An oscillator model better predicts cortical entrainment to music

Keith B. Doelling[a,1], M. Florencia Assaneo[a], Dana Bevilacqua[a], Bijan Pesaran[b], and David Poeppel[a,c]

[a]Department of Psychology, New York University, New York, NY 10003; [b]Center for Neural Science, New York University, New York, NY 10003; and [c]Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt am Main, Germany

Edited by Peter Hagoort, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and approved April 1, 2019 (received for review September 25, 2018)

A body of research demonstrates convincingly a role for synchronization of auditory cortex to rhythmic structure in sounds including speech and music. Some studies hypothesize that an oscillator in auditory cortex could underlie important temporal processes such as segmentation and prediction. An important critique of these findings raises the plausible concern that what is measured is perhaps not an oscillator but is instead a sequence of evoked responses. The two distinct mechanisms could look very similar in the case of rhythmic input, but an oscillator might better provide the computational roles mentioned above (i.e., segmentation and prediction). We advance an approach to adjudicate between the two models: analyzing the phase lag between stimulus and neural signal across different stimulation rates. We ran numerical simulations of evoked and oscillatory computational models, showing that in the evoked case,phase lag is heavily rate-dependent, while the oscillatory model displays marked phase concentration across stimulation rates. Next, we compared these model predictions with magnetoencephalography data recorded while participants listened to music of varying note rates. Our results show that the phase concentration of the experimental data is more in line with the oscillatory model than with the evoked model. This finding supports an auditory cortical signal that (*i*) contains components of both bottom-up evoked responses and internal oscillatory synchronization whose strengths are weighted by their appropriateness for particular stimulus types and (*ii*) cannot be explained by evoked responses alone.

oscillator | evoked response | MEG | music | computational models

Understanding the temporal dynamics of neural activity in the processing of rhythmic sounds is critical to uncovering a mechanistic explanation of speech and music perception. A considerable body of research has investigated such activity during processing of forward and backward speech (1–3), music (4), frequency- and amplitude-modulated noise (3, 5), and rhythmic tones (6). In each case, activity in auditory cortex synchronizes to rhythmic patterns in the stimulus. Critically, this synchrony correlates with intelligibility and comprehension of the stimulus itself, suggesting that the neural activity underpins behavioral processing (3, 7, 8). Synchronizing to acoustic rhythms may also support information transfer to other brain regions (9) and selection of information for attention (10).

These properties have been hypothesized to reflect the action of a neural oscillator in auditory cortex (e.g., ref. 11). A hypothesized oscillator would be the result of a population of neurons whose resting-state activity (with no external stimulation) fluctuates around an intrinsic natural frequency. Further, the oscillator frequency shifts to synchronize with the frequency of external stimulation only if that external frequency is within a range around the resting frequency. For a detailed discussion on such synchrony in weakly coupled oscillators see ref. 12.

The major alternative to the oscillatory entrainment hypothesis proposes that the auditory cortex shows a transient response to each acoustic input. Proponents of this model suggest that neural recordings are only rhythmic due to the rhythmic inputs they receive. In this case, the underlying mechanism is a stereotyped delayed-peak response to individual stimuli (e.g., syllables, notes, or other acoustic edges). As the stimuli occur periodically, the neural signal is also periodic. This model is similar to steady-state responses or frequency-tagging experiments which expect to find the frequency of a rhythmic input in the signal of the neural region processing it (e.g., ref. 13).

Distinguishing between evoked vs. oscillatory models and establishing which better explains the observed neural signals has been challenging using noninvasive human electrophysiological recordings such as magnetoencephalography (MEG) and EEG. This is, in part, due to the analytical tools used to identify significant frequency bands of activity which are sensitive to any type of rhythmic activity without distinguishing the generative mechanism. Further, showing resting-state oscillatory activity, a potentially distinguishing feature between the two models, has been difficult in human auditory cortex, presumably because the signal is weaker at rest. While resting intrinsic frequencies have been found more invasively in macaque auditory cortex (14), noninvasive human studies have required impressive but highly complex techniques to show the same effect (15, 16). A recent review (17) highlights the many methods used to tease apart the two models and their strengths and weaknesses.

