


Comparison of Matching Pursuit Algorithm with Other Signal Processing Techniques for Computation of the Time-Frequency Power Spectrum of Brain Signals

Subhash Chandran KS,¹ Ashutosh Mishra,^{4*} Vinay Shirhatti,^{2*} and  Supratim Ray^{1,3}

¹Department of Electrical Engineering, ²Indian Institute of Science (IISc) Mathematics Initiative, Department of Mathematics, and ³Centre for Neuroscience, Indian Institute of Science, Bangalore, India, 560012, and ⁴Indian Institute of Science Education and Research (IISER), Kolkata, India, 741246

Signals recorded from the brain often show rhythmic patterns at different frequencies, which are tightly coupled to the external stimuli as well as the internal state of the subject. In addition, these signals have very transient structures related to spiking or sudden onset of a stimulus, which have durations not exceeding tens of milliseconds. Further, brain signals are highly nonstationary because both behavioral state and external stimuli can change on a short time scale. It is therefore essential to study brain signals using techniques that can represent both rhythmic and transient components of the signal, something not always possible using standard signal processing techniques such as short time fourier transform, multitaper method, wavelet transform, or Hilbert transform. In this review, we describe a multiscale decomposition technique based on an over-complete dictionary called matching pursuit (MP), and show that it is able to capture both a sharp stimulus-onset transient and a sustained gamma rhythm in local field potential recorded from the primary visual cortex. We compare the performance of MP with other techniques and discuss its advantages and limitations. Data and codes for generating all time-frequency power spectra are provided.

Introduction

Signals recorded from the brain often show rhythms at various frequencies, which are associated with distinct behavioral states (Buzsáki and Draguhn, 2004; Buzsáki, 2006). For example, alpha rhythm (8–12 Hz) is associated with a relaxed awake state (for example, if eyes are closed), and gamma rhythm (center frequency between 30–80 Hz) is related to high-level cognitive processes (Buzsáki, 2006; Fries, 2009). Such rhythms can be readily studied in the spectral domain by using the Fourier transform, which decomposes a signal of a certain duration as a sum of sinusoids of different amplitudes and phases (Fig. 1; see details below). Brain signals, however, change rather rapidly. For example, once a visual stimulus is presented, all visual areas, including high-level areas such as inferotemporal cortex, become activated in less than 150 ms (Schmolsky et al., 1998; Tamura and Tanaka, 2001). Indeed, even during fixation, visual information changes with every microsaccade approximately three times every second (Bosman et al., 2009). Therefore, spectral analysis should be done only on short segments of the signal at a time, which is made possible using techniques such as short time fourier transform (STFT; see details below). However, a fundamental problem in

this case is that time and frequency components of a signal cannot be determined with arbitrary precision—improving the spectral resolution beyond a limit can only be achieved at the expense of temporal resolution (and vice versa). Several different techniques have been used to optimize the time–frequency trade-off. In this review, we first introduce different techniques using a common filter-bank interpretation (Allen and Rabiner, 1977; Portnoff, 1980; Smith and Barnwell, 1987; Vetterli, 1987), which allows a direct comparison of these methods in the time–frequency space. Next, we apply these techniques to the same signal so that the pros and cons of each can be analyzed directly. Finally, we introduce the matching pursuit (MP) algorithm and discuss some properties that allow us to improve the spectral and temporal resolution, and compare it with other methods.

We applied all of the discussed signal processing techniques to a local field potential (LFP) signal obtained from the primary visual cortex of a monkey while a large grating (radius of 2.4°, 4 cycles per degree, ~100% contrast, six different orientations separated by 30°; a total of 186 trials) was presented for 400 ms with an interstimulus interval of 600 ms (for task and methodological details, see Ray and Maunsell, 2011). Around each stimulus onset, we took a segment between –1.1475 and 0.9 s (total signal duration, 2.048 s; sampled at 2 kHz) and analyzed using five analysis techniques: STFT, multitaper method (MTM), wavelet transform (WT), MP, and Hilbert–Huang transform.

Fourier transform

Spectral analysis has traditionally been done using the Fourier transform (FT), which represents a signal as a sum of sinusoids at

Received Oct. 2, 2015; revised Nov. 19, 2015; accepted Dec. 18, 2015.

This work was supported by Wellcome Trust/DBT India Alliance (Intermediate Fellowship to S.R.), Tata Trusts Grant, and DBT-IISc Partnership Programme. We thank Dr. John Maunsell for providing data for analysis.

*A.M. and V.S. contributed equally to this work.

Correspondence should be addressed to Supratim Ray, Centre for Neuroscience, Indian Institute of Science, Bangalore, India, 560012. E-mail: sray@cns.iisc.ernet.in.

