

Voicexer

Why using mel spectrogram?

• **Minimizing human auditory sensitivity**

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

- Reduce the dimension of feature space

• Linear frequency 1025 Mel Frequencies 128

• **The advancement in investment program**

• Neural Vocoder



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

VoiceFixer

Why using mel spectrogram?

- Mimic human auditory sensitivity
 - High resolution at low frequency and low resolution at high frequency
- Reduce the dimension of feature space
 - Linear frequency 1025 → Mel Frequencies 128
- **The advancement in inverse mel spectrogram**
 - Neural Vocoder

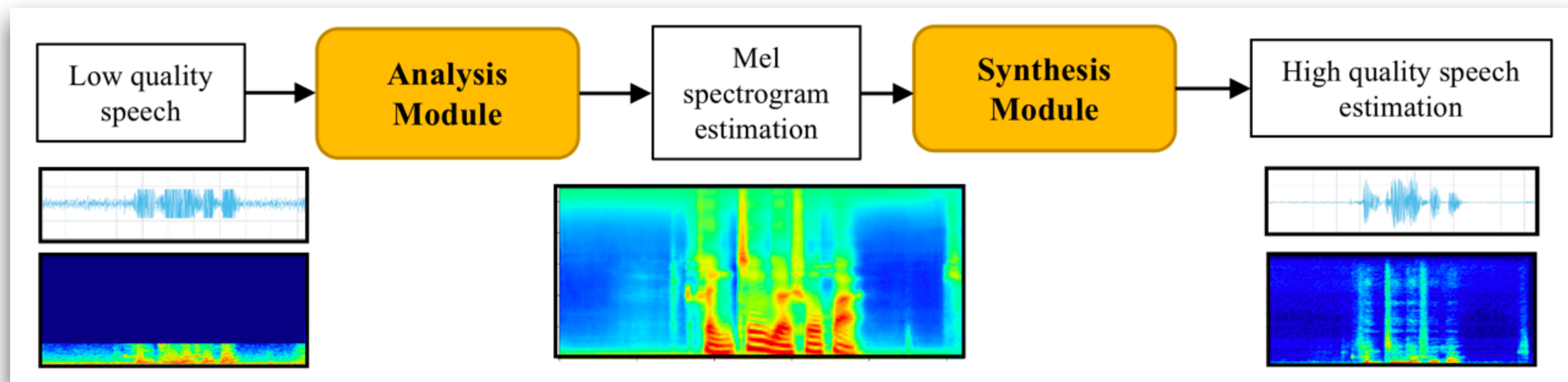
Table.2 MOS score on three evaluation set

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 perceptual quality!

VoiceFixer

Analysis Module



$$\left| \hat{S} \right|_{mel} = F(\left| X \right|_{mel}) \odot \left| X \right|_{mel}$$

$$L_{ana} = \left\| \left| \hat{X} \right|_{mel} - \left| S \right|_{mel} \right\|_1$$