Supplementary information

Relationship between spike shape and frequency

This section presents detailed explanations and visualizations to clarify the relationship between spike shape and frequency, building on the analysis discussed in the methods section. The focus is on understanding how frequency components, such as mean and main frequencies, influence the morphology of amperometric spikes, using both sine wave and Gaussian decay models.

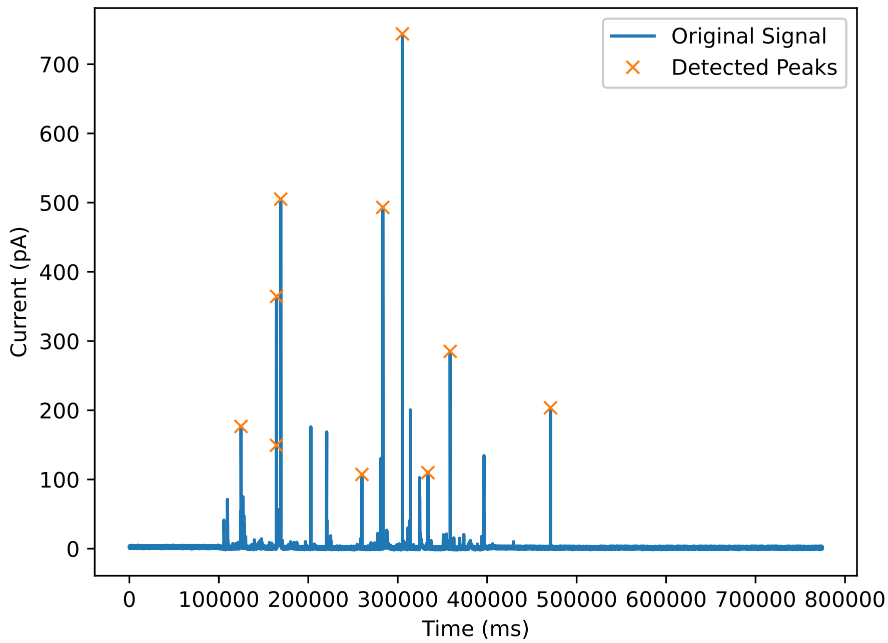


Figure 1. Experimental data with detected peaks: Sample data from the Hofmeister series, highlighting candidate spikes marked with orange crosses. These peaks represent key events in the dataset, selected for further analysis based on their prominence and shape characteristics.

The experimental spike data were analyzed by detecting peaks, followed by selecting three representative spikes for further examination as detailed in Figures 2, 3, and 4. The selection of these spikes was based on their prominence, height, and width, ensuring a range of spike shapes for comparison. A plot of the original dataset, marking all detected peaks and highlighting the three selected spikes, sets the context for the following analyses. By selecting spikes with a variety of heights and widths, it was ensured that different spike dynamics were captured, allowing for the evaluation of both typical and extreme cases within the dataset. This variety is important because spike morphology can vary widely depending on the physiological or experimental conditions, and thus, a diverse selection could provide more generalizable insights.

For each selected spike, a frequency analysis was performed by applying FFT to obtain the frequency spectrum of the original signal. This reveals the main and mean frequencies, representing the dominant frequency components influencing the spike shape. To visualize the effect of these frequency components, sine waves at the identified mean and main frequencies were superimposed onto the original spike signal. The phase and amplitude of these sine waves were adjusted to align with the peak of each spike, illustrating how oscillatory components at different frequencies contribute to the overall spike morphology. These sine wave overlays demonstrate that while these components capture certain aspects of spike shape, they are less effective in accurately representing the decay phase. The FFT analysis provides insight into the frequency content of the spike by decomposing the signal into its constituent frequencies. By comparing the main frequency, which represents the dominant oscillation, and the mean frequency, which averages the overall signal content, the relationship between frequency and spike shape becomes clearer. These frequency components, however, primarily capture the periodic aspects of the spike, particularly around its peak, leaving the complex dynamics of the decay phase less accurately modeled.

As discussed in the methods section on artificial data generation, modeling spikes with a linear rise and Gaussian decay provides a more accurate representation of amperometric signals compared to the sine wave approach. This model effectively

