WaveNet

WaveNet V1

- DeepMind splash in 2016

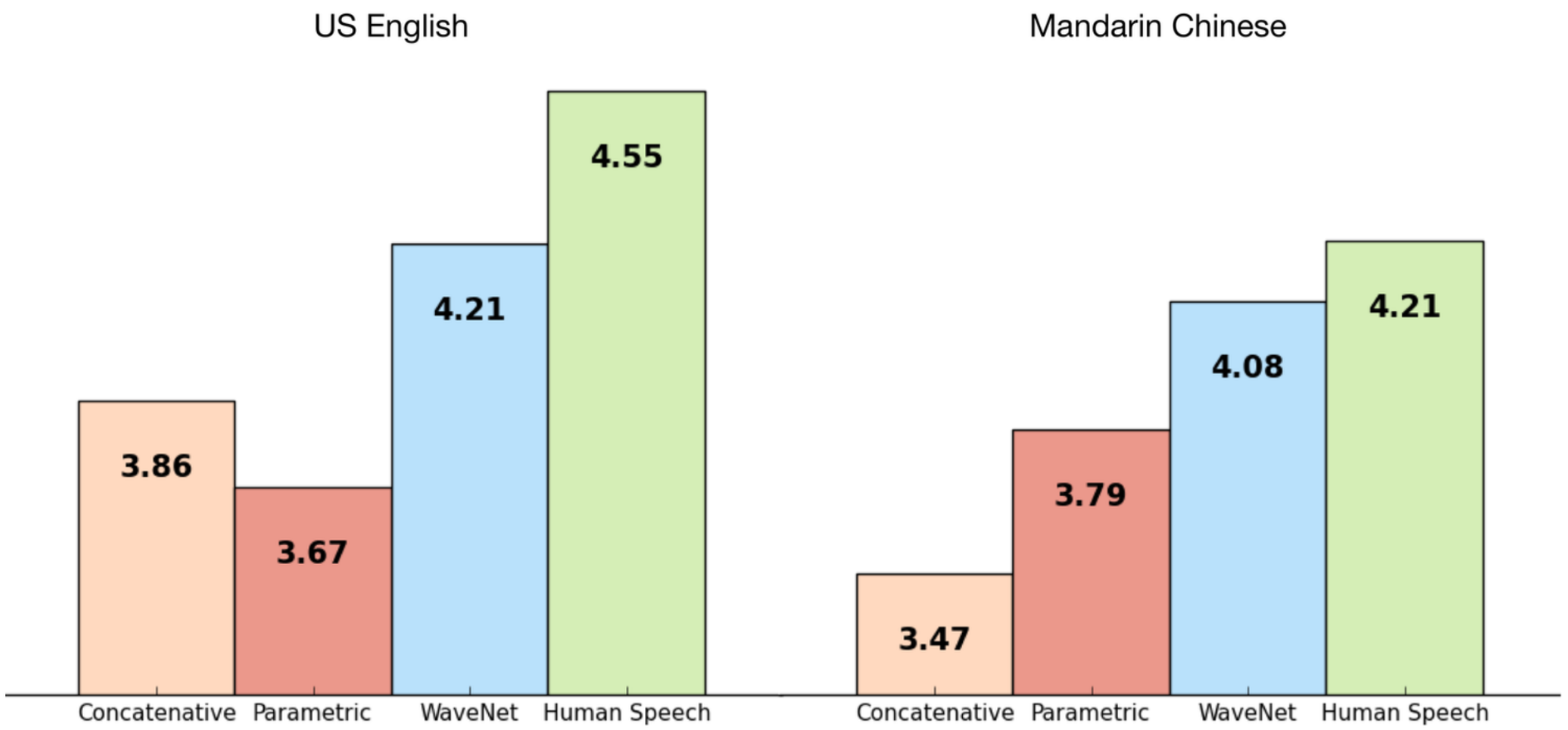


Image Source: Blog WaveNet: A Generative Model for Raw Audio

Let’s hear the difference

Blog WaveNet: A Generative Model for Raw Audio

- Parametric (RNN + LSTM + some staff)
- Concatenative (HMM)
- WaveNet

Audio Samples from Network

- Data output:
