The Bispectrum and Its Relationship to Phase-Amplitude Coupling

Christopher K. Kovach
Department of Neurosurgery
University of Iowa Carver College of Medicine
Iowa City, Iowa
E-mail: christopher-kovach@uiowa.edu

Hiroyuki Oya
Department of Neurosurgery
University of Iowa Carver College of Medicine
Iowa City, Iowa
E-mail: hiroyuki-oya@uiowa.edu

Hiroto Kawasaki
Department of Neurosurgery
University of Iowa Carver College of Medicine
Iowa City, Iowa
E-mail: hiroto-kawasaki@uiowa.edu

Summary.
Measures of nonlinear dependence in time series often take advantage of the fact that higher-order statistics manifest in the spectral domain as dependence across frequencies, related to so-called polyspectra. Of these, the most frequently applied are third-order measures derived from the bispectrum. Bispectral techniques have found occasional use in EEG literature, but other forms of cross-frequency dependence commonly regarded as easier to interpret, have recently attracted greater interest, in particular, phase-amplitude coupling (PAC). Here it is shown that standard measures of PAC are related to smoothings of the signal bispectrum, making them fundamentally bispectral estimators. Viewed this way, however, such measures exhibit some unfavorable qualities, such as poor bias properties, lack of correct symmetry and artificial constraints on the spectral range and resolution of the estimate. Moreover, information obscured by smoothing in measures of PAC, but preserved in standard bispectral estimators, may be critical for distinguishing nested oscillations from transient signal features and other non-oscillatory causes of “spurious” PAC. Future studies of phase-amplitude coupling should therefore include an explicit evaluation of the bispectrum. Beyond clarifying the relationship between the bispectrum and PAC, these observations motivate a more general framework for the interpretation of the bispectrum.

1. Introduction

Many interesting properties of signals in nature relate to nonlinear, non-Gaussian and non-stationary dynamics, which are poorly indexed by second-order measures such as power and cross spectra. Because the spectrum of a stationary Gaussian process lacks any statistical dependence across frequencies, measures of frequency-domain dependence can be extremely useful for gauging the presence and nature of higher order dynamics. For stationary signals, higher order spectra, “polyspectra,” which capture such dependence are the frequency-domain representations of higher moments [Nikias and Mendel, 1993]. The bispectrum is the third-order polyspectrum, making it an obvious place to start in the approach towards higher-order dynamics, at least from a statistical standpoint. Bispectral analysis has proved its practical worth in applications to EEG [Dumermuth et al., 1971; Sigl and Chamoun, 1994], most notably in gauging the depth of anesthesia [Barnett et al., 1971; Gan et al., 1997; Kearse et al., 1998; Myles et al., 2004]. Nevertheless, an important drawback for the non-statistician remains its lack
of an obvious connection to any simple physical interpretation (Fackrell et al., 1995). This point has been the source of some confusion in applied literature; for example, as recently reviewed in Hyafil (Hyafil et al., 2015), various authors have suggested incorrectly that bicoherence relates to phase entrainment across pairs of frequencies. While it is true that bispectral measures cannot easily be reduced to any single interpretation, a goal of the present work is to show that specific forms of dependence leave easily recognized signatures in the bispectrum, making bispectral measures invaluable for distinguishing between a variety of nonlinear phenomena.

A separate body of work has recently emerged from EEG literature, which examines the role of another form of cross-frequency coupling, phase-amplitude coupling (PAC) (Canolty et al., 2006; Hyafil et al., 2015). PAC refers to dependence between analytic amplitude at one frequency and analytic phase at another. Rather than any statistical first principle, interest in PAC has been motivated by the empirical discovery of PAC in signals recorded from the brain, alongside emerging computational and physiological models of how it arises within populations of interacting neurons (Akam and Kullmann, 2014). In this literature, the measurement of PAC is usually approached with second-order statistics, such as coherence, applied towards comparing analytic phase in one band with analytic amplitude extracted from another, regarding the two as separate signals (Schack et al., 2002; Canolty et al., 2006; Tort et al., 2010). The situation with PAC is therefore reversed from bispectral measures: its physical meaning is more evident than its relationship to any general body of statistical theory.

The present work aims to bridge this gap and resolve ambiguities of meaning in both directions. The central result is that second-order measures of PAC may be fundamentally understood as estimators of the bispectrum. In the same way that windowed stationary and time-varying estimates of the power spectrum can be equated to smoothings of the Wigner-Ville distribution, windowed bispectral estimators, which include those underlying measures of PAC, amount to different ways of smoothing and integrating over the third-order Wigner-Ville distribution (Gerr, 1988; Nikias and Raghvan, 1987; Swami, 1991; Hanssen and Scharf, 2003). In both cases, differences between estimators relate to properties of the respective smoothing kernels (Cohen, 1989). This observation demonstrates conclusively the meaning of PAC measures as they relate to the bispectrum and vice versa and establishes that second-order measures of PAC provide no unique information beyond what can be obtained from the bispectrum.

While PAC measures are fundamentally measures of the bispectrum, the reverse is not true; it is not correct to conclude that the bispectrum is principally a reflection of phase-amplitude coupling. Forms of “spurious” PAC, related, for example, to spectrally broad signal features, may be traced to the bispectral nature of PAC measures. One practical implication is that the superior resolution and lower bias of standard bispectral measures, in comparison to PAC measures, allows them to retain information critical for distinguishing between nested oscillations and other sources of apparent phase-amplitude coupling (Kramer et al., 2008). Following a brief review of bispectral and PAC estimation, we will observe how different regions of the bispectrum may be taken to reflect either phase-amplitude coupling or consistency of phase across a range of frequencies. It is shown that, properly interpreted, the bispectrum is highly useful for ascertaining the presence and origin of phase-amplitude coupling, which recommends its consideration in all future studies of PAC. Finally, we briefly consider how these points aid more generally in the interpretation of the bispectrum and its application towards isolating nonlinear signal features.

2. The Bispectrum

Let $X$ be a random realization of a harmonizable time-series process, obeying usual assumptions of boundedness and integrability, with moments defined up to order 3. In most cases, it will be assumed that $X$ is real-valued, unless otherwise specified. Frequency domain representations will be indicated with a tilde; for example, $\tilde{h}(\omega)$ is the Fourier transform of $h(t)$. Scaling and normalizing constants, such as $1/2\pi$ in the inverse Fourier transform, will also be suppressed for notational economy where doing is not expected to create confusion.
2.1. Definition

The bispectrum of the process generating \( X \) is given by the following expectation in the frequency domain:

\[
B(\omega_1, \omega_2) = E\left[\hat{X}(\omega_1)\hat{X}(\omega_2)\hat{X}^*(\omega_1 + \omega_2)\right]
\]  

(1)

Some insight into the meaning of this quantity comes from the inverse Fourier expansion of \( \hat{X} \):

\[
B = E\left[\int \int \int X(r)X(s)X(t)e^{-i\omega_1 (r-t)} - i\omega_2 (s-t) \, dr \, ds \, dt\right]
\]

(2)

which integrates a two-dimensional Fourier transform over time. Directly paralleling the ordinary power spectrum (Cohen, 1989), the integrand here may be understood as the 3rd-order Wigner-Ville distribution (Gerr, 1988; Hanssen and Scharf, 2003):

\[
W_3(\omega_1, \omega_2, t) = \int \int X(\tau_1 + t)X(\tau_2 + t)X(t)e^{-i\omega_1 \tau_1 - i\omega_2 \tau_2} \, d\tau_1 \, d\tau_2
\]

(3)

For a third-order stationary process, the expectation of the third-order moment depends only on the relative lags of the times:

\[
E\left[X(\tau_1 + t)X(\tau_2 + t)X(t)\right] = \mu_3(\tau_1, \tau_2)
\]

(4)

in which case

\[
B = \int \int \mu_3(\tau_1, \tau_2)e^{-i\omega_1 \tau_1 - i\omega_2 \tau_2} \, d\tau_1 \, d\tau_2
\]

(5)

In other words, for a third-order stationary process, \( B \) is the two-dimensional Fourier transform of the third moment, \( B = \tilde{\mu}_3(\omega_1, \omega_2) \). This relationship parallels the equivalence between the power spectrum and the second moment (i.e. auto-correlation), and extends likewise to higher polyspectra and moments (Nikias and Mendel, 1993; Hanssen and Scharf, 2003).

2.2. Properties of the Bispectrum

We next review a handful of properties that will be called upon in the following discussion.