DOI:10.1523/JNEUROSCI.3633-15.2016

Copyright © 2016 the authors 0270-6474/16/363399-10\$15.00/0

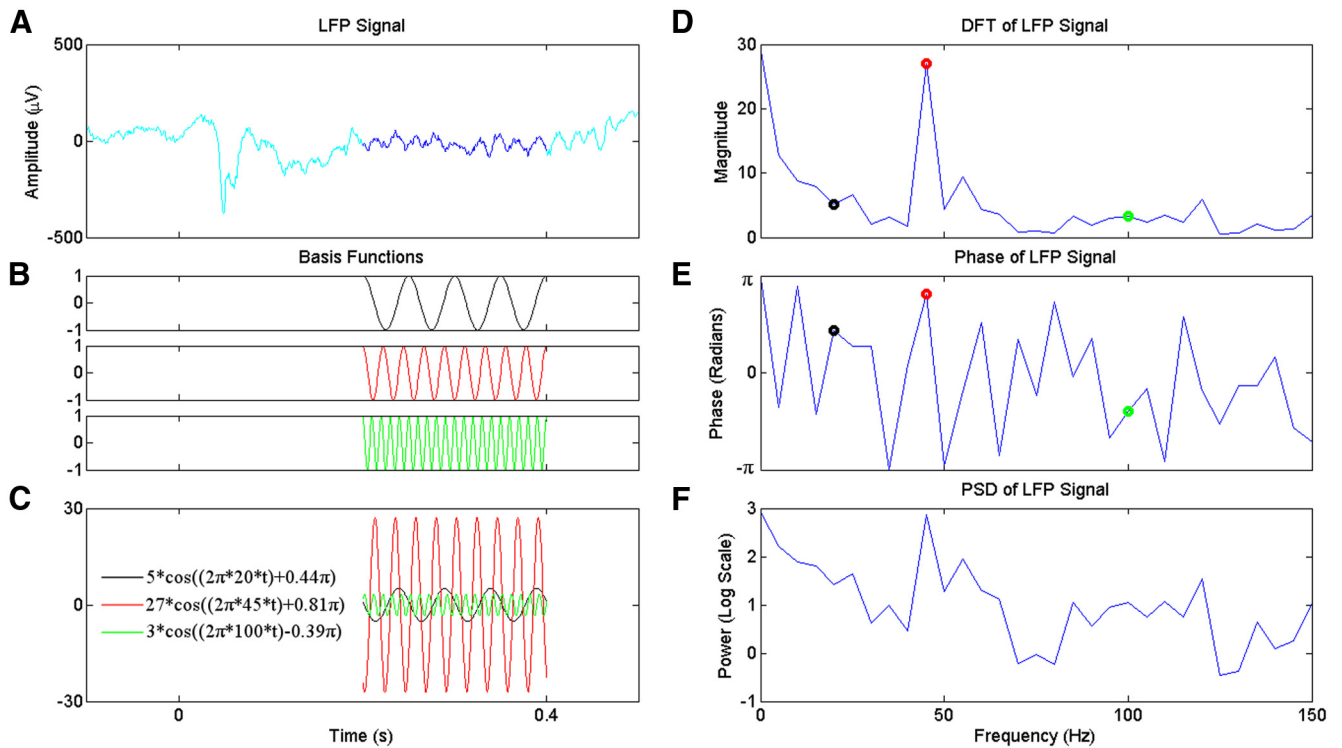


Figure 1. FT analysis of the LFP signal. **A**, A sample LFP signal recorded from primary visual cortex. The segment taken for FT analysis is highlighted in dark blue. **B**, Examples of basis functions available for the FT analysis at 20 (black), 45 (red), and 100 (green) Hz. **C**, Scaled and phase-shifted versions of the basis functions (shown in **B**). Scaling factors and phases are obtained from FT. **D**, **E**, Magnitude (**D**) and phase (**E**) of the FT calculated on the LFP segment highlighted in **A**. The magnitude and phase of the basis functions are marked in their corresponding colors and used in the equations shown in the legend of **C**. **F**, PSD of the LFP (in logarithmic scale). Note that we describe FT (and all other methods) in continuous time even though we perform the analyses on discrete signals (and therefore use DFT).

different frequencies. Equation 1 represents the FT of a time domain signal, $f(t)$, where ω is the angular frequency in radians per second (Cohen, 1995).

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int f(t)e^{-j\omega t} dt \tag{1}$$

For example, consider the signal shown in Figure 1A and suppose that we take the FT of the segment between 200 and 400 ms, which shows a prominent oscillation [shown in blue; note that because this signal is obtained by sampling the original continuous signal at some fixed rate (2000 samples/s in this case), we actually take the discrete Fourier transform (DFT) of the signal, but in this manuscript we discuss all the methods in continuous time domain only]. According to the theory of Fourier transform, any finite energy signal can be decomposed as a linear combination of sine and cosine functions at different frequencies, which are therefore the basis functions (by definition, any function in the function space can be represented as a linear combination of what are called “basis” functions of that space). Instead of using sines and cosines with their corresponding scaling factors, we can equivalently use a cosine function with a scaling factor and a phase. For example, Figure 1B shows basis functions at 20 (black), 45 (red), and 100 Hz (green), while Figure 1C shows these basis functions after appropriate scaling and phase shifting (such that the original signal is the sum of all the scaled and phase-shifted basis functions). The FT of a signal (Eq. 1) is a complex number that represents, at each frequency (ω), the corresponding scaling factor (Fig. 1D) and the phase (Fig. 1E). In formal terminology, if we define the inner or dot product of two complex functions $f(t)$ and $g(t)$ as $\langle f, g \rangle = \int f(t)g^*(t)dt$ (where $*$ is

the complex conjugate), then FT is simply an inner product (or a projection) of the signal with a complex exponential: $\langle f, e^{j\omega t} \rangle$, and the magnitude of this inner product at each frequency (Fig. 1D) gives the relative contribution of that frequency to the signal.

The energy of the signal, defined as $\int |f(t)|^2 dt$, is preserved after the FT operation (total signal energy = $\int |f(t)|^2 dt = \int |\hat{f}(\omega)|^2 d\omega$; Parseval’s theorem). The square of the absolute value of the FT amplitude, $|\hat{f}(\omega)|^2$, is called the power spectral density (PSD) and represents the power concentrated at different frequencies in the signal (Fig. 1F; shown in a log scale because otherwise the power at higher frequencies, which typically is only a small fraction of the total signal power, is difficult to observe).

FT is a powerful tool that can highlight the dominant frequency components present in a signal (for example, the oscillation in Fig. 1A is captured as a salient peak at ~ 45 Hz in Fig. 1D,F). However, as the FT computes the overall amplitude/phase at a frequency by integrating over the entire signal duration (200 ms in this case), it does not indicate how this amplitude or phase may be changing with time.

Short-time Fourier transform

To study the spectral properties of the signal as a function of time, we used STFT (Ackroyd, 1971; Allen, 1977; Allen and Rabiner, 1977; Portnoff, 1980; Cohen, 1995), where the signal is essentially broken down into short time segments of equal size and Fourier analysis is carried out on each segment separately. Breaking the signal into short segments is achieved through windowing, which involves multiplying translated versions of a window function with the signal. The STFT of the signal $f(t)$ is computed as