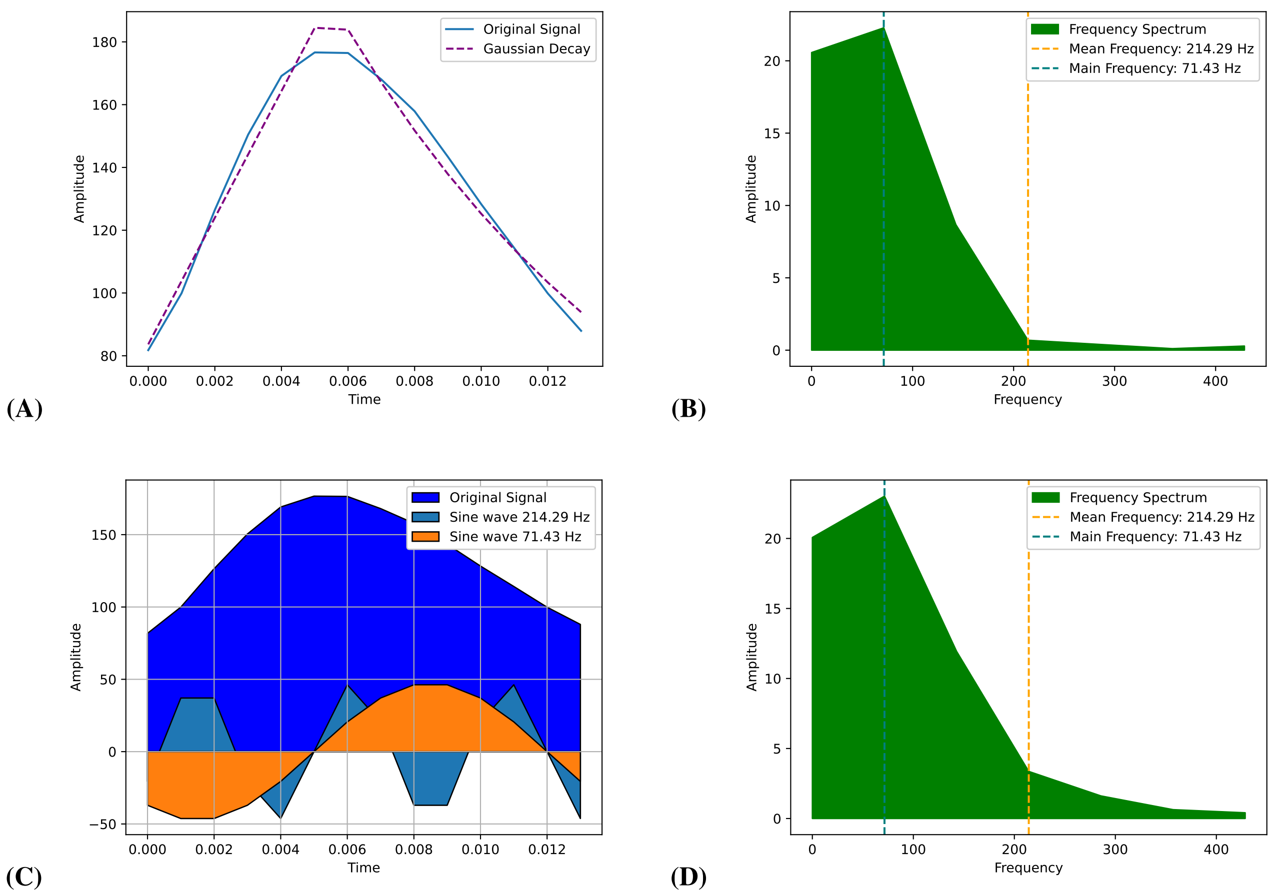


Figure 2. Relationship between spike shape and frequency for the first isolated spike from the experimental data: **(A)** The blue curve represents the original detected spike, while the purple dotted line shows the model fit using a linear rise and Gaussian decay model. **(B)** The frequency spectrum of the original spike, highlighting its mean and main frequencies. **(C)** The original spike signal overlaid with sine waves corresponding to its mean (214.29 Hz) and main (71.43 Hz) frequencies, illustrating the contributions of these oscillatory components. The phase and amplitude of the sine waves have been adjusted to align with the peak of the spike. **(D)** The frequency spectrum of the Gaussian decay model, highlighting its mean and main frequencies, for comparison with the original spike.