2.2.1. Symmetry

The third moment is unchanged under a permutation of the time delays, \( \tau_1 \) and \( \tau_2 \), over the three instances of \( X \) in Eq. (3), creating a six-fold symmetry, illustrated in Figure 1, with the following relations:

\[
\mu_3(\tau_1, \tau_2) = \mu_3(\tau_2, \tau_1) = \mu_3(-\tau_1, \tau_2 - \tau_1)
\]

(6)

Because the spectrum of a real-valued signal is conjugate symmetric about the origin, the 6-fold symmetry of the third moment becomes a 12-fold conjugate symmetry in the bispectrum under the following relations:

\[
\tilde{\mu}_3(\omega_1, \omega_2) = \tilde{\mu}_3(\omega_2, \omega_1) = \tilde{\mu}_3(-\omega_1 - \omega_2, \omega_2) = \tilde{\mu}_3^*(-\omega_1, -\omega_2)
\]

(7)

For a real signal, the full bispectrum may be recovered from estimates within any one of these region. These relations apply as well to the frequency dimensions of the third-order Wigner-Ville distribution, \( W_3 \).
2.2.2. Convolution

The bispectrum inherits a number of basic spectral properties, among which are those related to convolution. Because the convolution of two signals in time equates to multiplication in the spectral domain, it is directly apparent that the bispectrum of a signal formed from the convolution of two other signals is likewise the product of their separate bispectra, provided the signals are third-order independent or deterministic; that is,

$$B[X \circ Y] = B[X]B[Y]$$  \hspace{1cm} (8)

for third-order independent $X, Y$.

2.2.3. Multiplication

The reverse relationship, by which the product of two independent signals in time becomes a convolution in the frequency domain holds generally for the bispectrum only if at least one of the signals is also third-order stationary; that is

$$B[XY] = B[X] \odot B[Y]$$  \hspace{1cm} (9)

if $X$ and $Y$ are third-order independent and if $X$ or $Y$ is third-order stationary. Because of the stationarity requirement, this relation applies only in trivial cases (i.e. $X(t) = c$) when both signals are deterministic.

The relationship does, however, hold generally for the bispectral Wigner-Ville distribution; that is

$$W_3[XY] = W_3[X] \circ_\omega W_3[Y]$$  \hspace{1cm} (10)

for third-order independent $X$ and $Y$, where $\circ_\omega$ denotes convolution over the frequency dimension. Then

$$B[XY] = E \left[ \int W_3[X] \circ_\omega W_3[Y] \, dt \right]$$  \hspace{1cm} (11)

For third-order stationary $X$, time drops out of the expectation as $E[W_3[X]] = \tilde{\mu}_3[X]$, so that

$$B[XY] = \tilde{\mu}_3[X] \circ \int E[W_3[Y]] \, dt = \tilde{\mu}_3[X] \odot B[Y]$$  \hspace{1cm} (12)

if $X$ is stationary and third-order independent of $Y$.

2.2.4. Alternative Forms

The lags in (3) can be defined in different ways while preserving the essential properties of $W_3$ (Swami, 1991); it will at times be useful to apply a change of variables giving:

$$W_3(\nu_1, \nu_2, t) = \int \int X(t+\delta)X(t-\delta)X(t-\tau)e^{-i\nu_1\delta-i\nu_2\tau} \, d\tau \, d\delta$$  \hspace{1cm} (14)

where $\nu_1 \equiv \omega_1 - \omega_2$ and $\nu_2 \equiv \omega_1 + \omega_2$, $\delta \equiv (\tau_1 - \tau_2)/2$, and $\tau \equiv (\tau_1 + \tau_2)/2$. Where necessary, the distinction between these forms will be made through a similar abuse of notation in the arguments.

Similarly, in the spectral domain, it will later become useful to consider the change of variables $\omega'_1 = \omega_1$ and $\omega'_2 = \omega_1 + \omega_2/2$. This change effectively modifies the definition of the bispectrum to a form that is symmetrical around $\omega_2$:

$$B'(\omega'_1, \omega'_2) = E \left[ \tilde{X} (\omega'_1) \tilde{X} (\omega'_2 - \omega'_1/2) \tilde{X}^* (\omega'_2 + \omega'_1/2) \right]$$  \hspace{1cm} (15)
2.3. Bispectral Estimators

A central concern here is how to obtain an estimator of $B$ for a third-order stationary signal. The extension to non-stationary estimation parallels that of ordinary spectral estimation and will not be considered separately here. The most straightforward approaches follow the example set by conventional methods for computing Fourier power spectra (Nikias and Raghuveer, 1987; Birkelund et al., 2003). The so-called “indirect” method first estimates the third moment of the signal in the time domain and windows it, giving $\hat{\mu}_3^H(\tau_1, \tau_2) = \hat{\mu}_3(\tau_1, \tau_2) H(\tau_1, \tau_2)$, from which the smoothed bispectrum is obtained with a two-dimensional Fourier transform. The “direct” method first computes a series of windowed Fourier spectra, $\chi_j(\omega)$, and averages over the product $\chi_j(\omega_1)\chi_j(\omega_2)\tilde{\chi}_j^{*}(\omega_1 + \omega_2)$. Both approaches generate a largely equivalent family of estimators.

To minimize bias in the estimate and give an estimate with the proper symmetry, it is usually required that $H$ itself exhibit the same symmetry as the third moment, given in Eq. (6). Windows should also be real-valued in time and frequency and non-negative in the frequency domain (Sasaki et al., 1975). Two-dimensional windows fulfilling these requirements may be constructed in either the time or frequency domains from a window defined in one dimension, $h(t)$:

$$H(\tau_1, \tau_2) = h(\tau_1)h(\tau_2)h(\tau_1 - \tau_2) \quad \text{or} \quad \tilde{H}(\omega_1, \omega_2) = \tilde{h}(\omega_1)\tilde{h}(\omega_2)\tilde{h}(\omega_1 + \omega_2)$$

Although specific windows are shown to be optimal for minimizing bias of bispectral estimators (Sasaki et al., 1975), we will consider a more general case, which also includes windows for which symmetry relation may not hold. We will see that PAC measures can be equated to bispectral estimators that lack the proper symmetry constraint, which motivates dropping the constraint in the present discussion. It will still be assumed that the windows are non-negative and real-valued in the frequency domain and symmetric under a 180 degree rotation, although the complex conjugate will be indicated where appropriate for notational consistency.

The “direct” estimator most commonly begins with a series of windowed Fourier transforms, evenly spaced in time (here treated as continuous for simplicity), which may be regarded equivalently as the outcome of a bank of filters whose outputs are analytic signals shifted in frequency (i.e. complex demodulates) obtained by bandpass-filtering the original signal (Godfrey, 1965; Bingham et al., 1967; Kovach and Gander, 2016):

$$\chi_h(\omega, t) = \int h(s - t)X(s)e^{-i\omega s} \, ds$$

The corresponding change in the time-domain involves a modification of the lags in the 3rd moment, $\tau_1' = \tau_1$ and $\tau_2' = -\tau_2 - \tau_1/2$; adding $t = t' - \tau_1/2$ for the sake of symmetry gives

$$\mu_3'(\tau_1', \tau_2', t) = E\left[X(t' + \frac{\tau_1'}{2})X(t' - \frac{\tau_1'}{2})X(t' - \tau_2')\right]$$

(16)
An estimate of the bispectrum is obtained from this time-frequency representation as

\[
\hat{B}(\omega_1, \omega_2) = \frac{1}{T} \int T \chi_h(\omega_1, s) \chi_h(\omega_2, s) \chi^*_h(\omega_1 + \omega_2, s) ds
\]

\[
= \frac{1}{T} \iint X(t + \tau_1) X(t + \tau_2) X(t) e^{-i\omega_1 \tau_1 - i\omega_2 \tau_2} \int h(s + \tau_1) h(s + \tau_2) h(s) ds d\tau_1 d\tau_2 dt
\]

(19)

The last integrand defines the smoothing window:

\[
H(\tau_1, \tau_2) = \int h(s + \tau_1) h(s + \tau_2) h(s) ds
\]

\[
= \int \tilde{h}(\xi_1) \tilde{h}(\xi_2) \tilde{h}^*(\xi_1 + \xi_2) e^{i\xi_1 \tau_1 + i\xi_2 \tau_2} d\xi_1 d\xi_2
\]

(20)

Comparing this to Eq. (17) makes it clear that the resulting estimator fulfills the symmetry requirement of Eq. (6). Combining these results:

\[
B(\omega_1, \omega_2) = \iint W(\xi_1, \xi_2, t) \tilde{H}(\xi_1 - \omega_1, \xi_2 - \omega_2) d\xi_1 d\xi_2 dt
\]

(21)

which makes the smoothing property of the bispectral estimator explicit. Absent the integration over \(t\), this relation also gives a time-varying estimate of the bispectrum as a smoothing of the third-order Wigner-Ville distribution [Hanssen and Scharf 2003]. As in the second-order case, suppressing cross terms in the time-varying spectrum requires some additional smoothing over time.

