World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

QUANTIFYING THE TIME–FREQUENCY TRADE-OFF OF WINDOW LENGTH IN SHORT-TIME FOURIER TRANSFORM SPECTROGRAMS: A CONTROLLED STUDY WITH SYNTHETIC AND EEG SIGNALS

Emmanuel Ndunuju Amadiegwu [1,*] and Joy Ukibe [2]

[1] Department of Electronics and Computer Engineering, Faculty of Engineering, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
[2] Department of Computer Science, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
* Corresponding Author

ORCID Details
Emmanuel Ndunuju Amadiegwu: https://orcid.org/0009-0003-6523-2208
Joy Ukibe: https://orcid.org/0009-0002-8917-4543

World Journal of Advanced Research and Reviews, 2026, 32(02), 1378–1387

Publication history: Received on 16 November 2026; revised on 17 November 2026; accepted on 19 November 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.32.2.2946

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

A global magnitude or power spectrum derived from the discrete Fourier transform (DFT) summarizes frequency content without an explicit time axis, which limits its direct interpretation for non-stationary signals such as electroencephalography (EEG). This study derives the short-time Fourier transform (STFT) from the ordinary Fourier transform, clarifies how the analysis window shapes the measurement, and quantifies how window length controls the time and frequency errors of a spectrogram. The continuous and finite discrete STFT are developed from first principles, together with the windowing that turns an ideal spectral line into a main lobe and side lobes, the separation between zero-padding and resolution, and the time-frequency uncertainty bound $\sigma_t\sigma_f \geq 1/(4\pi)$. A reproducible C++20 implementation using FFTW is described. A controlled experiment then varies only the Hann window length, from 125 to 2000 ms at a 256 Hz sampling rate, while fixing the transform-length rule NFFT = L, the 75% overlap proportion, the PSD normalization convention, and the decision rules, using synthetic signals with known frequency, onset, and duration across 200 paired noisy trials at 0 dB signal-to-noise ratio. Peak-bin frequency error fell from 2.0 Hz at short windows to 0 Hz once the target frequency lay on the DFT grid, while burst-onset and burst-duration errors rose with window length, confirming the predicted opposing trends. A single-channel EEG recording illustrates the same representation on measured data without treating task labels as spectral ground truth. The results show that a spectrogram becomes a quantitative measurement only when its window, hop, normalization, estimator, and reference truth are stated explicitly.

Keywords: Short-Time Fourier Transform; Spectrogram; Time-Frequency Uncertainty; Spectral Leakage; Window Length; Electroencephalography

1 INTRODUCTION

The discrete Fourier transform of a length-N record,

$$X[k] = \Sigma_{n=0}^{N-1} x[n] \exp(-j2\pi kn/N), \qquad (1)$$

asks how strongly the entire record projects onto one complex oscillation. The word entire is the limitation. Consider two four-second signals sampled at $f_s = 256$ samples/s: a stationary tone $x_1(t) = \sin(2\pi \cdot 20t)$, and a signal $x_2(t)$ that equals $\sin(2\pi \cdot 20t)$ for $0 \le t < 2$ and $\sin(2\pi \cdot 60t)$ for $2 \le t < 4$. A global magnitude spectrum of x_2 reveals energy near 20 and 60 Hz, but it has discarded the answer to which component came first. Reversing the two halves changes phase while leaving the broad magnitude structure much the same. A magnitude spectrum organizes a record by frequency; it does not preserve a readable time axis [1].

This is not a failure of the Fourier transform. It follows from the question that was asked, because every global coefficient combines samples from the whole observation. To ask when a frequency occurs, the measurement must be local in time. Signals whose spectral character changes with time are called non-stationary. A tone switch is deliberately non-stationary; speech, rotating machinery during a speed change, and EEG may also contain transient or evolving structure. EEG records voltage differences between electrodes or between an electrode and a reference, so it reflects the combined electrical fields reaching the scalp rather than the activity of an isolated neuron. Because rhythmic and transient patterns may vary during a recording, the timing of spectral power can matter, which motivates a time-frequency display.

The objectives of this study are threefold. First, to derive the STFT from the ordinary Fourier transform and to interpret the window, the spectrogram, and the zero-padding grid as distinct components of one measurement. Second, to state the time-frequency uncertainty relation and separate it from practical resolution and bin spacing. Third, to measure how the analysis window length trades time error against frequency error under explicit decision rules, using synthetic signals whose answers are known and a public EEG recording as a realistic illustration.

2 MATERIALS AND METHODS

2.1 From a global transform to a local transform

Using ordinary frequency f in hertz, the continuous-time Fourier transform is $X(f) = \int x(\tau) \exp(-j2\pi f\tau)\, d\tau$, where the integration variable τ is distinguished from the location t of a local observation. Choosing a window w concentrated around zero and moving its centre to t gives $w_t(\tau) = w(\tau - t)$. Multiplication retains samples near t, and applying the same Fourier transform to the windowed signal defines one convention for the STFT:

$$X(t, f) = \int x(\tau)\, w(\tau - t) \exp(-j2\pi f\tau)\, d\tau. \qquad (2)$$

For fixed t this is the Fourier transform of the windowed signal; moving t produces another spectrum. The result has two independent coordinates, observation time t in seconds and frequency f in hertz. Some texts conjugate the window or shift the exponential by t, which changes phase but not magnitude; such conventions are compatible when stated and must not be mixed silently.

