

Available online at www.sciencedirect.com**ScienceDirect**Journal homepage: www.elsevier.com/locate/cortex

Commentary

Remarks on the analysis of steady-state responses: Spurious artifacts introduced by overlapping epochs

Lucas Benjamin, Ghislaine Dehaene-Lambertz and Ana Fló*

Cognitive Neuroimaging Unit, CNRS ERL 9003, INSERM U992, CEA, Université Paris-Saclay, NeuroSpin Center, Gif/Yvette, France

ABSTRACT

Periodic and stable sensory input can result in rhythmic and stable neural responses, a phenomenon commonly referred to as neural entrainment. Although the use of neural entrainment to investigate the regularities the brain tracks has increased in recent years, the methods used for its quantification are not well-defined in the literature. Here we argue that some strategies used in previous papers, are inadequate for the study of steady-state response, and lead to methodological artefacts. The aim of this commentary is to discuss these articles and to propose alternative measures of neural entrainment. Specifically, we applied four possible alternatives and two epoching approaches reported in the literature to quantify neural entrainment on simulated datasets. Our results demonstrate that overlapping epochs, as used in the original Batterink and colleagues articles, inevitably lead to a methodological artefact at the frequency corresponding to the overlap. We therefore strongly discourage this approach and encourage the re-analysis of data based on overlapping epochs. Additionally, we argue that the use of time–frequency decomposition to compute phase coherence at low frequencies to reveal neural entrainment is not optimal.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Neural entrainment

It is well established that a stable and rhythmic stimulation elicits a stable and rhythmic neural response (e.g., Picton et al., 2003; Regan, 1977), a phenomenon usually referred to

as neural entrainment or steady-state responses. An enhanced cortical response, measurable with electro or magneto-encephalography (EEG or MEG), appears at very specific frequencies in function of the periodicity of the stimulation. Although it is still discussed whether it results from multiple evoked responses to the stimulus train (Capilla et al., 2011) or from the alignment of intrinsic brain oscillation to the input (Doelling et al., 2019), this enhanced response

* Corresponding author. Cognitive Neuroimaging Unit, CNRS ERL 9003, INSERM U992, CEA, Université Paris-Saclay, NeuroSpin Center, Gif/Yvette, France.

E-mail address: ana.flo@cea.fr (A. Fló).

<https://doi.org/10.1016/j.cortex.2021.05.023>

0010-9452/© 2021 Elsevier Ltd. All rights reserved.

indicates which event the brain is tracking in a train of stimuli. It can be low-level properties of the sensory input (the “on-effect” of an image for example), or more abstract regularities, such as a face presented every four images (de Heering & Rossion, 2015). Therefore, this phenomenon receives more and more attention due to its theoretical implications (Buzsáki, 2006; Giraud & Poeppel, 2012), and also to its promising uses in special populations such as infants. For example, it has been used to study categorization in infants (de Heering & Rossion, 2015; Peykarjou et al., 2017), the tracking of linguistic structures in adults (Ding et al., 2015, 2017), and speech segmentation in adults (Batterink & Paller, 2017, 2019; Buiatti et al., 2009) and infants (Choi et al., 2020; Kabdebon et al., 2015).

To quantify neural entrainment, frequency-domain analyses are preferred to conventional time-domain analysis for two reasons. First, it is generally easy to predict the frequencies at which the increased response should occur (Zhou et al., 2016) reducing the number of statistical comparisons. Second, frequency-domain analysis provides a better signal-to-noise ratio (SNR) (Norcia et al., 2015). However, different methods can be used and have been used in the literature, some presenting a risk of bias when overlapping data are used (Batterink & Paller, 2017, 2019; Choi et al., 2020). Thus, our goal is to demonstrate the risk of bias when using overlapping data and provide artifact free alternatives. Additionally, we discuss the pros and cons of the different methods and then provide evidence with simulated data.

Although our point can be extended to all studies using neural entrainment, we decided to simulate a speech segmentation study. Neural entrainment is an adequate and extensively used tool to recover speech perception units within natural sentences (Ding et al., 2015, 2017) or artificial streams (Buiatti et al., 2009; Kabdebon et al., 2015). Because speech is a continuous stream of phonemes, syllables, and words, an important question notably concerns how infants succeed to chunk this stream in candidate words. It has been proposed that infants can take advantage of the probabilities of transitions between successive syllables that are constant within a word and drop at its boundaries (Saffran et al., 1996). Thus, if a stream is composed of four tri-syllabic words randomly concatenated without repetition, the transition probability is 1 within a word and .33 between words. Neural entrainment appears as a suitable approach to investigate this experimental paradigm. As the syllables have equal duration, the neural response is entrained at the syllable rate, as many segmentation experiments report (Batterink & Paller, 2017, 2019; Buiatti et al., 2009; Choi et al., 2020; Kabdebon et al., 2015). If the participant discovers the word regularity, the neural response should also be entrained at the word rate as also reported in the same studies.