### 2.4. Estimators with Different Analysis Filters

Eq. (19) constrains the analysis window for each of the three bands used in the estimate to be identical, which had the benefit of yielding an estimator with the proper symmetry. While the symmetry conditions in (6) are usually regarded as a criterion of admissibility for windows used in estimating the bispectrum, there is in fact no essential reason why one may not smooth the bispectrum with windows that do not obey these relations. Although the resulting estimator will be biased with respect to the symmetry of the bispectrum, it might conceivably have other useful properties that outweigh this consideration. For instance, one might select estimators that are better tuned to specific features of interest in the signal bispectrum, which we will later find illustrated with PAC measures. We therefore consider next the consequence of estimating the bispectrum using different analysis filters for each band, which leads to the following estimation window:

\[
H(\tau, \delta) = \int \tilde{h}_1(\nu_1 + \nu_2) \tilde{h}_2(\nu_2 - \nu_1) \tilde{g}^*(\nu_2) e^{i\nu_1 \delta + i\nu_2 \tau} d\nu_1 d\nu_2
\]

(22)

with separate analysis filters, \(h_1\), \(h_2\), and \(g\).

In general, because the smoothing occurs over the two-dimensional bispectral plane, the choice of third window is over-constrained. This is easily recognized in the limit as the bandwidth of one filter diverges from that of the other two, in which case the smoothing window factors over the two dimensions. For example, as \(\tilde{g} \to \delta\), one obtains:

\[
\tilde{H}(\nu_1, \nu_2) \to \tilde{h}_1(\nu_1) \tilde{h}_2\left(\frac{\nu_1}{2}\right) \tilde{g}^*(\nu_2)
\]

Whereas in the limit \(\tilde{h}_2 \to \delta\),

\[
\tilde{H}(\omega_1, \omega_2) \to \tilde{h}_1(\omega_1) \tilde{h}_2(\omega_2) \tilde{g}^*(\omega_1)
\]

and likewise for \(\tilde{h}_1 \to \delta\). In each of these cases the resulting window in one dimension is the product of the two broader filters. The estimator therefore produces the same result as when
Fig. 2. Relationship between analysis filter bandwidth and equivalent smoothing kernels for 4 bispectral estimators. Dashed lines indicate the axes given by $\nu_1 \equiv \omega_1 - \omega_2$ and $\nu_2 \equiv \omega_1 + \omega_2$. In each case, one of two Gaussian analysis filters is applied to each of the three bands used in estimating the bispectrum, ordered by center frequency: $\omega_1 \leq \omega_2 < \nu_2$. Frequency windows are Narrowband with standard width, $\sigma = 0.25$, or Broadband with standard width, $\sigma = \sqrt{2}$, and the estimators are indicated according to which window was applied to each ordered band: BBB, NNB, BBN and BNB. Top Left: BBB (also NNN under rescaling) Fixed analysis bandwidth. Each band is filtered using the same analysis window, $B$. Top Right: NNB Highest band is broad and the other two narrow, giving the asymptotically symmetric kernel, $\tilde{g}(\omega_1)\tilde{g}(\omega_2)$. Bottom Right: BBN Highest band is narrow and the remaining two are broad giving the asymptotic kernel, $|h(\nu_1/2)|^2 \tilde{g}(\nu_2)$. Bottom Left: NBB (also BNB under a transpose of axes) Second band is narrow and the remaining two broad, giving the asymptotic kernel, $\hat{h}(\omega_1)|g(\omega_2)|^2$. 
using two filters: the first given by a narrow filter and the second by the root product of the remaining two. Without loss of generality, the following sections therefore consider estimators for which two of the three analysis filters use the same window. We consider permutations of the bandwidth of the three analysis filters applied to the three bands sorted by frequency: $\omega_1 \leq \omega_2 < \nu_2$.

### 2.4.1. Two Filters: Narrow, Narrow, Broad

When $\tilde{g}$ is wider than $\tilde{h}$, $\tilde{H}$ becomes symmetric so that $W$ is smoothed equally along the $\nu_1$ and $\nu_2$ axes. As $\tilde{g}(\nu_2) \to 1$ we obtain the approximation:

$$\tilde{H}(\nu_1, \nu_2) \to \tilde{h} \left( \frac{\nu_1 + \nu_2}{2} \right) \tilde{h}^* \left( \frac{\nu_1 - \nu_2}{2} \right) = \tilde{h}(\omega_1) \tilde{h}^*(\omega_2) \quad (23)$$

An estimate of the bispectrum with radially symmetric smoothing is therefore recovered by:

$$B(\omega_1, \omega_2) = \int \chi_h(\omega_1, t)\chi_h(\omega_2, t)X(t)e^{i(\omega_1 + \omega_2)t} \, dt \quad (24)$$

Note that the radial symmetry here does not obey the symmetry relations of (6), so the estimate does not exactly recover the symmetry of the bispectrum. On the other hand, (24) suggests a less cumbersome approach to estimation which foregoes the summation over the product of three separate bands. The estimator is obtained instead from the frequency-domain covariance of the time-frequency decomposition after remodulation and weighting by the original signal. In practice, however, the time-frequency decomposition is often downsampled, so it will still be necessary to either apply an anti-aliasing filter to $X$ appropriate for the sampling rate and frequency range under consideration, or oversample $\chi_h$. Both options tend to negate any computational advantage that might otherwise be gained from such a simplification.

### 2.4.2. Two filters: Broad, Broad, Narrow

The estimator in Eq. (24) results in a smoothing window with radial symmetry, but the symmetry came about from greater smoothing in the $\nu_2$ direction entailing a sacrifice of resolution. One may consider the converse possibility, allowing the bandwidth of $g$ to be narrower than that of $h$. This produces a smoothing window with the limit $\tilde{g}(\nu_2) \to \delta(\nu_2)$. In the approach to this limit one obtains the following approximation:

$$\tilde{H}(\nu_1, \nu_2) \to \tilde{h} \left( \frac{\nu_1}{2} \right)^2 \tilde{g}^*(\nu_2) \quad (25)$$

Provided that $g$ is narrower than $h$, this estimator allows the smoothing along $\nu_1$ and $\nu_2$ to be separately controlled.

### 2.4.3. Two filters: Narrow, Broad, Broad

In the limit $\tilde{h}_1 \to \delta$ we have $g(\nu_2) \to g(\omega_2)$. Letting $h_2 = g$, without loss of generality, we have

$$\tilde{H}(\omega_1, \omega_2) \to \tilde{h}(\omega_1) \left| g(\omega_2) \right|^2 \quad (26)$$

The same result holds for the BNB estimator swapping the arguments. The smoothing kernel here is related to the previous examples through a 45 degree rotation in the bispectral plane. One important potential drawback of this family of estimators is that the outcome will not reflect the symmetry of the bispectrum about the diagonal, $\omega_1 = \omega_2$. In the following section we will see that measures of phase-amplitude coupling relate to bispectral estimators of this type.
2.5. Bicoherence

Each of the estimators considered above can be regarded as an inner product between two signals, the first being the higher analytic band, $\chi_3(\omega_1 + \omega_2, t)$ and the second obtained by multiplying the two lower bands $\chi_1(\omega_1, t)\chi_2(\omega_2, t)$. The degree of alignment between these two terms is often used as a normalized measure of dependence, which is obtained in the same way as for an ordinary correlation, by dividing the inner product with the magnitudes of the separate terms:

$$\beta(\omega_1, \omega_2) = \frac{\hat{B}(\omega_1, \omega_2)}{\sqrt{\int |\chi_1(\omega_1, t)\chi_2(\omega_2, t)|^2 dt} \sqrt{\int |\chi_3(\omega_1 + \omega_2, s)|^2 ds}}$$

(27)

This quantity is known as bicoherence.