For a real sequence sampled at f_s, let m be the frame index, n the position inside a frame, L the number of windowed samples, H the hop between frame starts, N_{FFT} the transform length with $N_{FFT} \ge L$, and k the DFT-bin index. Frame m starts at sample mH. Selecting L measurements, weighting them by the window, and zero-filling to N_{FFT} when requested yields the finite discrete STFT:

$$X[m, k] = \Sigma_{n=0}^{L-1} x[n + mH]\, w[n] \exp(-j2\pi kn/N_{FFT}). \qquad (3)$$

This is an ordinary DFT of a precisely defined finite vector. L is the number of actual signal samples allowed to influence one frame, while N_{FFT} is the number of equally spaced frequency coordinates at which the DFT is evaluated. If $N_{FFT} > L$, the windowed frame is followed by $N_{FFT} - L$ zeros, and the frame still contains only L measurements. The case $N_{FFT} < L$ would discard or fold data and is excluded here. The frame-centre time and the bin frequency are

$$t_m = (mH + (L-1)/2) / f_s, \quad f_k = k f_s / N_{FFT}. \qquad (4)$$

For a real signal, bins 0 through $\lfloor N_{FFT}/2 \rfloor$ carry the non-redundant one-sided representation. The grid spacing is f_s/N_{FFT}, but the resolving behaviour is controlled by the L measured, windowed samples.

2.2 A window is part of the measurement

Cutting a block from the record is already windowing. A length-L rectangular window $w_R[n]$ equals one for $0 \le n < L$ and zero otherwise. Multiplication in time becomes periodic convolution in frequency, so for an ideal complex sinusoid, whose spectrum is an impulse, convolution shifts a copy of the window spectrum to that frequency. A finite observation therefore replaces an ideal line by the window shape: a central main lobe and surrounding side lobes. Energy appearing away from the source frequency is spectral leakage, caused by finite observation and window shape rather than by aliasing. Summing the rectangular-window spectrum directly gives

$$W_R(\exp(j\omega)) = \exp(-j\omega(L-1)/2) \cdot \sin(L\omega/2) / \sin(\omega/2). \quad (5)$$

Its magnitude is L at $\omega = 0$, with nearest zeros at $\omega = \pm 2\pi/L$ bounding the main lobe; the smaller peaks beyond those zeros are side lobes. Convolution does not create new physical oscillators, because the shifted window spectrum is a property of the finite record, not a set of additional tones in the source.

A concrete case makes the mechanism explicit. For $x[n] = \exp(j\omega_0 n)$ over $0 \le n < L$ with an unpadded L-point DFT, the bin value is a geometric series whose magnitude is

$$|X[k]| = | \sin(L\delta_k/2) / \sin(\delta_k/2) |, \quad \delta_k = \omega_0 - 2\pi k/L. \quad (6)$$

If $\omega_0 = 2\pi k_0/L$ with integer k_0, then $X[k_0] = L$ and every other integer bin is a zero of the numerator, so all other projections cancel exactly. If k_0 is not an integer, the shifted zeros fall between the DFT sample locations and many bins are nonzero. As an illustration, at $f_s = 256$ Hz a 21 Hz sinusoid observed for L = 64 samples lasts 0.25 s and completes 5.25 cycles; with 4 Hz bin spacing it lies at bin coordinate 5.25, between bins 5 and 6. A 20 Hz tone completes exactly five cycles and aligns with bin 5. The periodic extension view, the non-integer cycle count, and the shifted window spectrum are three descriptions of the same finite-record calculation.

2.3 Rectangular and Hann windows make different compromises

The symmetric Hann window is

$$w_H[n] = \tfrac{1}{2}(1 - \cos(2\pi n/(L-1))), \quad 0 \le n < L. \quad (7)$$

The denominator L − 1 sets both endpoints to zero. A common periodic form uses denominator L; libraries differ in their defaults, and the implementation below uses the symmetric form [2]. Writing the cosine as exponentials, the Hann transform is a weighted sum of three shifted rectangular responses, with $\omega_H = 2\pi/(L-1)$:

$$W_H(\exp(j\omega)) = \tfrac{1}{2} W_R(\exp(j\omega)) - \tfrac{1}{4} W_R(\exp(j(\omega-\omega_H))) - \tfrac{1}{4} W_R(\exp(j(\omega+\omega_H))). \quad (8)$$

Near many side-lobe peaks these copies partially cancel, which lowers side lobes; near the centre they combine into a wider main lobe. Tapering the frame toward zero reduces the end-to-start mismatch and lowers side lobes relative to the rectangle, but it does not remove leakage, and the Hann main lobe is broader, about four DFT bins from null to null against two for the rectangle. The rectangle has the narrowest main lobe and is exact for a noiseless bin-centred sinusoid; Hann is a useful default when components are not bin-centred and side-lobe suppression matters. Window selection declares which spectral errors matter most.