$$\hat{f}_h(\tau, \omega) = \int f(t)h(t - \tau)e^{-j\omega t} dt \tag{2}$$

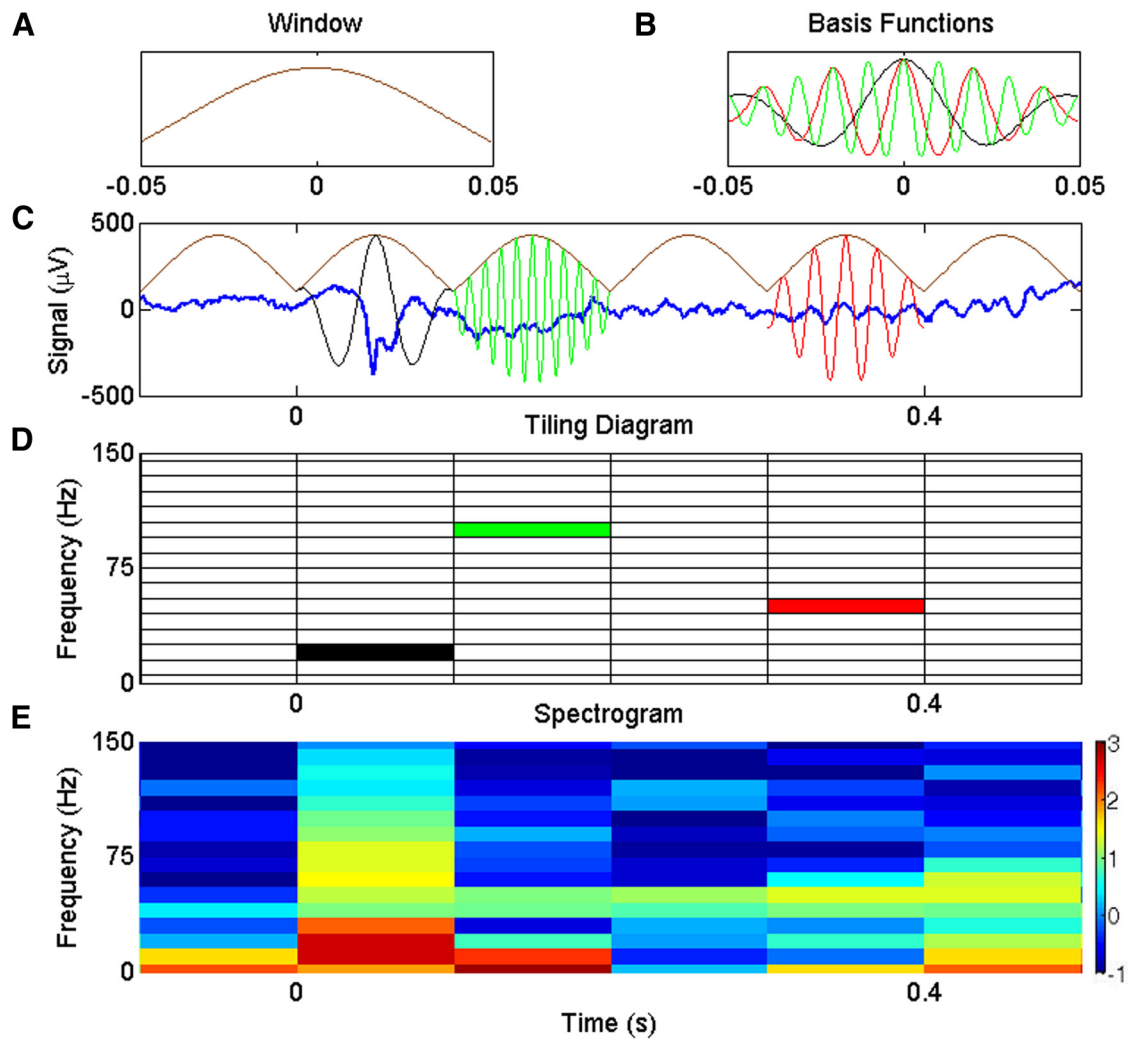


Figure 2. Signal decomposition using STFT. **A**, Zeroth-order Slepian taper of 100 ms duration. **B**, Three examples of basis functions available in the STFT analysis with center frequencies of 20 (black), 50 (red), and 100 (green) Hz. **C**, A sample of single-trial LFP signal (blue) recorded from primary visual cortex, along with the windowing operation using a 100-ms-long Slepian window (brown). The window is translated with a time period of 100 ms. Basis functions in **B** are shown in the corresponding colors (the window and the basis functions are not to scale). **D**, Time–frequency tiling diagram of STFT. The tiles corresponding to the three basis functions are shown in their respective colors. **E**, Spectrogram of LFP signal in using STFT with the windowing procedure represented in **C** (in logarithmic scale).

The time–frequency energy density spectrum, also called the spectrogram, is obtained as the magnitude-square of the STFT, $|\hat{f}_h(\tau, \omega)|^2$. The energy of STFT of the signal is the product of the energies of the window function and the signal (Cohen, 1995), so this transformation satisfies energy conservation if the window has unit energy. However, the spectrogram is directly influenced by the spectrum of the window function, as explained below.

Filter-bank interpretation

Equation 2 can be rewritten in the inner product form as

$$\hat{f}_h(\tau, \omega) = \langle f(t), h(t - \tau)e^{j\omega t} \rangle \quad (3)$$

In this form, the STFT can be thought of as an inner product between the signal and a series of windowed sinusoids given by $h(t - \tau)e^{j\omega t}$, which are the basis functions onto which the signal is projected. For example, Figure 2A shows a zero-order Slepian (also known as a discrete prolate spheroidal sequence, or DPSS) window (Slepian, 1978; Moore and Cada, 2004) of 100 ms duration (this particular window is used in MTM, as explained in more detail below), while Figure 2B shows the corresponding

basis functions with three different center frequencies at 20 (black), 50 (red), and 100 Hz (green). The spectrum of such windowed sinusoids is obtained by convolution of the spectra of the sinusoids (which is a delta function) and the window function, which, by the sifting theorem, is the spectrum of the window centered at ω . Since the energy of the windowed sinusoid is concentrated in a narrow band around ω , it can be thought of as a bandpass filter (a filter that only allows frequencies in a particular band to pass) centered at ω . Overall, STFT can be thought of as a series of bandpass filtering operations (Allen, 1977; Allen and Rabiner, 1977; Portnoff, 1980). Each filter has the same bandwidth since window length remains unchanged, so it is also called uniform filter-bank analysis.

For example, Figure 2C shows the same sample LFP signal as Figure 1A, along with translated versions of the window function (brown) and example copies of the basis functions from Figure 2B. The window is shifted with a time period of 100 ms (i.e., τ is in multiples of 100 ms), so there is no overlap between successive windowed sections in this case, although this is just for the sake of illustration (see below).

We can represent these filters in the time–frequency plane as tiles, as illustrated in Figure 2D (the three basis functions in Fig. 2C are highlighted in their corresponding colors). The position of the tile on the frequency axis is decided by the frequency of the sinusoidal component (ω) and on the time axis by the parameter τ in Equation 2, which is also called a translation parameter because it determines how much the window function is translated (or shifted). The width and height of each tile remains the same on the entire time–frequency plane for STFT. The spectrogram (on a logarithmic scale), which is simply the square of the magnitude of the projection on these basis functions, is shown in Figure 2E. The transient activity related to stimulus onset can be seen between 0 and 100 ms. Note that although we show only a window-modulated cosine basis function in Figure 2C, there are actually two orthogonal basis functions—a sine and a cosine, inside each tile. The amplitude and phase of the STFT are obtained from the projections of the signal on both these functions.

Time–frequency uncertainty principle

Time–frequency uncertainty principle puts a lower limit on the area of the tiles shown in Figure 2D: shortening the width horizontally (improvement in time resolution) necessarily leads to an increase in the vertical length (degradation of frequency resolution), and vice versa.

The area of each time–frequency tile, at first glance, appears to depend only on the signal length over which FT is computed. For example, if the analysis window is 100 ms long, Fourier coefficients are obtained for every 10 Hz ($1/T$, where T is the duration of the signal in seconds; also called the Rayleigh frequency), such that each tile has an area of 1, regardless of the window being used. The frequency resolution is such that the underlying sinusoids become orthogonal (sinusoid of $1/T$ Hz has exactly one cycle in duration T , $2/T$ Hz has two cycles and so on; therefore, sinusoids of frequencies 0, $1/T$, $2/T$, etc. form an orthogonal set).

The time–frequency uncertainty limit, which refers to the uncertainty in the localization of the windowed basis function in both spectral and temporal domains, depends on the shape of the window as well. Formally, given any signal $s(t)$, the total energy of the signal can be computed as either $\int |s(t)|^2 dt$ or $\int |S(\omega)|^2 d\omega$ (Parseval's theorem). If we normalize the signal such that it has unit energy, $|s(t)|^2$ and $|S(\omega)|^2$ can be thought of as probability distribution functions representing how the signal energy is spread in time and frequency domains. Standard deviation or duration of a signal, σ_t , and root mean square bandwidth, σ_ω , can be computed as

$$\sigma_t = \sqrt{\langle t^2 \rangle - \langle t \rangle^2}, \sigma_\omega = \sqrt{\langle \omega^2 \rangle - \langle \omega \rangle^2} \quad (4)$$

In Equation 4, $\langle t \rangle = \int t |s(t)|^2 dt$, $\langle \omega \rangle = \int \omega |S(\omega)|^2 d\omega$, $\langle t^2 \rangle = \int t^2 |s(t)|^2 dt$, and $\langle \omega^2 \rangle = \int \omega^2 |S(\omega)|^2 d\omega$. The time–frequency uncertainty principle states that the time–frequency bandwidth product can never be less than one-half; that is, $\sigma_t \sigma_\omega \geq \frac{1}{2}$ (Cohen, 1989, 1995).

Regarding the resolution, there are three points worth noting. First, the size of the tile depends on the choice of the window. Time–frequency bandwidth product reaches the theoretical limit of one-half only for a Gaussian window (Gaussian tiles are smallest; Cohen, 1995). However, the energy of the basis function inside a tile is not restricted to the tile alone—some of the energy spreads beyond the tile (formally called “spectral leakage”). Different windows are, therefore, used to satisfy different resolution and spectral leakage constraints (Oppenheim and Schaffer, 2013). For example, given a spectral bandwidth σ_ω , the Slepian window

shown in Figure 2A has the highest spectral concentration within the bandwidth. In other words, even though the Slepian tile is slightly larger than a Gaussian tile, higher percentage of the total window energy is inside the tile. Most of the commonly used windows have good spectral concentration, so the choice of window is not as critical a factor as the window length as far as resolution is concerned. We have used a Slepian window for STFT computation only for comparison with MTM; other windows give similar results.