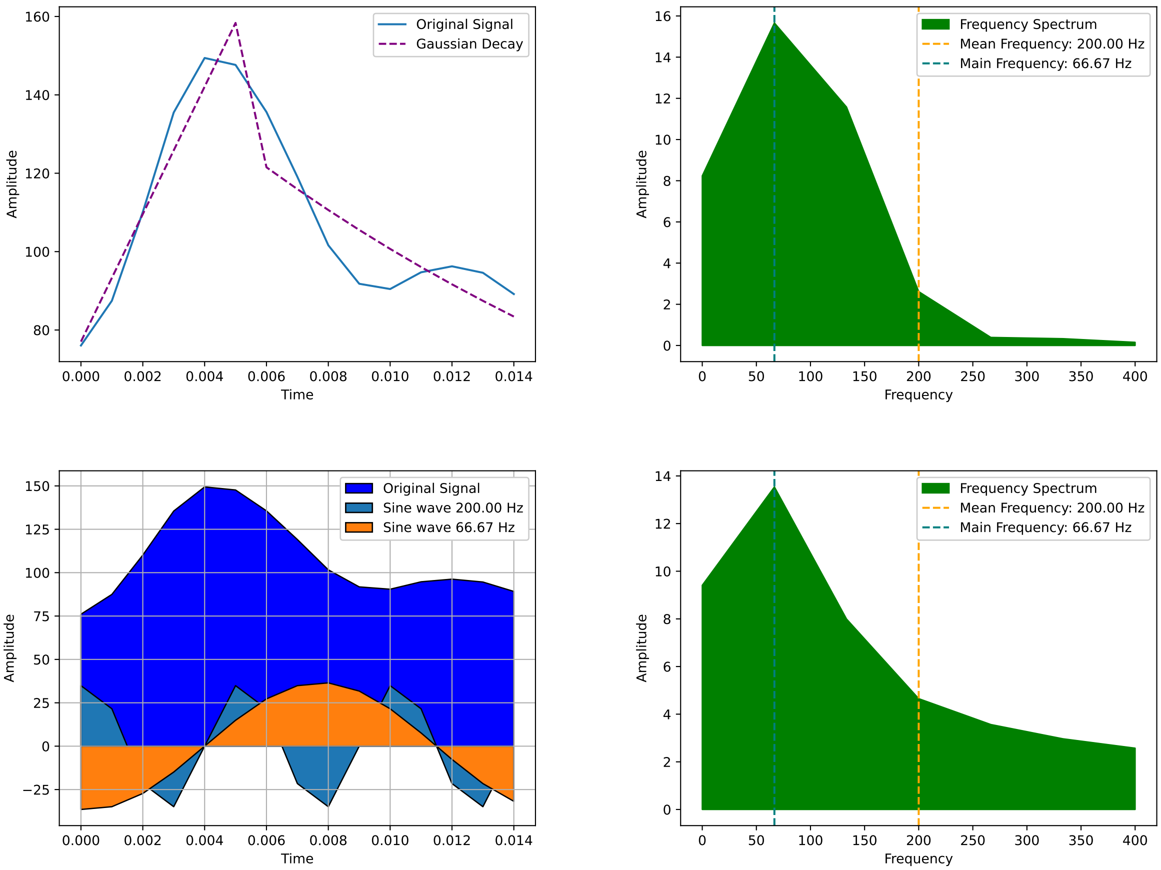


Figure 3. Relationship between spike shape and frequency for the first isolated spike from the experimental data: **(A)** The blue curve represents the original detected spike, while the purple dotted line shows the model fit using a linear rise and Gaussian decay model. **(B)** The frequency spectrum of the original spike, highlighting its mean and main frequencies. **(C)** The original spike signal overlaid with sine waves corresponding to its mean (200 Hz) and main (66.67 Hz) frequencies, illustrating the contributions of these oscillatory components. The phase and amplitude of the sine waves have been adjusted to align with the peak of the spike. **(D)** The frequency spectrum of the Gaussian decay model, highlighting its mean and main frequencies, for comparison with the original spike.

captures both the rapid rise and the exponential decay phases, better simulating the physical characteristics of the signal dynamics. For each selected spike, comparisons between the original signal and the Gaussian decay model, as shown in figures 2, 3, and 4, demonstrate that the Gaussian model offers a closer match to the actual signal shape, particularly in the decay phase. This highlights its effectiveness in representing the complex temporal behavior of amperometric signals. The effectiveness of the Gaussian decay model over the sine wave approach lies in its ability to reflect the underlying biophysical processes of spike generation. Amperometric spikes exhibit a rapid rise, typically driven by the influx of ions or neurotransmitters, followed by a gradual decay as the system returns to equilibrium. The Gaussian decay model captures this smooth, exponential decay more accurately than sine wave models, which are less suited for representing the falling phase of the spike. This precise alignment with the actual signal shape is critical for accurate biophysical interpretations.

Frequency spectra of the Gaussian decay model, when compared to those of the original signals, reveal a more comprehensive representation of the frequency components of the signal, particularly in the lower-frequency range associated with the decay phase. While sine wave overlays provide some insight into dominant frequencies, the Gaussian decay model offers a more accurate and holistic portrayal of the overall shape and frequency content of the signal. This alignment between the experimental data and the Gaussian model further validates its suitability for capturing spike dynamics. The broader range of frequency components captured by the Gaussian decay model, especially in the slower decay phase, is crucial for understanding the temporal dynamics of the signal. The model accounts for both the fast, high-frequency oscillations characteristic of the rising phase and the slower, lower-frequency components that govern the decay. The close match between the experimental and modeled spectra underscores the robustness of the Gaussian decay model in accurately representing the full frequency profile of amperometric signals.