2.4 Resolution, zero-padding, and the time-frequency trade-off

Setting $N_{FFT} = L$, a window of L samples has nominal duration $T_w = L/f_s$ and grid spacing

$$\Delta f_{bin} = f_s/L = 1/T_w. \quad (9)$$

A short window multiplies the signal by something narrow in time, so it can separate events at different times, but its broad window spectrum smears a sinusoid over frequency. A long window accumulates many cycles and narrows the spectral response, but each coefficient combines a longer span of the signal, so a sudden onset is smeared across the frames whose windows straddle it. Zero-padding changes the ruler, not the observation: the padded DFT samples the same finite-frame response at more locations, which supports peak interpolation but cannot narrow a lobe or separate two already-merged components. Table 1 lists the grid parameters used in the experiment.

Table 1: STFT grid parameters for the synthetic experiment at $f_s = 256$ Hz. Each duration doubles the previous one, so observation time and unpadded bin spacing change in controlled factors of two.

Window duration	Samples L	Hop H = L/4	Bin spacing f_s/L
125 ms	32	8	8 Hz
250 ms	64	16	4 Hz
500 ms	128	32	2 Hz
1000 ms	256	64	1 Hz
2000 ms	512	128	0.5 Hz

2.5 The time-frequency uncertainty relation

The qualitative trade-off has a precise continuous counterpart. For a square-integrable, energy-normalized window with time spread σ_t and angular-frequency spread σ_ω, centring the window and applying integration by parts, the Cauchy-Schwarz inequality, and Parseval's relation gives $\sigma_t\sigma_\omega \geq 1/2$. Because $\omega = 2\pi f$, so that $\sigma_\omega = 2\pi\sigma_f$,

$$\sigma_t \, \sigma_f \geq 1/(4\pi), \quad (10)$$

with equality for a Gaussian window up to translation, modulation, scale, and constant phase; a Hann window does not attain the bound [3]. Four quantities must be kept separate: the theoretical product $\sigma_t\sigma_f$, which relates energy-weighted spreads of a Fourier pair; practical resolution, which describes an analysis system distinguishing components under a stated rule; DFT-bin spacing, which is where a finite transform is evaluated; and localization error, which compares an estimator against known ground truth. The identity $(L/f_s)(f_s/L) = 1$ is only the product of nominal window duration and unpadded bin spacing; it is not the uncertainty equality.

2.6 Hop size and overlap

The window need not advance by its full length. If each frame contains L samples and starts H samples after the previous one, the overlap and frame count are

$$N_{overlap} = L - H, \quad \text{Overlap}(\%) = (L-H)/L \times 100, \quad (11)$$

$$M = 1 + \lfloor (N_x - L)/H \rfloor \quad \text{for } N_x \geq L. \quad (12)$$

Overlap reuses the same measured samples in more than one local spectrum; nothing is interpolated between samples. It provides a denser sampling of the time-varying STFT, which makes heatmaps less blocky and reduces the chance that a transient falls only near a window edge, but it does not shorten the window, narrow its spectral response, or evade uncertainty. The experiment uses a fixed 75% overlap (H = L/4) so that frame coordinates are dense while the hop proportion is held constant across window lengths.

2.7 From complex coefficients to a spectrogram

The coefficient X[m,k] is complex. Its magnitude measures local coefficient size and its argument records phase. Common displays are the magnitude spectrogram $A = |X[m,k]|$, the power spectrogram $P = |X[m,k]|^2$, and power in decibels $P_{dB} = 10 \log10(P/P_{ref})$, where a small positive floor is applied before the logarithm as a display choice. For a real, one-sided frame, a power spectral density (PSD) estimate is

$$\hat{S}_{xx}[m, k] = c_k \, |X[m, k]|^2 \, / \, (f_s \, \Sigma_{n=0}^{L-1} \, w^2[n]), \quad (13)$$

with $c_k = 2$ for interior positive-frequency bins and $c_k = 1$ at DC and, for even N_{FFT}, at Nyquist. Its units are signal-unit squared per hertz. The squared-window normalization is essential when comparing window lengths, because raw $|X|^2$ grows with both the amount of data and the window weights. A conventional spectrogram maps time horizontally, frequency vertically, and power to colour; a stationary tone forms a horizontal ridge, a short burst a ridge limited in time but blurred at its edges, and a linear chirp a sloped ridge.