  - 16 KHz (24 KHz) rate with 8-bit (16-bit) resolution

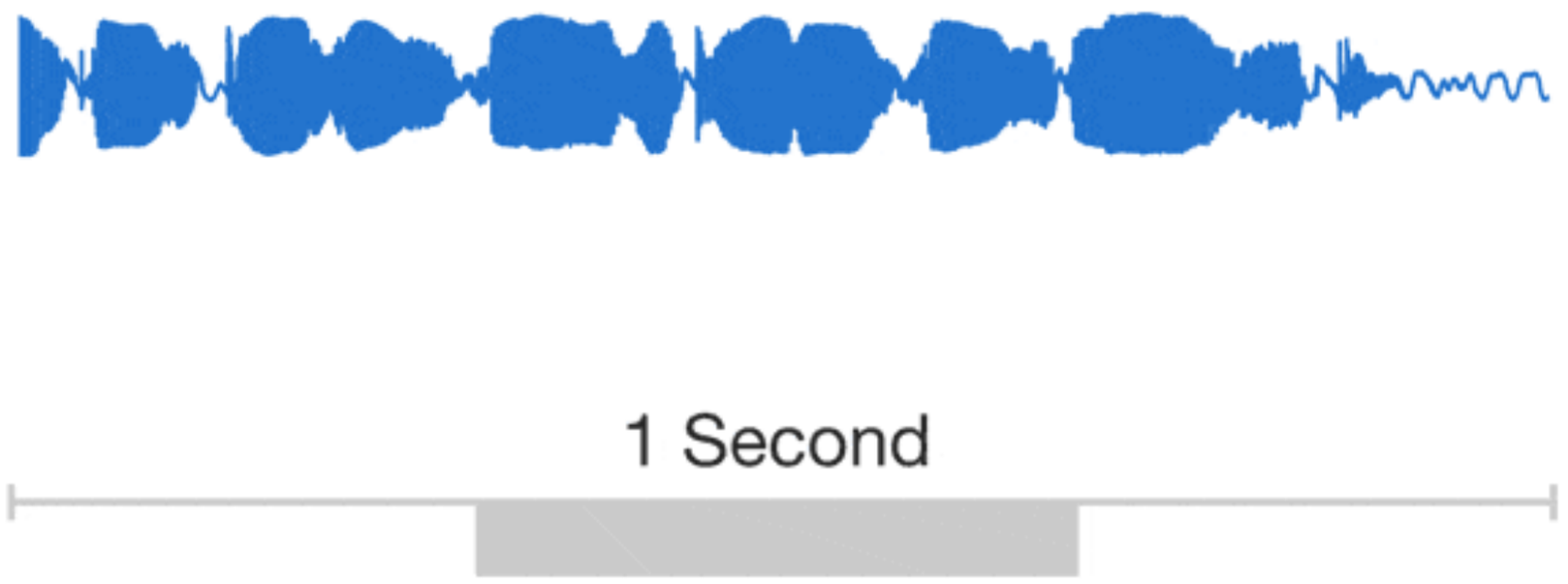


Image Source: Blog WaveNet: A Generative Model for Raw Audio

CNN instead of RNN

Note the «Linear Footprint»

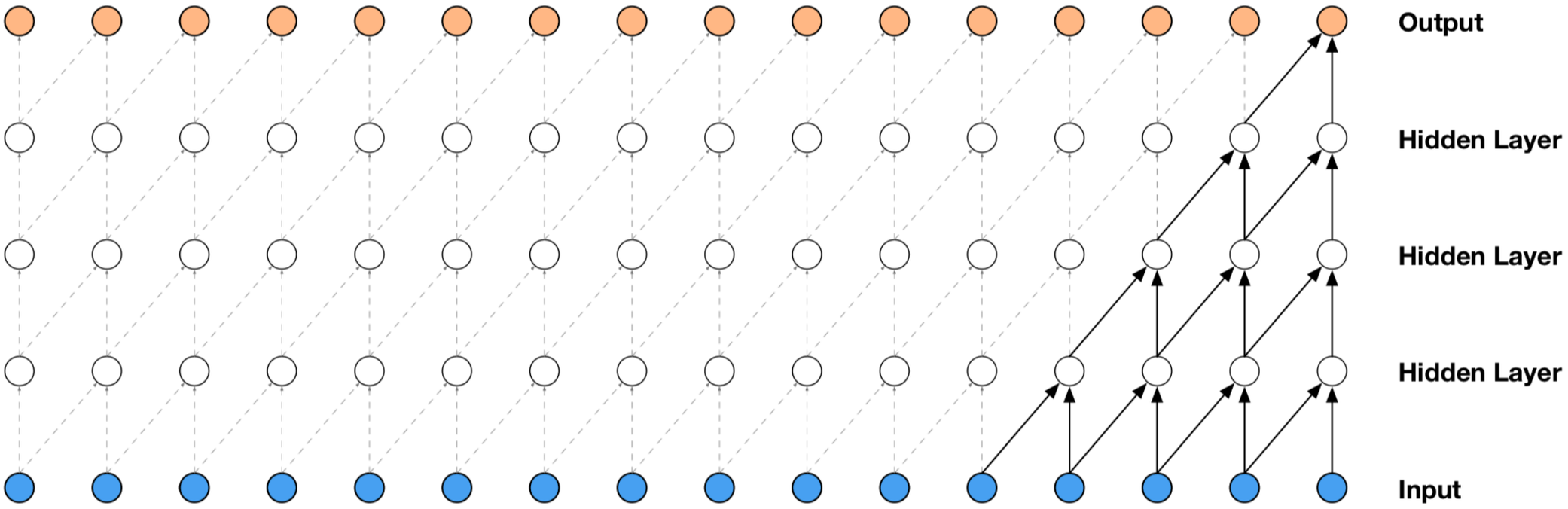


Image Source: Paper WaveNet: A Generative Model for Raw Audio

Dilated CNN

Note the «Exponential Footprint»