1.2. Methodological considerations

When using rapid periodic stimulation, a sustained oscillatory activity is expected to emerge at the stimulation frequencies. The most direct analytical approach would be to transform the entire dataset in the frequency domain by applying, for example, the Fast Fourier Transformation (FFT) algorithm. However, this procedure is not the most commonly used

because the signal's stationary nature allows obtaining a better SNR by dividing the signal into shorter segments aligned with the stimulation. In brief, having many data segments with aligned activity enables separating phase-locked from non-phase-locked activity. Therefore, two constraints oppose achieving a correct analysis of neural entrainment: on the one hand, to have long segments and, on the other hand, to have many of them.

Each epoch should be long enough for the signal and noise to be well represented, and thus, to obtain an optimal description in the frequency domain. The epoch's length determines the frequency resolution ($f_{res} = 1/\text{epoch length}$), and the frequency resolution has to be at least half the frequency of the slower steady-state response to be able to see it in the frequency domain. In other words, the slower response must repeat at least once to be detected in the frequency domain. Moreover, longer epochs also provide a better SNR because, with a higher frequency resolution, the noise is spread over more frequency bins, while the signal remains restricted to specific frequencies (Norcia et al., 2015).

Conversely, dividing the recording into sub-segments aligned with the stimulus onset allows focusing on the coherent activity and provides a better SNR. One way of doing so is by averaging the responses across epochs. The averaging process enhances the phase-locked activity and reduces non-phase locked activity, such as in classical event-related potentials (ERP). Subsequently, entrainment appears as an increase in power at specific frequencies in the average signal. The amplification of phase-locked activity by averaging allows the detection of entrainment even when the steady-state responses are much weaker than the background noise. Another way is by directly computing the phase coherence across the epochs. If activity phase-locked to the stimuli onset exists, the phases' distribution will not be uniform at certain frequencies, but phases will be coherent across epochs. The measurement of this coherence is called Inter Trial Coherence (ITC). Many trials are needed to estimate the phase distribution and to measure the ITC properly. Therefore, in both computations, the same opposite constraints exist. Long segments provide a better frequency resolution. Many epochs allow a better enhancement of phase-locked activity or to better estimate the phases' distribution.

While most studies use multiple epochs to estimate the power spectrum or the phase coherence across them, the computations are differently implemented in practice. We have isolated two epoching approaches and four different calculations to estimate neural entrainment (see Fig. 1). We applied these methods to simulated data to evaluate their performance and potential biases.

1.3. Methods used in the literature

As mentioned before, the definition of epochs plays a crucial role in these analyses. Epochs have to be long enough to properly represent signal and noise, and at the same time, the more epochs, the better to detect phase-locked activity. In most of the papers, the segment's length is calculated to include 10–25 repetitions of the slower response. Although the epoch's length may affect sensitivity, it does not lead to artifactual measures of entrainment. Most papers have used

independent epochs (e.g., de Heering & Rossion, 2015; Ding et al., 2015; Kabdebon et al., 2015), but some have used overlapping epochs to increase the total number of epochs (Batterink & Paller, 2017, 2019; Buiatti et al., 2009; Choi et al., 2020) (Fig. 1). We believe that using overlapping epochs entails a methodological artifact; so, we decided to test both approaches.

The four methods described in the literature further differ in the measures to estimate neural entrainment (power spectrum or phase coherence) and the procedure used to compute those measures (order of operations or whole epochs vs time–frequency analysis). In Method 1, the power spectrum is computed for single epochs and then averaged across epochs to obtain an averaged power spectrum. Buiatti and colleagues used this approach in a speech segmentation task in adults (Buiatti et al., 2009). In this case, the entrainment measure depends on the presence of induced and evoked activity that is strong enough to emerge from the noise level. In Method 2, the average across epochs is obtained first, and its power spectrum is computed afterward. Here, non-phase locked activity is first reduced by the averaging step, which increases sensitivity to phase-locked activity. Method 2 is the most common approach used in numerous papers (e.g., de Heering &

Rossion, 2015; Ding et al., 2015). Notice that averaging the signal before computing the power spectrum makes Method 2 phase-dependent and similar to methods directly quantifying the ITC across all epochs, as Method 3 does. However, because Method 3 is exclusively phase-based, it is less dependent on amplitude changes of the signal than power-based methods. Method 3 is also widely used in the literature (e.g., Kabdebon et al., 2015). Method 4 has been implemented by Batterink and colleagues in speech segmentation experiments (Batterink & Paller, 2017, 2019; Choi et al., 2020). It also measures ITC, but through time–frequency analysis. Time–frequency analysis can be performed using different algorithms, notably by wavelet decomposition or by applying the Fourier transform to the data in short sliding time windows. While here we specifically discuss the use of wavelet decomposition (as implemented by Batterink and colleagues), both algorithms give similar results and entail the same type of properties and limitations. We argue that while time–frequency analysis is in theory applicable, it is less suitable for this particular context.