2.8 Reproducible C++20 implementation

The transform was implemented in C++20 using the FFTW real-to-complex transform. One plan is created and reused with aligned input and output buffers for every frame. For each complete frame the code extracts L samples, applies the symmetric Hann window, zero-fills to N_{FFT}, executes the FFT, retains the $N_{FFT}/2 + 1$ non-redundant complex bins, and

computes a one-sided PSD using the normalization of equation (13). Incomplete final frames are discarded, since padding them would add a boundary assumption and extra frames. FFTW uses the same negative-exponent, unnormalized DFT convention as this article, so an inverse pipeline must divide by N_{FFT} [4]. The listing below shows the core loop; the complete program, EDF+ reader, build script, fixed seeds, data, and figures are available in the project repository [5], and the reported values were generated by that code with FFTW 3.3.11 rather than read from the plots.

```
const auto window = symmetric_hann(cfg.window_length);
double window_energy = 0.0;
for (double v : window) window_energy += v * v;
const std::size_t bins = cfg.fft_size / 2 + 1;
fftw_plan plan = fftw_plan_dft_r2c_1d(
    (int)cfg.fft_size, input, output, FFTW_ESTIMATE);

for (std::size_t start = 0;
     start + cfg.window_length <= signal.size();
     start += cfg.hop_size) {
    for (std::size_t n = 0; n < cfg.window_length; ++n)
        input[n] = signal[start + n] * window[n];
    for (std::size_t n = cfg.window_length; n < cfg.fft_size; ++n)
        input[n] = 0.0;              // zero-pad: finer grid only
    fftw_execute(plan);
    for (std::size_t k = 0; k < bins; ++k) {
        const std::complex<double> X{output[k][0], output[k][1]};
        const bool dc = (k == 0);
        const bool nyq = (cfg.fft_size % 2 == 0) && (k == cfg.fft_size / 2);
        const double one_sided = (dc || nyq) ? 1.0 : 2.0;
        psd[k] = one_sided * std::norm(X)
                 / (cfg.sample_rate_hz * window_energy);
    }
}
```

2.9 Experimental design

The central experiment asks what changes when only the Hann window length changes. To make that answer identifiable, the transform-length rule NFFT = L, sample rate, overlap proportion, normalization convention, signal duration, and decision rules are held fixed. The numerical FFT length and hop therefore scale with L rather than remaining constant, as summarized in Table 2. Synthetic sources are generated from equations, so their frequency, onset, offset, amplitude, and noise distribution are known; their purpose is measurement rather than imitation of EEG.

Table 2: Experimental settings and window-length sweep for the eight-second synthetic records.

Fixed choice	Reason
Eight-second synthetic record	Leaves time before and after the burst and room for several two-second frames.
f_s = 256 Hz	Gives a 128 Hz Nyquist above every component and exact power-of-two sample counts.
125, 250, 500, 1000, 2000 ms	Each duration doubles the previous, changing time and bin spacing by factors of two.
Symmetric Hann window	One fixed, explicitly defined taper for every comparison.
N_{FFT} = L	Prevents zero-padding density from being confused with a longer measured window.
H = L/4 (75% overlap)	Dense frame coordinates with a common hop proportion.
One-sided PSD	Real sources; energy normalization permits comparison across lengths.
200 paired trials	A distribution of errors with the same realization faced by every length.

Four deterministic sources are defined over the eight-second record:

$$x_{tone}(t) = \sin(2\pi \cdot 30t),$$

$$x_{two}(t) = \sin(2\pi \cdot 30t) + 0.8 \sin(2\pi \cdot 34t),$$

$$x_{burst}(t) = 1_{[3.0,\,3.5)}(t) \cdot \sin(2\pi \cdot 40t),$$

$$x_{chirp}(t) = \sin(2\pi(10t + (40/16)\, t^2)),$$

where the indicator $1_{[3.0,3.5)}(t)$ equals one during that half-second interval and zero elsewhere. The stationary tone tests peak-frequency estimation, the two-tone signal tests a fixed 4 Hz separation at unequal amplitude, the burst supplies exact onset, offset, and duration, and the chirp has instantaneous frequency $f_i(t) = 10 + 5t$ Hz for visualizing ridge averaging. At 125 ms, the 30 Hz tone lies at bin coordinate 3.75, between 24 and 32 Hz, with 32 Hz the nearest bin. At 250 ms, it lies at coordinate 7.5, halfway between 28 and 32 Hz. At 500 ms and longer it lies exactly on the grid. These cases distinguish grid quantization from lobe interpolation. The second tone at 34 Hz gives a known 4 Hz separation that straddles the approximate Hann null-to-null main-lobe width of $4f_s/L$. The burst carrier at 40 Hz lies on every grid in the set, so its timing experiment is not confounded by a carrier that moves on and off bins.

