PNAS
www.pnas.org

Supplementary Information for

Modeling cortical entrainment to music: the evoked response is not enough

Keith B. Doelling, M. Florencia Assaneo, Dana Bevilacqua, Bijan Pesaran and David Poeppel

Keith Doelling.
E-mail: keith.doelling@gmail.com

This PDF file includes:

Supplementary text
Figs. S1 to S2
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1816414116

Supporting Information Text

Participant kernels in the Evoked Model. The Evoked Model uses a kernel which is convolved with the stimulus envelope to generate an output. Our model uses the grand average response to single tones, combining repetitions over trials and subjects to yield the best kernel shape and thus the most typical phase pattern across note rates. However, M100 responses clearly vary across participants. We, therefore, plot phase patterns based on Evoked models using the kernels from individual participants to see how the evoked model behaves under such variance. **Figure S1** shows a remarkably consistent phase pattern, in which phase lag increases as note rate increases, so much so that in several participants (e.g. Participant 4) the 8 nps condition even overlaps with the lower note rate stimuli.

Analysis of the Evoked Response. A critical assumption of our analysis is that the temporal profile of the evoked response to a single note remains largely the same despite different contexts. This is an assumption we know to be false. The evoked response can change with varying stimulus rates (1–3). More recent studies have shown that the evoked response shape can change due to the content of the input, such as different phonemes in speech (4).

Therefore, we must investigate the evoked responses to the individual notes within our stimuli to understand how our various conditions affected these responses. In Experiment 2, we created the stimuli note by note and therefore have readily available access to the timing of each note relative to the stimulus onset. We were therefore able to use these timings to re-segment the data to analyze the evoked response to individual notes, rather than to the entire clip.

We investigated the evoked response on two dimensions: sharp vs smooth attack, and note rate. **Figure S2A** shows an example of a sharp and smooth note and give a clear example of the type of transformation used. We designed these to significantly reduce the evoked response amplitude. Each note of the assigned pitch were added at the correct time point and duration to build up a musical stimulus. **Figure S2B** shows the evoked response to these individual notes. Here, we removed the 5 and 8 nps conditions as the notes were too close together such that each evoked response is contaminated by the preceding and proceeding note. We found that the response to the smooth stimulus is significantly reduced relative to the sharp stimulus at the M100 and M200 responses. More specifically, there were two clusters of significant difference from 100 - 130 ms (summed $t = -23.25$, $p = 0.038$) and from 160 - 240 ms (summed $t = 32.00, p = 0.0098$)

We are also interested to know how the note rate affected the response to individual notes. **Figure S2CD** shows the average response to individual notes as a function of the stimulus note rate. Immediately clear is that the amplitude of the response decreases as a function of note rate. We ran a Spearman correlation of the M100 peak amplitude for each subject against the stimulus rate. We found a significant negative correlation in both hemispheres (Left: $\rho = -0.690$, $p < 0.0001$, Right: $\rho = -0.682$, $p < 0.0001$). Of critical interest to us is whether the peak lag is affected by note rate. We identified the M100 peak for each condition (marked on the figure as a large dot and line), for each subject and ran permutation statistics to test for such a relationship. We found a significant positive correlation between stimulus note rate and peak lag (Left: $\rho = 0.296, p = 0.010$ Right: $\rho = 0.527, p < 0.0001$). This relationship suggests that the faster the note rate, the *later* the response. Note that an effect in this direction would *decrease* the effect of phase concentration in the MEG data relative to evoked model. As stimulus rate increases the response lag makes up a larger proportion of the cycle length. Therefore, this affect of lag cannot explain the increased phase concentration in the MEG data relative to the evoked model.

Methods.

Data Re-segmentation and Baseline. To analyze the evoked response to individual notes, we first had to re-segment the data to be organized by note rather than musical piece. In experiment 2, we had created the stimuli note by note and therefore were directly aware of the timing of each note. Armed with this information, we first segmented the data by our triggers which indicated the beginning of each piece. We then baselined corrected the clips by the mean of the 500 ms preceding the entire clip. After baselining, we re-segmented the data based on the timing of unique note onsets.

Permutation Testing. To test statistically for differences in the ERP for sharp vs smooth attack, we used a permutation cluster test (5). For each subject, we had an average response for each condition. We randomly shuffled the labels of these average responses and used a T-test at each time point to identify significant differences at $\alpha = 0.05$. We then identified consecutive time points of significance and summed the t-values across these time-points. We conducted this shuffling over 10,000 iterations and for each iteration took the maximum summed t-value. We compared this random distribution against the true sharp vs smooth comparison and used as a p-value the proportion of iterations whose magnitude exceeded that of the true comparison.