Second, moving the window in small steps increases the number of time points in the STFT, but it does not improve the temporal resolution. This is because the underlying basis functions are not orthogonal when they are slightly shifted from each other, so that the projections are highly correlated. In effect, it only smooths the STFT. The temporal resolution is determined by window properties (length and type) only.

Finally, zero padding (a procedure in which a series of zeros are appended to a signal to increase its length) before computing the spectrum does not improve the frequency resolution. Even though the signal is made longer by adding zeros and appears to have higher spectral resolution, the resulting spectrum is that of a different signal (composed of the original signal and a series of zeros). Specifically, the zero-padded signal is the product of the true signal (of infinite duration, whose PSD we are interested in) and a “rect” function (which takes a value of one for the duration of the observed signal and zero outside). The resulting spectrum is therefore the convolution of the true spectrum and the spectrum of a rect function (a sinc function). At best, zero padding only smooths the spectrogram (Kay and Marple, 1981).

Figure 3, A and B, show STFT analysis using short (50 ms) and long (200 ms) windows, respectively (with $\tau = 1$ ms). The first column shows two example basis functions of center frequencies 100 and 50 Hz, respectively, while the second column shows the tiling diagram [for simplicity, in all the tiling diagrams used in this review, we have used the window length (T) as the tile width and the corresponding Rayleigh frequency ($1/T$) as the tile height, regardless of the window used, such that the area of each tile is 1]. The plots on the right show time–frequency power spectrum in logarithmic scale for a single trial (third column), average spectrum across 186 trials (fourth column), and the change in power from baseline (0.3 to 0.1 s before stimulus onset; fifth column). The spectrogram computed using the short window shows the transient activity with high precision, but gamma rhythm and the artifact due to second harmonic of the line noise (120 Hz) are weak or absent (Fig. 3A). However, spectrogram computed using the long window shows a salient gamma rhythm (Fig. 3B). Even the 120 Hz line noise artifact can be seen in the trial-averaged spectrogram. The transient activity is, however, spread out in time. Thus, capturing both the transient and rhythmic activities simultaneously with good resolution is not possible with STFT.

Multitaper method

The spectrogram of a signal (as in Fig. 2E) yields a single value for each tile that corresponds to the square of the inner product of its basis function and the signal (magnitude is represented by the color), but this value has high variability (Brillinger, 1972; Jarvis and Mitra, 2001; Babadi and Brown, 2014). Therefore, typically, a large number of signal repetitions are required to get a more reliable estimate of the true spectrogram (for example, 186 spectrograms of the type shown in the third columns of Fig. 3, A and B, and averaged to get the fourth columns). However, sometimes

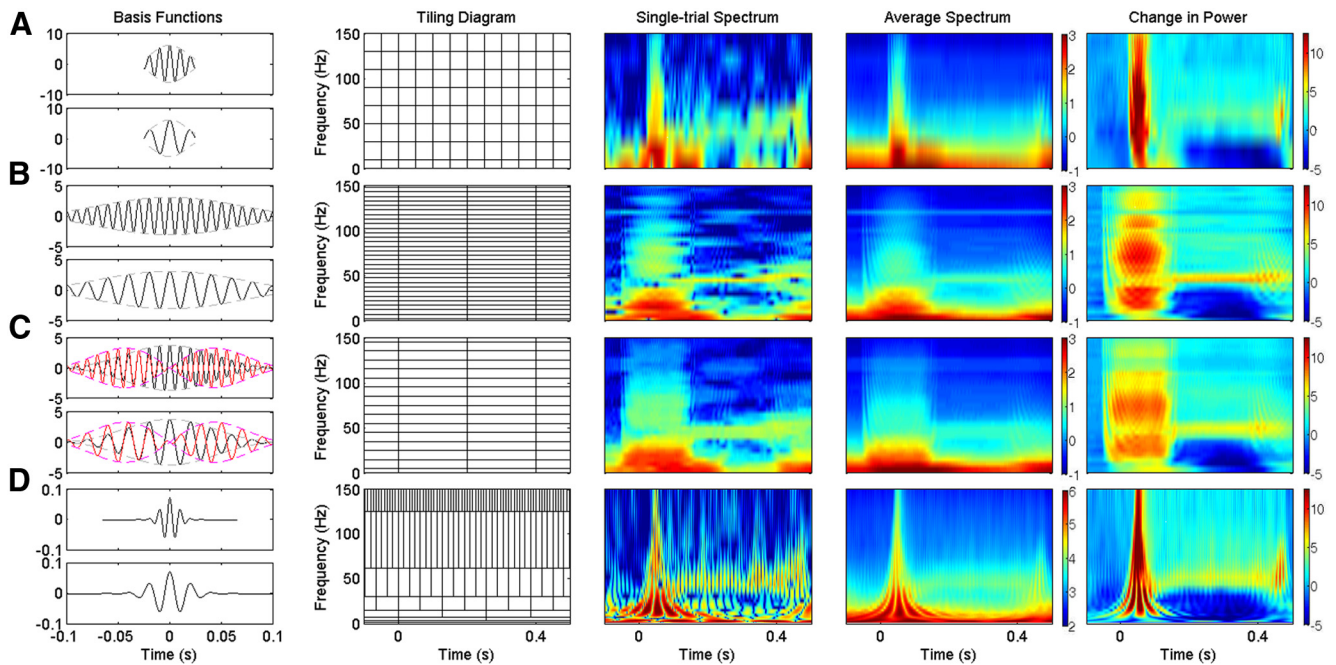


Figure 3. Comparison of different spectral analysis techniques. From left to right; basis functions at 100 (top) and 50 (bottom) Hz, time–frequency tiling diagram, and time–frequency power spectra (in logarithmic scale) of a single-trial LFP signal, averaged across 186 trials, and change in power from the baseline period. **A–D**, STFT with short zeroth-order Slepian window (50 ms; **A**), STFT with long zeroth-order Slepian window (200 ms; **B**), MTM with three tapers of 200 ms duration (**C**), and WT using Morlet wavelet (**D**). See text for details. Note that the absolute values of the spectra (third and fourth columns) are different for STFT/MTM and WT because of differences in implementation software, but the range is the same (4 orders of magnitude). The change in power (fifth column), measured in decibels, is in the same range.

only a few signal repetitions are available, yielding a noisy estimate of the true spectrogram.

MTM (Thomson, 1982; Park et al., 1987; Bokil et al., 2010; Babadi and Brown, 2014) is a technique that allows us to increase the size of the tile such that more than one basis function can be placed inside each tile. These functions are orthogonal, so they provide independent estimates of the inner product. For example, if the size of the tile is increased by a factor of 3, 5 different functions can be placed inside the tile (with energy localized reasonably well inside). We can therefore trade-off the resolution (size of the tile) with the reliability of the estimate (number of basis functions that get averaged).

We performed MTM analysis using three windows (called tapers) corresponding to a time–frequency bandwidth product (TB) of 2 (Fig. 3C; for a formal description, see Methodological details, below), with the same length of window as used in STFT analysis (200 ms). Examples of basis functions are shown in the first column of Figure 3C, which are sinusoids modulated by the tapers used (lightly shaded envelopes). Black and red waveforms show basis functions corresponding to zeroth- and first-order Slepian tapers with center frequencies of 100 Hz (top) and 50 Hz (bottom), respectively. The time–frequency energy spectrum using MTM has poorer spectral resolution than the corresponding plots using STFT (Fig. 3, compare **C** with **B**) that makes it harder to see the gamma rhythm or the line noise artifact at 120 Hz. The MTM spectrogram has lower variability than STFT (the single-trial estimate shown in the third column is closer to average estimate shown in the fourth column for MTM), although it is difficult to observe this in these plots.