2.5.1. Alternative Definition

It is worth noting that this common definition bicoherence does not share the full set of symmetry properties with the bispectrum because the normalizing term is not likewise symmetric. Whereas $\beta(\omega_1, \omega_2) = \beta(\omega_2, \omega_1)$, it is not generally true that $\beta(\omega_1, \omega_2) = \beta(-\omega_1 - \omega_2, \omega_2)$. Some alternative definitions of bicoherence do preserve the appropriate symmetry. The following defines an alternative index [Hagihira et al., 2001]:

$$\beta_{\phi}(\omega_1, \omega_2) = \frac{\hat{B}(\omega_1, \omega_2)}{\int |\chi_1(\omega_1, t)\chi_2(\omega_2, t)\chi_3^{*}(\omega_1 + \omega_2, t)| dt}$$

(28)

This index may be interpreted as a weighted-average vector strength over the phase component of the bispectrum, thus it gives a measure of phase locking [Kovach In Press].

3. Phase-Amplitude Coupling

Phase-amplitude coupling refers to dependence between the analytic envelope of an oscillatory signal component within one band and phase within another. For the envelope of the first component to fluctuate at the scale of the second, the bandwidth of the first component must be at least as great as the center frequency of the second and its center frequency correspondingly higher, for which reason the first component is a “fast oscillation” (FO), while the second is the “slow oscillation” (SO).

Measures of PAC relate the analytic signal in a band encompassing the SO to the analytic envelope from the band of the FO, treating the two as separate signals, typically using a second-order statistic such as coherence, phase-locking or weighted mean vector strength. A number of elaborations of these second-order measures have been described; for example, Tort’s “modulation index,” computes a pseudo-entropy by treating the mean amplitude as though it were a probability distribution over phase [Tort et al., 2010]. These extensions improve sensitivity to more general forms of dependence, but the essential qualities of PAC measures are well represented by simpler second-order metrics.

While it is also most common to use signal amplitude in quantifying PAC, one might use signal power (squared amplitude) in place of amplitude in the same way. We will observe that measures using analytic power are easy to relate directly to the bispectrum. Because conventional measures that use amplitude are related to those that use squared amplitude by a simple scalar transformation of the filtered input signals, they reflect the signal bispectrum to a first approximation and can otherwise be interpreted as bispectral estimates computed on suitably transformed input signals. Similarly, measures that apply alternative methods of quantifying second-order dependence, such as with Kullback-Leibler divergence or other entropy-related metric over the distribution of phase, can be translated to the bispectrum with relatively little effort. A detailed consideration of these alternative methods of quantifying the dependence is beyond the scope of the present work, but the essential points described in the following sections remain valid for all measures of PAC derived from the joint distribution of analytic amplitude and phase across frequency bands.
3.1. Relationship of PPC to the Bispectrum

The following section works out the mathematical equivalence between phase-power coherence (PPC) and the NBB class of bispectral estimators. PPC is calculated from the cross spectrum between the squared analytic envelope at one frequency and the original signal at another. The squared envelope within a band centered at $\gamma$ filtered with the analysis window, $h$, is given by

$$P(\gamma, t) = \int \int h(r - t) h^*(u - t) X(r) X^*(u) e^{-i \gamma (r - u)} \, dr \, du$$

(29)

PPC is calculated from the cross spectrum between $P$ and $X$, obtained with a second analysis filter, $g$, which will be of a narrower band than $h$, centered on the frequency, $\theta$:

$$\phi(\theta, \gamma) = \int \int \int g(t - \tau) g^*(s - \tau) X(s) P^*(\gamma, t) e^{-i \theta (s - t)} \, ds \, dt \, d\tau$$

(30)

$$= \int \int g^{(2)}(s - t) X(s) P^*(\gamma, t) e^{-i \theta (s - t)} \, ds \, dt$$

where $g^{(2)}$ denotes the autocorrelation of $g$. Expanding $P$, we have

$$\phi(\theta, \gamma) = \int \int \int \int g^{(2)}(s - t) h(r - t) h(u - t) X(s) X(r) X^*(u) e^{-i \gamma (r - u) - i \theta (s - t)} \, ds \, dt \, du \, dr$$

(31)

The term that integrates over $t$ independently of the signal yields a smoothing kernel:

$$A = \int g^{(2)}(s - t) h(r - t) h(u - t) e^{-i \theta (s - t)} \, dt$$

$$= \int |g(\xi_3)|^2 \tilde{h}(\xi_1) \tilde{h}^*(\xi_2) e^{i \xi_1 \theta + i \xi_2 u + i \xi_3 t} \, d\xi_1 \, d\xi_2 \, d\xi_3$$

(32)

$$= \int |\tilde{g}(\xi_2 - \xi_1 + \theta)|^2 \tilde{h}(\xi_1) \tilde{h}^*(\xi_2) e^{i (r - s) \xi_2} \, d\xi_1 \, d\xi_2$$

$$= \int |\tilde{g}(\theta + 2 \lambda_2)|^2 \tilde{h} \tilde{h}^* \tilde{h}(\lambda_1 + \lambda_2) \tilde{h}^*(\lambda_1 - \lambda_2) e^{i \lambda_1 (r - u) + i \lambda_2 (r + u) + i 2 \lambda_2 s} \, d\lambda_1 \, d\lambda_2$$

from which, setting $r = s + \tau_2$ and $u = s + \tau_1$:

$$A = \int |\tilde{g}(\theta + 2 \lambda_2)|^2 \tilde{h} \tilde{h}^* \tilde{h}(\lambda_1 + \lambda_2) \tilde{h}^*(\lambda_1 - \lambda_2) e^{-i \lambda_1 \lambda_2 \tau_2 - i \lambda_2 \tau_2} \, d\lambda_1 \, d\lambda_2$$

(33)

Substituting back into (31), one arrives at

$$\phi(\theta, \gamma) = \int \int \int |\tilde{g}(\theta + 2 \lambda_2)|^2 \tilde{h} \tilde{h}^* \tilde{h}(\lambda_1 + \lambda_2) \tilde{h}^*(\lambda_1 - \lambda_2) W(\lambda_2 + \lambda_1 - \gamma, \lambda_2 - \lambda_1 + \gamma) \, d\lambda_1 \, d\lambda_2 \, d\tau_2$$

(34)

which with another change of variables, $\omega_1 \equiv -2 \lambda_2$ and $\omega_2 \equiv \lambda_2 - \lambda_1 + \gamma$ and applying the symmetry relation from Eq. (7), $W(-\omega_2 - \omega_1, \omega_2, z) = W(\omega_1, \omega_2, z)$, this becomes

$$\phi(\theta, \gamma) = \int \int \int |\tilde{g}(\omega_1 - \theta)|^2 \tilde{h} \tilde{h}^* \tilde{h}(\omega_2 + \omega_1 - \gamma) \tilde{h}(\omega_2 - \gamma) W(\omega_1, \omega_2, z) \, d\omega_1 \, d\omega_2 \, dz$$

(35)

Finally, because in measuring phase-amplitude coupling, the SO filter, $g$, is selected with a narrower bandwidth than the FO filter, $h$, approaching the limit $\tilde{g} \to \delta$, we may use the following good approximation

$$\phi(\theta, \gamma) \approx \int \int \int |\tilde{g}(\omega_1 - \theta)|^2 \tilde{h} \tilde{h}^* \tilde{h}(\omega_2 + \theta - \gamma) \tilde{h}(\omega_2 - \gamma) W(\omega_1, \omega_2, z) \, d\omega_1 \, d\omega_2 \, dz$$

(36)
3.1.1. Comparison of PAC and Standard Bispectral Kernels

Eq. (36) very nearly describes an NBB smoothing kernel like the one in (26). The only nontrivial difference lies in the appearance of \( \theta \) as an argument in one of the broad window, \( h \), terms. The effect of this term is easily seen for the Gaussian window of bandwidth \( \sigma_w \):

\[
\tilde{h}^* (\omega_2 - \gamma + \theta) \tilde{h} (\omega_2 - \gamma) = e^{-\frac{1}{2\sigma_w^2}[2(\omega_2-\gamma+\theta/2)^2-\theta^2/2]} \tag{37}
\]

For the Gaussian case, the result is the same as smoothing with an NBB window, only shifted along \( \omega_2 \) by \( \theta/2 \) and attenuated as a function of the phase-providing frequency by \( e^{-\theta^2/2} \) without any other change in the size or shape of the smoothing window. More generally, if \( h \) is symmetric, then \( \tilde{h}^* (\omega_2 - \gamma + \theta) \tilde{h} (\omega_2 - \gamma) \) is symmetric and centered at \( \gamma - \theta/2 \).

