Adam Stammer
Technical Writing
Report Proposal
Feb. 23rd, 2018

Topic: Voice Synthesis

Game programming has been a hobby of mine for quite some time, and one of the most limiting aspects of an independent game, is voice narration. Finding a person with the right equipment, the right voice, and the right price is the bane of many indie developers. Voices synthesized by a computer is a way around this problem. I was researching neural nets when I found Deepmind's Wavenet, a machine learning based implementation of voice synthesis that is proving to be far more effective than any of the preexisting methods. It obviously has many other applications, but that is what first got me interested.

Voice synthesis has been used in many applications for decades, but it is constantly getting better. New methods are being developed right now using machine learning and the results are closer to human speech than ever before. Voice interfacing devices such as personal assistants (i.e. cortana, siri) could benefit greatly from more immersive person interactions. These technologies also greatly affect voice recognition, a form of computer interaction that is seeing increasing popularity, and likely will continue to. Other new technologies being developed may eventually be applied to voice synthesis as well.

I'm hoping to gain a better understanding of how computers perceive voice. I am most interested in the machine learning based methods, but an understanding of the older methods is helpful too. The wavenet neural network architecture has already been applied to other fields, like image generation/recognition, text generation, etc. This will only continue to expand as I near entering the workforce, and I do view machine learning as a potential career path. Understanding voice synthesis and recognition is something that I am likely to face eventually and having a good base understanding of it will certainly be helpful.