Estimators and metrics are declared before results. For the stationary tone, the frequency estimate is the frequency coordinate of the largest bin of the frame-averaged one-sided PSD in 20 to 45 Hz, with ties assigned to the lower frequency, and the error is $e_f = |\hat{f} - 30\ \text{Hz}|$; peak-bin estimation includes grid quantization, while a separately reported quadratic interpolation refines the lobe centre without adding resolving power. The two tones are declared resolved only if two local maxima occur in 26 to 38 Hz, each within 2 Hz of a different true tone, with a valley at least 3 dB below the smaller peak. For the burst, PSD is integrated over a fixed 36 to 44 Hz band per frame, and onset is the earliest frame centre whose band power exceeds a per-length threshold for two consecutive frames; for each calibration record, the statistic is the maximum across adjacent frame pairs of the smaller of their two band powers. For each window length, 1000 independent noise-only eight-second records are generated at the test-noise standard deviation, separately from test trials. The threshold is the 990th value of the 1000 sorted record maxima (zero-based index 989), an empirical 99th-percentile threshold calibrated at the record level. It targets a 1% record-level false-alarm probability without assuming frame independence; the finite calibration sample does not guarantee an exact 1% rate. Noise is added at 0 dB whole-record signal-to-noise ratio, every seed is reused across all five window lengths so comparisons are paired, and four statistics are reported across independent records:

Let a[m] indicate that the band power in frame m exceeds the threshold. The onset index is the first m for which a[m] and a[m+1] are both true. Starting at j = m+1, increment j while another frame exists and either a[j] or a[j+1] is true. The offset is the centre time of frame j, so a single below-threshold frame is bridged and the search stops at the first of two consecutive below-threshold frames, or at the final available frame. Estimated duration is offset minus onset. Errors are measured against the known 3.0 s onset, 3.5 s offset, and 0.5 s duration; records without an onset are counted as missed detections and excluded from timing-error summaries.

$$\text{bias} = (1/R)\, \Sigma_{r=1}^{R} (\hat{\theta}_r - \theta), \quad \text{MAE} = (1/R)\, \Sigma_r |\hat{\theta}_r - \theta|, \quad \text{RMSE} = \sqrt{((1/R)\, \Sigma_r (\hat{\theta}_r - \theta)^2)}. \qquad (14)$$

Sample standard deviation is computed from signed trial errors $e_r = \hat{\theta}_r - \theta$ as $s = \sqrt{[\Sigma_r (e_r - \bar{e})^2/(R-1)]}$, where $\bar{e}$ is their mean. The full CSV reports bias, MAE, RMSE, and sample standard deviation for frequency and timing errors; Table 4 presents selected MAEs and resolution counts.

2.10 EEG data source and preprocessing

The practical demonstration uses a public recording from the PhysioNet EEG Motor Movement/Imagery Dataset (EEGMMIDB), version 1.0.0, specifically Subject S001, Run 04, a motor-imagery run supplied as EDF+ with event annotations and sampled at 160 Hz [6]. In this run the annotation codes mark rest and left or right hand-imagery cues; they describe the experimental cue, not a guaranteed mental state or the onset of a specific cortical rhythm. Channel C3, a standard scalp position over the left-central sensorimotor region, was selected in advance of inspecting the spectrograms, which avoids choosing whichever channel makes the trade-off look most dramatic; the sensorimotor literature motivates the location without implying an isolated cortical source, since every scalp channel is a referenced voltage affected by distributed fields and volume conduction [7]. The same C3 interval is analyzed with all five window lengths. At 160 Hz the requested durations give the exact sample counts in Table 3.

Table 3: EEG window sample counts and grid parameters at the native 160 Hz sampling rate of S001 Run 04.

Requested duration	Samples	Bin spacing ($N_{FFT} = L$)	75% overlap hop
125 ms	20	8 Hz	5 samples
250 ms	40	4 Hz	10 samples
500 ms	80	2 Hz	20 samples
1000 ms	160	1 Hz	40 samples
2000 ms	320	0.5 Hz	80 samples

For a transparent comparison, the implementation selects C3 in advance, subtracts only the mean of the displayed interval, and applies no band-pass, notch filter, artifact rejection, normalization, or resampling, passing the microvolt samples into the same C++ STFT pipeline. Displays use a one-sided PSD in $\mu V^2/Hz$ with fixed decibel colour limits across window lengths, so that weak structure in one panel cannot be stretched to look as intense as strong structure in another. Mean removal does not remove slow drift, mains interference, or eye and muscle artifacts, and it is not a substitute for a documented filter. The annotations provide task timing, not exact oscillatory ground truth, so the EEG panels are compared for visible concentration and smearing, while the synthetic signals carry the quantitative truth.

3 RESULTS AND DISCUSSION

3.1 Measured synthetic results

The experiment used 200 paired Gaussian-noise trials at 0 dB whole-record signal-to-noise ratio, with every window length receiving the same clean signal and noise realization in a given trial. Table 4 reports the frequency and timing errors.

Table 4: Measured synthetic results over 200 paired trials at 0 dB SNR. Peak-bin frequency error reflects the DFT grid; timing error grows with window length.

Window	Bin spacing	Freq. MAE	Interp. MAE	Two tones resolved	Onset MAE	Duration MAE
125 ms	8 Hz	2.000 Hz	0.098 Hz	0/200	0.004 s	0.035 s
250 ms	4 Hz	2.000 Hz	0.012 Hz	0/200	0.005 s	0.076 s
500 ms	2 Hz	0.000 Hz	0.007 Hz	0/200	0.120 s	0.368 s
1000 ms	1 Hz	0.000 Hz	0.003 Hz	200/200	0.253 s	0.751 s
2000 ms	0.5 Hz	0.000 Hz	0.002 Hz	200/200	0.502 s	1.513 s

The first three frequency results follow the grid calculation made before the run. At 125 ms the bins are 8 Hz apart and the nearest grid value to 30 Hz is 32 Hz; at 250 ms, 30 Hz lies midway between 28 and 32 Hz. Both peak-bin conditions therefore report a 2 Hz MAE. At 500 ms and longer, 30 Hz lies exactly on the grid, so the raw-bin error is zero in that narrow sense, which does not imply zero error for an arbitrary frequency. Quadratic interpolation reduces the coarse-grid error for short windows but adds no resolving power.

