

Statistical analysis of periodic data in neuroscience

Daniel H. Baker

13/12/2020

Abstract

Background

A widely used paradigm in many branches of neuroscience is to drive the nervous system using periodic stimuli. This entrains neural responses at the stimulation frequency, resulting in high signal-to-noise ratios relative to single stimulus presentations. These periodic responses, often called the *steady-state* or *frequency following* response, can be recorded from single neurons (Enroth-Cugell and Robson 1966) and local field potentials (Morrone, Burr, and Speed 1987) using invasive methods, or with non-invasive electroencephalography (EEG) and magnetoencephalography (MEG) systems, both in humans (Norcia et al. 2015) and in diverse animal species including insects (Afsari et al. 2014), birds (Porciatti et al. 1990), rodents (Hwang et al. 2019) and primates (Nakayama and Mackeben 1982). Steady-state methods are used to measure early sensory responses in vision (Regan 1966), hearing (Rees, Green, and Kay 1986) and somatosensation (Snyder 1992), and closely related paradigms have been developed to target specific stimulus features such as orientation (Braddick, Wattam-Bell, and Atkinson 1986), and facial expression (Gray et al. 2020) and identity (Liu-Shuang, Norcia, and Rossion 2014). In fMRI research, *travelling wave* methods (Engel et al. 1994; Sereno et al. 1995) are used to map the retinotopic responses of early visual cortex, again using stimuli that change periodically, this time in spatial position. Finally, physiological reflexes such as the pupillary response to light can be entrained in a similar way (Spitschan et al. 2014).

A convenient way to analyse the data from periodic stimulation experiments is to take the Fourier transform of the measured signal. The amplitude of the response at the stimulation frequency (and its harmonics - integer multiples of the stimulation frequency) is a convenient, and well-isolated index of the brain's response. Fourier spectra comprise both amplitude and phase information, which can be equivalently expressed as complex numbers with real and imaginary components. In many studies the phase information is routinely discarded, and statistical comparisons are performed on the amplitude data only. However an alternative is to use multivariate statistics, which take into account both the amplitude and phase information (represented as real and imaginary components). Multivariate methods have the advantage that they are more sensitive to weak signals, and therefore offer increased statistical power relative to univariate methods.

For pointwise and pairwise comparisons, Hotelling's T^2 statistic (Hotelling 1931) is a multivariate extension of the t-test. For the one-sample case, the test statistic is defined as:

$$T^2 = N(\bar{x} - \mu)'C^{-1}(\bar{x} - \mu), \quad (1)$$

where N is the number of observations, \bar{x} is the multivariate sample mean, μ is the point of comparison, C^{-1} is the inverse covariance matrix, and $'$ denotes vector transposition. Conceptually, the T^2 statistic extends the univariate t-statistic by incorporating the covariance between the dependent variables. Two-sample and paired variants are also available, and the test can be applied to an arbitrary number of dependent variables (though here we will consider only the bivariate case).

More recently, Victor and Mast (1991) proposed a simpler version of T^2 , called T_{circ}^2 . The T_{circ}^2 statistic makes the strong assumption that the dependent variables (real and imaginary values) are uncorrelated and have equal variance. When these conditions are met, the test statistic is defined as:

$$T_{circ}^2 = (N - 1) \frac{|\bar{x} - \mu|^2}{\sum |x_j - \bar{x}|^2} \quad (2)$$

where x_j denotes the j -th observation of the dependent variables, and all other terms retain their previous meanings. Notice that no covariance term is present in equation (2), because of the independence assumption. This makes the statistic simpler to calculate, but causes problems when the assumption is violated.