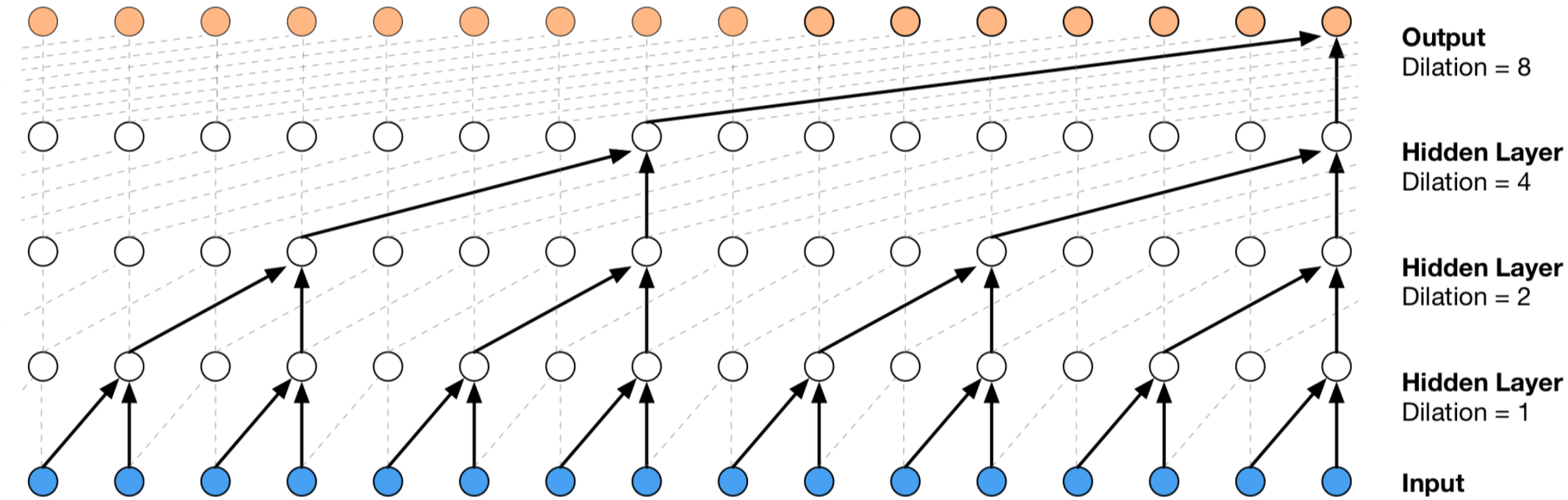
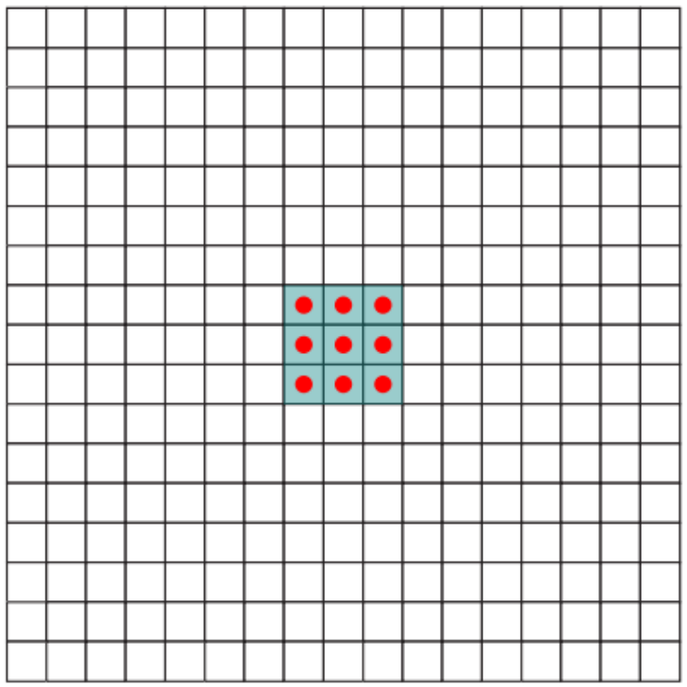


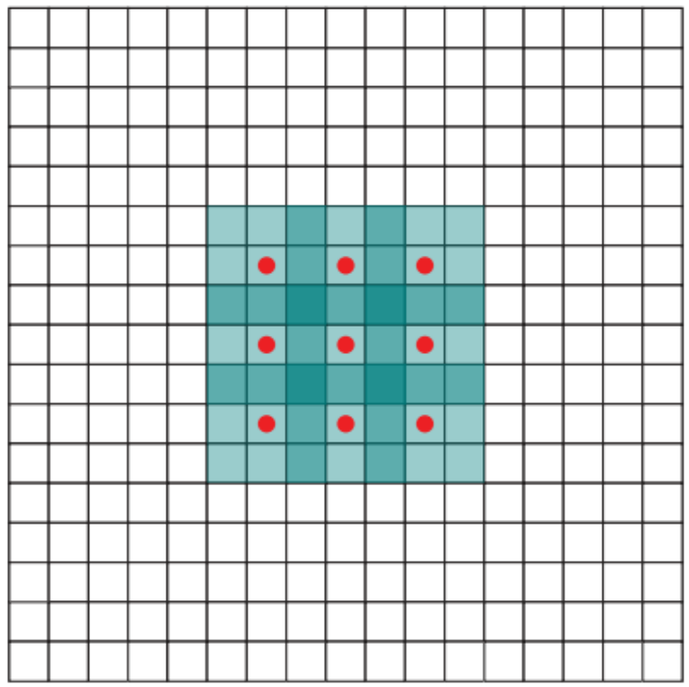
Image Source: Paper WaveNet: A Generative Model for Raw Audio