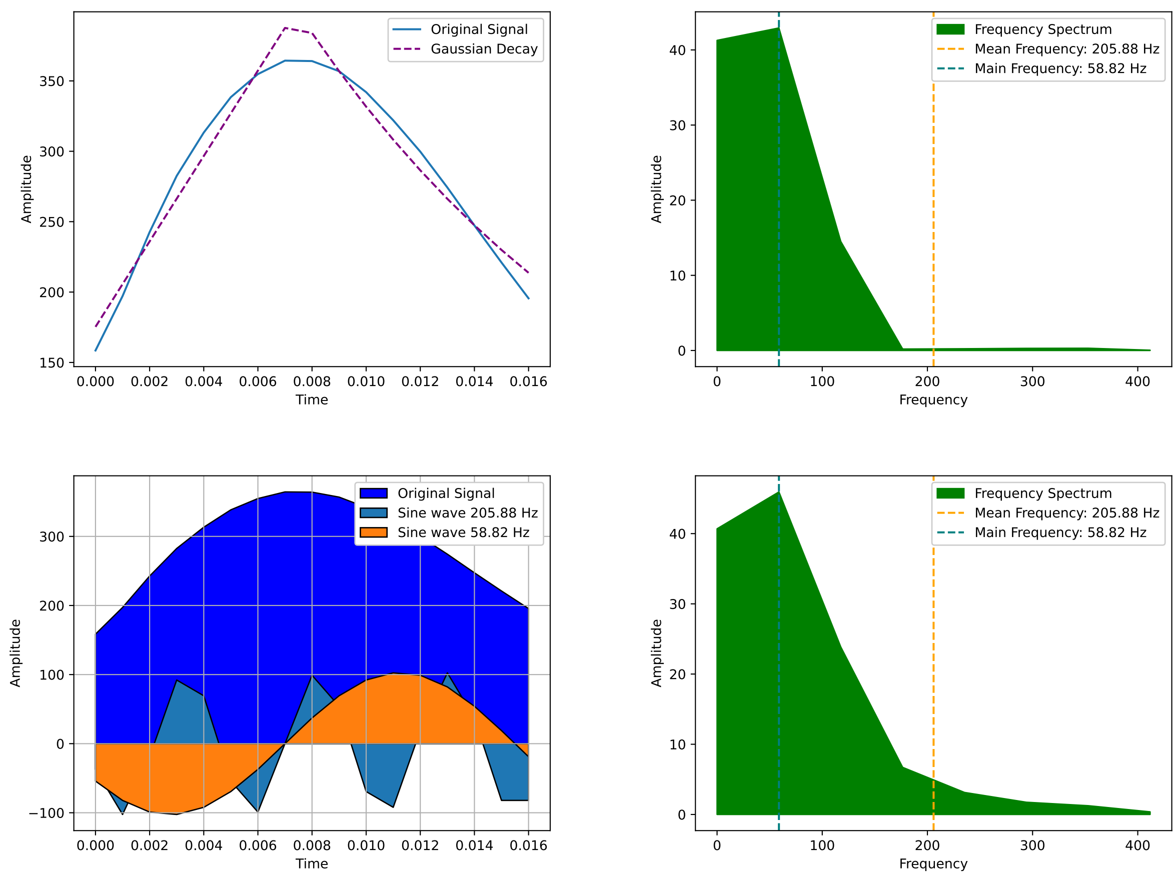


Figure 4. Relationship between spike shape and frequency for the first isolated spike from the experimental data: **(A)** The blue curve represents the original detected spike, while the purple dotted line shows the model fit using a linear rise and Gaussian decay model. **(B)** The frequency spectrum of the original spike, highlighting its mean and main frequencies. **(C)** The original spike signal overlaid with sine waves corresponding to its mean (205.88 Hz) and main (58.82 Hz) frequencies, illustrating the contributions of these oscillatory components. The phase and amplitude of the sine waves have been adjusted to align with the peak of the spike. **(D)** The frequency spectrum of the Gaussian decay model, highlighting its mean and main frequencies, for comparison with the original spike.

On the standard error of the mean of median of $t_{1/2}$

For each experimental condition or dataset, the median of $t_{1/2}$ values is calculated to provide a robust measure of central tendency. The mean of these median values is then computed across different experimental repetitions or groups to summarize the central tendency across conditions. To estimate the precision of this summary measure, we calculate the standard error of the mean of the medians. This approach ensures that the central tendency of non-normally distributed $t_{1/2}$ values is represented accurately, while the standard error reflects the variability across experimental trials.

The use of the median instead of the mean for individual $t_{1/2}$ distributions is justified by the non-normality of the data. By computing the mean of the medians, we obtain a central tendency across experiments, and the standard error of this measure quantifies its precision. This approach is particularly useful in biological data, where variability and non-normality are common.

Effect of filtering

To minimize the impact of noise and artifacts, low-pass filtering was applied to the raw data. The cutoff frequency was carefully chosen to ensure that the biologically relevant frequency components of the spikes remained intact, with the filtering primarily removing high-frequency noise. Sensitivity tests confirmed that the spike mean frequency was largely unaffected by filtering, provided that the cutoff frequency was set above the dominant frequency range of the spikes.

Sensitivity to spike boundaries

Spike boundaries were defined using a thresholding method to detect significant deviations from the baseline, ensuring that the full duration of each spike was captured. We conducted sensitivity tests to assess the impact of small variations in spike boundaries on the mean frequency. The results showed that mean frequency is robust to minor boundary adjustments as long as the full spike, including its rising and falling phases, is captured. However, significant misalignment of boundaries could affect the frequency analysis, underscoring the importance of accurately defining the spike limits.

Analytical Verification

Calculating FFT over the entire time series (including baseline) would show no characteristic frequencies apart from a peak at around 0 Hz since the slowing varying baseline dominates the signal, as an example, see Fig. 5.

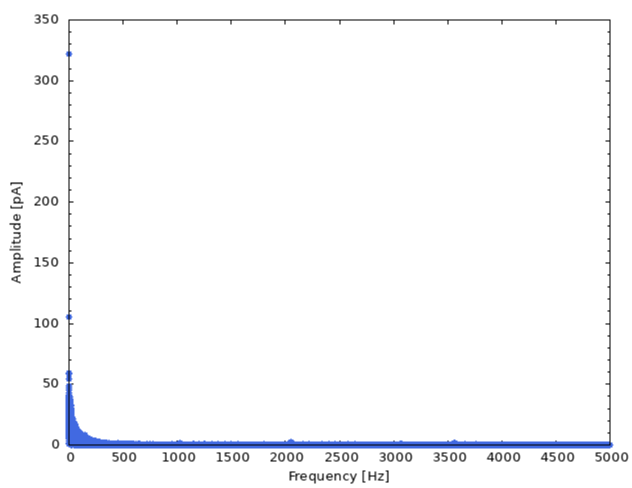


Figure 5. Fast Fourier Transform of a Br^- time series. All other data also show a peak frequency component near 0 Hz.

Therefore, the FFT should be applied on a spike-by-spike basis to extract the spike-specific mean frequencies. As motivated in the previous sections, the frequency analysis method is introduced based on the hypothesis that a relation between spike shape and the mean frequency exists. We here show the proof of this hypothesis for two simple waveforms, i.e. a sinusoidal waveform and a symmetric triangle waveform.