In the present paper, we develop best practice guidelines for performing statistical tests on multivariate Fourier components derived from periodic stimulation paradigms. We first demonstrate why parametric univariate statistics are inappropriate for such data, because amplitudes for weak signals are not normally distributed. We then investigate conditions under which we should use either the T^2 or T_{circ}^2 statistic. The range of sample sizes and effect sizes where T_{circ}^2 is more sensitive is identified. We develop a novel method for testing the assumptions of the T_{circ}^2 statistic, based on calculating the condition index of a multivariate data set. We then discuss appropriate methods for identifying outliers and calculating error bars. Finally, we demonstrate the proposed techniques on two example SSVEP data sets, and recommend some best practice guidelines for analysis decisions. All data and scripts used to produce this manuscript are available at: <https://github.com/bakerdh/FourierStats>. These include R functions to implement one-sample Hotelling's (1931) T^2 statistic, Victor & Mast's (1991) T_{circ}^2 statistic, and the condition index statistic proposed in this paper.

Fourier amplitudes violate parametric assumptions of univariate statistics

Many studies use parametric t-tests and analysis of variance (ANOVA) to analyse periodic data. Specifically, the amplitude component of the Fourier spectrum at the stimulation frequency is used as the dependent variable, discarding the phase information. This is problematic, because the amplitude is an absolute quantity, and can never fall below zero. Distributions of amplitudes for weak signals are therefore likely to be positively skewed, and in general will violate the assumption of normality.

The upper row of Figure (fig:amphists) shows scatterplots of simulated Fourier components, expressed using real (x) and imaginary (y) components. The amplitudes are the lengths of the lines joining each grey point to the origin. The lower row in Figure (fig:amphists) shows distributions of amplitudes for the same set of signal strengths. These distributions only approach normality when the signal strength is more than twice the standard deviation (Cohen's $d > 2$; Cohen's d is the mean difference scaled by the standard deviation, see Cohen (1988)). One consequence of this is that t-tests will typically have an inflated Type 1 error (false positive) rate for many signals encountered empirically.

Typical solutions, such as log-transforming the data, are unlikely to be equally applicable to all conditions. For example, if one wishes to compare a baseline where no stimulus was presented with a condition involving a strong signal, the former will be skewed and the latter normal. Applying a transform to both conditions is therefore problematic. Non-parametric statistics are a potential option, but these have generally lower statistical power than their parametric equivalents. Instead, the bivariate statistics discussed in the introduction avoid these issues and have greater statistical power, as we will demonstrate in the next section.

Conditions under which T_{circ}^2 is more sensitive than T^2

Using the T^2 and T_{circ}^2 allow us to retain the phase information, and therefore have greater power than univariate T-tests, as well as avoiding problems caused by using absolute amplitude values. Victor & Mast (1991) report simulations showing situations where T_{circ}^2 has greater power than T^2 . Here we replicate and extend these simulations (see Figure (fig:powerfig)), and demonstrate that the regime where T_{circ}^2 has