Dilated CNN

dilation rate (or factor) of 1



dilation rate (or factor) of 2



dilation rate (or factor) of 4

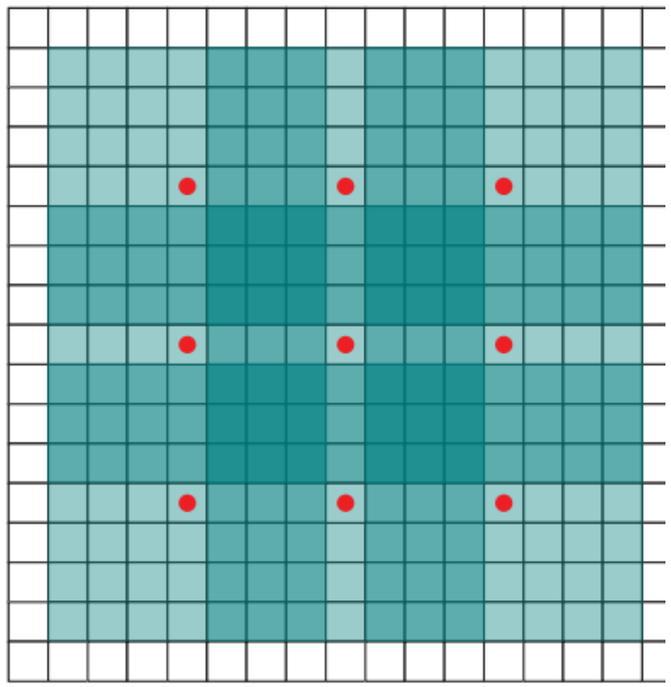
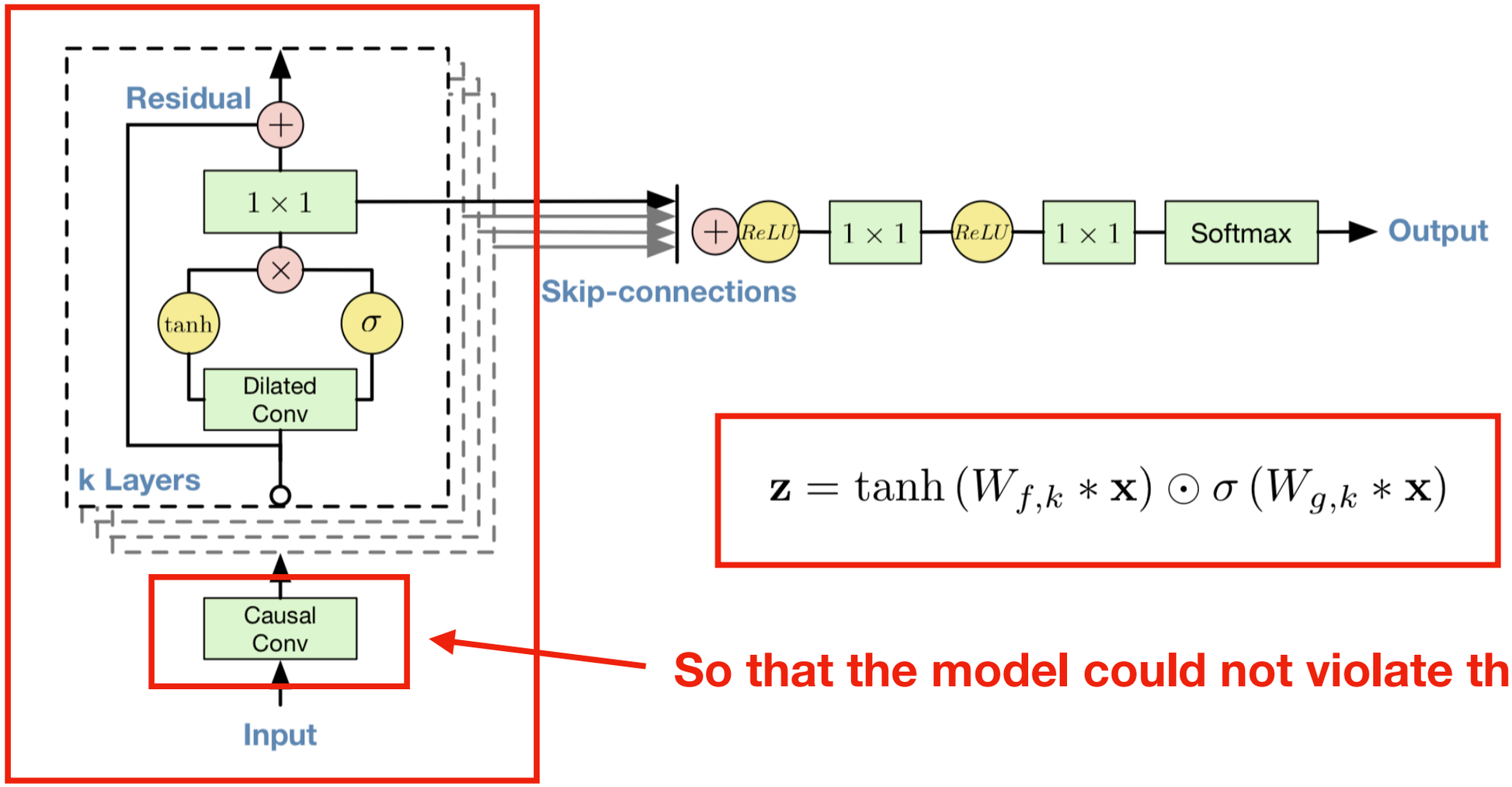


Image Source: http://sergeiturukin.com/2017/03/02/wavenet.html

CNN PRO/CON

- ADVANTAGES:
  - In general CNNs are faster to train
  - The «look back» is exponential
- DISADVANTAGES:
  - No «next sample» scheme

Sigmoid Gate of Tanh Units



$$\mathbf{z} = \tanh\left(W_{f,k} * \mathbf{x}\right) \odot \sigma\left(W_{g,k} * \mathbf{x}\right)$$

So that the model could not violate the ordering

Each CNN node includes Gating and ResNet

Output of Distribution

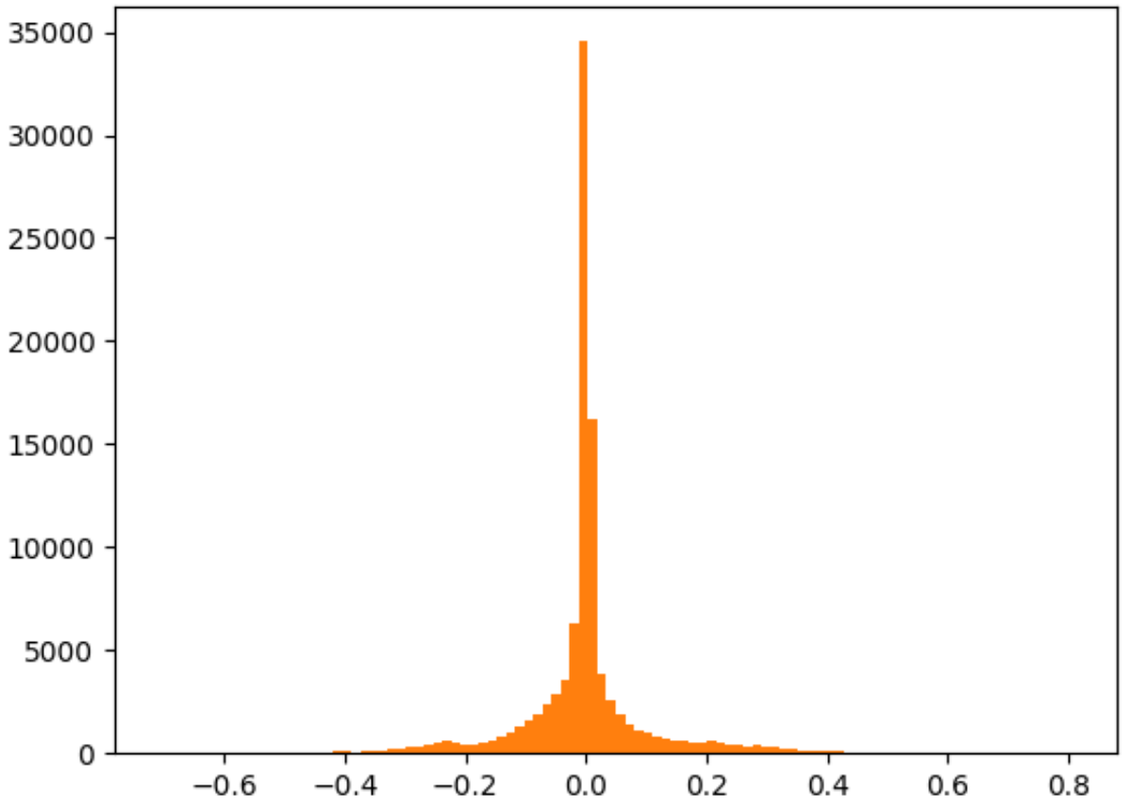
- No pooling layers — the output has the same time dimensionality as the input.
- Output — categorical distribution for each time step.