Because a large number of stimulus repetitions (186) are available, there is little benefit of using MTM as compared with STFT in this case. Indeed, given the transient nature of brain signals and technical difficulties in improving frequency resolution, it is ad-

visable, wherever possible, to collect a large number of repeats of a stimulus or behavioral condition rather than sacrificing frequency resolution to improve the spectral estimate. Moreover, since MTM is also a uniform filter-bank approach, all the tiles covering the time–frequency space have the same shape and cannot properly represent both transient and rhythmic components. The key for improving the resolution is to use tiles of different shapes to cover the time–frequency space, as discussed below.

Wavelet transform

If a signal has transients at high frequencies and sustained oscillations at low frequencies, it might be better to use different types of tiles at low and high frequencies. Such a representation is made possible through the WT (Daubechies, 1990; Vetterli and Herley, 1992; Mallat, 2008), where we can use tiles with good frequency resolution at lower frequencies and good time resolution at higher frequencies, as shown in the tiling diagram in Figure 3D (for a formal description, see Methodological details, below).

To illustrate the spanning of the time–frequency plane by wavelet basis functions, we used the Morlet wavelet (Gaussian modulated sinusoids). Basis functions corresponding to a center frequency of 100 and 50 Hz, which are scaled versions of the same function, are shown in the top and bottom panels in the first column of Figure 3D. Like STFT and MTM, WT is also filter-bank analysis, but the bandwidth of each filter varies with center frequency.

In our WT analysis of the LFP signal, we used a dyadic decomposition scheme where we chose desired center frequencies of the wavelet that were equal to the sampling frequency divided by powers of two (i.e., $F_s/2$, $F_s/4$, $F_s/8$, ..., where F_s is the sampling frequency) and computed the appropriate scale values for which the wavelet coefficients were calculated (see the Matlab code available at <https://github.com/supratimray/SpectralAnalysis>

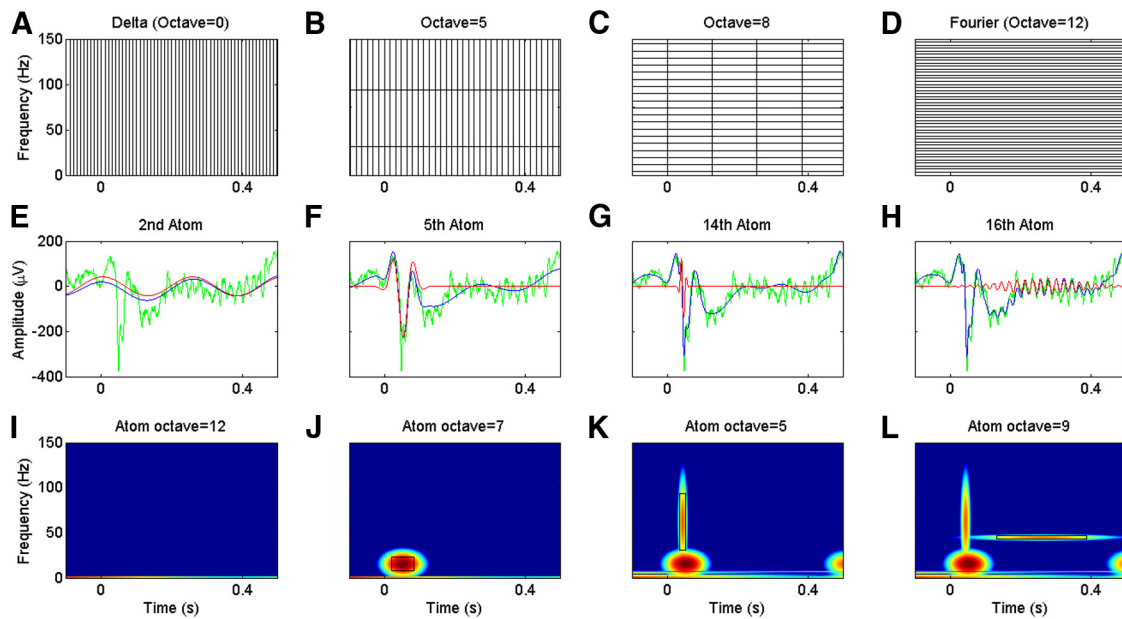


Figure 4. Explanation of the MP algorithm. **A–D**, Time–frequency tiling diagram of Dirac representation (**A**), Gabor function with a scale value of 32 (**B**), Gabor function with a scale value of 256 (**C**), and Fourier representation (**D**). All of these functions are members of the dyadic dictionary used in the MP analysis. **E–H**, Green traces are a single-trial LFP signal (same in all plots); red traces are the atoms selected at second (**E**), fifth (**F**), 14th (**G**), and 16th (**H**) iterations; blue traces are partial signal approximation up to that iteration. **I–L**, The WVD-derived time–frequency spectra up to second, fifth, 14th, and 16th iterations of MP algorithm in logarithmic scale. The selected atoms are highlighted by a black rectangle.

Codes for details). In the time–frequency power spectra, higher-frequency components of the transient related to the stimulus onset are fairly well localized in time, but lower-frequency components are not. Gamma rhythm is observed at ~ 50 Hz, but the frequency resolution is poorer. The 120 Hz noise is completely missing in these plots.

Even though different tiles are available in WT, their position in the spectrum is fixed with respect to frequency, such that it is not possible to capture transient activity at low frequencies or rhythmic activity at high frequencies (such as the line noise harmonic at 120 Hz). Thus, we need an adaptive spectral analysis technique that provides a customized tiling scheme to represent LFP signals effectively in the time–frequency plane.

Matching pursuit algorithm

Figure 4, **A–D**, shows four different ways of tiling the time–frequency space. Figure 4**A** involves decomposition using delta functions, which has the best temporal resolution but no frequency information, while Figure 4**D** is a Fourier decomposition, which gives the highest spectral resolution but provides no temporal information. Figure 4, **B** and **C**, correspond to STFT decomposition using short and long windows. As discussed earlier, while each one of these is a valid way of dividing the time–frequency space, both transient activity and rhythmic activity cannot be adequately captured with any of these decompositions.

MP (Mallat and Zhang, 1993) is a technique that allows us to potentially use any tile from a large number of different types of STFTs, including the extreme cases of delta and Fourier transform, to represent the signal. This involves first creating a large and redundant dictionary of tiles of different sizes, including all the tiles shown in Figure 4, **A–D**, as well as other tiles of intermediate shapes. Then the tile that explains the maximum energy of the signal (has the largest inner product) is picked up. The explained portion is removed from the signal and the process is repeated. Eventually this method covers the time–frequency

space with customized tiles of different shapes depending on the properties of the signal itself.

We performed 500 iterations of the MP algorithm (for a formal description, see Methodological details, below). In the time–frequency representation, 11 types of tiles with different scales (ranging from 2 to 2048 samples in the time domain) were available, out of which two are shown in Figure 4, **B** and **C** (corresponding to scales of 2^5 and 2^8 samples, or 16 and 128 ms). Delta and Fourier decompositions (Fig. 4**A**, **D**), which can be thought of as extreme cases with scales of 1 and 4096, were also available as basis functions. Unlike WT, these tiles can occupy any position on the time–frequency plane and their bandwidth does not depend on the center frequency [note that we have assumed that the width of the tile is equal to the scale (s) in the time domain and its inverse ($1/s$) in the frequency domain to be consistent with the tiling scheme used in STFT].