The two-tone result concerns lobe width rather than grid position. Both 30 and 34 Hz are exact bins at 500 ms, yet the pair stays unresolved in every trial because the Hann responses overlap too strongly under the 3 dB-valley rule. At 1000 ms the main-lobe width has narrowed to the 4 Hz separation and all 200 trials satisfy the rule. Appearance on the DFT grid is therefore insufficient evidence that two components are distinguishable. The timing results move in the opposite direction: onset MAE is approximately 5 ms or less for the two shortest windows, partly because one frame centre of the 125 ms condition falls near the fixed 3.0 s onset, and rises to about half a second at 2000 ms, while duration error grows because a long window stays above threshold while it overlaps samples on both sides of the half-second burst. All 200 bursts were detected at every length, so these summaries hide no missed trials. Figure 2 shows the opposing trends on a single axis pair.

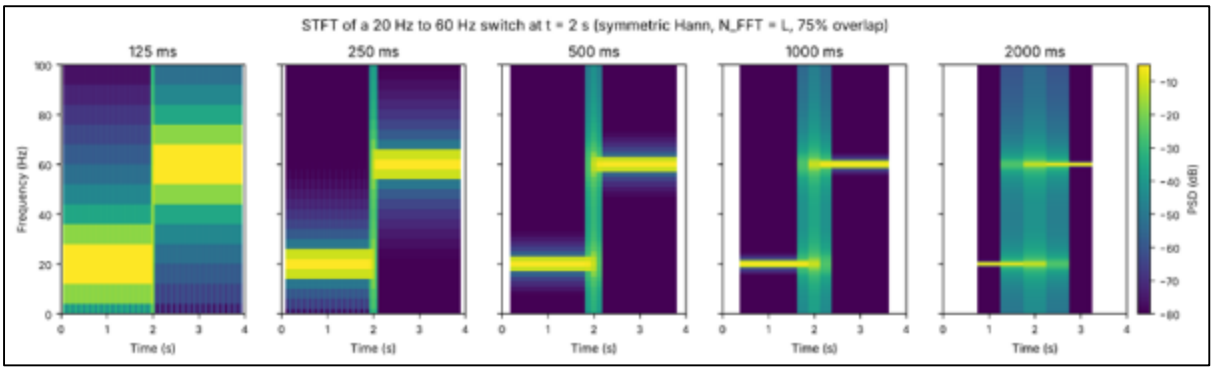


Figure 1: Spectrograms of a 20 Hz to 60 Hz switch at t = 2 s, computed with a symmetric Hann window, N_{FFT} = L, 75% overlap, and identical decibel colour limits. Short windows change close to two seconds but spread energy in frequency; long windows give narrow frequency bands but spread the switch across time.

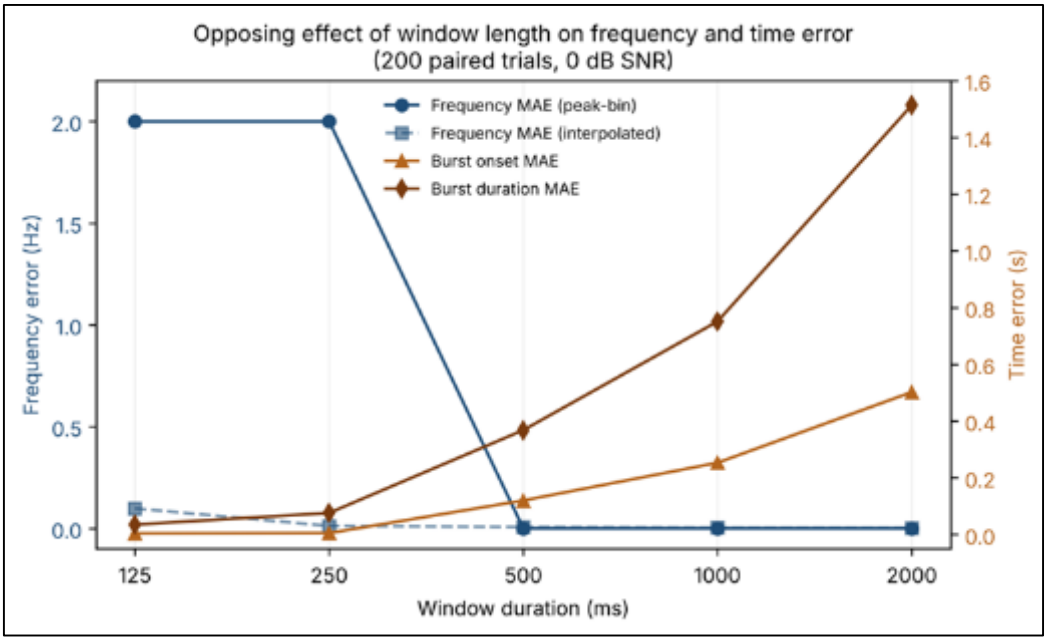


Figure 2: Frequency error (left axis) and timing error (right axis) against window duration, from the 200-trial results of Table 4. Peak-bin frequency error falls as the target reaches the grid, while burst onset and duration errors rise with window length.