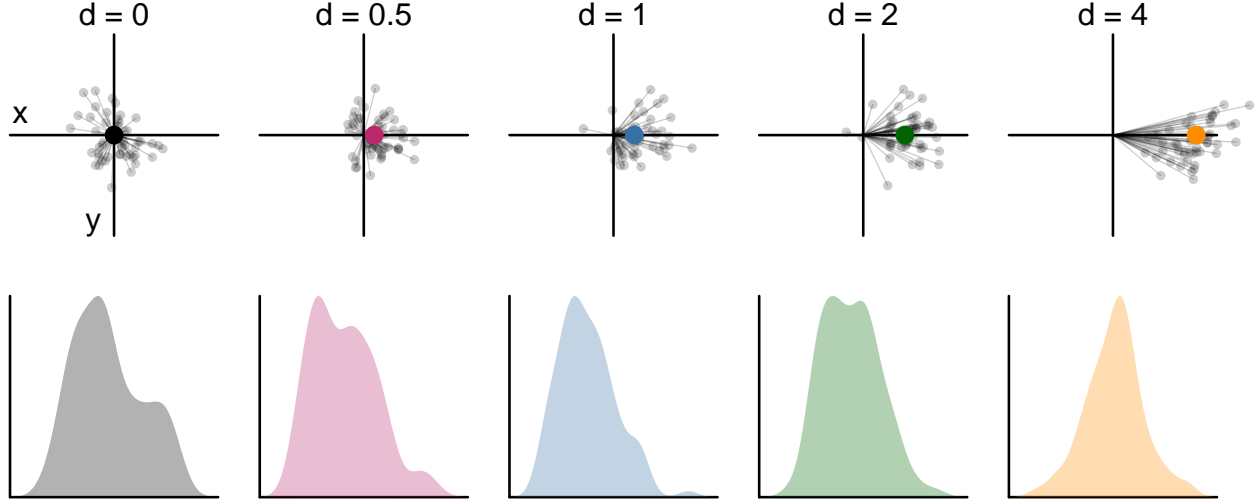


Figure 1: Demonstration of skew in absolute Fourier amplitudes for weak signals. Signal strength is quantified as Cohen's d , defined as the ratio of the mean to the standard deviation of the sample. The upper row shows samples of 50 grey points, and the population mean (coloured points). The lower row shows kernel density functions generated from 100,000 amplitude values. Note that the mean phase of the signal is irrelevant for these simulations, and is shown in the positive x -direction for consistency.

greater power occurs particularly for large effect sizes and small sample sizes (see Figure @ref(fig:powerfig)f). However, for effect sizes around $0.5 < d < 1$, T_{circ}^2 is more sensitive even with around 16 observations. This advantage is lost for large sample sizes ($N > 32$) and large effect sizes ($d > 2$ when $N > 8$)

Limitations of T_{circ}^2 when assumptions are violated

A novel method to test the assumptions of T_{circ}^2

Despite the severe consequences of violating the assumptions of the T_{circ}^2 statistic, there is currently no accepted test for those assumptions that could be applied to real data. Victor and Mast (1991) suggest that their test should be applicable to multiple repetitions of a stimulus condition collected from a single individual, whereas data from multiple participants may be less likely to exhibit independence of the real and imaginary components (Pei et al. 2017).

One convenient way to test the assumptions of T_{circ}^2 is to assess the *condition index* of a data set, which describes the ratio of eigenvalues for a cloud of points. The eigenvectors are the major and minor axes of the bounding ellipse (the straight lines in the example icons at the foot of Figure @ref(fig:falsealarms)a,b). Conventionally, the condition index is calculated as the square root of the longest/shortest eigenvector length. For uncorrelated random numbers the distribution of condition indices is positively skewed, with a minimum of 1 (Edelman 1988). This is because two independent samples of numbers from the same underlying distribution will still by chance have unequal eigenvectors. Edelman (1988) provides an equation (Eq 14) for the probability density function of condition indices (x) as a function of sample size (n):

$$pdf = (n-1)2^{n-1} \frac{x^2-1}{(x^2+1)^n} x^{(n-2)}$$

Attempts to validate this by simulation suggest that for small sample sizes ($n < 10$) a closer approximation is given by:

$$pdf = (n-2)2^{n-2} \frac{x^2-1}{(x^2+1)^{(n-1)}} x^{(n-3)}$$

In Figure @ref(fig:distcomparison)a we show an example distribution derived from the second expression (black curve), the inverse cumulative density function (red curve), and simulations from 100000 random