To test for differences in peak lag and amplitude as a function of note rate, we ran a similar permutation test. In this case, for each subject, we had an average response for each note rate. We randomly shuffled the labels for each subject and extract the peak lag and amplitude between 70 and 150 ms post note onset. We then ran Spearman rank correlations to test for a monotonic relationship between note rate and both peak lag and amplitude for each subject. We averaged the resulting ρ value across subjects and repeated this over 10,000 iterations. We compared these random distributions against the true average rho values across subjects and used as p-values the proportion of iterations whose magnitude exceeded that of the true correlation.

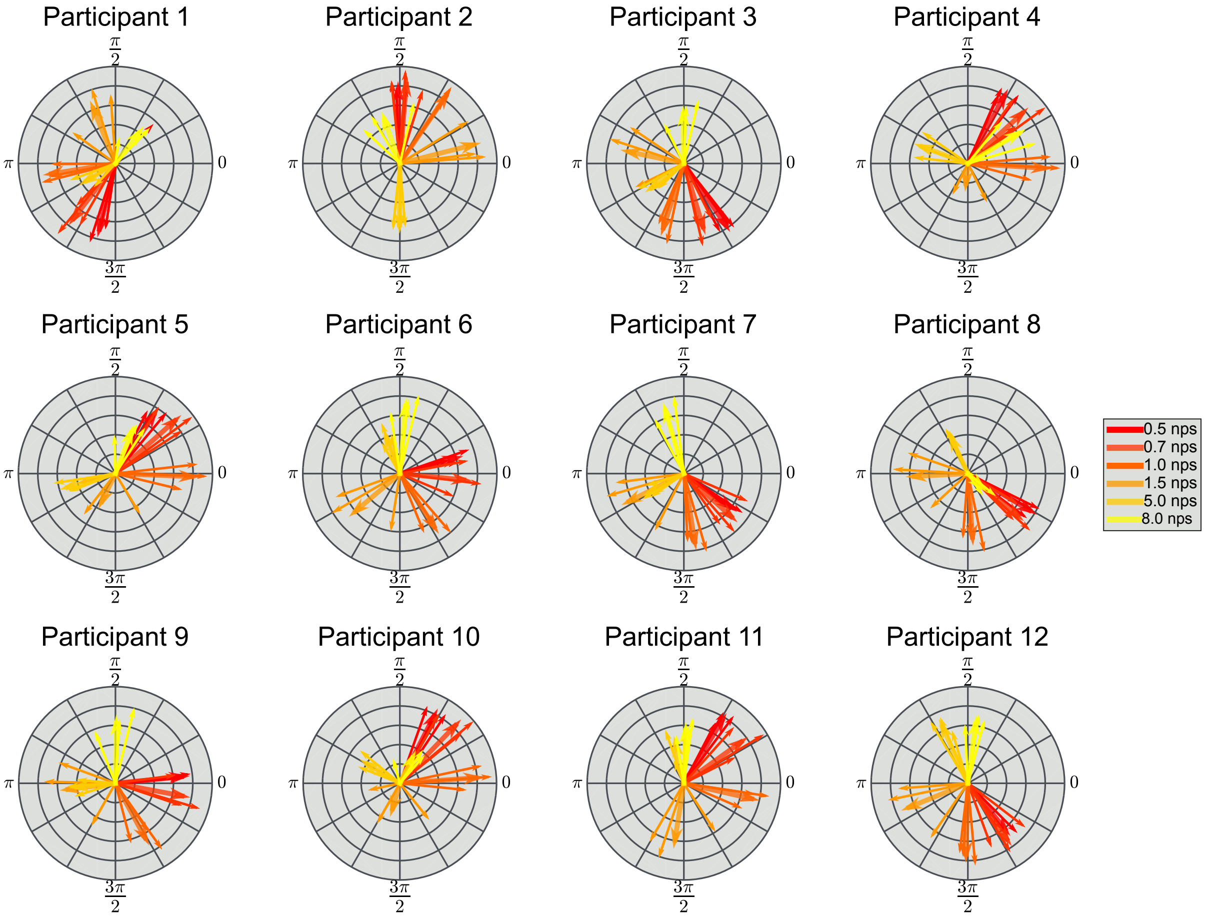

Fig. S1. *Phase patterns using single subject kernels in evoked model.* The Evoked model derives a grand average response to a single tone based on the response of individual participants to be convolved with the stimulus envelope. Here, we plot the phase patterns derived from an Evoked model which uses each participant's response as an individual kernel. The shifts in phase pattern are due to slightly varying shapes in the participant's evoked response. However, the overall pattern (increasing $\Delta\phi$) as a function of note rate persists in all cases.