The Matlab scripts of those different methods can be downloaded at <https://github.com/neurokidslab/Neural-Entrainment>.

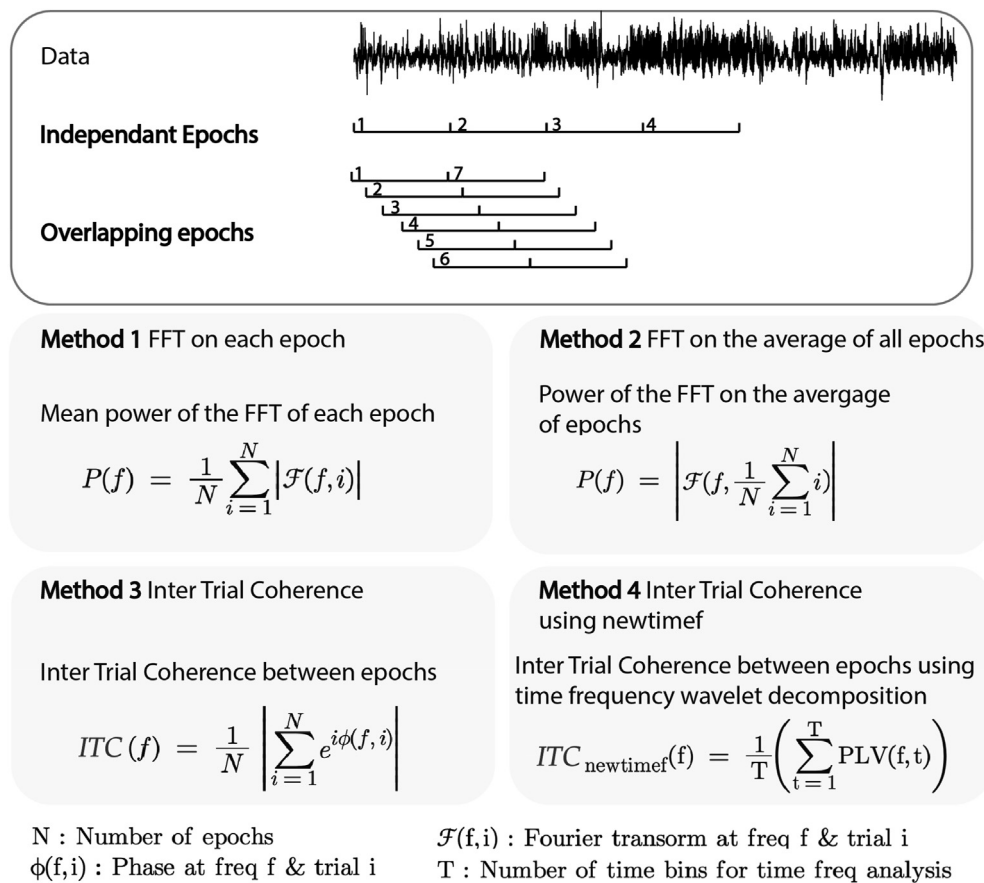


Fig. 1 – Description of the two epoching approaches and the four methods used to estimate neural entrainment. Note that methods 3 and 4 use the same approach (ITC) but on the whole epoch (method 3) or on the time average of time frequency analysis (method 4).

2. Methods

2.1. Creation of the stimulated data

To compare the different methods and their risks of bias, we calculated the results of the four methods described above with independent and overlapping epochs on simulated data. To have datasets with a realistic noise structure, we used real EEG data, to which we added (or not) simulated steady-state responses.

We used EEG data from 27 asleep 5-month-old infants without stimulation. The data were recorded using a 128-electrode net (Electrical Geodesics, Inc.) referred to the vertex with a sampling frequency of 500 Hz, then bandpass-filtered [.2, 40] Hz. The segments containing motion artefacts were removed. Only data from one frontal electrode (Fz) was kept for the simulations. For each infant, we had between 25 and 48 min of artefacts-free data. From these artefacts-free data, we recreated 40 recordings of 720 sec each (12 mn).