$$F(x) = \operatorname{sgn}(x)\frac{\ln(1+\mu|x|)}{\ln(1+\mu)} \qquad -1 \le x \le 1$$

μ-law

Output of Distribution

Before μ-law



After μ-law

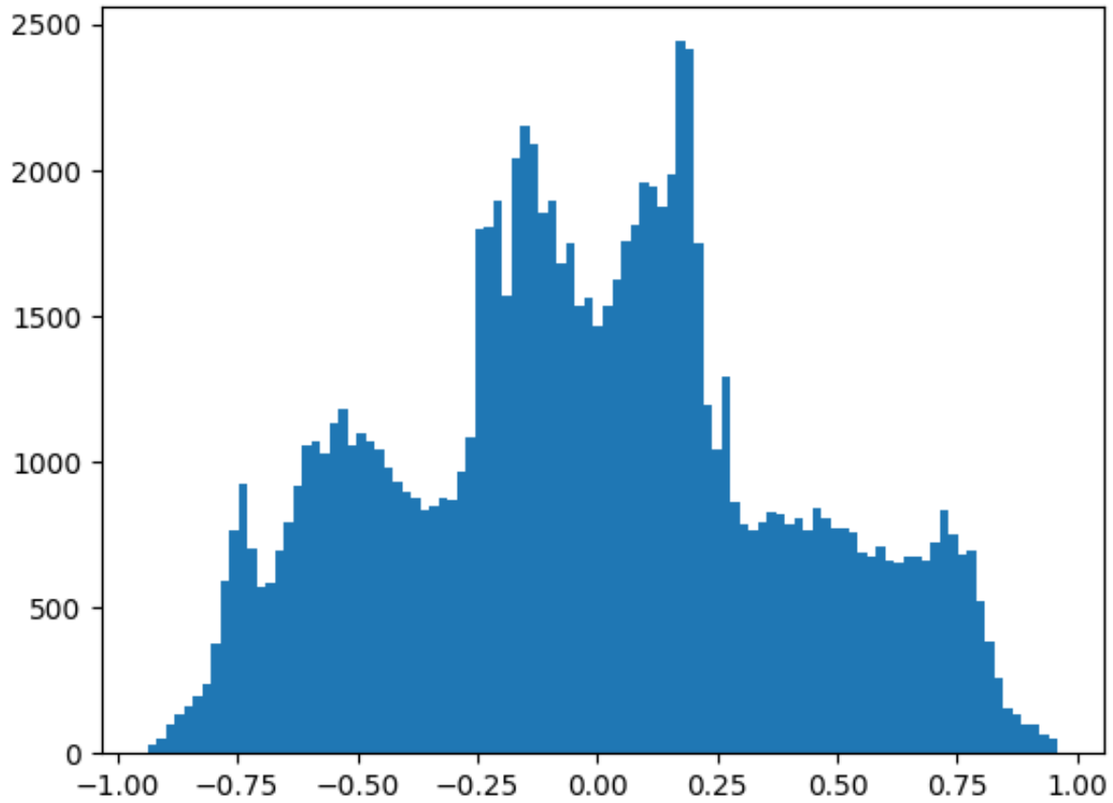


Image Source: http://sergeiturukin.com/2017/03/02/wavenet.html

Computational Burden

- Training can be QUICK:
  - Parallel training at all time steps.
- Actual running is SLOW:
  - 1 sec of output = 1-2 minute of GPU