Figure 4, **E–H**, show the LFP signal (green), the selected basis function (also called an “atom”; red), and the reconstructed signal at the n^{th} iteration of MP (blue). Corresponding time–frequency energy distributions up to the n^{th} iteration are shown in Figure 4, **I–L**. Here, the tile corresponding to the n^{th} atom is shown in a black rectangle (the octave number is shown on top). The atom shown in Figure 4, **G** and **K**, captures part of the stimulus onset transient, while the atom in Figure 4, **H** and **L**, captures part of the gamma rhythm.

Figure 5, **A–C**, show the time–frequency spectra obtained using MP for a single trial, averaged across all 186 trials and change in power from baseline, respectively. In these plots, gamma, transient activity, and the artifact at 120 Hz can be seen with high precision. The artifact due to flickering of the monitor (100 Hz) can also be seen clearly in Figure 5**B**, which is not captured by any of the other methods described above. This is because the noise and monitor artifacts are represented by sinusoids that are 2.048 s long and therefore have a resolution of ~ 0.5 Hz.

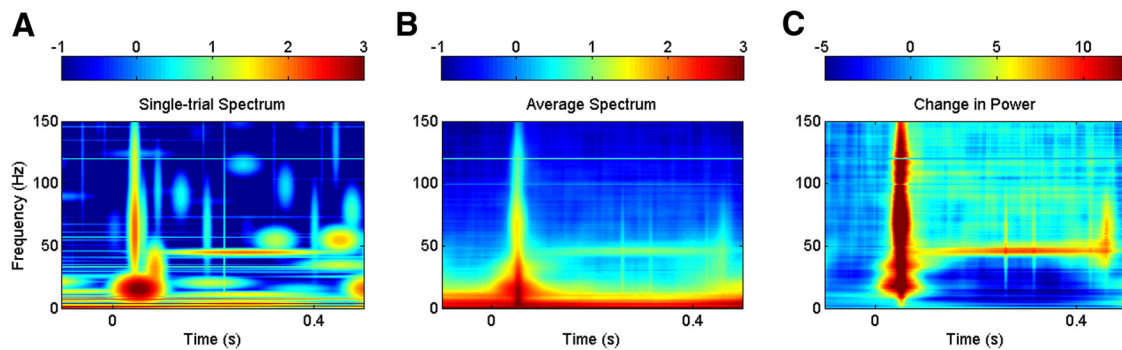


Figure 5. The time–frequency power spectra using MP decomposition in logarithmic scale. **A–C**, The spectrum of single-trial LFP signal (**A**), averaged across 186 trials (**B**), and change in power from the baseline period (**C**).

Apart from using customized tiles, there are two other forms of customization that improve resolution. First, in STFT, MTM, and WT, the same set of tiles is used for all the stimulus repeats. Therefore, when we average across repeats, we only average the values in each tile. In MP, because each repeat is potentially associated with a different set of tiles, averaging across repeats also averages out the shapes of the tiles. Second, in other methods, smoothing is achieved by taking small steps in the time domain (as explained above), but this form of smoothing depends on the shape of the window and their temporal overlap, which are fixed parameters. In MP, if the number of iterations exceeds the number allowed by the time–frequency uncertainty principle, it also leads to smoothing, but here even the smoothing is customized. These customizations allow us to capture stimulus onset transients, rhythms, and stimulus artifacts at high resolution without any violation of the time–frequency uncertainty principle.

There are some important issues and limitations to MP. First, the choice of the dictionary elements is left to the user and therefore the dictionary elements should be selected so as to suit the signal to be analyzed. In an earlier study, we compared the performance of MP with different types of dictionaries and showed that the averaged time–frequency energy spectrum does not critically depend on the choice of dictionary, at least for the data used in that study (Ray et al., 2003). Second, since MP chooses atoms in a greedy fashion, if the algorithm chooses a wrong atom in an iteration, it goes on correcting this mistake in the subsequent iterations (this problem is discussed with an example in Chen et al., 2001). Algorithms like basis pursuit (Chen and Donoho, 1994; Chen et al., 2001), which is a non-greedy approach, can be explored to address this issue. Third, MP requires a large convergence time (Pati et al., 1993). Finally, if two simultaneously recorded signals are decomposed using MP, it is difficult to compare phase coupling across frequencies because a different set of atoms are used in the decomposition of each signal. This is particularly important in neurophysiological studies, where the consistency of phase relationships at a particular frequency across repeats (which can be quantified using a metric called coherence) is often of primary importance. In such cases, either methods such as MTM could be used, or MP can be used to first decompose the signal in different frequency bands by reconstructing parts of the signal from atoms that have center frequency in a specified range (Ray et al., 2008), followed by an estimation of the phase consistency of the filtered components (which are now in the same frequency range). Another option is to perform MP simultaneously on multiple channels to find common signal structures (which would be represented by a common

atom) present in multiple channels (called multichannel matching pursuit; Durka et al., 2005).

The methods described so far use predefined basis functions and the projections of the signal (or the residue, as in MP) on them in the process of signal decomposition. We show that MP is particularly well suited to represent the various temporal and spectral characteristics of the LFP signal. In the documents provided with the codes and data (see Methodological details, below), we discuss an empirical method, the Hilbert–Huang transform, which does not use predefined basis functions and the corresponding signal projections, but instead decomposes signals using an adaptive technique.

Conclusion

This study compares popular spectral analysis methods with MP algorithm in studying signals like LFP, which have components occurring at different instants and at varied time scales (or frequency). To extract these features faithfully, a method with good localization capability (or resolution), flexibility in resolution across the time–frequency plane, and an adaptive way of decomposing the signal is required. MP, which combines the robust theoretical structure of Fourier-based methods with an adaptive technique to select the basis functions, satisfies these properties and yields meaningful time–frequency representation of the signals.

The main strength of MP is its ability to represent sharp transients in the signal. Since such transients have been difficult to capture using traditional signal-processing methods, a paradigm where the behavioral state of the subject is held constant for several seconds (for example, by asking the subject to maintain fixation at a point on the screen while visual stimuli are presented for long durations) is typically used, and the first few hundred milliseconds after stimulus onset are simply excluded from analysis to avoid response transients. While this paradigm is useful for studying sustained oscillatory responses, it does not capture the complexity of natural vision where we make eye movements every ~ 300 ms (that presents a new image on the retina; Bosman et al., 2009) and significant processing occurs within the first few hundred milliseconds of stimulus onset. MP has been shown to be very useful in the analysis of gamma oscillations generated by stimuli that are presented for short durations in rapid succession (Ray et al., 2013). Another situation where MP is particularly useful is in the study of spike–LFP interactions, because MP can capture the sharp transient associated with a spike by a Gaussian with a width of a few milliseconds (Ray et al., 2008, 2011; Ray, 2015). However, when focusing on the relationship between the phases of sustained oscillations in different brain

areas, which requires estimation of coherence, methods such as MTM are more useful. Overall, adaptive methods such as MP are better suited for analyzing highly nonstationary signals, and development of such adaptive techniques that have less *a priori* assumptions regarding the nature of the signal is key towards understanding cognitive functions from brain signals.