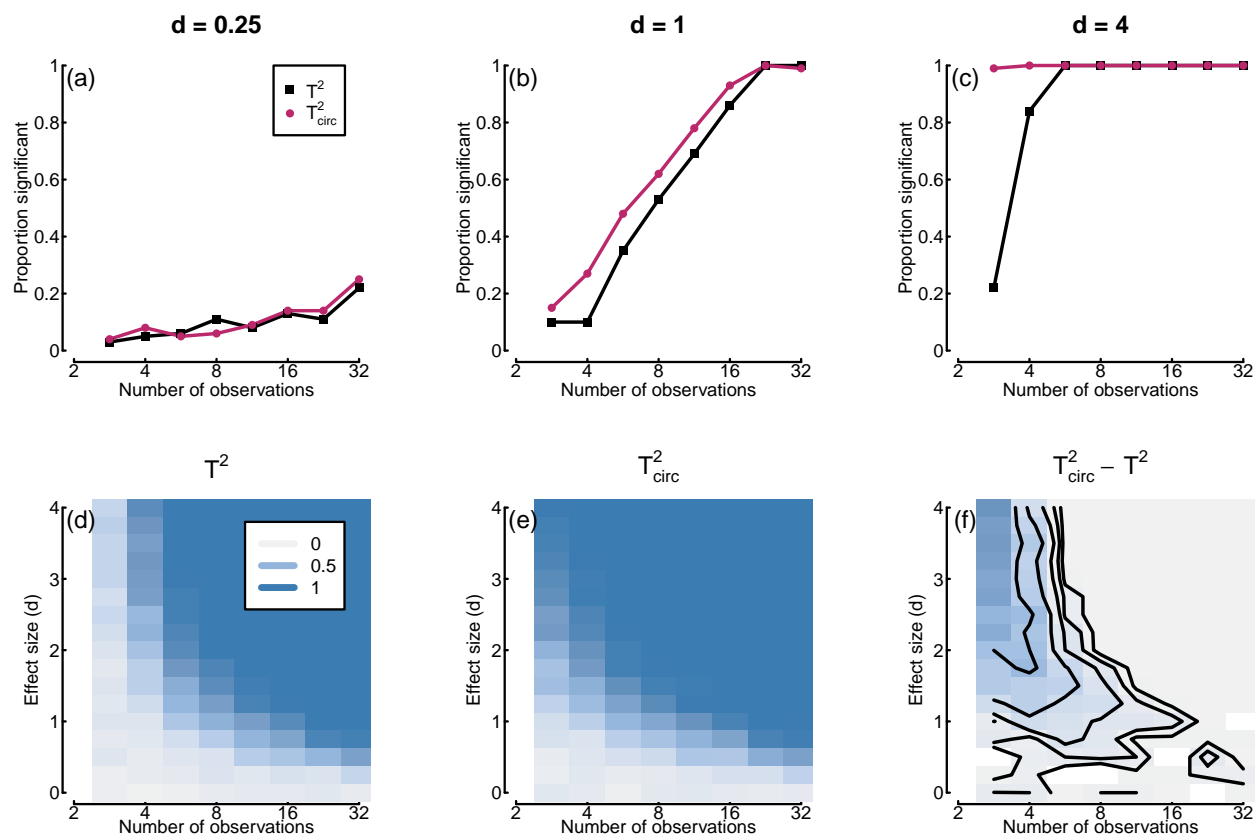


Figure 2: .