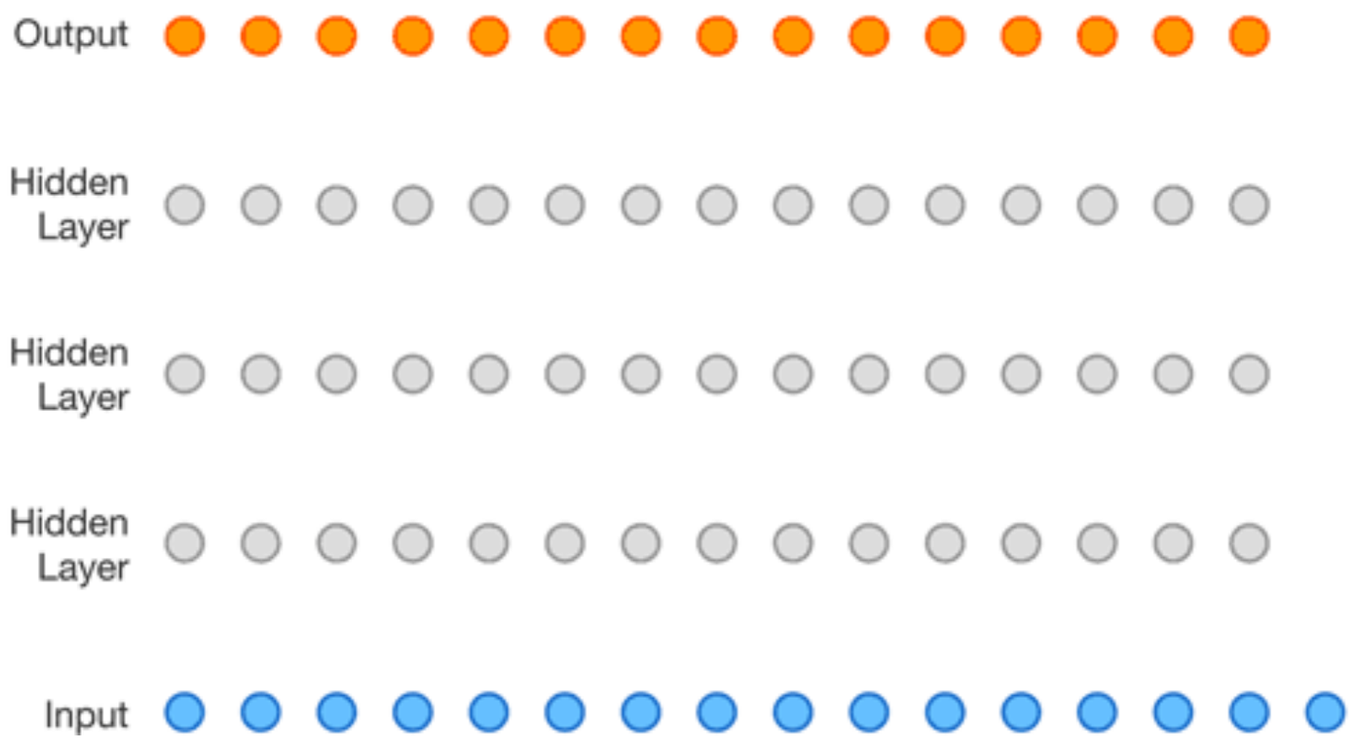


Image Source: Blog WaveNet: A Generative Model for Raw Audio

Fast WaveNet

Use queues for smarter calculations

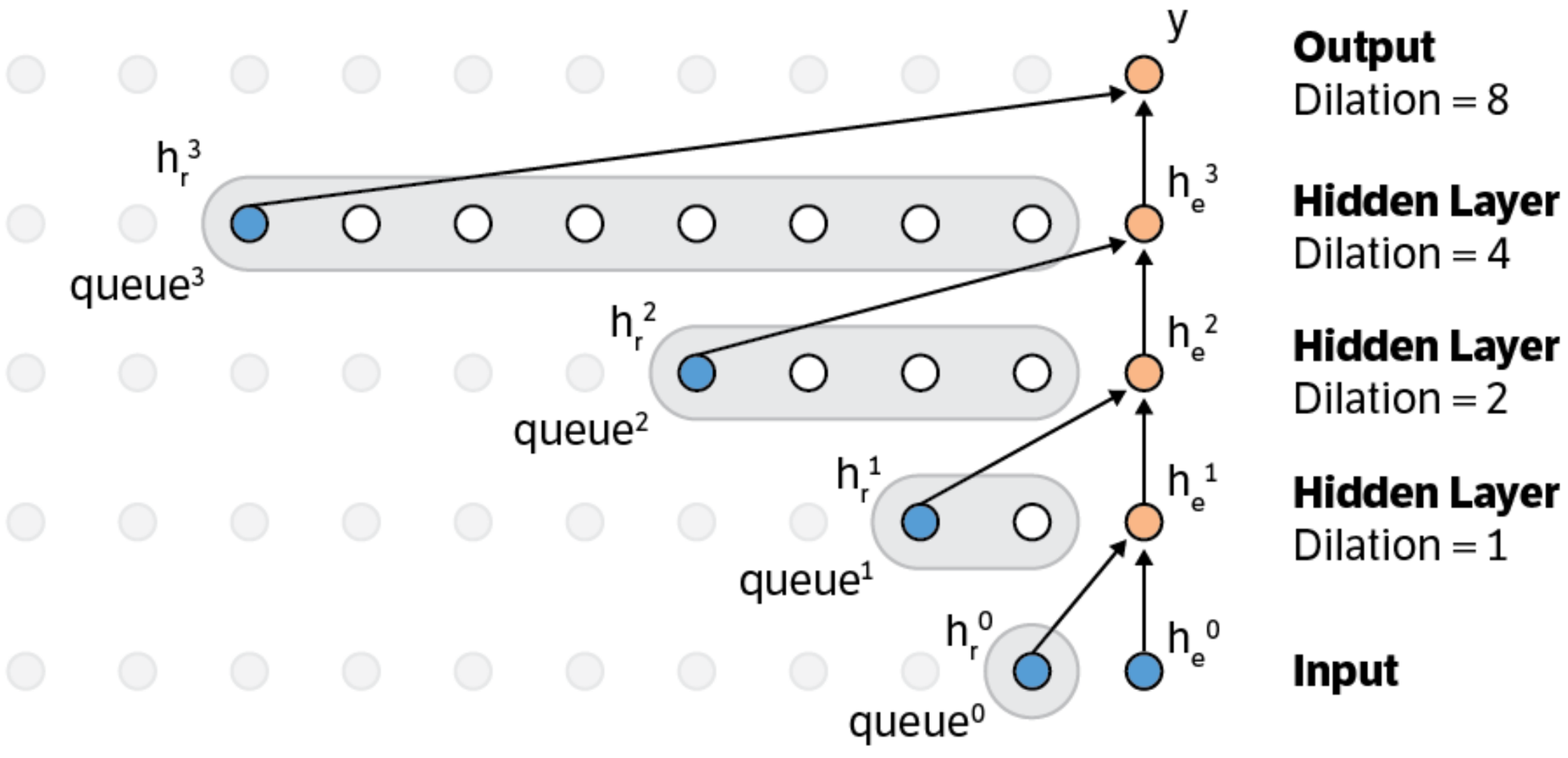


Image Source: https://github.com/tomlepaine/fast-wavenet

Parallel WaveNet

Image Source: https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/

Parallel WaveNet