PAC measures require the FO filter to have a bandwidth at least as great as the range of SO frequencies of interest, because the spectral range of fluctuations in the power amplitude is limited to the bandwidth of the filter used to extract them, so that \( \theta \) must remain small relative to the bandwidth of \( h \) and \( |h(\omega + \theta)| \approx |h(\omega)| \). Precisely within this range, PAC estimates may be treated directly as NBB estimators of the bispectrum, which use the window:

\[
H_{PAC} = \left| \tilde{g}(\omega_1) \tilde{h}(\omega_2) \right|^2 \tag{38}
\]

This constraint that limits the spectral range of the SO according to the bandwidth of the FO seems at first glance to be a fundamental consideration in measuring PAC, but the preceding analysis shows it to be in fact a dispensable property of the estimator unrelated to anything inherent in the underlying measured quantity.

One approach to circumventing the spectral limitation of PAC estimators modifies the FO filters used according to the SO frequency. For example, the problem is sometimes addressed by scaling the bandwidth of the FO filter, \( h \), by the center frequency of the SO filter (Berman et al., 2012): \( \tilde{h}_\theta(\omega) = \tilde{h}(\frac{\omega}{\theta}) \), giving

\[
\tilde{h}_\theta^* (\omega_2 + \theta - \gamma) \tilde{h}_\theta (\omega_2 - \gamma) = \tilde{h}^* \left( \frac{\omega_2 - \gamma}{\theta} + \frac{1}{2} \right) \tilde{h} \left( \frac{\omega_2 - \gamma}{\theta} \right) \approx \left| \tilde{h} \left( \frac{\omega_2 - \gamma}{\theta} \right) \right|^2 \tag{39}
\]

This adjustment removes the attenuation with SO frequency by ensuring that the bandwidth of the FO filter is appropriate for modulation frequencies in the range of \( \theta \). It still results in diminishing resolution of the FO with increasing SO, which is not a necessary constraint on bispectral estimators.

An apparent disadvantage of standard bispectral estimators might be noted at this point, which is the shift of the FO spectrum by \( \theta/2 \). The reason for the shift becomes apparent if one considers the most basic example of PAC: a signal composed of an amplitude modulated sinusoid \( (1 - \cos(\theta t))e^{i\omega t} \), added to the modulating sinusoid \( e^{i\theta t} \). The spectrum of this signal contains 4 peaks, at \( \omega = \theta \), \( \omega = \gamma - \theta \), \( \omega = \gamma \) and \( \omega = \gamma + \theta \):

\[
\tilde{X}(\omega) = \delta(\omega - \theta) + 2\delta(\omega - \gamma) + \delta(\omega - \gamma - \theta) + \delta(\omega - \gamma + \theta)
\]

In the bispectrum, two unique peaks appear where the product \( \tilde{X}(\omega_1)\tilde{X}(\omega_2)\tilde{X}(\omega_1+\omega_2) \) does not vanish, at \( (\omega_1 = \theta, \omega_2 = \gamma - \theta) \) and \( (\omega_1 = \theta, \omega_2 = \gamma) \). The midpoint of these peaks is therefore centered at the SO frequency, \( \theta \) along \( \omega_1 \) but shifted towards zero by \( \theta/2 \) from the center of the FO band. This shift from the FO frequency is however not essential to the bispectral estimator and can be removed through the simple change of variables given in Eq. (15).

It should be reemphasized that these differences between PPC and bispectral estimators involve rather arbitrary properties of the respective smoothing kernels and do not reflect any essential difference in the quantity measured. It might be argued that the limit on the combined range and resolution of the estimator imposed by PAC helps tune it to bispectral characteristics of PAC. Such tuning may improve signal-to-noise sensitivity to PAC as well as computational
Fig. 3. Comparison of bicoherence and PAC for two test signals, one modeling a periodic broad-spectrum transient and the other, a nested oscillation with no entrainment of phase over frequencies. **Top Row:** The first test signal was designed to contain nested oscillations by adding 30-80 Hz filtered white noise amplitude modulated according to the phase in a second 6-10 Hz band of filtered noise, which was also added, along with a background of 0 dB $1/f$ noise (top left panel; signal with noise is gray line). **Bottom Row:** For the second test signal, a periodic train of transients was simulated as $\exp[10\cos(\phi(t))]$ where $\phi(t)$ is analytic phase from 6-10Hz filtered white noise, to which 0 dB $1/f$ noise was added (bottom left panel). **Middle Column:** Bicoherence clearly reveals the presence of nested oscillations in the first test signal (top middle panel). For the second test signal, the origin of cross-frequency coupling in sharp transients is reflected by harmonic structure and terms on the diagonal (bottom middle panel). **Right Column:** Phase-power coherence computed with 1 Hz bandwidth for phase and 40 Hz for amplitude also reveals the presence of nested oscillations in the first test signal (top right panel). Anisotropic smoothing in PAC obscures the symmetry of the 3rd order polyspectrum under exchange of axes, but the effect of symmetric terms can still be discerned as a slight increase in coupling between low frequency amplitude high-frequency phase (arrow). For the second test signal, the harmonic structure is reflected in the bands within the phase-providing frequencies, but the relative importance of diagonal terms is obscured by smoothing along the amplitude-frequency ($\omega_2$) axis (bottom right panel).

efficiency, but it nevertheless forces some *a priori* assumptions about the bandwidth and spectral range of PAC. More importantly, smoothing may sacrifice information vital for distinguishing between alternative forms of cross-frequency coupling.

### 3.2. Interregional PAC and Cross Bicoherence

The foregoing discussion extends naturally to bivariate cross-frequency coupling involving two signals, for example, obtained from two separate recording channels. The bivariate extension of PAC measures describe the dependence between amplitude in one signal with phase in another. Using the same argument as in the single-channel case, measures of interregional PAC may be equated to the cross bispectrum of the form:

$$B_{ij}(\omega_1, \omega_2) = E\left[\tilde{X}_i(\omega_1)\tilde{X}_j(\omega_2 - \frac{\omega_1}{2})\tilde{X}_j^*(\omega_2 + \frac{\omega_1}{2})\right]$$  \hspace{1cm} (40)

where in the language of PAC, $i$ indexes the “phase-providing” signal and $j$ the “amplitude-providing” signal.
4. Bispectral Signatures of PAC

We have seen that phase-power coherence and related measures of PAC are fundamentally bispectral quantities, which differ from conventional bispectral estimators only in the shape of the smoothing quantities. While past authors have noted some similarities between bispectral and PAC measures (Kramer et al., 2008; Hyafil et al., 2015), the formal equivalence between them appears not to have been previously shown. Next we consider how the bispectrum might be applied towards identifying PAC, followed by a discussion of some broader implications for both measures.

4.1. Preliminary Definition

First, we need a more precise statement of what is meant by PAC. An often-cited example of what is not meant are cases of “spurious” PAC that arise from recurring nonlinear transient signal features, such as spectrally broad “spikes” or sharp-edged waves, which may or may not be periodic (see Figure 3) (Kramer et al., 2008). Such false PAC will tend to exhibit a consistent phase relationship across a range of frequencies, in particular, along the harmonics of any fundamental periodicity. “True” PAC is taken here to mean nested oscillations with the following characteristics:

(a) For a fast oscillation (FO) to be “nested,” within a slow oscillation (SO), its amplitude must vary at a time scale around the period of the slow oscillation.
(b) The FO should be concurrent with the SO.
(c) The phase of the FO must either lack any consistent phase relationship with that of the SO or it must fall within a band that is isolated from the surrounding harmonics of the FO.

The first point implies that each burst of the FO occupies a bandwidth wider than the center frequency of the slow oscillation in which it is embedded, giving a characteristically smooth spectrum at the scale of the SO. The second point excludes non-concurrent responses; for example, a fast oscillation followed by a slower one. The third point excludes spectrally broad features associated with sharp transients and ensures that any oscillation nested in the SO with a consistent phase relationship is genuinely oscillatory. These points justify bispectral criteria for identifying PAC outlined in the following sections. More generally, they provide an entry point for characterizing nonlinear signal features through the bispectrum.

4.2. Signal Model

The following section presents a basic signal model that will serve as a starting point in describing bispectral features associated with phase-amplitude coupling and other third-order signal features. It describes the case when recurring transient features with characteristic but unknown spectra lie embedded at unknown times in the observed record. The generating process can be separated into two parts: a (1) point process, $N$, whose increments determine the times, $\tau_i$, at which (2) some random or deterministic transient feature, $f_i(t)$ is embedded in the signal. The first process is concerned only with timing and the second only with the emitted waveform. The spectral representation of such a signal is

$$\tilde{X}(\omega) = \sum_i \tilde{f}_i(\omega) e^{-i\omega \tau_i}$$

(41)

The following sections consider the bispectrum that results from this process when $f_i$ is independent of time and of $N$ and, in the first case, deterministic, and, in the second, i.i.d. random.

