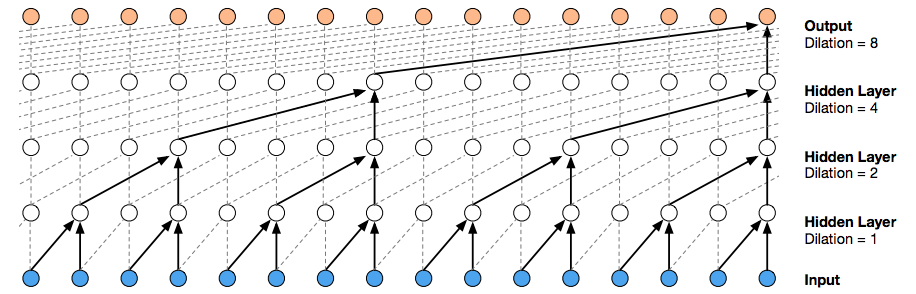
**Wavenet** is about generating audio waveforms for text-to-speech and more. Text-to-speech basically means that we have a voice reading whatever we have written down .The difference in this work is that it can synthesize some samples in someone's voice provided that we have training samples of this person speaking. It also generates waveforms sample by sample which is particularly perilous because we typically need to produce these at the rate of 16 or 24 thousand samples per second, and as we listen to the TV, radio and talk to each other several hours a day, the human ear and brain is particularly suited to processing this kind of signal .if the result is off by only the slightest amount, we immediately recognize it. It is not using a recurrent neural network, which is typically suited to learn sequences of things, and is widely used for sound synthesis. It is using a convolutional neural network, which is quiet surprising because it is not meant to the process sequences of the data that change in time. However, this variant contains an extension that is able to do that.



Training a convolution neural network is a walk in the park compared to a recurrent neural network. And the results beat all existing widely used techniques by a large margin. One of these is the concatenative technique, which builds sentences from a huge amount of small speech fragments and this have seen a ton of improvements during the years, but the output are still robotic and it is noticeable that we are not listening to human but a computer. The deep mind guys also report that: notice that non- space sounds, such as breathing and mouth movements, are also sometimes generated by wavenet , this reflects the greater flexibility of a row audio model. It also experimented with music generation, and the results are just stunning. I don’t know what to say. somewhere it also seen someone pitching the idea of creating audiobooks automatically with such a technique.it also reported that the algorithm currently takes 90 minutes to synthesize one second of sound waveforms.