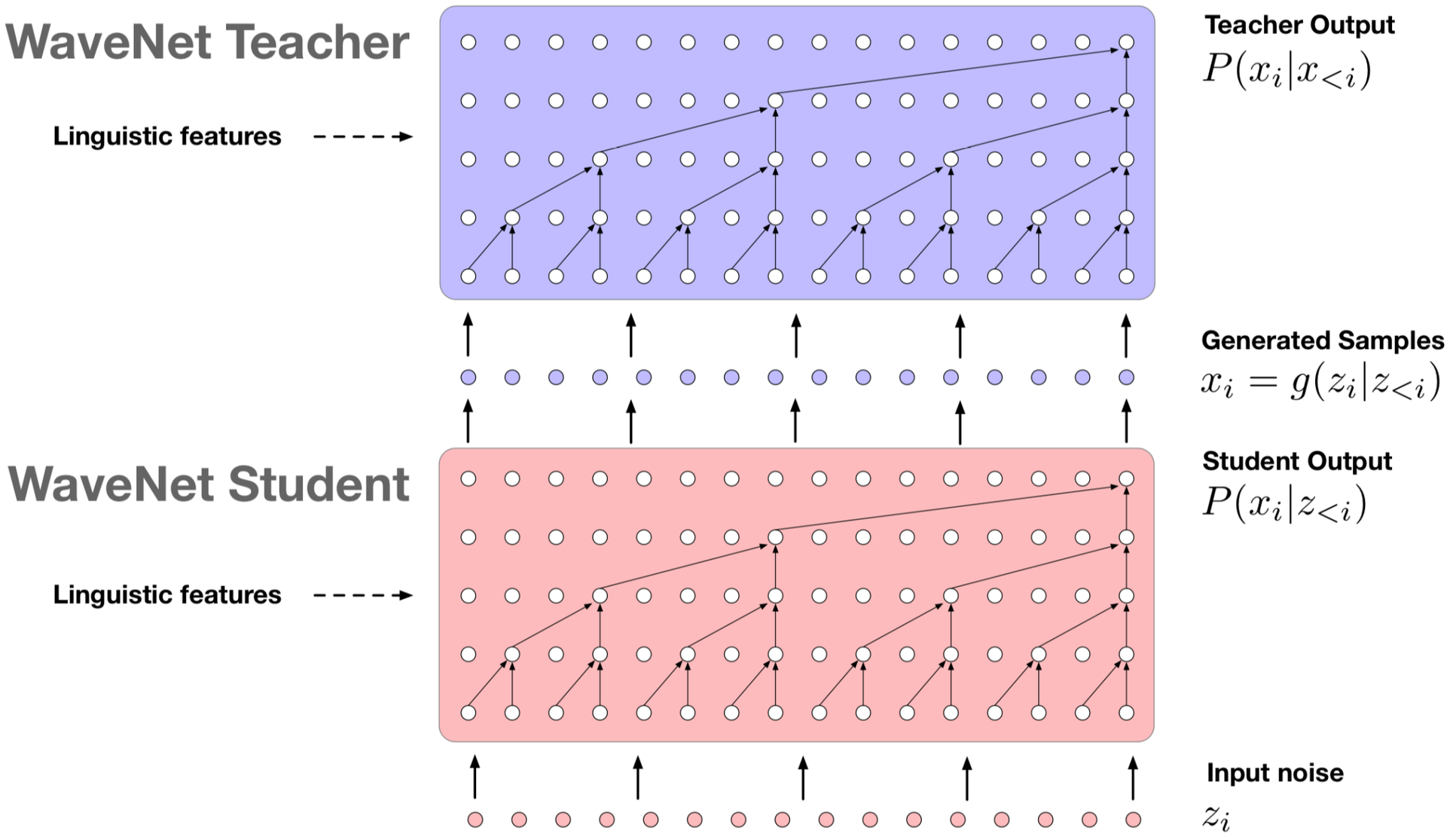


Image Source: https://arxiv.org/pdf/1711.10433.pdf

Now It Can Be Used

Image Source: https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/

WaveNet launches in the Google Assistant

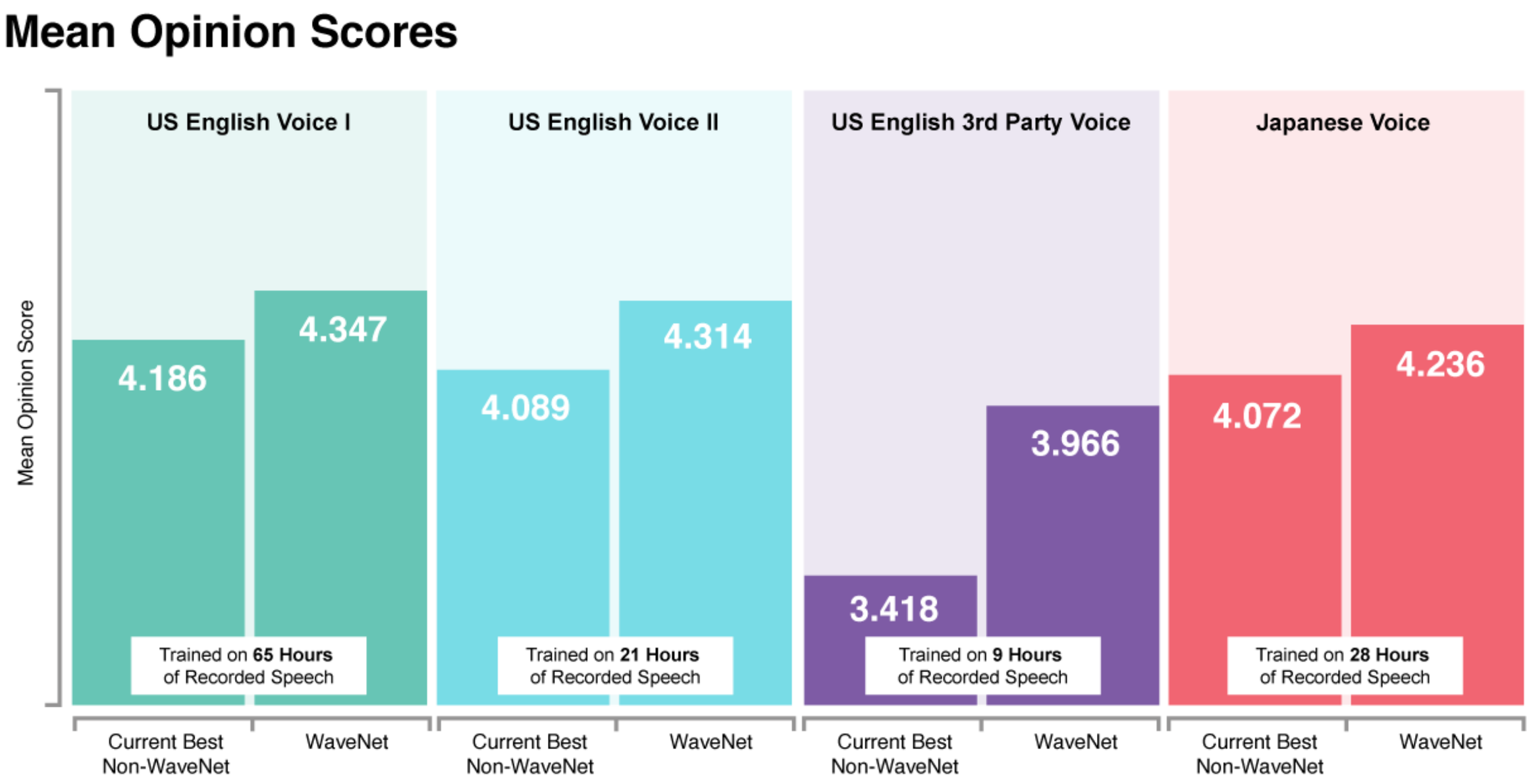


Image Source: https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/

NSynth: Neural Audio Synthesis