4.2.1. Deterministic Features

If the $f_i$ are deterministic with $f_i = f$, it follows from the convolution property (Eq. 8)

$$B_X(\omega_1, \omega_2) = \tilde{f}(\omega_1) \tilde{f}(\omega_2) \tilde{f}^*(\omega_1 + \omega_2) B_N(\omega_1, \omega_2)$$

(42)
with
\[
B_N(\omega_1, \omega_2) = \iiint e^{-i\omega_1 r - i\omega_2 s + i(\omega_1 + \omega_2)t} E[dN(r) dN(s) dN(t)]
\] (43)

For a “simple” point process, events coincide with vanishing probability, so that \(E[\langle dN(t) \rangle^k] = E[dN(t)]\). For a process with the first three moments defined, \(\mu_1, \mu_2, \mu_3\) (Bartlett 1963; Daley and Vere-Jones 2003), the expectation in Eq. (43) yields
\[
\begin{align*}
E[dN(r) dN(s) dN(t)] &= [\mu_1(t)\delta(r-t)\delta(s-t) \\
&+ \mu_2(r-t)\delta(s-t) + \mu_2(s-t)\delta(r-t) + \mu_2(s-t)\delta(r-s) + \mu_3(r-t, s-t)] dr ds dt
\end{align*}
\] (44)

The bispectrum for this process is then
\[
B_N(\omega_1, \omega_2) = \lambda [1 + \tilde{\mu}_2(\omega_1) + \tilde{\mu}_2(\omega_2) + \tilde{\mu}_2(\omega_1 + \omega_2) + \tilde{\mu}_3(\omega_1, \omega_2)]
\] (45)

where \(\lambda\) is the average rate \((\lambda / T \rightarrow \lambda)\).

For a stationary homogeneous Poisson driving process (having constant \(\mu(t) = \lambda\)),
\[
B_N(\omega_1, \omega_2) = \lambda + \lambda^2 [\delta(\omega_1) + \delta(\omega_2) + \delta(\omega_1 + \omega_2)] + \lambda^3 \delta(\omega_1)\delta(\omega_2)
\] (46)

so that, neglecting the probability mass at the origins and along \(\omega_1 + \omega_2 = 0\),
\[
B_X = \lambda \tilde{f}(\omega_1) \tilde{f}(\omega_2) \tilde{f}^*(\omega_1 + \omega_2)
\] (47)

This quantity is nonzero when \(f\) contains harmonics or is otherwise spectrally broad, so that its support covers some \(\omega_1, \omega_2\), and \(\omega_1 + \omega_2\). In fact, because \(f\) is transient by assumption, this condition is already given: \(f\) must decay to zero within some finite time window, \(\Delta T\), and therefore its spectrum contains at a minimum a main lobe and side lobes spaced at \(2\pi / \Delta T\). But if \(f\) has zero mean, a high center frequency and is highly oscillatory with an otherwise narrow spectrum, the product may still be negligible. The bispectrum will contain larger peaks when \(f\) has a broad spectrum on the order of 2/3 its center frequency, non-zero mean, or contains one or more harmonic complexes.

4.2.2. Random Features
Extending the preceding analysis to the case when each \(f_i\) is itself the realization of a random process independent of \(N\),
\[
B_X(\omega_1, \omega_2) = \sum_{i,j} E \left[ \tilde{f}_i(\omega_1) \tilde{f}_j(\omega_2) \tilde{f}^*_i(\omega_1 + \omega_2) e^{i(\omega_1(\tau_i) + \omega_2(\tau_i))} \right]
\]
\[
= E \left[ \sum_i \tilde{f}_i(\omega_1) \tilde{f}_i(\omega_2) \tilde{f}^*_i(\omega_1 + \omega_2) \\
+ \sum_{i \neq j} \tilde{f}_i(\omega_1) \tilde{f}_j(\omega_2) \tilde{f}^*_j(\omega_1 + \omega_2) e^{-i\omega_1(\tau_i - \tau_j)} \\
+ \sum_{i \neq j} \tilde{f}_i(\omega_2) \tilde{f}_j(\omega_1) \tilde{f}^*_j(\omega_1 + \omega_2) e^{-i\omega_2(\tau_i - \tau_j)} \\
+ \sum_{i \neq j \neq k} \tilde{f}_i(\omega_1) \tilde{f}_j(\omega_2) \tilde{f}^*_k(\omega_1 + \omega_2) e^{-i\omega_1(\tau_i - \tau_k) - \omega_2(\tau_j - \tau_k)} \right]
\] (48)
Fig. 4. Regions of the bispectrum used by the inside and outside criteria. Phase-amplitude coupling is characterized by energy in the “outside” regions given by the support of \( \tilde{f}_{SO} f_{FO} \) (gray boxes) while harmonic features related to signal transients also tend to create energy in the “inside” regions (white boxes).

If the process generating \( f_i \) is independent of \( N \) and time, so that \( \tilde{f}_i \sim \tilde{f}_j \), for all \( i, j \), and \( N \) simple, with the first three moments defined, \( \mu_1, \mu_2, \mu_3 \) (Bartlett [1963] Daley and Vere-Jones [2003]):

\[
B_X(\omega_1, \omega_2) = \lambda \left[ \langle \tilde{f}(\omega_1) \tilde{f}(\omega_2) \tilde{f}^*(\omega_1 + \omega_2) \rangle + \langle \tilde{f}(\omega_1) \rangle \langle \tilde{f}(\omega_2) \tilde{f}^*(\omega_1 + \omega_2) \rangle \tilde{\mu}_2(\omega_1) \\
+ \langle \tilde{f}(\omega_2) \rangle \langle \tilde{f}(\omega_1) \tilde{f}^*(\omega_1 + \omega_2) \rangle \tilde{\mu}_2(\omega_2) \\
+ \langle \tilde{f}(\omega_1) \rangle \langle \tilde{f}(\omega_2) \rangle \langle \tilde{f}^*(\omega_1 + \omega_2) \rangle \tilde{\mu}_3(\omega_1, \omega_2) \right]
\]  

(49)

with the brackets denoting expectation, and a normalization by the cumulative event count, \( N \), is left implicit. This reduces to \( B_X = B_f B_N \) if the error of \( f_i \) involves independent additive Gaussian noise.

This division of labor allows one to separate the spectral contributions of local features described by \( f \) from those of the large-scale driving process; these respective contributions can be distinguished in the bispectrum. In the case of phase-amplitude coupling, \( f \) will be treated as the sum of two components: a single burst of fast oscillations, \( f_{FO} \), whose spectrum is characteristically broad and smooth, and a slow oscillation, \( f_{SO} \), with a narrow spectrum. In constructing the full signal, narrowband large-scale features can be viewed as the consequence of a filter applied to the train of impulses generated by the point process, where \( f_{SO} \) plays the role of filter function.

†The Poisson formulation adopted here for the driving point process is mildly restrictive in that it assumes the event count within a given interval follows a Poisson distribution. Periodicity in the driving process therefore does not entail a train of evenly spaced events, but instead a train of evenly spaced event clusters with Poisson-distributed size. An alternative point-process model that conditions on history, such as a renewal process, might handle periodicity more naturally in this setting, but elucidating the spectra of such processes is more technically involved, making them less useful for the present purpose. One way to address the case of simple uniform periodicity with the Poisson model, which leaves the foregoing spectral analysis unchanged, is to allow \( \lambda \) to take the form of transient bursts of infinitesimal duration, each burst generating a cluster with Poisson-distributed size. The resulting signal then contains \( f \) at the cluster times modulated by cluster size. The variance of amplitude in \( f \) introduced by this weighting with cluster sizes can be made arbitrarily small by scaling \( \lambda \) so as to increase cluster size, with the inverse scaling applied \( f \) to keep amplitude fixed.
4.3. Bispectral Definition of PAC

If the mean of \( f, \bar{f}(0) \), is nonzero, it is easily seen that the bispectrum of \( f \) contains its own power spectrum, \( \bar{f}(0) |\bar{f}(\omega)|^2 \), along the axes \( \omega_1 = 0 \) and \( \omega_2 = 0 \). More generally, suppose \( f \) is the sum of two components, \( f = f_{SO} + f_{FO} \), the first of which, \( f_{SO} \), occupies a narrower bandwidth than the second at center frequency, \( \xi \), while the amplitude of the second rises and falls transiently at a time scale shorter than the period of the first. This modulation implies that \( f_{FO} \) is effectively windowed at a time scale less than \( 2\pi\xi^{-1} \). In the spectral domain, this equates to: the spectrum of \( f_{FO} \) is smooth at the scale of \( \xi \), which justifies the following approximation

\[ \tilde{f}_{FO}(\omega)\tilde{f}_{FO}(\omega + \xi) \approx \tilde{f}_{FO}(\omega)|^2 e^{-i\xi\tau} \quad (50) \]

where \( \tau \) accounts for some arbitrary time delay. The term \( e^{-i\xi\tau} \) describes the phase of the amplitude modulation relative to the emission time. We may partition the bispectral plane according to the spectral ranges of \( f_{SO} \) and \( f_{FO} \), as shown in Fig. [4]. Where energy falls within this partitioning gives the first set of criteria for distinguishing PAC.