First, we simulated a study in which infants listen to randomly concatenated syllables with a fixed duration of 300 msec; thus, the neural signal should be entrained at 3.33 Hz. We did so by adding a half-sinusoidal wave (frequency 2.5 Hz) every 300 msec with amplitude 1 (Fig. 2) to each of the 40 recordings. Second, we simulated a study in which infants discover the three-syllabic words (900 msec) embedded in the stream; thus, the neural response is entrained by both the syllable regularity and the word regularity (i.e., expected neural entrainment at 3.33 and 1.11 Hz). We created the simulated data by adding a half-sinusoidal wave (frequency .8 Hz) every 900 msec with amplitude 2 (Fig. 2) to each of the 40 recordings. The signal to noise ratio was 1/40 for the "syllabic" steady-response (300 msec), and 1/20 for "word" steady-response (900 msec) (the signal to noise ratio was estimated as the signal amplitude divided by the standard deviation of the original EEG data). We chose to use half-sinusoidal waves, because this is the simplest way of simulating evoked ERPs (Yeung et al., 2004). Moreover, half-sinusoidal waves result in the presence of harmonics in the frequency domain (Zhou et al., 2016), which are usually observed in this type of experiments (e.g., Ding et al., 2015).

Thus, we analyzed three sets of data, each one with 40 recordings: (i) original data, (ii) Simulated steady-state response at 3.33 Hz and (iii) Simulated steady-state response at 3.33 Hz and 1.11 Hz. Note that the third set corresponds to Batterink and colleagues' studies (Batterink & Paller, 2017, 2019; Choi et al., 2020). In their studies, the syllabic rate was 3.33 Hz and the word rate 1.11 Hz.

2.2. Data preprocessing

The three sets of data were segmented in epochs of 10.8 sec that either did not overlap (66 epochs), or have an overlap of 11/12 of their length (788 epochs), as in Batterink and colleagues' studies (Batterink & Paller, 2017, 2019; Choi et al., 2020). Notice that 10.8 sec corresponds to 12 times the slower expected steady-state response (i.e., 900 msec).

The analysis for each method was as follows. In Method 1, the FFT was applied to the data, the power spectrum was

calculated for single epochs, and finally, the average power spectrum was obtained. Afterwards, the power spectrum at each frequency bin was normalized by the six adjacent frequency bins. In Method 2, the FFT was applied to the data, the average across epochs was computed, and the power spectrum was computed. As in method 1, the power spectrum at each frequency bin was normalized by the six adjacent frequency bins. In Method 3, the FFT was applied to the epochs, and the ITC was computed. The ITC was normalized by computing at each frequency bin the difference with the ITC at the adjacent six. In Method 4, a Morlet wavelet transformation was used to estimate the time–frequency ITC across epochs, by using the *newtimef* function of EEGLAB (Delorme & Makeig, 2004). The time dimension was then averaged to obtain a single ITC value at each frequency bin. We used the same parameters as Batterink and colleagues (Batterink & Paller, 2017, 2019; Choi et al., 2020) (.1 Hz step, with 1 cycle at .2 Hz linearly increasing until 45 cycles at 20.2 Hz) to faithfully reproduce their method and characterize the risk of artifact associated with it.

3. Results

3.1. Original dataset

In the original dataset, no indices of neural entrainment are expected. It is indeed the case when non-overlapping epochs are used whatever the method: the response is flat across frequency bins (blue lines in Fig. 3 left column for methods 1 to 4). By contrast, using overlapping epochs (blue lines in Fig. 3 right column), creates a pronounced peak at the frequency corresponding to the overlap, for methods 2, 3 and 4, demonstrating a methodological artefact of the approach. The origin of the methodological artefacts stems from the multiple uses of the same data. Averaging the epochs (method 2) creates an artifactual periodic signal at the overlap frequency because the same data appears periodically. Thus, the mean signal power spectrum shows a peak at the frequency 1/overlap and its harmonics. A similar effect occurs when the phase coherence is estimated by using the FFT over the whole epoch or time–frequency analysis by computing the wavelet transform. Comparing the phase between epochs that share a large amount of data only shifted by a fixed value will result in an artificially high phase coherence at the frequency (and harmonics) of this overlap shift value. Interestingly, the presence of artifact when the time–frequency approach is used depends on the wavelet width and the amount of overlapping epochs. If the width of the wavelet does not cover the epochs' overlap, the results will be artifact-free. In the particular case of method 1 there are no visible artefacts when using overlapping epochs, because power is estimated over single epochs; therefore, the phase information is lost before the data of different epochs are averaged.