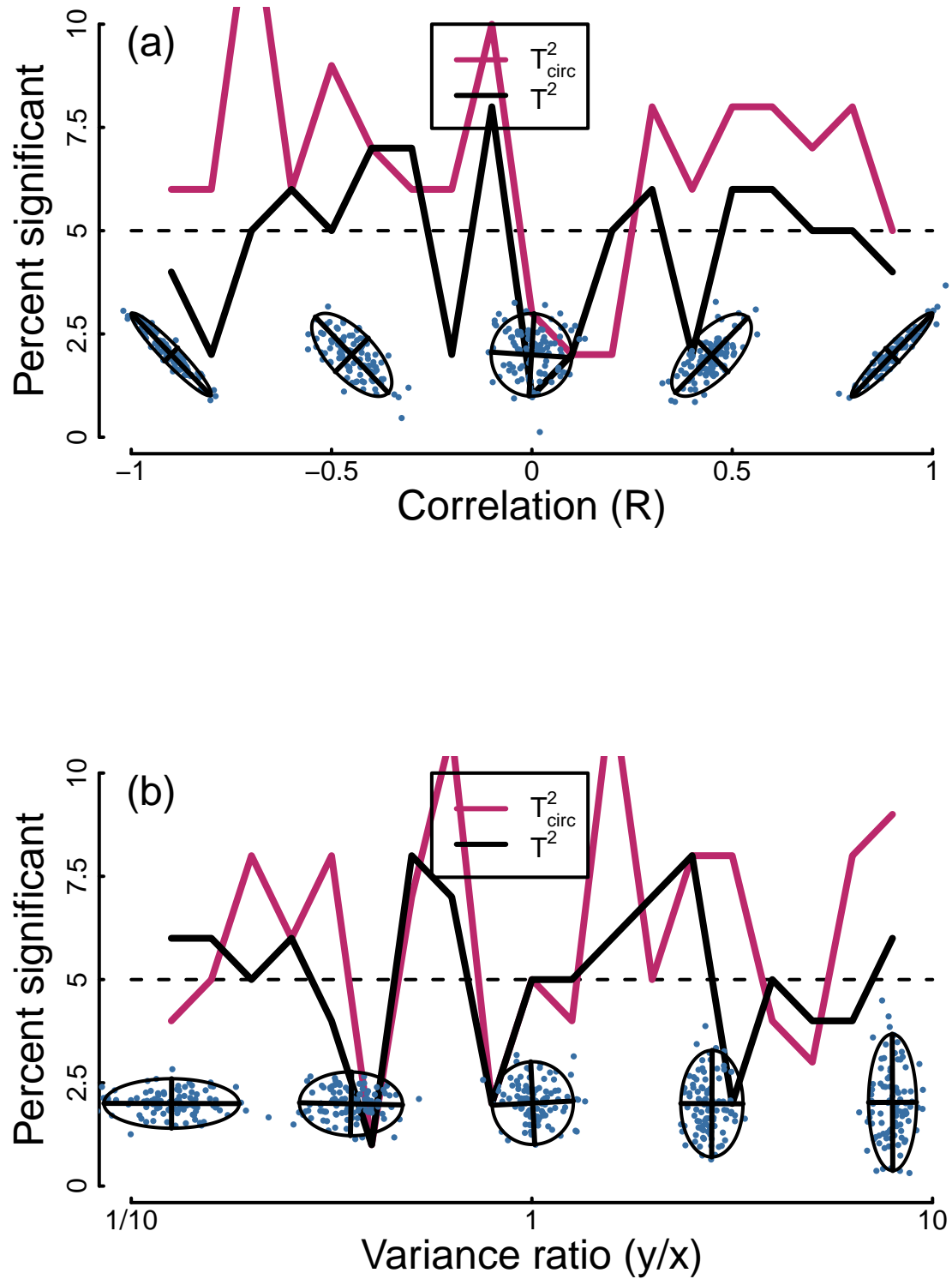


Figure 3: Simulations showing the Type I error rate for both tests as a function of the correlation between two variables (a) and the ratio of variances (b). Estimates are for 100000 simulations per condition, with $N=10$ observations. The icons at the foot of each panel show example scatterplots with bounding ellipses and eigenvectors.

data sets for $n=10$ (blue shading). The vertical lines show the critical (95%) threshold for the analytic and simulated results. Figure @ref(fig:distcomparison)b shows how these thresholds change as a function of the number of observations, and it is clear that the modified expression (red) most closely approximates the simulation results (blue).

