Scale-Invariant Signal to Distortion Rate (SI-SDR)

Scale-Invariant SDR: $SI\text{-}SDR = 10 \log_{10} \left(\dfrac{\left| \frac{\tilde{y}_{t,k}^T y_{t,k}}{|y_{t,k}|} y_{t,k} \right|^2}{\left| \frac{\tilde{y}_{t,k}^T y_{t,k}}{|y_{t,k}|^2} y_{t,k} - \tilde{y}_{t,k} \right|^2} \right)$

Downscale-Dependent Signal to Distortion Ration (DsDSDR): $SNR = 10 \log_{10} \left(\dfrac{|y_{t,k}|^2}{|y_{t,k} - \tilde{y}_{t,k}|^2} \right)$

$$L_{down} = SNR + 10 \log_{10} \left(\frac{\tilde{y}_{t,k}^T y_{t,k}}{|y_{t,k}|^2} \right)^2$$

$$SDSDR = \min(SNR, L_{down})$$

By Heitkaemper ~ "Demystifying tasnet"; $SI\text{-}SDR = L_{og} - L_2 = 10 \frac{1}{k} \sum_k \log_{10} \sum_t |y_{t,k} - \tilde{y}_{t,k}|^2$

"The main idea behind compressing the loss with a log function is that distortions with low amplitude may be perceptually unpleasant as high amplitude ones. With compression, the model is encouraged to focus on low amplitude distortion as well."

Signal Noise Ratio: $SDR = \dfrac{P_{signal}}{P_{noise}} = \left(\dfrac{A_{signal}}{A_{noise}} \right)^2$

$$SNR_{dB} = 10 \log_{10}(SNR)$$