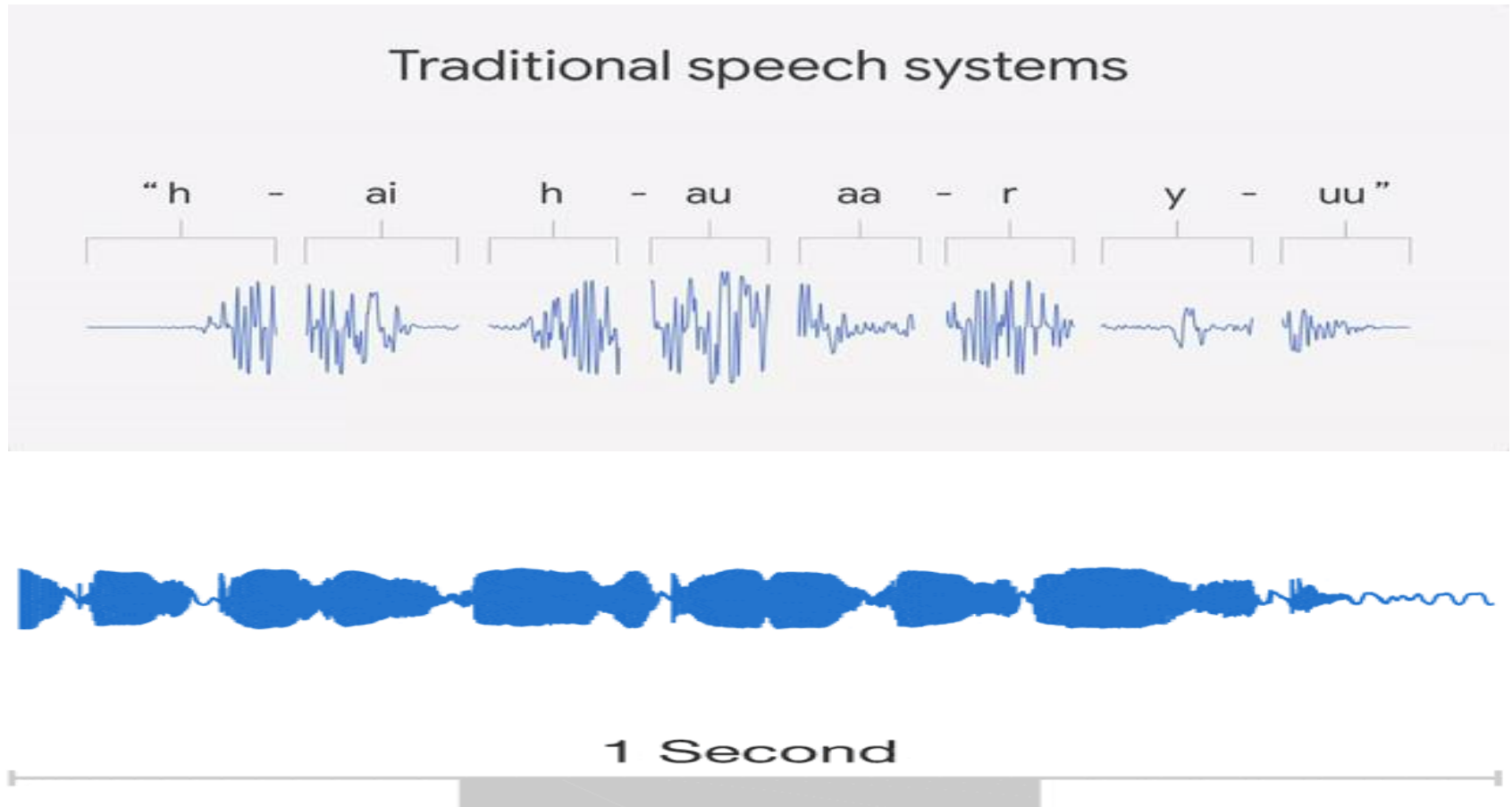


Google WaveNet

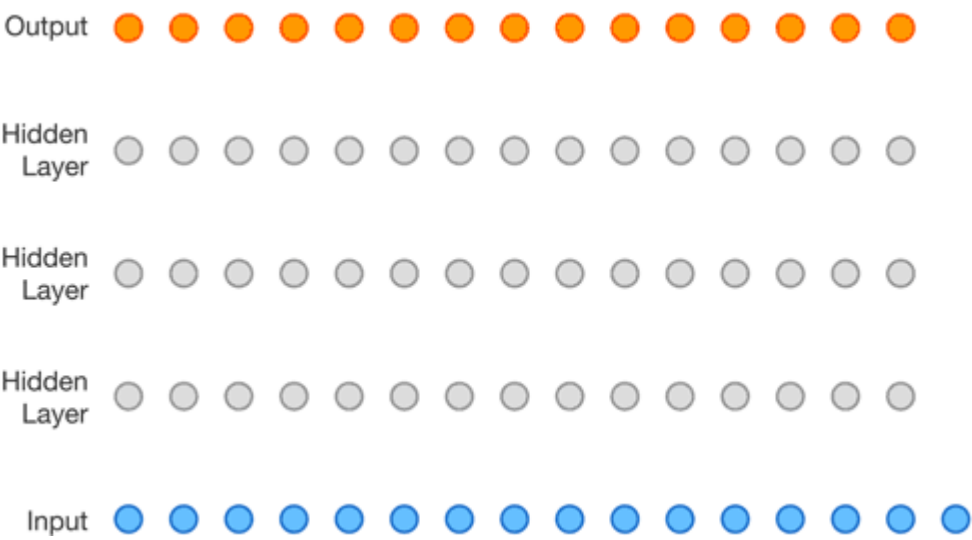


WaveNet is a powerful **predictive technique** that uses multiple Deep Learning (DL) strategies from Computer Vision (CV) and Audio Signal Processing models and applies them to **longitudinal (time-series) data**.

Google WaveNet: A generative model for raw audio, 2016

deepmind.com/blog/article/wavenet-generative-model-raw-audio

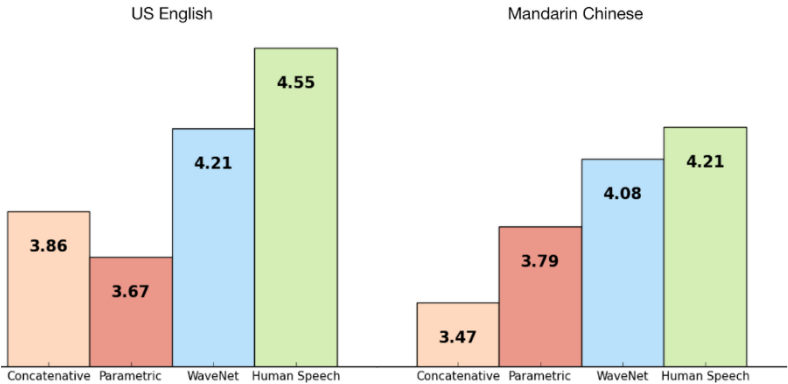
Google WaveNet mit einer CNN-Architektur



A **CNN** that **grows**. It was **trained** on a **large dataset of speech samples**. During training, the network determined the underlying **structure** of the speech and what **waveforms** were **realistic**. The trained network then **synthesised a voice** one sample at a time, with each generated sample taking into account the properties of the previous sample. The resulting voice contained **natural intonation** and other features such as **lip smacks**.

Speech samples	Subjective 5-scale MOS in naturalness	
	North American English	Mandarin Chinese
LSTM-RNN parametric	3.67 ± 0.098	3.79 ± 0.084
HMM-driven concatenative	3.86 ± 0.137	3.47 ± 0.108
WaveNet (L+F)	4.21 ± 0.081	4.08 ± 0.085
Natural (8-bit μ -law)	4.46 ± 0.067	4.25 ± 0.082
Natural (16-bit linear PCM)	4.55 ± 0.075	4.21 ± 0.071

Table 1: Subjective 5-scale mean opinion scores of speech samples from LSTM-RNN-based statistical parametric, HMM-driven unit selection concatenative, and proposed WaveNet-based speech synthesizers, 8-bit μ -law encoded natural speech, and 16-bit linear pulse-code modulation (PCM) natural speech. WaveNet improved the previous state of the art significantly, reducing the gap between natural speech and best previous model by more than 50%.



Google WaveNet: A generative model for raw audio, 2016

deepmind.com/blog/article/wavenet-generative-model-raw-audio

Google WaveNet



Google Wavenet | Google I/O 2018
www.youtube.com/watch?v=JjK8apEishQ