Methodological details

Multitaper method

STFT spectral estimates have high variability (which does not decrease even if the window length is increased), because essentially we are trying to estimate a continuous function (an infinite set of points) with finite data (Brillinger, 1972; Jarvis and Mitra, 2001; Babadi and Brown, 2014). MTM, introduced by David J. Thomson in 1982 (Thomson, 1982), provides a trade-off between spectral resolution and variance. Further, the window (taper) used in MTM has very good spectral concentration, which reduces the bias in the spectral estimate due to spectral leakage.

In MTM, multiple spectrograms are created from the signal using a set of orthogonal window functions. The average of these spectra gives the Multitaper time-frequency energy spectrum. The multitaper spectral estimate ($S_{MT}(\tau, \omega)$) of a signal $f(t)$ is defined as

$$S_{MT}(\tau, \omega) = \frac{1}{K} \sum_{m=1}^K | \int f(t) h_m(t - \tau) e^{-j\omega t} dt |^2 \quad (5)$$

In Equation 5, K is the number of windows used in the analysis and h_m is the m^{th} window. We start by setting the time–frequency bandwidth product to a number TB (greater than or equal to 1; $TB = 1$ reduces to STFT), thus using tiles that are TB times larger in the frequency domain than STFT tiles (Fig. 3C). The window functions are chosen to maximize the spectral concentration within the chosen bandwidth, which yields DPSS or Slepian windows (or tapers). The zeroth-order Slepian taper looks like a Gaussian window function, but as the order increases, the tapers become more oscillatory (Fig. 3C, first column). Since the Slepian tapers are orthogonal to each other, MTM averages mutually independent spectral estimates, which in turn reduces the variance of the spectrum. The number of tapers (K) that have good spectral concentration and can be used in the MTM analysis is $K = 2TB - 1$. Informally, this represents the maximum number of functions that can be placed inside each tile with their energy reasonably localized inside the tile.

Wavelet transform

Unlike STFT and MTM, WT uses time–frequency tiles with different heights and widths to build the spectrum of the signal, $f(t)$. The basis functions used in this method are scaled and translated versions of a single function, known as the mother wavelet. A function $\psi(t)$ is called a wavelet if it has a zero average ($\int \psi(t) dt = 0$), normalized energy ($\|\psi\| = 1$), and is sufficiently well localized in time (Mallat, 2008).

The continuous wavelet transform (CWT) of a signal $f(t)$ is then defined as

$$W_f(u, s) = \int f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t - u}{s} \right) dt \quad (6)$$

Here, $\psi(t)$ is the mother wavelet function. A family of wavelet functions is formed through scaling the mother wavelet by a fac-

tor s and translating by a factor u ($\frac{1}{\sqrt{s}}$ is the energy normalization constant). CWT is the projection of the signal on this wavelet family. The time-scale power distribution (Scalogram) of the signal $f(t)$ is obtained by squaring the CWT, $|W_f(u, s)|^2$.

Since a wavelet is a zero average function, the amplitude of its spectrum at origin, i.e., at zero frequency, is zero. So the spectrum is equivalent to that of a bandpass filter. Bandwidth of this filter is determined by the scaling parameter. The filter has large bandwidths for lower values of scales (and vice versa). In fact, each member in the wavelet family, $\frac{1}{\sqrt{s}} \psi \left(\frac{t - u}{s} \right)$ has the same Q-factor (ratio of center frequency to bandwidth). Consequently, in the time–frequency plane, low-frequency wavelets have a better frequency resolution than the high-frequency ones (which, conversely, have better temporal resolution). Using different scales for the mother wavelet covers the entire frequency axis. By converting scales to the corresponding frequency values, the time-scale power spectrum can be transformed into a time–frequency power spectrum.

Matching pursuit

MP is a greedy algorithm that decomposes a signal into a linear combination of waveforms called atoms that are well localized in time and frequency. An example of such waveforms is the Gabor function (Gaussian-modulated sinusoidal function), which has the least time–frequency bandwidth product. A dictionary of time–frequency atoms is created by shifting (u), scaling (s), and modulating (ξ) a single atom. One can thus form an over-complete or a redundant dictionary, and MP then decomposes the signal using atoms picked iteratively from this redundant dictionary as per certain selection criteria. Let the index γ represent the scale, translation, and modulation parameters, i.e., $\gamma = (s, u, \xi)$. Then the dictionary elements can be defined as,

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g \left(\frac{t - u}{s} \right) e^{j\xi t} \quad (7)$$

Here, $\frac{1}{\sqrt{s}}$ is the normalization constant that guarantees that the energy of the atom is unity. MP initializes the decomposed signal to zero and the residue (error after approximation) to the signal itself (given a signal f , the first residue is $R^0 f = f$). At each iteration, MP picks up an atom from the dictionary that maximizes the inner product with the residue. Specifically, at the $(n + 1)^{\text{th}}$ iteration, the atom is chosen as follows:

$$R^n f = \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} + R^{n+1} f$$

$$g_{\gamma_n} = \arg \max_{g_{\gamma_i} \in D} | \langle R^n f, g_{\gamma_i} \rangle | \quad (8)$$

Here, D refers to the dictionary used for MP decomposition. After m iterations, the signal is approximated as

$$f = \sum_{n=0}^{m-1} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} + R^m f \quad (9)$$

Although the selected atoms are not orthogonal to each other, the residue after each iteration is orthogonal to the atom selected at that iteration (in Eq. 8, g_{γ_n} is orthogonal to $R^{n+1} f$). This is an important property, which ensures that the energy of the residue approaches zero as m approaches infinity, and the signal energy is equal to the sum of the square of the inner products (for formal proofs, see Mallat and Zhang, 1993).

The time–frequency energy distribution is derived from the Wigner–Ville distribution (WVD) (Cohen, 1989, 1995) of the atoms selected in the MP decomposition (Mallat and Zhang, 1993). WVD of a signal, $f(t)$, is defined as

$$W_f(t, \omega) = \frac{1}{2\pi} \int f^* \left(t - \frac{\tau}{2} \right) f \left(t + \frac{\tau}{2} \right) e^{-j\omega\tau} d\tau \quad (10)$$

If $s(t) = a(t) + b(t)$, the WVD of $f(t)$ is $W_s(t, \omega) = W_{a,a}(t, \omega) + W_{b,b}(t, \omega) + W_{a,b}(t, \omega) + W_{b,a}(t, \omega)$, where

$$W_{a,b}(t, \omega) = \frac{1}{2\pi} \int a^* \left(t - \frac{\tau}{2} \right) b \left(t + \frac{\tau}{2} \right) e^{-j\omega\tau} d\tau \quad (11)$$

Here, $W_{a,b}$ and $W_{b,a}$ are called the cross-terms, and $W_{a,a}$ and $W_{b,b}$ are self-terms. Due to the energy conservation property of the decomposition obtained by MP (see above), the self-terms completely represent the energy in the original signal. Therefore, we can ignore the cross-terms of the traditional WVD. This property is very desirable, since the representation now becomes devoid of unnecessary cross-terms of WVD (Mallat and Zhang, 1993) and gives a cleaner and physically more relevant time–frequency energy spectrum

$$E_f(t, \omega) = \sum_{n=0}^{\infty} |\langle R^n f, g_{\gamma n} \rangle|^2 W_{g_{\gamma n}}(t, \omega) = \|f\|^2 \quad (12)$$

where $\|f\|^2$ is the energy of the signal. In our analysis, we have used the dyadic Gabor dictionary using real atoms, as originally proposed by Mallat and Zhang (1993). This dictionary is defined as

$$g_{\gamma}(t) = k(\gamma) e^{-\pi \left(\frac{t-u}{s} \right)^2} \cos(\xi(t-u) + \phi) \quad (13)$$

where $k(\gamma)$ is the normalization constant. Given a signal f of length N samples, a dyadic dictionary is created with $\gamma = \{a^j, pa^j\Delta u, ka^{-j}\Delta\xi\}$ where $a = 2$, $\Delta u = 0.5$, $\Delta\xi = \pi$, $0 < j < \log_2 N$, $0 \leq p < N2^{-j+1}$, and $0 \leq k < 2^{j+1}$ (Mallat and Zhang, 1993; Ray et al., 2003). The parameter j is also called the octave of an atom that varies from 1 to 11 for a signal of length of 4096. For real atoms, the signal phase (ϕ) that is otherwise hidden in the complex exponentials becomes an explicit parameter to be optimized, although in the algorithm provided by Mallat and Zhang (1993) it is estimated by taking the inner product with complex Gabors and recovering the phase from the coefficients. In addition, delta and Fourier atoms are included.

