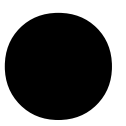
VoiceFixer

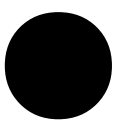
TFGAN Vocoder - Training - Frequency Domain losses



Loss

Function:

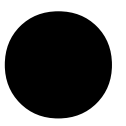
$$L_{syn} = L^T + L^F + \lambda_1 L^D$$



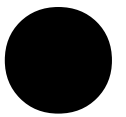
Frequency

Domain

Losses:



$$L^F = \lambda_2 L^{mel} + \sum_k L_k^f$$



$$L_k^f(\hat{s}, s) = \lambda_3 L_k^{sc}(\hat{s}, s) + \lambda_4 L_k^{mag}(\hat{s}, s)$$

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

Training

Vocoder


TFGAN

Domain

Frequency

losses

spectral


mel

loss,


loss


Mel

magnitude

on


loss:

convergence

- Capture

and

information:

Loss

domain

log


Loss

scale:


on

linear


scale:

$$L^{mel}(\hat{s}, s) = \left\| \, |\hat{S}|_{mel} - |S|_{mel} \, \right\|_2$$

$$L^{sc}(\hat{s}, s) = \frac{\left\| \, |\hat{S}| - |S| \, \right\|_F}{\left\| \, |\hat{S}| \, \right\|_F}$$

$$L^{mag}(\hat{s}, s) = \left\| log(|\hat{S}|) - log(|S|) \right\|_1,$$

- Loss

- Frequency

$$L_{syn} = L^T + L^F + \lambda_1 L^D$$

$$L^F = \lambda_2 L^{mel} + \sum_k L_k^f$$

$$L_k^f(\hat{s}, s) = \lambda_3 L_k^{sc}(\hat{s}, s) + \lambda_4 L_k^{mag}(\hat{s}, s)$$

parameter


STFT

Table.4


each

for

Losses:


Domain

Function:

- Loss

- Frequency