



**Voicewriter**

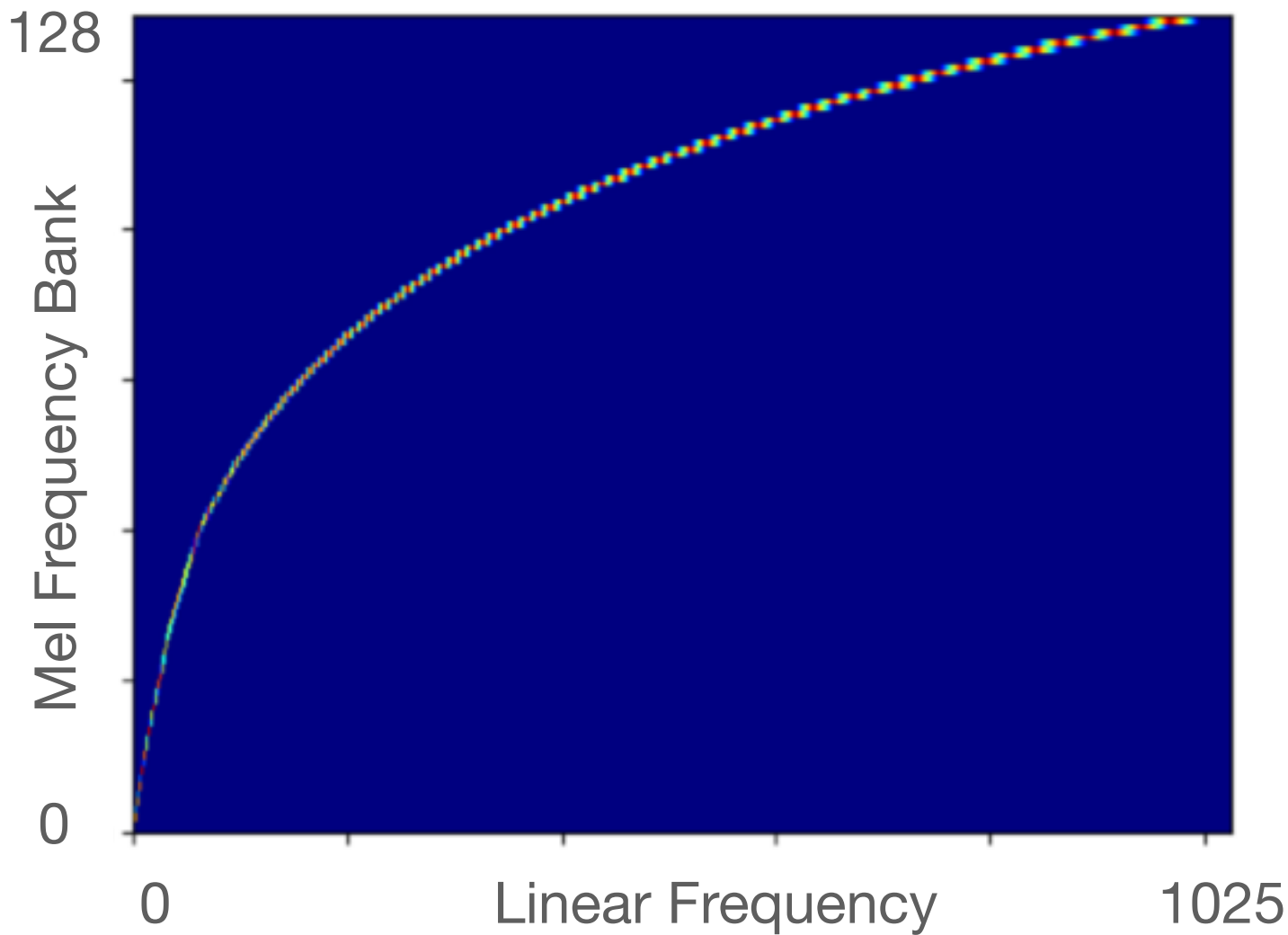
**Why using mel spectrogram?**

• **Minimizing human auditory sensitivity**

- High resolution at low frequency and low resolution at high frequency

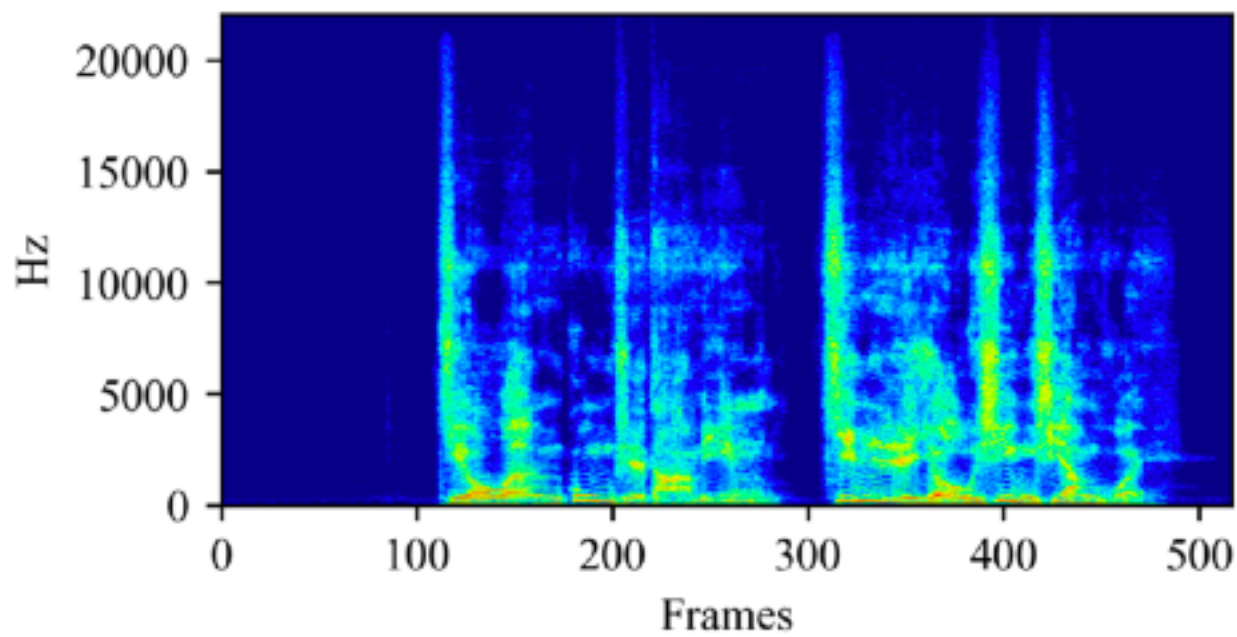
• **Low-dimensional feature**

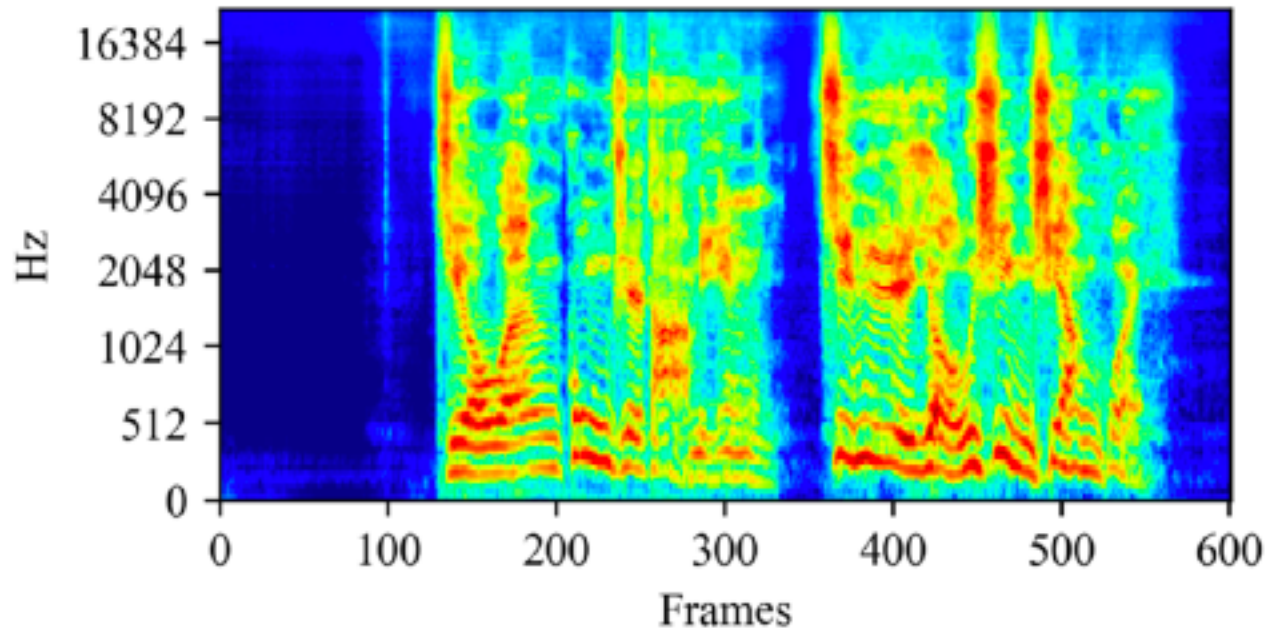
• Linear frequency 1025 Mel Frequencies 128











# LogScaleLinearSpectrogram

# LogscaleMelSpectrogram

resolution at high frequency

• High resolution at low frequency and low

frequency



- Reduce the dimension of feature space

• High resolution at low frequency and low resolution at high

• Linear frequency 1025 Mel Frequencies 128

• **Minimizing human auditory sensitivity**

Testset	Vocoder (TFGAN)	Original
ALL-GSR	3.74	3.95
VCTK-Demand	3.64	3.69
VCTK-Test	3.47	3.51

Almost the same on paper as quality!

Table 22 Most common threat evaluation set