The same 20 to 60 Hz switch was transformed with every length under identical colour limits, as shown in Figure 1. The short-window panels change close to the two-second boundary but spread each tone across frequency, while the long-window panels resolve the two tones sharply yet carry energy across a wider region around the switch. These two figures are the visual counterpart of the numerical trade-off: no single window minimizes both time and frequency error for this class of signal.

3.2 EEG demonstration

The practical demonstration processed the first 30 seconds of C3 from S001 Run 04 at the native 160 Hz, with the only preprocessing being subtraction of the interval mean. This interval contains 4800 samples over $0 \le t < 30$ s, with the final sample at 29.99375 s. Figure 3 shows the mean-removed voltage input before Hann windowing; connecting samples provides a readable trace rather than additional measurements. The waveform shows voltage variation over time, while the spectrograms show the distribution of local spectral power. The five spectrograms in Figure 4 use fixed decibel colour limits so that window length is the only changed variable. Short windows produce broad, rapidly varying patches, while long windows produce narrower and more persistent bands while mixing activity across a longer portion of each annotated interval. This is a statement about the analysis representation of one selected recording, not evidence that an annotation caused a spectral feature. The annotations provide task timing rather than exact oscillatory ground truth, and a single subject and run cannot support a population or clinical claim; the synthetic experiment supplies the quantitative ground truth that the EEG recording cannot.

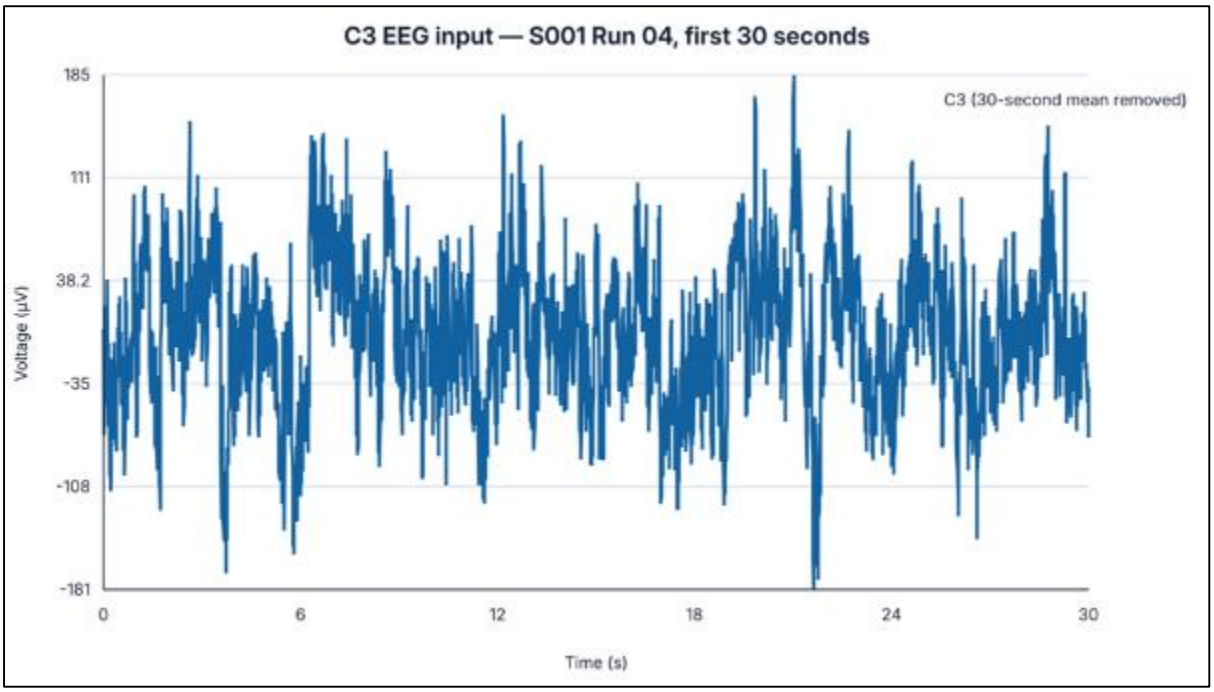


Figure 3: Time-domain C3 EEG input from PhysioNet EEGMMIDB S001 Run 04: first 30 s at 160 Hz (4800 samples), in µV after subtraction of the 30-second sample mean. No filtering, resampling, artifact rejection, or amplitude normalization is applied. The same mean-removed samples are supplied to all five STFT conditions in Figure 4.