3.2. Simulated entrainment

When simulated steady-state responses are added, neural entrainment is detected at the syllable and word frequencies by methods 2, 3, and 4 (green and red lines in Fig. 3). The peaks

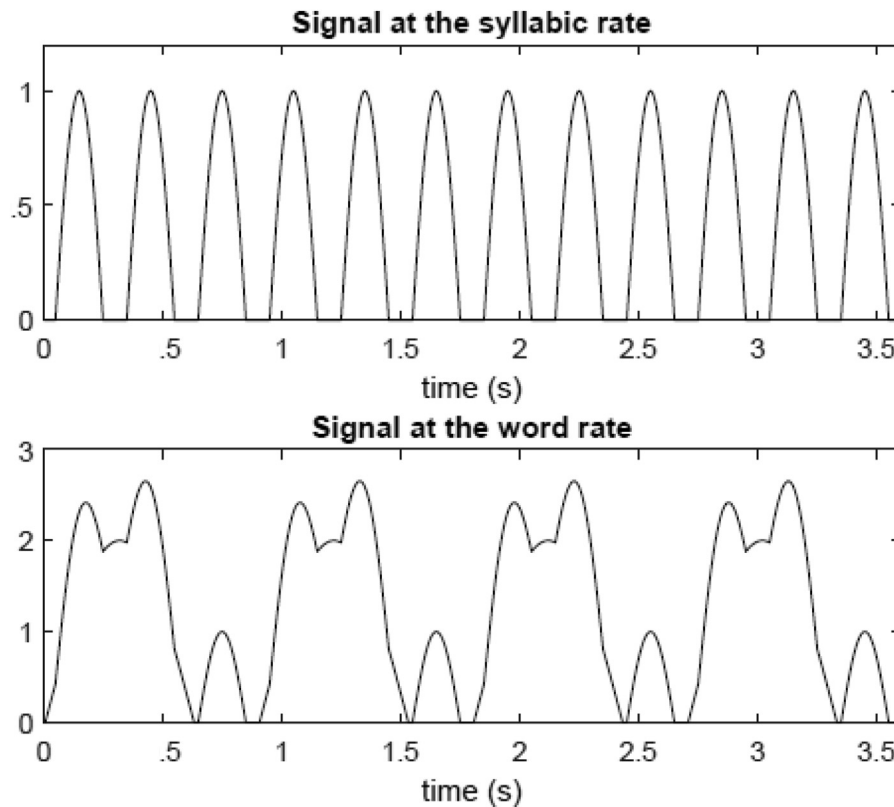


Fig. 2 – Simulated steady-state response to the syllables (one every 300 msec) (top), and to the syllables and the words (one every 900 msec) (bottom).

clearly appear relatively to the no-stimulation condition (blue line) in the case of independent epochs. When using overlapping epochs, the entrainment is added to the intrinsic bias due to overlapping epochs.

Method 1 was not sensitive enough to detect neural entrainment, which can be understood if we consider that the power is computed on single epochs and then averaged. Thus, the steady-state response has to be strong enough relative to the noise to emerge at the single epoch level. In other words, the method does not take advantage of the phase coherence across trials of the steady-state response. The sensitivity is slightly increased when the number of epochs is increased by using overlapping epochs but this method remains sub-optimal.

The ITC computed by performing a time–frequency analysis (method 4) shows a particular behavior. Entrainment appears at the frequency of the overlap and its harmonics, similarly to ITC computed on full epochs (method 3), but the signal is smoother than with the other methods. The smoothness of the signal affects both the estimate of neural entrainment resulting from the simulated steady-state response and the artefact due to overlapping epochs. This phenomenon is a consequence of the uncertainty principle in signal processing. The time–frequency approach allows investigating modulations either in the intensity or in the synchronization of the activity across trials over different frequency bands and at different time points. However, the

uncertainty principle exposes a trade-off between frequency and time resolution ($dt \cdot dw \geq 1/2$), that results in a smoothing of the measure of entrainment. With the parameters used here, at 3.33 Hz the width of the wavelet is 2.36 sec, and at 1.11 Hz its width is 2.71 sec, in both cases much smaller than the duration of a whole epoch (10.8 s). The gain in time resolution, which has no interest here, comes with a loss in frequency resolution explaining the signal's smoothing.