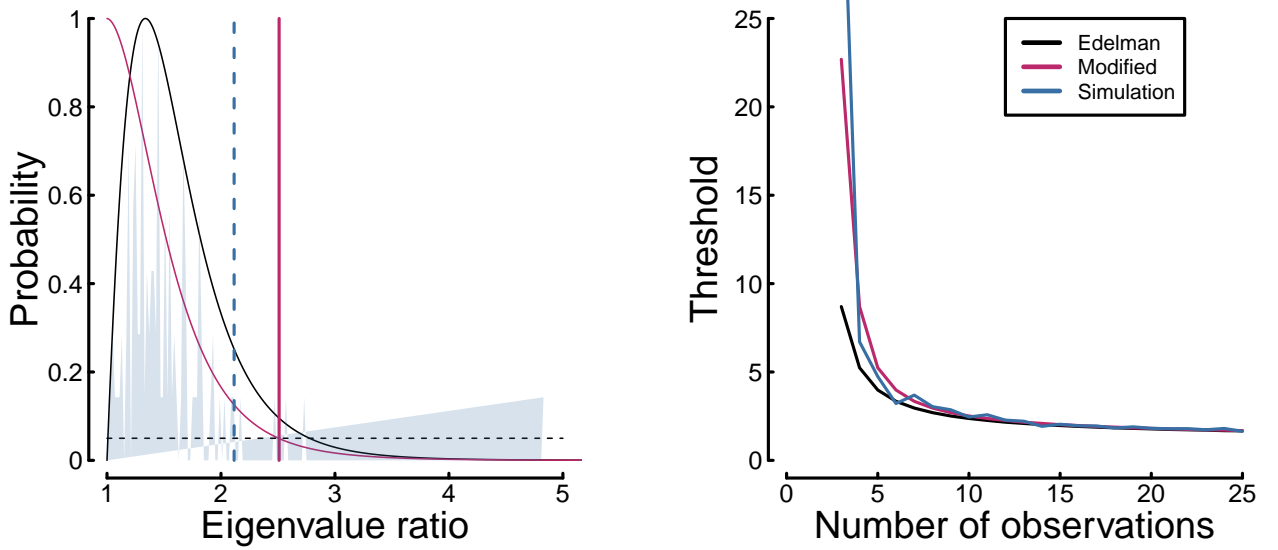


Figure 4: d

The eigenvalue ratio can be used as a test of the assumptions of T_{circ}^2 . If we observe a condition index above the critical threshold for the number of observations then the data set can be said to significantly violate the assumption of equal eigenvalues. Because our modified equation provides an inverse density function (red curve in Figure @ref(fig:distcomparison)a), we can use this to calculate a p-value for the test. If the test is non-significant, one can proceed with T_{circ}^2 ; if it is significant, T^2 should be used instead.

Identifying and removing outliers using the Mahalanobis distance

If a data set produces a significant result using the condition index test, this could be due to the presence of one or more outliers. The Mahalanobis distance (Mahalanobis 1936) is a useful metric for identifying such multivariate outliers so that they can be excluded. It calculates the Euclidean distance between each data point and the centroid, and scales it by the variance in the direction of the vector that joins the two points. This means that any correlations in the data set are taken into account when calculating the distance metric, D .

The effectiveness of this approach to outlier exclusion can be assessed by simulation using the condition index test. Figure @ref(fig:outlierplot) shows the proportion of significant condition index tests as a function of the Mahalanobis distance of a single outlier, for a range of sample sizes (curves). In all cases, the functions depart from the Type 1 error rate ($\alpha = 0.05$; horizontal dashed line in Figure @ref(fig:outlierplot)) when the Mahalanobis distance exceeds a value around 3. This seems a reasonable heuristic for outlier exclusion, and is the multivariate equivalent of excluding data points more than 3 standard deviations from the mean. Following this heuristic should reduce the likelihood that outliers will invalidate the assumptions of the T_{circ}^2 test.

