**Dataset info:**

**Total samples:** 24,444,800

**Duration:** 30,556 seconds

**Sample rate:** 8000 Hz

**Frame parameters calculation:**

**frame\_length =** 0.032 × 8000 = 256 samples (32ms window)

**hop\_length\_frame = frame\_length/2 =** 128 samples (50% overlap)

**hop\_length\_frame\_noise =** 128 samples (matching frame hop)

**nb\_samples calculation:**

**total\_frames = (total\_samples - frame\_length) / hop\_length\_frame**

= (24,444,800 - 256) / 128 ≈ 190,972 possible frames

**recommended\_nb\_samples =** 15000 (about 8% of total frames)

**Standard frame size for speech processing**

**50% overlap for good reconstruction**

**nb\_samples=15000 provides:**

**Sufficient training data**

**Manageable training time**

**Good representation of dataset**

**~12000 training samples (80%)**

**~3000 validation samples (20%)**