The time-varying information provided by time–frequency approaches like wavelet decomposition is valuable in many contexts, but not meaningful in most steady-state stimulation experiments. In a steady-state experiment, the stimulus appears repeatedly, and the data are divided into sub-segments containing multiple instances of each stimulus. Therefore, looking for a time dimension within epochs is meaningless when all epochs are analyzed together. The strength of steady-state stimulation is that the neural response is periodic and stable, justifying a frequency domain analysis. To obtain a measure of phase coherence at each frequency using time frequency method, the measure must be averaged across the time dimension (Batterink & Paller, 2017, 2019; Choi et al., 2020). This step directly shows the lack of meaning of the time dimension in this context. Additionally, the computational cost of the time–frequency analysis on our simulated data set relative to the whole epoch approach was of the order of 104 times higher.

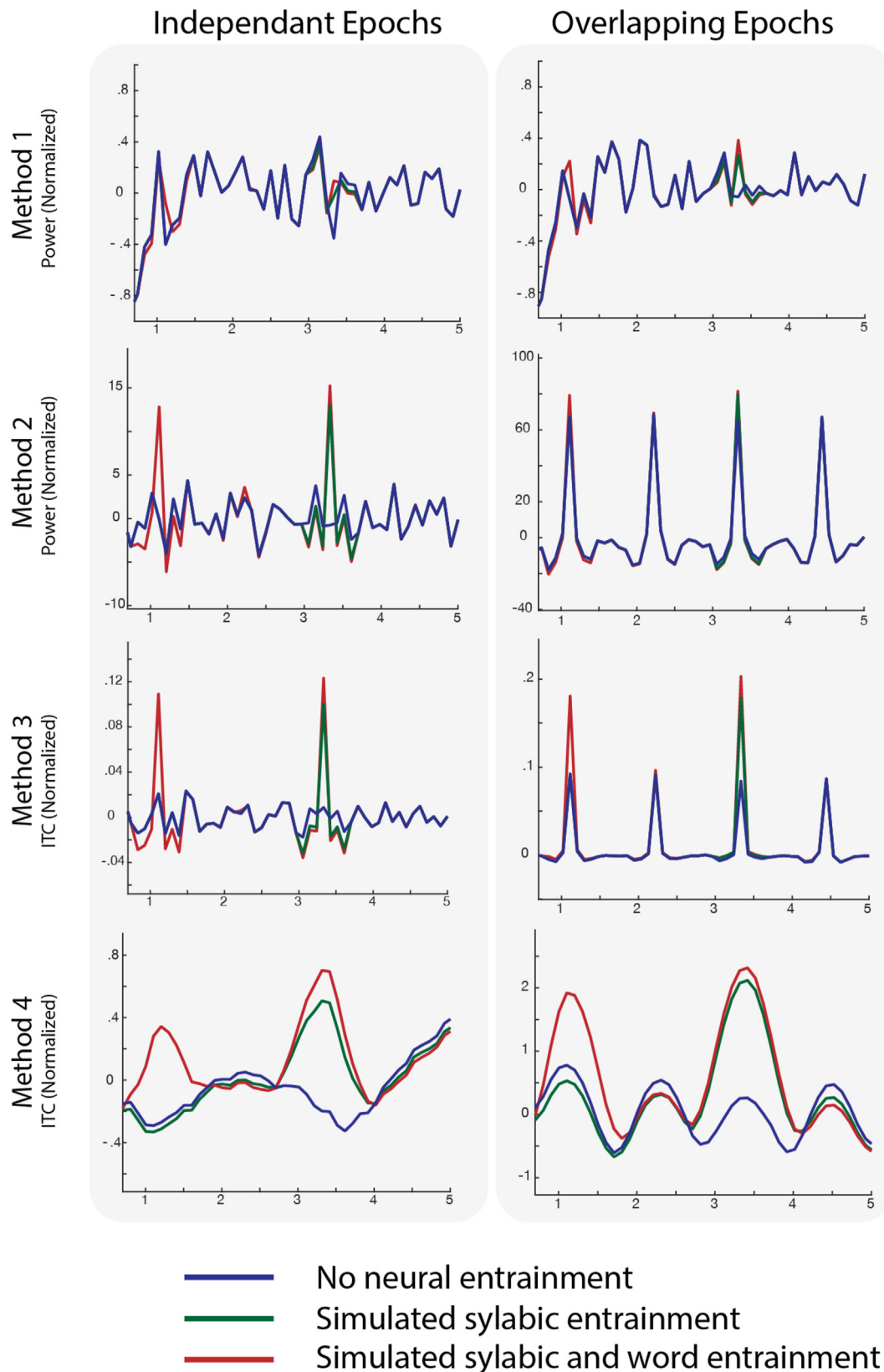


Fig. 3 – Results of the analysis with the different methods and simulations. The expected peaks at the syllable (3.33 Hz) and word (1.11) frequencies are clearly distinguishable from the control situation (blue line) when epochs are independent (left column, line 2 to 4). When overlapping epochs are used, artifactual peaks are present (right column). The time frequency approach used in method 4 (last line) inadequately smooths the frequency profiles.