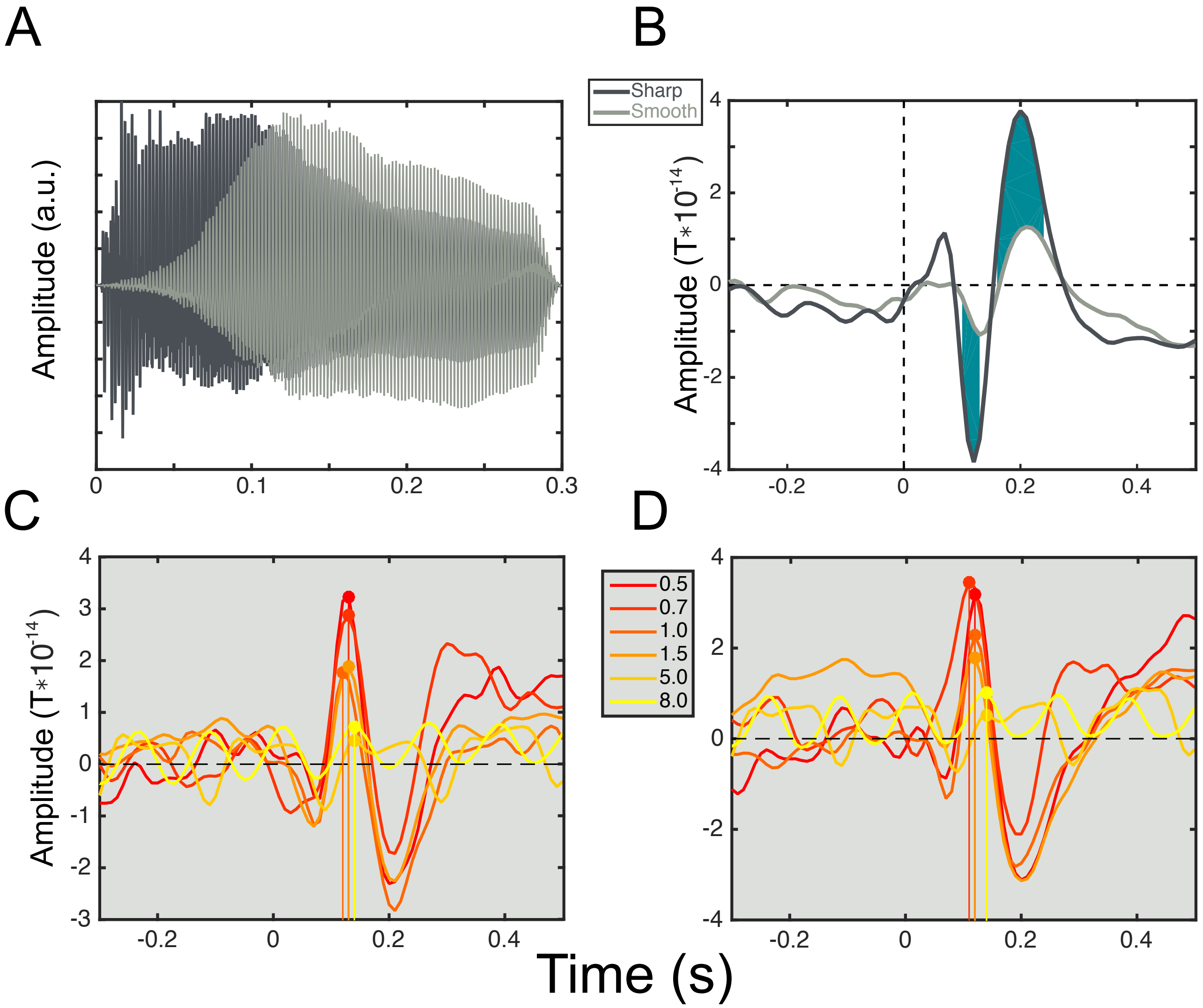

Fig. S2. *ERP response to notes. A.* Musical note A4 (440 Hz) under the sharp attack (dark gray) and smooth attack (light gray) stimulus conditions. *B.* Average response from auditory MEG channels to all sharp (dark gray) and smooth (light gray) notes from clips with note rates from 0.5 - 1.5 nps. Clusters of significant difference are highlighted in teal. *CD.* Average response from auditory MEG channels to individual notes by note rate in Left (*C*) and Right (*D*) Hemisphere. Peak time of the M100 response is highlighted for each note rate.

References

1. Hari R, Kaila K, Katila T, Tuomisto T, Varpula T (1982) Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. *Electroencephalogr Clin Neurophysiol* 54(5):561–9.
2. Michie P, et al. (2000) Duration and frequency mismatch negativity in schizophrenia. *Clinical Neurophysiology* 111(6):1054 – 1065.
3. Wang AL, Mouraux A, Liang M, Iannetti GD (2008) The enhancement of the n1 wave elicited by sensory stimuli presented at very short inter-stimulus intervals is a general feature across sensory systems. *PloS one* 3(12):e3929.
4. Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology* 25(19):2457–2465.
5. Maris E, Oostenveld R (2007) Nonparametric statistical testing of eeg- and meg-data. *J Neurosci Methods* 164(1):177–90.