Sinusoidal Waveform

For a given signal, $y(t)$, the signal can be transformed to the frequency domain, $\mathscr{F}(y(t)) = Y(\xi)$ using the Fourier Transform. Next, to compute the mean frequency, f_{mean}, one needs to calculate the ratio of the first moment of $|Y(\xi)|^2$ and the expectation of $|Y(\xi)|^2$, which is given by,

$$f_{\text{mean}}(\omega) = \frac{\int_0^\infty |Y(\xi)|^2 \, \xi \, d\xi}{\int_0^\infty |Y(\xi)|^2 \, d\xi} \tag{1}$$

77 where the $|Y(\xi)|$ is given by,

$$|Y(\xi)| = \sqrt{\mathrm{Re}(Y(\xi))^2 + \mathrm{Im}(Y(\xi))^2}. \tag{2}$$

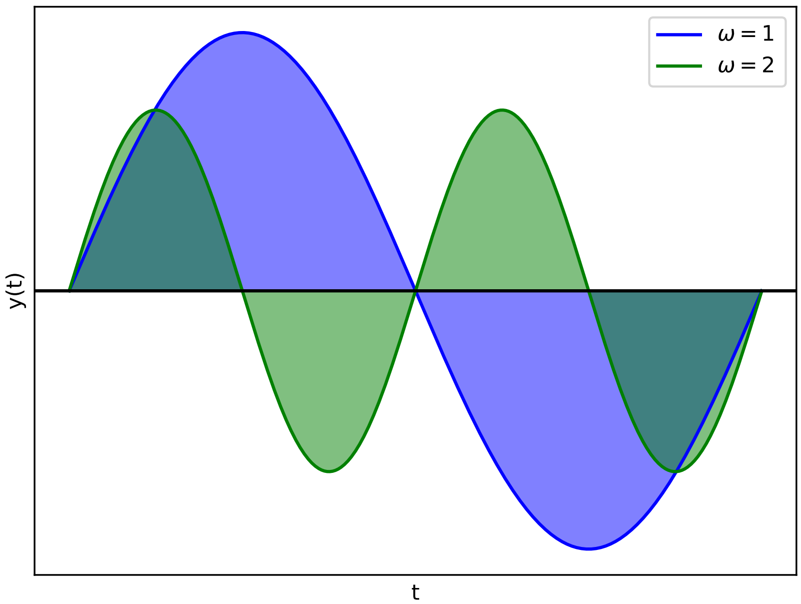


Figure 6. Illustration of sine waves with different frequencies in time domain.

78 Let us consider an example of the waveform, $\sin(\omega t)$ as illustrated in Fig. 6,

$$\begin{aligned} Y(\xi) &= \int_0^{2\pi} y(t)\, e^{-2\pi i t \xi}\, dt \\ &= \int_0^{2\pi} \sin(\omega t)\, e^{-2\pi i t \xi}\, dt \\ &= \frac{1}{2i}\Big[\delta\big(\xi - \frac{\omega}{2\pi}\big) - \delta\big(\xi + \frac{\omega}{2\pi}\big)\Big] \end{aligned} \tag{3}$$

$$\begin{aligned} |Y(\xi)| &= \sqrt{\mathrm{Re}(Y(\xi))^2 + \mathrm{Im}(Y(\xi))^2} \\ &= \frac{1}{2}\Big[\delta\big(\xi - \frac{\omega}{2\pi}\big) - \delta\big(\xi + \frac{\omega}{2\pi}\big)\Big] \end{aligned} \tag{4}$$

$$|Y(\xi)| = \delta\big(\xi - \frac{\omega}{2\pi}\big) \tag{5}$$

79 Mean frequency can then be calculated as,

$$\begin{aligned} f_{\mathrm{mean}}(\omega) &= \frac{\int_0^\infty |Y(\xi)|^2\, \xi\, d\xi}{\int_0^\infty |Y(\xi)|^2\, d\xi} \\ &= \frac{\int_0^\infty \delta(\xi - \frac{\omega}{2\pi})\, \xi\, d\xi}{\int_0^\infty \delta(\xi - \frac{\omega}{2\pi})\, d\xi} \\ f_{\mathrm{mean}}(\omega) &= \frac{\omega}{2\pi} \end{aligned} \tag{6}$$

As expected, we get a formulation for the mean frequency as being directly proportional to ω.

Triangle Waveform

Now we can consider a waveform that represents a symmetric amperometry spike with a linear rise and fall. A triangular wave or triangle wave is a periodic, piecewise linear, continuous real function (also referred to as the hat function) such as that illustrated in Fig. 7.

$$y(t) = \begin{cases} 1 - \frac{|t|}{a}, & |t| \leq a \\ 0, & \text{otherwise} \end{cases}$$

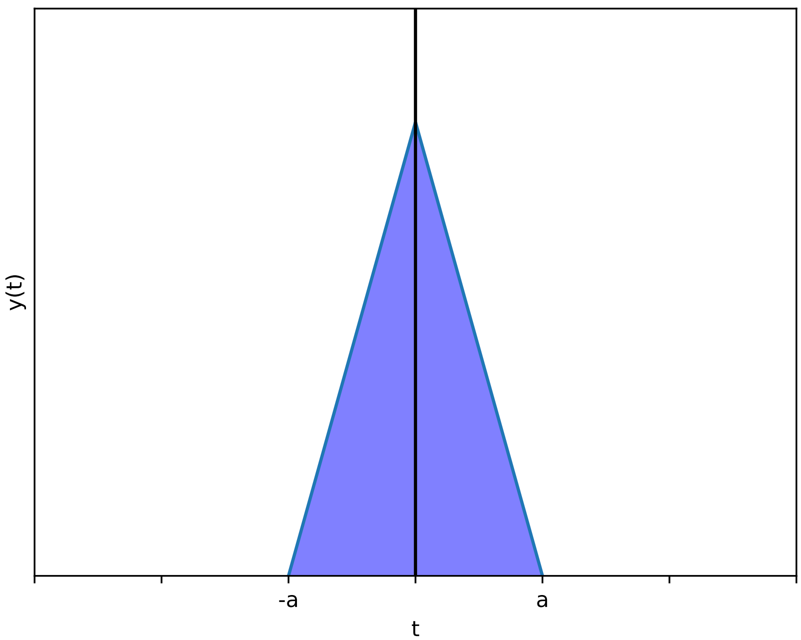


Figure 7. Illustration of a symmetric hat function in the time domain with base length $2a$.

