Mathematical Necessity of Fixed Sampling Rate in Quantum-Audio Pipelines

1 Introduction

The integrity of the processing pipeline—comprising the Short-Time Fourier Transform (STFT), padding, Continuous Wavelet Transform (CWT), and Quantum Walk evolution—relies fundamentally on the assumption that all audio segments inhabit the same vector space.

For the pipeline to function correctly, all generated spectrograms must possess an identical number of frequency bins and temporal frames. This uniformity is mathematically impossible unless all input audio is sampled at a strictly identical rate prior to the STFT operation.

This document outlines the four technical constraints that mandate a fixed sampling rate (f_s).

2 STFT Frequency Domain Consistency

The frequency resolution (Δf) of an STFT spectrogram is derived from the sampling rate and the FFT window size (N_{FFT}). It is defined as:

$$\Delta f = \frac{f_s}{N_{FFT}} \tag{1}$$

If the sampling rate f_s varies between segments:

- The spacing between frequency bins (Δf) changes.
- The k-th bin in one segment represents a different physical frequency (in Hz) than the k-th bin in another segment.

Consequently, the segments effectively inhabit different frequency axes. It becomes invalid to compare feature matrices or treat them as basis states in the same dimensional space.

3 Temporal Resolution and Frame Alignment

The temporal duration of a single STFT hop (time resolution) is determined by the hop length (L_{hop}) and the sampling rate:

$$t_{hop} = \frac{L_{hop}}{f_s} \tag{2}$$

A variation in f_s leads to inconsistent time spacing. For two audio segments of identical physical duration (e.g., 4 musical beats), a difference in sampling rate results in a disparate total number of frames (e.g., 110 frames vs. 136 frames). This prevents consistent padding or trimming, prohibiting the alignment required for batch processing.

4 Wavelet Scale Invariance

In the Continuous Wavelet Transform (CWT), the wavelet scales (a) correspond to physical frequencies (f_a) via the relationship:

$$f_a = \frac{f_c \cdot f_s}{a} \tag{3}$$

where f_c is the center frequency of the mother wavelet. If f_s is not constant, a specific scale a maps to different physical frequencies across different samples. This inconsistency destroys the comparability of wavelet-energy vectors, rendering downstream analysis unreliable.

5 Hilbert Space and Basis Definition

The Quantum Walk algorithm relies on a well-defined Hilbert space $\mathcal{H}$. The basis states:

$$|s_0\rangle, |s_1\rangle, \ldots, |s_N\rangle$$

must represent vectors within the same space to allow for valid linear algebraic operations.

Varying sampling rates result in different definitions for the STFT bins and time steps, effectively creating distinct vector spaces for each segment. This precludes the valid application of:

- Cosine similarity measures.
- Machine Learning embeddings.
- Construction of transition matrices.
- Unitary evolution operators.

6 Summary

Enforcing a fixed sampling rate guarantees that every audio segment maps to a congruent frequency–time grid. This uniformity provides equal-dimensional representations for all segments, thereby establishing a valid, shared Hilbert space essential for reliable quantum evolution and analysis.