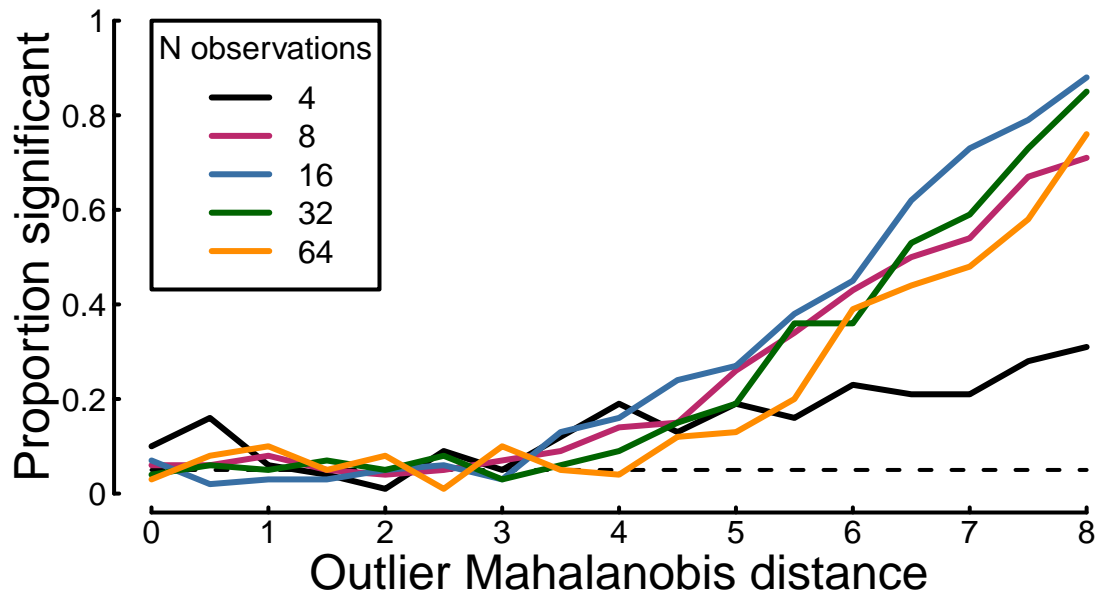


Figure 5: d

Methods for calculating error bars for Fourier amplitudes

Applying these methods to real data sets

Human visual steady-state data

Mouse auditory steady-state data

<https://www.nature.com/articles/s41597-020-00621-z> <https://doi.gin.g-node.org/10.12751/g-node.e5tyek/>

Recommendations for analysing periodic data

Multi/Bivariate statistics should be used for phase-locked Fourier data

Always remove outliers with a Mahalanobis distance > 3

Always test assumptions if using T_{circ}^2

No benefit to using tsqcirc for sample sizes > 32

With more than two conditions use MANOVA

Use bootstrapping to calculate error bars

Further considerations

Mention about non phase-locked responses

Greater power with coherent averaging

Could develop ANOVA-circ

References

- Afsari, Farinaz, Kenneth V Christensen, Garrick Paul Smith, Morten Hentzer, Olivia M Nippe, Christopher J H Elliott, and Alex R Wade. 2014. “Abnormal Visual Gain Control in a Parkinson’s Disease Model.” *Hum Mol Genet* 23 (17): 4465–78. <https://doi.org/10.1093/hmg/ddu159>.
- Braddick, O J, J Wattam-Bell, and J Atkinson. 1986. “Orientation-Specific Cortical Responses Develop in Early Infancy.” *Nature* 320 (6063): 617–9. <https://doi.org/10.1038/320617a0>.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum, Hillsdale, New Jersey.
- Edelman, Alan. 1988. “Eigenvalues and Condition Numbers of Random Matrices.” *SIAM Journal on Matrix Analysis and Applications* 9 (4): 543–60. <https://doi.org/10.1137/0609045>.
- Engel, S A, D E Rumelhart, B A Wandell, A T Lee, G H Glover, E J Chichilnisky, and M N Shadlen. 1994. “fMRI of Human Visual Cortex.” *Nature* 369 (6481): 525. <https://doi.org/10.1038/369525a0>.
- Enroth-Cugell, C, and J G Robson. 1966. “The Contrast Sensitivity of Retinal Ganglion Cells of the Cat.” *J Physiol* 187 (3): 517–52. <https://doi.org/10.1113/jphysiol.1966.sp008107>.
- Gray, Katie L H, Tessa R Flack, Miaomiao Yu, Freya A Lygo, and Daniel H Baker. 2020. “Nonlinear Transduction of Emotional Facial Expression.” *Vision Res* 170 (May): 1–11. <https://doi.org/10.1016/j.visres.2020.03.004>.
- Hotelling, H. 1931. “The Generalization of Student’s Ratio.” *The Annals of Mathematical Statistics* 2 (3). Institute of Mathematical Statistics: 360–78. <http://www.jstor.org/stable/2957535>.
- Hwang, Eunjin, Ritchie E Brown, Bernat Kocsis, Tae Kim, James T McKenna, James M McNally, Hio-Been Han, and Jee Hyun Choi. 2019. “Optogenetic Stimulation of Basal Forebrain Parvalbumin Neurons Modulates the Cortical Topography of Auditory Steady-State Responses.” *Brain Struct Funct* 224 (4): 1505–18. <https://doi.org/10.1007/s00429-019-01845-5>.
- Liu-Shuang, Joan, Anthony M Norcia, and Bruno Rossion. 2014. “An Objective Index of Individual Face Discrimination in the Right Occipito-Temporal Cortex by Means of Fast Periodic Oddball Stimulation.” *Neuropsychologia* 52 (January): 57–72. <https://doi.org/10.1016/j.neuropsychologia.2013.10.022>.
- Mahalanobis, P.C. 1936. “On the Generalised Distance in Statistics.” *Proceedings of the National Academy of Sciences of India* 2: 49–55.
- Morrone, M C, D C Burr, and H D Speed. 1987. “Cross-Orientation Inhibition in Cat Is Gaba Mediated.” *Exp Brain Res* 67 (3): 635–44. <https://doi.org/10.1007/BF00247294>.
- Nakayama, K, and M Mackeben. 1982. “Steady State Visual Evoked Potentials in the Alert Primate.” *Vision Res* 22 (10): 1261–71. [https://doi.org/10.1016/0042-6989\(82\)90138-9](https://doi.org/10.1016/0042-6989(82)90138-9).
- Norcia, Anthony M, L Gregory Appelbaum, Justin M Ales, Benoit R Cottareau, and Bruno Rossion. 2015. “The Steady-State Visual Evoked Potential in Vision Research: A Review.” *J Vis* 15 (6): 4. <https://doi.org/10.1167/15.6.4>.
- Pei, Francesca, Stefano Baldassi, Jeffrey J Tsai, Holly E Gerhard, and Anthony M Norcia. 2017. “Development of Contrast Normalization Mechanisms During Childhood and Adolescence.” *Vision Res* 133 (April): 12–20. <https://doi.org/10.1016/j.visres.2016.03.010>.
- Porciatti, V, G Fontanesi, A Raffaelli, and P Bagnoli. 1990. “Binocularity in the Little Owl, *Athene Noctua*. II. Properties of Visually Evoked Potentials from the Wulst in Response to Monocular and Binocular Stimulation with Sine Wave Gratings.” *Brain Behav Evol* 35 (1): 40–48. <https://doi.org/10.1159/000115855>.
- Rees, A, G G Green, and R H Kay. 1986. “Steady-State Evoked Responses to Sinusoidally Amplitude-Modulated Sounds Recorded in Man.” *Hear Res* 23 (2): 123–33. [https://doi.org/10.1016/0378-5955\(86\)90009-2](https://doi.org/10.1016/0378-5955(86)90009-2).