Following the steps outlined in the previous subsection, we can find the relationship between mean frequency and length of the base of triangle as,

$$\begin{aligned} \mathscr{F}(y(t)) &= \int_{-\infty}^{\infty} e^{-2\pi i \xi t} dt \\ &= \int_{-a}^{0} \left(1 + \frac{t}{a}\right) e^{-2\pi i \xi t}\, dt + \int_{0}^{a} \left(1 - \frac{t}{a}\right) e^{-2\pi i \xi t}\, dt \\ \mathscr{F}(y(t) &= a^2 \operatorname{sinc}^2(a\, \xi) \end{aligned} \tag{7}$$

$$|y(\xi)| = 2a \operatorname{sinc}^2(a\xi) \tag{8}$$

Mean frequency,

$$\begin{aligned} f_{\text{mean}} &= \frac{\int_0^{\infty} |Y(\xi)|^2\, \xi\, d\xi}{\int_0^{\infty} |Y(\xi)|^2\, d\xi} \\ &= \frac{3}{a} \ln 2\pi \end{aligned} \tag{9}$$

Therefore, we can see that larger the a then wider is the spike, and hence lower is its mean frequency, as one might expect.

Code

The `frequency-analysis` package is implemented in Python3.2.0 and the most recent version can be downloaded from: https://github.com/JRC-COMBINE/SSFAAT. The package consists of four modules explained in Fig. 8.

params $\longrightarrow$ visualization $\longrightarrow$ automator $\longrightarrow$ main $\longleftarrow$ utils

Figure 8. Structure of the spike-based frequency analysis Package. All programs fetch parameters set by the use in the `params` module.

The `main` module uses the helper functions defined in `utils` to do spike-by-spike mean frequency evaluation for each category. Further, it shows a visualization of the distribution of mean frequencies. The user could change these parameters to analyze their data.

The `params` module defines the global parameters, such as the paths to raw data, interim data, sampling frequency and output path as outlined below in Fig. 9.

The `automator` module identifies spikes based on a pre-defined threshold and extracts spikes. category_list refers to a user-defined input that specifies the different experimental groups or conditions under comparison. This list is required to organize and label the data accordingly, allowing for the systematic analysis of spike properties across different categories (e.g., different experimental conditions or treatments).

A summary of flow of the program in the form of a pseudocode is outlined as follows:

Algorithm 1 The `automator` module executes the method `peak_detector_automated` using user-inputs including list of label categories, path to raw data, path to output files and peak threshold set in `params` module.

```
procedure PEAK_DETECTOR_AUTOMATED(category_list, raw_path_list, target_path_list, peak_threshold)
    for each category in category_list do
        raw_path_category = raw_path_list[category]
        for every file in raw_path_category do
            if data file not in exclusion_list then
                load raw data
                find standard deviation of baseline from first 30 data points
                calculate peaks based on peak_threshold; get peak properties and store
                if peaks exist then
                    calculate normalized cutoff position and apply time series filtering
                    extract t_max, I_start, I_end of the spikes
                    determine t_1/2, t_rise and t_fall window
                end if
                Write all extract spike parameters to file in target_path_list and store
            end if
        end for
    end for
    return output_file
end procedure
```

The `utils` module contains the helper functions for the FFT analysis and performs the following functions: 1. given the raw time series and sampling frequency, `meanfreq` evaluates the mean frequency 2. for each time series together with the pre-computed characteristic spike features of a given category, `analyze_meanfreq` evaluates the mean frequency of each individual spike event w.r.t. a given interested window (e.g. $t_{1/2}$ window) 3. The mean frequency is evaluated within the $t_{1/2}$ time window, which includes the most critical portion of the spike where it rises from its peak amplitude and falls back. This ensures that the frequency analysis reflects the core dynamics of the spike. 4. for a given simulated signal, `analyze_meanfreq` evaluates the mean frequency of each spike w.r.t. a given interested window and returns the median of mean frequencies 5. `resample` function resamples each spike from its original length to a new length scaled by a factor 6. `create_spike_train` generates an artificial spike train of given length and sampling frequency, wherein the spike geometry is specified in the supplementary section on artificial data generation. 7. The 'resample' function is used to extend the time range of the spikes to five times their original duration. This resampling is necessary to enhance the resolution in the frequency domain when applying FFT. By increasing the time range, we improve the granularity of the frequency components, ensuring a more detailed and accurate frequency analysis, particularly for spikes with lower frequencies or longer durations. In

```
// General parameters
num_categories  # number of categories or labels
category_list         # list of strings with label ids
path_raw_category    # path to raw data of each category
target_path_category # write path of the excel sheet that contain
                       spike features for each category

// FFT parameters
f_samp                          # sampling frequency

// Peak detection parameters
thrs                           # threshold for peak detection

// Other
exclusion_list  # list of file names with dubious data
```