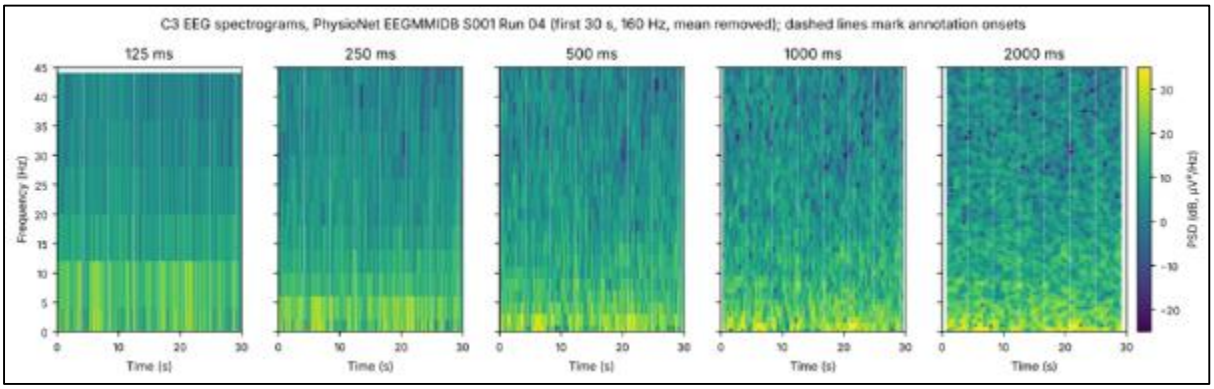


Figure 4: C3 EEG spectrograms from PhysioNet EEGMMIDB S001 Run 04 (first 30 s at 160 Hz, mean removed) at window lengths of 125, 250, 500, 1000, and 2000 ms, with identical decibel colour limits in $\mu V^2/Hz$ and dashed lines at the annotation onsets. Short windows give broad, rapidly varying structure, while long windows give narrower frequency bands and mix activity across a longer portion of each interval.

4 CONCLUSION

The short-time Fourier transform is not a different species of transform. It is the Fourier transform applied repeatedly after a moving multiplication, and that multiplication both enables time localization and necessarily convolves the spectrum with the window response. Four controls act on different parts of the result: the window duration L/f_s sets how much measured time enters one coefficient; the window spectrum sets main-lobe and side-lobe behaviour; the hop H/f_s sets the spacing between frame coordinates rather than intrinsic resolution; and the grid f_s/N_{FFT} sets the evaluated frequencies, while padding adds no observations. Normalization determines what the displayed magnitudes mean and whether different windows can be compared. The uncertainty relation gives a lower bound for Fourier-pair spreads, not a prediction of a particular peak detector, so practical resolution must be declared and measured. The controlled experiment confirmed the predicted opposing trends: shorter windows reduced timing error and increased frequency error, longer windows did the reverse, under the fixed 75% overlap rule; timing errors reflect both window smearing and the increasing frame-centre spacing H/fs. A spectrogram is persuasive because it is visual, but it becomes an

engineering measurement only when its window, hop, scaling, boundary policy, estimator, and ground truth are stated explicitly.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors acknowledge the PhysioNet EEG Motor Movement/Imagery Dataset and its contributors for making the EEG recording publicly available, and the FFTW project for the fast Fourier transform library used in the implementation.

Disclosure of conflict of interest

The authors declare no financial or non-financial conflicts of interest.

Statement of ethical approval

This study is a secondary analysis of a publicly available, de-identified EEG recording distributed by PhysioNet under its data-use terms. No new experiments involving human participants or animals were conducted by the authors, and no additional ethical approval was required. All synthetic signals were generated computationally.

Funding: This research received no external funding.

Informed consent: Not applicable to this secondary analysis of publicly available, de-identified EEG data. No new participants were recruited for this study.

REFERENCES

[1] Rowell D. Sampling and the discrete Fourier transform. Cambridge (MA): MIT OpenCourseWare; 2008.

[2] SciPy developers. scipy.signal.windows.hann [Internet]. [cited 2026 Nov 12]. Available from: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.windows.hann.html

[3] Cohen L. Time-frequency analysis. Englewood Cliffs (NJ): Prentice Hall; 1995.

[4] Frigo M, Johnson SG. FFTW 3.3.10 manual [Internet]. 2023 [cited 2026 Nov 12]. Available from: http://www.fftw.org/

[5] ByteBard15. signal_processing_from_first_principles [Internet]. GitHub; [cited 2026 Nov 22]. Available from: https://github.com/ByteBard15/signal_processing_from_first_principles

[6] Schalk G. EEG Motor Movement/Imagery Dataset, version 1.0.0 [Internet]. PhysioNet; 2009 [cited 2026 Nov 12]. Available from: https://physionet.org/content/eegmmidb/1.0.0/

[7] McFarland DJ, Miner LA, Vaughan TM, Wolpaw JR. Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topogr. 2000;12(3):177-186.