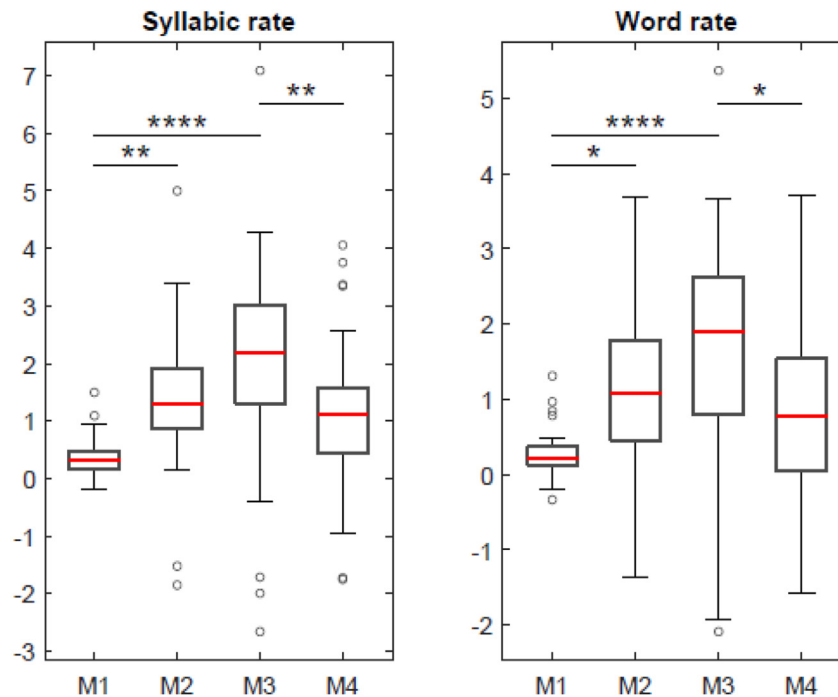


Fig. 4 – Comparison of the different methods to measure neural entrainment at the syllabic and word rate. The measure was computed as the difference between the entrainment when syllabic and word responses were added with the noise level of entrainment. Values were normalized by the standard deviation of the entrainment across frequencies in the range .5–5 Hz.

4. Comparison of the different methods

To compare the sensitivity of the different methods more directly, we used the results obtained from the real data (dataset 1) and the real data plus the simulated signal at the syllabic and word rate (dataset 3). We assessed sensitivity as $S = (E3(f) - E1(f)) / \text{std}(E1)$, where $E3$ is the normalized Power/ITC obtained from the dataset 3 (signal at the syllable and word frequencies), $E1$ is the normalized Power/ITC obtained from dataset 1 (no signal added), f is the frequency of interest. The measure was divided by the standard deviation of $E1$ in the frequency-range .5–5 Hz, to estimate the signal relative to the noise level. We limited the analysis to the use of non-overlapping epochs.

We run 1-way ANOVAs with method as an independent factor to compare the methods' sensitivity (Fig. 4). At the syllabic rate we observed a main effect of method ($F(1,157) = 8.758$, $p = .00356$). Pairwise comparison using Bonferroni correction showed that the effect was due to a difference between method 1 and 2 ($p = .0018$), method 1 and 3 ($p = 8.7 \times 10^{-8}$), and method 3 and 4 ($p = .005$). All the other comparisons were not significant ($p > .05$). Results at the word rate were similar. We observed a main effect of method ($F(1,157) = 6.018$, $p = .0153$). Also in this case, pairwise comparison using Bonferroni correction showed that the effect was due to a difference between method 1 and 2 ($p = .013$), method 1 and 3 ($p = 1.5 \times 10^{-5}$), and method 3 and 4 ($p = .038$). All the other comparisons were not significant ($p > .05$).

These results explicitly demonstrate the lower sensitivity of method 1 for detecting phase locked activity. Using time–frequency analysis to estimate the ITC also resulted in lower sensitivity than using the whole epoch, probably due to the loss of frequency resolution. Measuring neural entrainment based on the power of the average response (method 2) or the ITC (method 3) show comparable sensitivity. This can be understood, because, as mentioned earlier, both approaches rely on the presence of phase locked activity and both are obtained through FFT of the whole epoch.

