(a) Ground truth   (b) RFWave   (c) RFWave-Universal

(d) BigVGAN-base   (e) BigVGAN   (f) BigVGAN-v2

(g) Vocos   (h) PriorGrad   (i) FreGrad

Figure A.5: Examples of spectrograms from the MUSDB18 HQ dataset, specifically from Al James - Schoolboy Fascination vocals, illustrate the differences between GAN-based and diffusion-based methods. GAN-based methods tend to produce horizontal lines in the high-frequency regions of the spectrogram, resulting in unnatural artifacts. In contrast, diffusion-based methods, particularly our RFWave, generate more natural and coherent high-frequency spectrograms. Notably, except for RFWave-Universal and BigVGAN-v2, which were trained on a mixed dataset, the other models were trained on LibriTTS and are being evaluated in an out-of-distribution test.