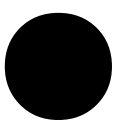
VoiceFixer

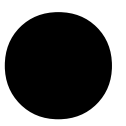
TFGAN Vocoder - Training - Time Domain losses



Loss

Function:

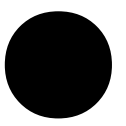
$$L_{syn} = L^T + L^F + \lambda_1 L^D$$

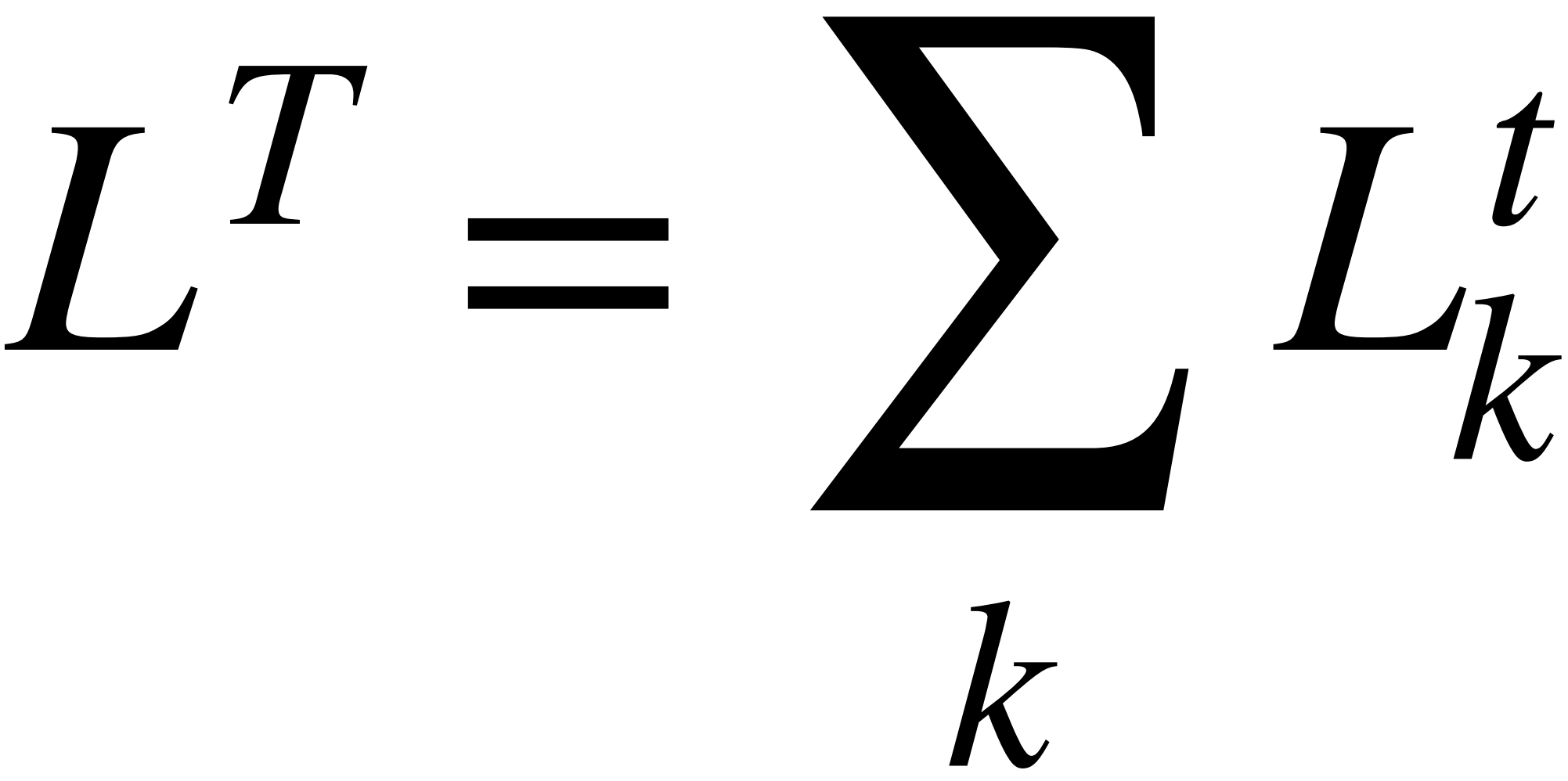


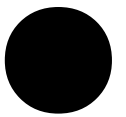
Time

Domain

Losses:







$$L_k^t(\hat{s}, s) = \lambda_5 L_k^{energy}(\hat{s}, s) + \lambda_6 L_k^{phase}(\hat{s}, s) + \lambda_7 L_k^{time}(\hat{s}, s)$$