Figure 9. Overview of parameters that requires user intervention in the `params` module.

the following, we will focus on the core method in this module, `analyze_meanfreq`:

Algorithm 2 The `utils` module executes the method `analyze_meanfreq` using user-inputs including the path to raw data directory and path to excel sheet.

procedure ANALYZE_MEANFREQ(path_raw, path_excel)
  initialize an empty 2-D array to write out mean_freq
  for every file in path_raw **do**
    read in time series from file
    read in spike parameters from path_excel
    read in start and end time of the interested window for each spike, t_start and t_end
    convert time parameters into array index, index_start and index_end
    determine window_data for each spike
    for each extracted spike **do**
      execute `meanfreq(window_data, f_samp)` method
        and store into mean_freq[file_index][spike_index]
    end for
  end for
  return `mean_freq`
end procedure

On raw data

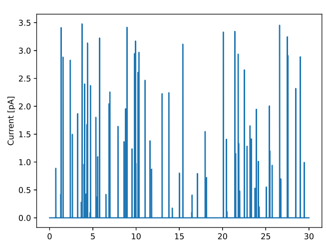
A summary of the attributes of the candidate datasets used in our analysis is given in Table 1. A short summary of the procedure adopted to generate these datasets is summarized in the following subsections:

Hofmeister Series Dataset

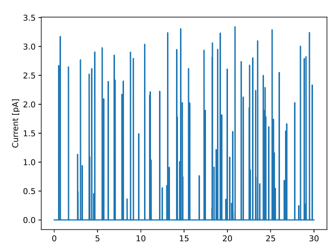
Bovine adrenal glands were obtained from a local slaughterhouse and the cells were kept at 37°C in isotonic solution during the whole experimental process. Electrochemical recordings from single chromaffin cells were performed on an inverted microscope in a Faraday cage. The working electrode was held at +700 mV versus an Ag/AgCl reference electrode and the output was filtered at 2.1 kHz and digitized at 10 kHz. For single-cell exocytosis, the micro-disk electrode was moved slowly by a patch-clamp micromanipulator to place it on the membrane of a chromaffin cell without causing any damage to the surface. Ten seconds after the start of recording, 30 mM K^+ stimulating solution in a glass micropipette was injected into the surrounding of the chromaffin cells with a single 30-s injection pulse.

Attribute	Hofmeister Series	Electrodes	DMSO
Method	SCA	VIEC	IVIEC
Cell/ vesicle type	Adrenal chromaffin cells	chromaffin vesicles	chromaffin cells
Conditions	K^+ and anions stimulation	Dipping Method	Control stimulated with Ba^{2+}, additionally incubated with 0.6% DMSO
Categories	Hofmeister Ions	Au, Pt, C	Control, DMSO
Length (sec)	40 – 126	105 – 1459	795 – 1544
Sampling Frequency (kHz)	10	10	10
# Samples	158	22	21
# Samples per category	• Br^- - 26 • Cl^- - 29 • ClO_4^- - 30 • NO_3^- - 31 • SCN^- - 27	• C - 4 • Pt - 7 • Au - 7	• Control - 16 • DMSO - 21

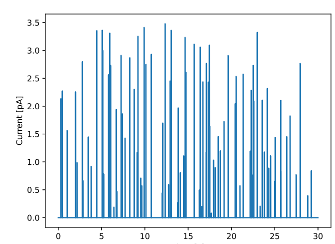
Table 1. Summary of attributes of candidate datasets



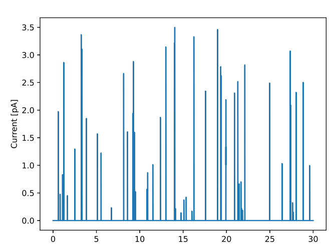
(a) Spike trains with Br^- stimulation



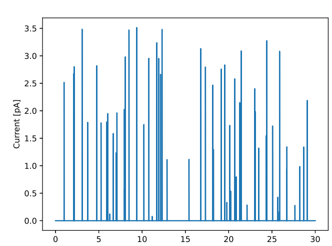
(b) Spike trains with Cl^- stimulation



(c) Spike trains with ClO_4^- stimulation



(d) Spike trains with NO_3^- stimulation

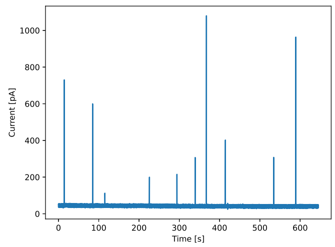


(e) Spike trains with SCN^- stimulation

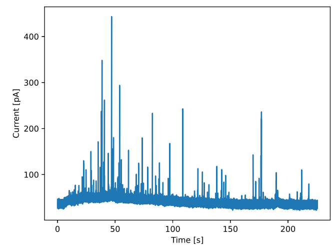
Figure 10. Amperometric traces of chromaffin cells under different ion stimulations obtained through SCA experiments

DMSO Dataset

Bovine chromaffin cells were isolated from the adrenal medulla by enzymatic digestion and the cells were kept at 37°C. Electrochemical recordings from single cells were performed on an inverted microscope, in a Faraday cage. The working electrode was held at +700 mV versus an Ag/AgCl reference electrode and the output was filtered at 2 kHz by using a Bessel filter. For cytometry recording, the tip of the nanoelectrode was inserted through the cell membrane with a patch-clamp micromanipulator. For exocytosis experiments, the nanotip electrode was positioned on top of the cell. Each cell was stimulated once with 2 mM Ba^{2+} for 5 seconds through the micropipette coupled to a microinjection system.