- Regan, D. 1966. "Some Characteristics of Average Steady-State and Transient Responses Evoked by Modulated Light." *Electroencephalogr Clin Neurophysiol* 20 (3): 238–48. [https://doi.org/10.1016/0013-4694\(66\)90088-5](https://doi.org/10.1016/0013-4694(66)90088-5).
- Sereno, M I, A M Dale, J B Reppas, K K Kwong, J W Belliveau, T J Brady, B R Rosen, and R B Tootell. 1995. "Borders of Multiple Visual Areas in Humans Revealed by Functional Magnetic Resonance Imaging." *Science* 268 (5212): 889–93. <https://doi.org/10.1126/science.7754376>.
- Snyder, A Z. 1992. "Steady-State Vibration Evoked Potentials: Descriptions of Technique and Characterization of Responses." *Electroencephalogr Clin Neurophysiol* 84 (3): 257–68. [https://doi.org/10.1016/0168-5597\(92\)90007-x](https://doi.org/10.1016/0168-5597(92)90007-x).
- Spitschan, Manuel, Sandeep Jain, David H Brainard, and Geoffrey K Aguirre. 2014. "Opponent Melanopsin and S-Cone Signals in the Human Pupillary Light Response." *Proc Natl Acad Sci U S A* 111 (43): 15568–72. <https://doi.org/10.1073/pnas.1400942111>.
- Victor, J D, and J Mast. 1991. "A New Statistic for Steady-State Evoked Potentials." *Electroencephalogr Clin Neurophysiol* 78 (5): 378–88. [https://doi.org/10.1016/0013-4694\(91\)90099-p](https://doi.org/10.1016/0013-4694(91)90099-p).