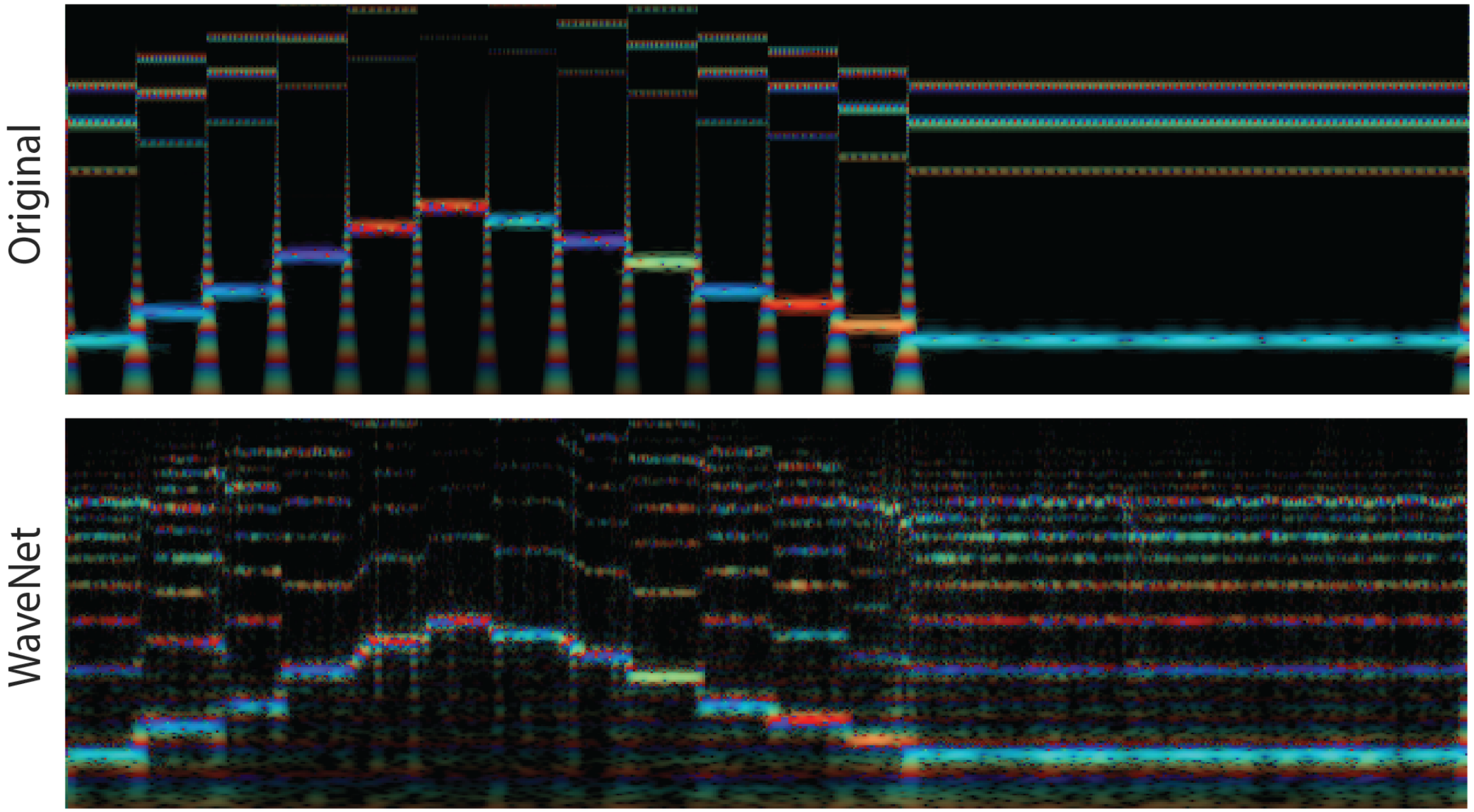


Image Source: https://magenta.tensorflow.org/nsynth

Sample RNN

- WAVENET: A GENERATIVE MODEL FOR RAW AUDIO
- FAST WAVENET GENERATION ALGORITHM
- Parallel WaveNet: Fast High-Fidelity Speech Synthesis
- https://deepmind.com/
- Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders
- NSynth: Neural Audio Synthesis