(a) Spike trains without DMSO incubation



(b) Spike trains with DMSO incubation

Figure 11. Amperometric traces of chromaffin cells under control conditions (no DMSO incubation) and with DMSO incubation obtained through IVIEC experiments

Electrodes Dataset

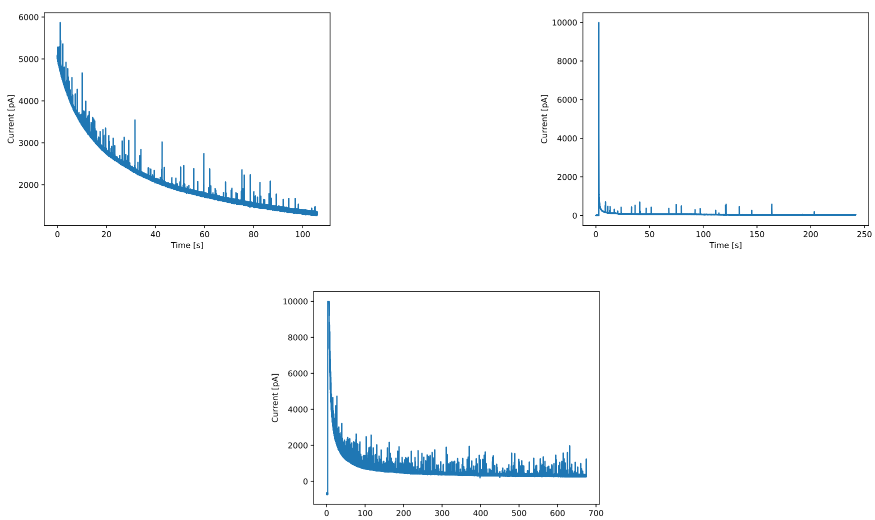


Figure 12. Amperometric traces of isolated chromaffin vesicles recorded at Au (top left), Pt (bottom) and Carbon (top right) disk microelectrodes at E = +700 mV vs Ag/AgCl obtained through VIEC experiments.

Bovine adrenal glands were obtained from a local slaughterhouse and transported in a cold Locke's buffer. Glands were trimmed of surrounding fat and rinsed through adrenal vein with Locke's solution. The medulla was detached from the cortex with a scalpel and then mechanically homogenized in ice-cold homogenizing buffer. The homogenate was centrifuged at 1000×g for 10 minutes to eliminate non-lysed cells and cell debris. After that, the supernatant was subsequently centrifuged at 10000×g for 20 minutes to pellet vesicles. All centrifugation was performed at 4°C. The final pellet of chromaffin vesicles was resuspended and diluted in homogenizing buffer for VIEC measurements.

For the VIEC experiments, the electrodes were first dipped in a vesicle suspension for 30 minutes at 4°C and then placed in homogenizing buffer for 20 minutes at 37°C for experimental recording. During the measurements, a constant potential of +700 mV vs. Ag/AgCl reference electrode was applied to the working electrode using a low current potentiostat (Axopatch

200B, Molecular Devices, Sunnyvale, CA, U.S.A). The signal output was filtered at 2 kHz using a 4-pole Bessel filter and digitized at 10 kHz using a Digidata model 1440A and Axoscope 10.3 software (Axon Instruments Inc., Sunnyvale, CA, U.S.A.).

For preparation of microelectrodes, a carbon fiber with 33 μm diameter was aspirated into a borosilicate capillary (1.2 mm O.D., 0.69 mm I.D., Sutter Instrument Co., Novato, CA, U.S.A.). The capillaries were subsequently pulled using a micropipette puller (Narishige Inc., London, U.K) and the carbon fiber was cut at the glass junction. The gap between the carbon fiber and glass was sealed by dipping the pulled tip in epoxy. The glued electrodes were placed in an oven at 100°C overnight to complete the sealing step. The sealed electrodes were beveled at 45° angle (EG-400, Narishige Inc., London, U.K.). A similar procedure was utilized for gold and platinum disk microelectrode fabrication. Here, either a 1-cm length of 125-μm-diameter Au wire or 100-μm-diameter Pt wire (Goodfellow, Cambridge Ltd. U.K.) that was connected to a longer piece of a conductive wire (silver wire, 10 cm) using silver paste was inserted into the pulled capillary and similarly sealed with epoxy and beveled at a 45° angle.

Artificial Data Generation

An artificial dataset to test our frequency analysis hypothesis for the Hofmeister series dataset was created in the following manner. First all time series in our artificial dataset are assigned a fixed length of $300,000$, i.e. 30 seconds recording time assuming 10 kHz sampling frequency. For each spike train, the number of spikes is randomly selected between $[50,100]$, and each such spike is assigned a random width depending on the artificial ion type, i.e. artificial $\mathbf{Cl}^-$ in $[10,20]$, artificial $\mathbf{Br}^-$ in $[20,30]$, artificial $\mathbf{NO}_3^-$ in $[30,40]$, artificial $\mathbf{ClO}_4-$ in $[40,50]$, artificial $\mathbf{SCN}^-$ in $[50,60]$, meant to mimic the observations in the real Hofmeister series dataset. Similar to real spike shapes, the artificial spikes consist of a steep linear rising segment and an exponentially decaying part. Finally, FFT was applied to each spike train and the statistics were pooled. For each category, 25 time series were generated.