4.3.1. Defining PAC: The Outside Criterion

The first criterion for PAC is that the time scale over which the amplitude of the nested fast oscillation, \( f_{FO} \), varies must be less than the period of the slow oscillation, \( f_{SO} \), in which it is embedded. When this condition applies, the power spectrum of \( f_{FO} \) appears parallel to the \( \omega_2 \) axis within the support of \( \tilde{f}_{SO} \) because within this range

\[ \bar{f}(\omega_1)\bar{f}(\omega_2)\bar{f}^*(\omega_1 + \omega_2) \approx \tilde{f}_{SO}(\omega_1)|\tilde{f}_{FO}(\omega_2)|^2 e^{i\omega_1\Delta\tau} \quad (51) \]

and likewise for the \( \omega_1 \) axis under symmetry, where \( \Delta\tau \) is the time delay between the \( f_{SO} \) and \( f_{FO} \) components. It can be seen that the power spectrum of \( f_{FO} \) is reproduced parallel to the \( \omega_1 \) axis at \( \xi \), while in the orthogonal direction, the spectrum of \( f_{SO} \) is reproduced parallel to \( \omega_2 \) within the support of \( \tilde{f}_{FO}(\omega_1) \). Examples of this effect can be found in Figure [3](middle panels) for both the nested-oscillation and transient test signal.

Extending this analysis to the stochastic case, suppose that the spectrum of \( f_{FO} \) overlaps negligibly with \( f_{SO} \), so that \( \tilde{f}_{SO}(\omega)\tilde{f}_{FO}(\omega) \approx 0 \), then within the support of \( \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2) \) we are left with

\[ E \left[ \tilde{X}(\omega_1)\tilde{X}(\omega_2)\tilde{X}^*(\omega_1 + \omega_2) \right] = \right. \]
\[ \left. + \left\langle \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle \mu_2(\omega_1) \right. \]
\[ + \left\langle \tilde{f}_{FO}(\omega_2)\tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle \mu_2(\omega_2) \right. \]
\[ + \left\langle \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle \mu_2(\omega_1 + \omega_2) \right. \]
\[ + \left\langle \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle \mu_3(\omega_1, \omega_2) \right. \] \quad (52)

When \( f_{SO} \) and \( f_{FO} \) are deterministic this reduces to Eq. (12). At the other extreme, suppose the phase of \( f_{FO} \) is random such that \( \tilde{f}_{FO} = \langle \tilde{f}_{FO} \rangle = 0 \), then only the first two terms of Eq. (52) remain:

\[ E \left[ \tilde{X}(\omega_1)\tilde{X}(\omega_2)\tilde{X}^*(\omega_1 + \omega_2) \right] \]
\[ = \left\langle \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle + \left\langle \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle \mu_2(\omega_1) \]
\[ \approx \left\langle \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle + \left\langle \tilde{f}_{SO}(\omega_1)\tilde{f}_{FO}(\omega_2)\tilde{f}_{FO}(\omega_1 + \omega_2) \right\rangle e^{-i\omega_1\Delta\tau} \mu_2(\omega_1) \] \quad (53)
Fig. 5. Distinguishing the origin of PAC in human ECoG data with bicoherence. Twenty minutes of ECoG data were obtained from a patient undergoing invasive clinical monitoring for epilepsy while the patient watched a television show. Contact 216 over occipital cortex (bottom row) recorded frequent interictal spikes, while the neighboring contact, 222 (top row) contained few spikes. **Left column**: A 5 second window showing a typical series of discharges in channel 216 (bottom left panel). **Middle column**: Both channels exhibit strong apparent phase-amplitude coupling as measured by phase-power coherence, in channel 216, between 50-150 Hz gamma power and 10 Hz phase (bottom middle panel) and 100-200 Hz power and 2-5 Hz phase in channel 222 (top middle panel). **Right column**: NNB Bicoherence reflects the phase-amplitude coupling in both cases but also reveals the presence of harmonic diagonal and off-diagonal terms in channel 216 (bottom right panel, white line indicates $\omega_1 = \omega_2$), suggesting that PAC in this case originates from the sharp-edged ictal discharges. In channel 222 (top right panel), bicoherence reveals no harmonic structure and little energy along the diagonal and off-diagonal region, implying that PAC cannot likely be explained by sharp-edged signal features. Insets show symmetric bicoherence over the full range of frequencies with the region in the adjoining panels indicated by the box.

The first term does not vanish when there is a consistent relationship of phase between $f_{SO}$ and the amplitude modulation of $f_{FO}$. The second term remains when there is also a consistent lag between both terms and the emission of the point process, which is reflected accordingly in the weighting by the spectrum of the point process, $\tilde{\mu}_2(\omega_1)$. But because we are free to define the emission times of the point process such that $\tau_{SO} = 0$, the second term merely reflects the power spectral contribution of the driving point process. The bispectral estimate within the support of $f_{SO}(\omega_1)f_{FO}(\omega_2)$ therefore depends on the timing of the amplitude of $f_{FO}$ relative to the phase of $f_{SO}$, but not the phase of $f_{FO}$.

This form of of cross-frequency dependence fulfills sensible criteria for phase-amplitude coupling: it implies a correlation between the amplitude of the amplitude-providing signal component, $f_{FO}$, and an underlying oscillation with a longer period than the time-scale of the amplitude modulation of $f_{FO}$, which is what is meant conventionally by phase-amplitude coupling. Reflecting its time scale, the spectrum of $f_{FO}$ must be both relatively broad and smooth, but because the power spectrum discards phase information in $f_{FO}$, the phase of $f_{FO}$ does not require any consistent relationship with $f_{SO}$, which is also a frequent characteristic of nested oscillations in phase-amplitude coupling. This condition does not, however, rule out a contribution from spectrally broad transients (see example in Fig. 3, lower middle panel), a possibility that motivates the second and third criteria, next.
Fig. 6. Recovering signal features from the bispectrum. Top Left: The test signal (black line) simulates a series of high-frequency oscillation (FO) followed by a slower transient response, resembling a physiological evoked response embedded in Gaussian 1/f noise at random times (gray line, signal with noise). Top Right: Bicoherence reflects the association between the FO and the following slow response. Bottom Row: The impulse response obtained from an inverse Fourier transform along $\omega_1$ (Right) (See Eq. 54) approximately recovers the slow response over the support of the FO along $\omega_2$ (Left).

4.3.2. Defining PAC: The Inside Criterion

The central part of the bispectrum covers the regions of support for $\tilde{f}_{SO}(\omega_1)\tilde{f}_{SO}(\omega_2)$ and $\tilde{f}_{FO}(\omega_1)\tilde{f}_{FO}(\omega_2)$ (white boxes in Fig. 4). Non-vanishing terms in the bispectrum within these regions tend to reflect spectrally broad features associated with sharp-edged transients or harmonic structure when the driving process is periodic. Harmonic terms will generate a lattice of peaks along axes given by $m\omega_1 = n\omega_2$ where $f(n\omega)f(m\omega)f^*(n+m\omega)$ is non-vanishing; an example of this is found in the lower middle panel of Fig. 3. When the phase of $f_{FO}$ remains in a consistent relationship with that of $f_{SO}$ any of the third, fourth and fifth terms in Eq. (53) may become non-vanishing; for a periodic driving process, these terms create harmonic banding within the support of $f_{FO}$ and peaks at the aforementioned lattice points.

