doesn’t require the voice agent per se, the speed delay is due primarily to voice recognition. Reducing delay in future technology will be paramount to usability. We also found that users like to use their clothing and immediate items to indicate stress, such as stretching, grasping, and crumpling. Thus we have created prototypes to see if yarn fabric can be used as digital anxiety sensors for the VR system. In this project, we created fabric weaved sensors that uses separated conductive yarn material to detect human gestures like grasping, stretching, and stroking. This “Weavable Tech” strategy further increase the ability to detect subjective stress, providing additional sensor data for ML that can improve VR treatment over time.

Brundage, S. B., Brinton, J. M., and Hancock, A. B. “Utility of Virtual Reality Environments to Examine Physiological Reactivity and Subjective Distress in Adults Who Stutter.” *Journal of Fluency Disorders*. 2016: 10.

Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N. K., Lee, B. “Detection of Stress Levels from Biosignals Measured in Virtual Reality Environments Using a Kernel-Based Extreme Learning Machine.” *Sensors.* 2017: 17(10) 2435.

Davidson, G. C. “Systematic Desensitization as a Counterconditioning Process.” *Journal of Abnormal Psychology*. 1968: 73(2).

Difede, J., Cukor, J., Jayasinghe, N., Patt, I., Jedel, S., Spielman, L., Giosan, C., and Hoffman, H. G. “Virtual Reality Exposure Therapy for the Treatment of Posttraumatic Stress Disorder Following September 11, 2001.” *J. Clin. Psychiatry*. 2007: 68(11).

Fitzpatrick, K. K., Darcy, A., and Vierhile, M. “Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial.” *JMIR Mental Health*. 2017: 4(2).

Harris, S. R., Kemmerling, R. L., and North, M. M. “Brief Virtual Reality Therapy for Public Speaking Anxiety.” *CyberPsychology & Behavior*. 2004: 5(6).

Hebb, D. O. *The Organization of Behavior*. New York: Wiley. 1949.

Iverach, L., O’Brian, S., Jones, M., Block, S., Lincoln, M., Harrison, E., Hewat, S., Mezies, R. G., Packman, A., and Onslow, M. “Prevalence of Anxiety Disorders Among Adults Seeking Speech Therapy for Stuttering.” *Journal of Anxiety Disorders*. 2009: 23(7).

Kenjo, M. “Speech Therapy using Role Plays for a Child with Severe Stuttering.” *Japan Journal of Logopedics and Phoniatrics*. 2003: 44(2).

Lahiri, U. *Virtual-Reality Based Gaze-Sensitive Adaptive Response Technology for Children with Autism Spectrum Disorder*. Nashville: Vanderbilt University. 2011.

Minsky, M., and Papert, S. *Perceptrons: An Introduction to Computational Geometry*. Cambridge: MIT Press. 1969.

Russell, S., and Norvig, P. *Artificial Intelligence: A Modern Approach*. Prentice Hall: 2009, 3rd ed.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. “Learning Representations by Back-propagating Errors.” *Nature*. 1986: 323.

Wiederhold, B. K., Gevirtz, R., and Wiederhold, M. D. “Fear of Flying: A Case Report Using Virtual Reality Therapy with Physiological Monitoring.” *CyberPsychoogy & Behavior*. 2009: 1(2).