VoiceFixer

TFGAN Vocoder - Training - Time Domain losses

- Loss Function: $L_{syn} = L^T + L^F + \lambda_1 L^D$
- Time Domain Losses:
  - $L^T = \sum_k L_k^t$
  - $L_k^t(\hat{s}, s) = \lambda_5 L_k^{energy}(\hat{s}, s) + \lambda_6 L_k^{phase}(\hat{s}, s) + \lambda_7 L_k^{time}(\hat{s}, s)$
- **Time loss, energy loss and phase loss:**
  - Loss on time sample: $L^{time}(\hat{s}, s) = \left\| v(\hat{s}) - v(s) \right\|_1$
  - Capture energy information: $L^{energy}(\hat{s}, s) = \left\| v(\hat{s}_w^{\,2}) - v(s_w^2) \right\|_1$
  - Remove metallic effect: $L^{phase}(\hat{s}, s) = \left\| \Delta v(\hat{s}_w^{\,2}) - \Delta v(s_w^2) \right\|_1$,

$v(s)_{1\times w} = (m(s_0), m(s_1), \ldots, m(s_w))$, $m(\cdot)$ is the mean function

Table.3 Windowing parameter for each k

k	1	2	3	4
frame-length	1	240	480	960
hop-length	1	120	240	480

VoiceFixer

TFGAN Vocoder - Training - Frequency Domain losses

Table.4 STFT parameter for each k

k	1	2	3	4	5	6	7
win-length	4096	2048	1024	512	256	128	64
hop-length	2048	1024	512	256	128	64	32
fft-size	8192	4096	2048	1024	512	256	128