5. Conclusions

Our simulations show that overlapping epochs unavoidably lead to artifactual entrainment at the overlap frequency. Nevertheless, this does not mean that overlapping epochs prevent the measurement of brain oscillations. The simulated neural entrainment is added to the methodological artefact. Thus, the main results of previous works using this approach might be valid especially when there was a difference with a control condition exposed to the same bias. In principle, if an adequate control condition is used, the entrainment derived from brain activity could be quantified. This is the case of one adult segmentation experiment (Batterink & Paller, 2017), where the entrainment at the word and syllabic rates during the presentation of a structured stream of syllables was compared with the entrainment during a random stream of

syllables. In a second paper, the authors investigated the effect of top-down attention by comparing speech segmentation while manipulating the attention to a competitive task in two groups of subjects (Batterink & Paller, 2019). While there was no within-subjects control condition, we could assume that the methodological artefact is comparable between groups; thus, the comparison across groups would remain valid. In an infant segmentation task experiment using the same methodological approach (Choi et al., 2020), the authors report an increase in entrainment to the word rate relatively to the syllabic rate during the familiarization. While the experiment did not have a control condition, in principle, the increase cannot arise from the methodological artefact.

Based on the above considerations, we do not question the validity of the data sets of Batterink and colleagues and their main results (Batterink & Paller, 2017, 2019; Choi et al., 2020). However, we believe that overlapping epochs should be avoided because they introduce an artefact that compromises the interpretation of the results while not providing a clear advantage. For data transparency, we encourage the re-analysis of the dataset. Additionally, the time–frequency approach has an enormous computational cost compared to whole epoch methods, and the useless time dimension comes with a loss of frequency resolution. Therefore, we see no theoretical advantage in using time–frequency decomposition for a classical analysis of steady-state experiments. For the analysis of classical steady-state experiments, we recommend using whole epoch approaches such as method 2 or 3 on non-overlapping epochs.

Declarations of interest

None.

Author contribution

A.F., G.D., and L.B. conceptualized the study. A.F. and L.B. conducted the analyses. A.F. and L.B. wrote the paper with revisions from G.D.

Acknowledgments

This research has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 695710). We would like to thank Maria Chait for comments on the last version of the manuscript.

REFERENCES

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, 90, 31–45. <https://doi.org/10.1016/j.cortex.2017.02.004>

- Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, 115, 56–71. <https://doi.org/10.1016/j.cortex.2019.01.013>
- Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, 44(2), 509–519. <https://doi.org/10.1016/j.neuroimage.2008.09.015>
- Buzsáki, G. (2006). *Rhythms of the brain*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195301069.001.0001>
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., & Gross, J. (2011). Steady-state visual evoked potentials can be explained by Temporal Superposition of transient event-related responses. *Plos One*, 6(1), e14543. <https://doi.org/10.1371/journal.pone.0014543>
- Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word Patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*. <https://doi.org/10.1177/0956797620933237>, 0956797620933237.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience*, 11. <https://doi.org/10.3389/fnhum.2017.00481>
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1), 158–164. <https://doi.org/10.1038/nn.4186>
- Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, 116(20), 10113–10121. <https://doi.org/10.1073/pnas.1816414116>
- Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. <https://doi.org/10.1038/nn.3063>
- de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *ELife*, 4(June), 1–14. <https://doi.org/10.7554/eLife.06564>
- Kabdebon, C., Pena, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*, 148, 25–36. <https://doi.org/10.1016/j.bandl.2015.03.005>
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottareau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, 15(6). <https://doi.org/10.1167/15.6.4>, 4–4.
- Peykarjou, S., Hoehl, S., Pauen, S., & Rossion, B. (2017). Rapid categorization of human and Ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Scientific Reports*, 7(1), 12526. <https://doi.org/10.1038/s41598-017-12760-2>
- Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *International Journal of Audiology*, 42(4), 177–219. <https://doi.org/10.3109/14992020309101316>
- Regan, D. (1977). Steady-state evoked potentials. *Journal of the Optical Society of America*, 67(11), 1475. <https://doi.org/10.1364/JOSA.67.001475>

- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). *Statistical learning by 8-month-old infants* (Vol. 274). <https://doi.org/10.1126/science.274.5294.1926>
- Yeung, N., Bogacz, R., Holroyd, C. B., & Cohen, J. D. (2004). Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods. *Psychophysiology*, 41(6), 822–832. <https://doi.org/10.1111/j.1469-8986.2004.00239.x>
- Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: Periodicity, Fundamental frequency, and harmonics. *Frontiers in Human Neuroscience*, 10(June), 1–8. <https://doi.org/10.3389/fnhum.2016.00274>

Received 22 February 2021

Reviewed 2 April 2021

Revised 10 April 2021

Accepted 10 May 2021

Published online xxx