While performing MP decomposition, it is important to take a signal that is much longer than the duration of interest, because the edges of the signal contain artifacts arising from an inherent periodicity assumption of DFT (to compute DFT of a signal of length N , we construct an infinitely long periodic signal of period N by concatenating copies of the original signal and taking its discrete Fourier series). Note that this is true for any decomposition technique, but the artifacts can be reduced by multiplying by a window such that the edges of the signal are close to zero. Because we do not use any explicit window in MP, it is advisable to use a signal at least three times longer than the duration of interest and study only the middle section.

The codes and data used for this paper can be found online. The MP code for Windows, Mac OSX, and Linux is available at <https://github.com/supratimray/MP>; the data and codes used to generate Figures 3 and 5, as well as the spectral analysis using the Hilbert–Huang transform, are available at <https://github.com/supratimray/SpectralAnalysisCodes>.

References

- Ackroyd MH (1971) Short-time spectra and time–frequency energy distributions. *J Acoust Soc Am* 50:1229–1231. [CrossRef](#)
- Allen JB (1977) Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Trans Acoust Speech Signal Process* 25:235–238. [CrossRef](#)
- Allen JB, Rabiner L (1977) A unified approach to short-time Fourier analysis and synthesis. *Proc IEEE* 65:1558–1564. [CrossRef](#)
- Babadi B, Brown EN (2014) A review of multitaper spectral analysis. *IEEE Trans Biomed Eng* 61:1555–1564. [CrossRef](#) [Medline](#)
- Bokil H, Andrews P, Kulkarni JE, Mehta S, Mitra PP (2010) Chronux: a platform for analyzing neural signals. *J Neurosci Methods* 192:146–151. [CrossRef](#) [Medline](#)
- Bosman CA, Womelsdorf T, Desimone R, Fries P (2009) A microsaccadic rhythm modulates gamma-band synchronization and behavior. *J Neurosci* 29:9471–9480. [CrossRef](#) [Medline](#)
- Brillinger DR (1972) The spectral analysis of stationary interval functions. In: *The regents of the University of California*.
- Buzsáki G (2006) *Rhythms of the brain*. Oxford: Oxford UP.
- Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. *Science* 304:1926–1929. [CrossRef](#) [Medline](#)
- Chen S, Donoho D (1994) Basis pursuit. In: 1994 conference record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers, Vol. 1, pp 41–44. Washington, DC: IEEE.
- Chen S, Donoho D, Saunders M (2001) Atomic decomposition by basis pursuit. *SIAM Rev* 43:129–159. [CrossRef](#)
- Cohen L (1989) Time–frequency distributions—a review. *Proc IEEE* 77:941–981. [CrossRef](#)
- Cohen L (1995) *Time–frequency analysis*. Upper Saddle River, NJ: Prentice Hall.
- Daubechies I (1990) The wavelet transform, time–frequency localization and signal analysis. *IEEE Trans Inf Theory* 36:961–1005. [CrossRef](#)
- Durka PJ, Matysiak A, Montes EM, Sosa PV, Blinowska KJ (2005) Multichannel matching pursuit and EEG inverse solutions. *J Neurosci Methods* 148:49–59. [CrossRef](#) [Medline](#)
- Fries P (2009) Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu Rev Neurosci* 32:209–224. [CrossRef](#) [Medline](#)
- Jarvis MR, Mitra PP (2001) Sampling properties of the spectrum and coherence of sequences of action potentials. *Neural Comput* 13:717–749. [CrossRef](#) [Medline](#)
- Kay SM, Marple JSL (1981) Spectrum analysis—a modern perspective. *Proc IEEE* 69:1380–1419. [CrossRef](#)
- Mallat S (2008) *A wavelet tour of signal processing: the sparse way*, 3rd ed. New York: Academic.
- Mallat SG, Zhang Z (1993) Matching pursuits with time–frequency dictionaries. *IEEE Trans Signal Process* 41:3397–3415. [CrossRef](#)
- Moore IC, Cada M (2004) Prolate spheroidal wave functions, an introduction to the slepian series and its properties. *Appl Comput Harmon Anal* 16:208–230. [CrossRef](#)
- Oppenheim AV, Schaffer RW (2013) *Discrete-time signal processing*. Upper Saddle River, NJ: Pearson Education.
- Park J, Lindberg CR, Vernon FL (1987) Multitaper spectral analysis of high-frequency seismograms. *J Geophys Res Solid Earth* 92:12675–12684. [CrossRef](#)
- Pati YC, Rezaifar R, Krishnaprasad PS (1993) Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: 1993 Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, vol. 1, pp 40–44. Washington, DC: IEEE.
- Portnoff MR (1980) Time–frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Trans Acoust Speech Signal Process* 28:55–69. [CrossRef](#)
- Ray S (2015) Challenges in the quantification and interpretation of spike-LFP relationships. *Curr Opin Neurobiol* 31:111–118. [CrossRef](#) [Medline](#)
- Ray S, Maunsell JH (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9:e1000610. [CrossRef](#) [Medline](#)
- Ray S, Jouny CC, Crone NE, Boatman D, Thakor NV, Franaszczuk PJ (2003) Human ECoG analysis during speech perception using matching pursuit: a comparison between stochastic and dyadic dictionaries. *IEEE Trans Biomed Eng* 50:1371–1373. [CrossRef](#) [Medline](#)

- Ray S, Hsiao SS, Crone NE, Franaszczuk PJ, Niebur E (2008) Effect of stimulus intensity on the spike-local field potential relationship in the secondary somatosensory cortex. *J Neurosci* 28:7334–7343. [CrossRef](#) [Medline](#)
- Ray S, Ni AM, Maunsell JH (2013) Strength of gamma rhythm depends on normalization. *PLoS Biol* 11:e1001477. [CrossRef](#) [Medline](#)
- Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal timing across the macaque visual system. *J Neurophysiol* 79:3272–3278. [Medline](#)
- Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: the discrete case. *Bell Syst Tech J* 57:1371–1430. [CrossRef](#)
- Smith MJT, Barnwell ITP (1987) A new filter bank theory for time-frequency representation. *IEEE Trans Acoust Speech Signal Process* 35:314–327. [CrossRef](#)
- Tamura H, Tanaka K (2001) Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb Cortex* 11:384–399. [CrossRef](#) [Medline](#)
- Thomson DJ (1982) Spectrum estimation and harmonic analysis. *Proc IEEE* 70:1055–1096. [CrossRef](#)
- Vetterli M (1987) A theory of multirate filter banks. *IEEE Trans Acoust Speech Signal Process* 35:356–372. [CrossRef](#)
- Vetterli M, Herley C (1992) Wavelets and filter banks: theory and design. *IEEE Trans Signal Process* 40:2207–2232. [CrossRef](#)