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

Table.3 Windowing parameter for each k

Domain


TFGAN


Time

Training

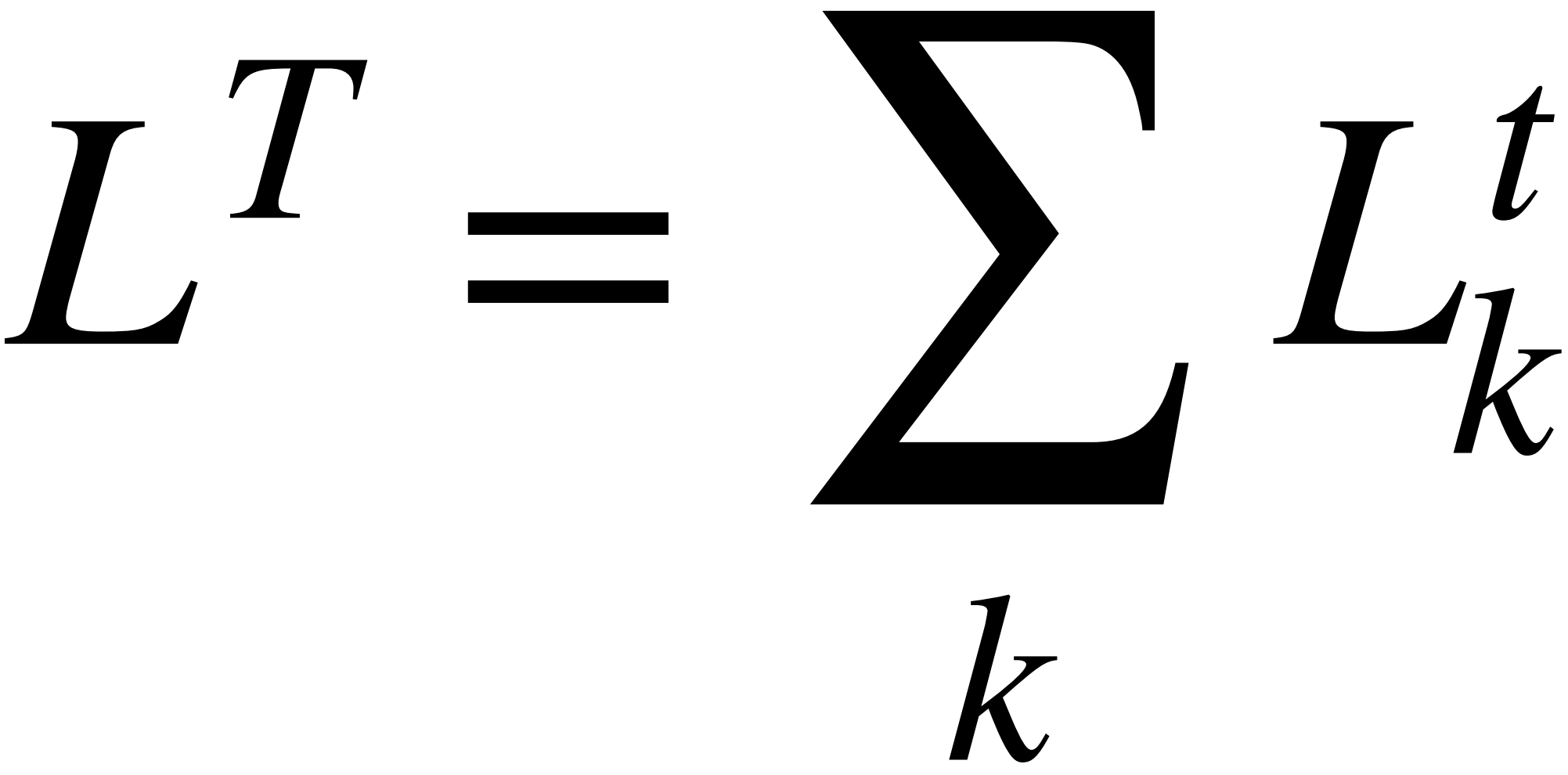
losses

Vocoder

- Loss

- Time

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



$$L_k^t(\hat{s}, s) = \lambda_5 L_k^{energy}(\hat{s}, s) + \lambda_6 L_k^{phase}(\hat{s}, s) + \lambda_7 L_k^{time}(\hat{s}, s)$$


energy


and

loss:

o Loss

on

- Time

loss,


phase


loss


time


effect:

is

the


energy

metallic

- Capture

mean

- Remove

sample:

information:

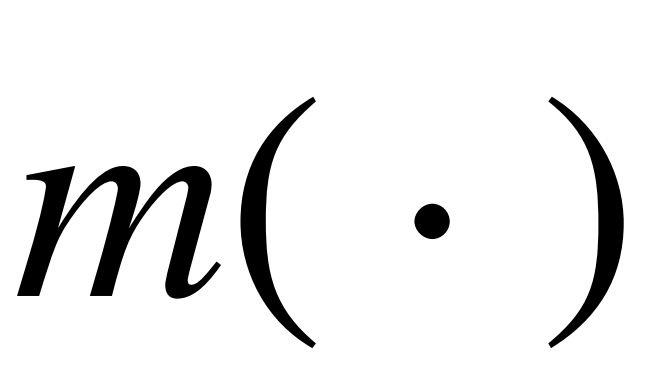
function

$$L^{time}(\hat{s}, s) = \left\| v(\hat{s}) - v(s) \right\|_1$$

$$L^{energy}(\hat{s}, s) = \left\| v(\hat{s}_w^{\,2}) - v(s_w^2) \right\|_1$$

$$L^{phase}(\hat{s}, s) = \left\| \Delta v(\hat{s}_w^{\,2}) - \Delta v(s_w^2) \right\|_1,$$

$$v(s)_{1\times w} = (m(s_0), m(s_1), \ldots, m(s_w))$$



parameter

each

for

Windowing

Table.3

Domain

Function:

Losses:

- Loss

- Time