4.3.3. Defining PAC: The Time-Delay Criterion

Eq. (51) implies that the bispectrum contains potentially useful information about the relative timing of SO phase and FO amplitude modulation. Information about timing can be recovered with an inverse Fourier transform along one of the frequency dimensions, as shown in Fig. 6. For the homogenous point process with constant $\lambda$, in the presence of pure PAC:

$$I(\tau, \omega_2) = \int B(\omega_1, \omega_2) e^{i\omega_1 \tau} d\omega_1 \approx f_{SO}(\tau + \Delta \tau) \left| \tilde{f}_{FO}(\omega_2) \right|^2 + \tilde{f}_{SO}(\omega_2) e^{i\omega_2 \Delta \tau} f_{FO}^{(3)}(\tau)$$ (54)
Because $f_{SO}$ and $f_{FO}$ occupy non-overlapping spectral ranges, the result is a two-dimensional function with two bands, the first around the center frequency of $f_{SO}$, containing the autocorrelation of $f_{FO}$ scaled by $\tilde{f}_{SO}(\omega)$ and the second within the support of $\tilde{f}_{FO}$, containing $f_{SO}$ scaled by $|\tilde{f}_{FO}(\omega)|^2$, both shifted by the time delay between them. For a periodic $N$, the terms are reduplicated at the corresponding periodicity.

5. Conclusion

With respect to the relationship between PAC and the bispectrum, we have shown:

(a) Common measures of PAC based on analytic phase and amplitude are in fact bispectral estimators, and as such provide no unique information beyond what is recovered by standard bispectral estimators.

(b) PAC measures are severely biased with respect to the symmetry of the bispectrum and introduce artificial constraints on the range and resolution of the estimator.

These limitations provide a clear rationale for favoring standard bispectral estimators in the evaluation of phase-amplitude coupling. We have further given a more detailed framework by which to evaluate the presence and nature of PAC from the bispectrum.

Beyond clarifying the interpretation of common measures of PAC, some more broadly applicable insights into the meaning of the bispectrum may be drawn from this discussion. In standard second-order spectral analysis it is often useful to regard the signal as the output of a linear system driven by an input of Gaussian white noise. A similar treatment for higher-order spectra must introduce higher-order dependence by way of the driving input or through nonlinearities of the system (Nikias and Mendel, 1993). The former strategy was applied in section 4.2, where the driving point process supplied a flat nonzero bispectrum, while the signal bispectrum was obtained from a linear process, in direct analogy to the second-order case (an approach which may be directly extended to other higher spectra as well). This model lends itself to a particularly straightforward interpretation: it describes a signal containing some recurring feature embedded at times determined by the driving point process. Such a view of the problem is particularly suited to settings that call for the blind identification of unknown features based solely on their stable recurrence over time. In section 4.3.3 we observed how such a feature might be recovered from the bispectrum (see Fig. 6 and Eq. 54) through a simple inverse Fourier transform. More general algorithms for recovering the spectrum of the generating signal may be applicable as well (Bartelt et al., 1984).

Of greatest interest is the application to signals in which multiple or variable features are embedded at unknown times. The system generating such a signal is in general no longer linear, but it might be modeled as a mixture of linear systems. An interesting topic for further consideration is whether and how one might decompose the bispectrum to identify the components of such a mixture.

References

Akam, T. and D. M. Kullmann (2014). Oscillatory multiplexing of population codes for selective communication in the mammalian brain. Nature Reviews Neuroscience 15(2), 111–122.

Barnett, T., L. Johnson, P. Naitoh, N. Hicks, and C. Nute (1971). Bispectrum analysis of electroencephalogram signals during waking and sleeping. Science 172(3981), 401–402.

Bartelt, H., A. W. Lohmann, and B. Wirnitzer (1984). Phase and amplitude recovery from bispectra. Applied Optics 23(18), 3121–3129.

Bartlett, M. (1963). The spectral analysis of point processes. Journal of the Royal Statistical Society. Series B (Methodological), 264–296.
Berman, J. I., J. McDaniel, S. Liu, L. Cornew, W. Gaetz, T. P. Roberts, and J. C. Edgar (2012). Variable bandwidth filtering for improved sensitivity of cross-frequency coupling metrics. *Brain connectivity* 2(3), 155–163.

Bingham, C., M. Godfrey, and J. Tukey (1967). Modern techniques of power spectrum estimation. *IEEE Transactions on Audio and Electroacoustics* 15(2), 56–66.

Birkelund, Y., A. Hanssen, and E. J. Powers (2003). Multitaper estimators of polyspectra. *Signal Processing* 83(3), 545–559.

Canolty, R. T., E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *science* 313(5793), 1626–1628.

Cohen, L. (1989, Jul). Time-frequency distributions—a review. *Proceedings of the IEEE* 77(7), 941–981.

Daley, D. J. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods* (2nd ed.). Springer Science & Business Media.

Dumermuth, G., P. Huber, B. Kleiner, and T. Gasser (1971). Analysis of the interrelations between frequency bands of the eeg by means of the bispectrum a preliminary study. *Electroencephalography and clinical neurophysiology* 34(2), 137–148.

Fackrell, J., P. White, J. Hammond, R. Pinnington, and A. Parsons (1995). The interpretation of the bispectra of vibration signals: I. theory. *Mechanical Systems and Signal Processing* 9(3), 257–266.

Gan, T. J., P. S. Glass, A. Windsor, F. Payne, C. Rosow, P. Sebel, and P. Manberg (1997). Bispectral index monitoring allows faster emergence and improved recovery from propofol, alfentanil, and nitrous oxide anesthesia. *The Journal of the American Society of Anesthesiologists* 87(4), 808–815.

Gerr, N. L. (1988, Mar). Introducing a third-order wirger distribution. *Proceedings of the IEEE* 76(3), 290–292.

Godfrey, M. D. (1965). An exploratory study of the bi-spectrum of economic time series. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 14(1), 48–69.

Hagihira, S., M. Takashina, T. Mori, T. Mashimo, and I. Yoshiya (2001). Practical issues in bispectral analysis of electroencephalographic signals. *Anesthesia & Analgesia* 93(4), 966–970.

Hanssen, A. and L. L. Scharf (2003). A theory of polyspectra for nonstationary stochastic processes. *IEEE Transactions on Signal Processing* 51(5), 1243–1252.

Hyafil, A., A.-L. Giraud, L. Fontolan, and B. Gutkin (2015). Neural cross-frequency coupling: connecting architectures, mechanisms, and functions. *Trends in neurosciences* 38(11), 725–740.

Kearse, L. A., C. Rosow, A. Zaslavsky, P. Connors, M. Dershwitz, and W. Denman (1998). Bispectral analysis of the electroencephalogram predicts conscious processing of information during propofol sedation and hypnosis. *The Journal of the American Society of Anesthesiologists* 88(1), 25–34.

Kovach, C. (In Press). A biased look at phase locking: Brief critical review and proposed remedy. *IEEE Transactions on signal processing*.

Kovach, C. K. and P. E. Gander (2016). The demodulated band transform. *Journal of neuroscience methods* 261, 135–154.
Kramer, M. A., A. B. Tort, and N. J. Kopell (2008). Sharp edge artifacts and spurious coupling in eeg frequency comodulation measures. *Journal of neuroscience methods 170*(2), 352–357.

Myles, P., K. Leslie, J. McNeil, A. Forbes, M. Chan, B.-A. T. Group, et al. (2004). Bispectral index monitoring to prevent awareness during anaesthesia: the b-aware randomised controlled trial. *The lancet 363*(9423), 1757–1763.

Nikias, C. L. and J. M. Mendel (1993). Signal processing with higher-order spectra. *IEEE signal processing magazine 10*(3), 10–37.

Nikias, C. L. and M. R. Raghuveer (1987). Bispectrum estimation: A digital signal processing framework. *Proceedings of the IEEE 75*(7), 869–891.

Sasaki, K., T. Sato, and Y. Yamashita (1975). Minimum bias windows for bispectral estimation. *Journal of Sound and Vibration 40*(1), 139–148.

Schack, B., N. Vath, H. Petsche, H.-G. Geissler, and E. Möller (2002). Phase-coupling of theta–gamma eeg rhythms during short-term memory processing. *International Journal of Psychophysiology 44*(2), 143–163.

Sigl, J. C. and N. G. Chamoun (1994). An introduction to bispectral analysis for the electroencephalogram. *Journal of clinical monitoring 10*(6), 392–404.

Swami, A. (1991, Apr). Third-order wigner distributions: definitions and properties. In *[Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, pp. 3081–3084 vol.5.

Tort, A. B. L., R. Komorowski, H. Eichenbaum, and N. Kopell (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of Neurophysiology 104*(2), 1195–1210.