These two competing hypotheses are conceptually distinct. The oscillator model suggests that the spectral characteristics of the neural signal are due in large part to the specific

Significance

Previous work in humans has found rhythmic cortical activity while listening to rhythmic sounds such as speech or music. Whether this activity reflects oscillatory dynamics of a neural circuit or instead evoked responses to the rhythmic stimulus has been difficult to determine. Here, we devised a metric to tease apart the two hypotheses by analyzing phase lag across many stimulation rates. We test this phase concentration metric using numerical simulations and generate quantitative predictions to compare against recorded magnetoencephalography data. Both previously recorded and new data were better predicted by a model of oscillatory dynamics than evoked responses. This work, therefore, provides definitive evidence for the presence of an oscillatory dynamic in auditory cortex during processing of rhythmic stimuli.

Author contributions: K.B.D., B.P., and D.P. designed research; K.B.D. and D.B. performed research; K.B.D. and M.F.A. contributed new reagents/analytic tools; K.B.D. and D.B. analyzed data; and K.B.D., M.F.A., B.P., and D.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence should be addressed. Email: keith.doelling@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1816414116/-/DCSupplemental.

Published online April 24, 2019.

neural circuitry. While the evoked model is capable of prioritizing certain timescales depending on the frequency content of its evoked response, the output is generally a weighted reflection of the input. The oscillator model has a number of theoretical advantages for auditory processing: (*i*) The cycles of the oscillators generate windows in time for a simple mechanism of stream segmentation (18, 19); (*ii*) the ability to adapt to a range of frequencies based on previous input can support more robust temporal prediction (6); and (*iii*) processing can be made more efficient by aligning the optimal phase of the oscillator to the moments in the acoustics with the most information (11). Given these computational differences, it is important to know which of these mechanisms is actually implemented.

There are sharp disagreements both on the role of oscillatory behavior in the brain and, most critically, on what counts as evidence for oscillatory activity in noninvasive electrophysiological recordings. See, for example, the recent discussion on enhanced neural tracking of low-frequency sounds (20–22). Here, our goal is to present a methodology to effectively resolve these disagreements.

Our method focuses on analyzing how the phase lag between neural response and acoustic periodic input varies as a function of the stimulus rate. To illustrate the logic of this analysis, we present the behavior of two toy models in Fig. 1. The evoked model (Fig. 1*A*) generates a neural response with a fixed time lag for each incoming stimulus. As the stimulus rate increases, the lag stays largely fixed, becoming an increasing portion of the stimulus cycle. According to this model, the stimulus position changes relative to the two surrounding peaks of the model response. In contrast, the oscillator model (Fig. 1*B*) generates an oscillation at the stimulus frequency. The shape of its cycle shifts so that the phase of the model output at stimulus onset remains similar across stimulus rates that exist within its range of synchrony (Fig. 1*B*).

$$\Delta\phi = \frac{2\pi(\Delta t)}{\lambda} \qquad [1]$$

The relation between time and phase is governed by Eq. **1**, where $\Delta\phi$ is the phase lag between the signals, Δt is the time lag, and λ is the cycle length. In the evoked model, Δt is fixed: As λ decreases, $\Delta\phi$ increases. In the oscillator model, Δt adjusts with the cycle length to maintain a relatively constant proportion and thus a constant phase.

The rose plots shown in Fig. 1 *C* and *D* are typical of this study. The vector angle represents the phase lag between two rhythmic signals ("east": 0 rad, fully aligned; "west": π rad, fully opposed) and the length typically represents the consistency of that phase across time (i.e., phase locking value). Phase locking is high in both cases. However, in the evoked model, the phase difference is widely spread around the cycle, whereas the oscillator model shows greater phase consistence. We term this consistency high phase concentration and estimate its value using the phase concentration metric (PCM). PCM is calculated as follows: First the length of red, yellow, and orange vectors is normalized, then the mean vector is computed. The length of this mean vector, the PCM, shown in teal (evoked) and violet (oscillator) distinguishes between the models.

The PCM analysis is mathematically quite simple and similar to intertrial phase coherence (ITPC; refs. 23 and 24). However, it is conceptually distinct. While ITPC is sensitive to similar phase patterns in neural data across repetitions of the same stimulus, the PCM compares phase patterns to different stimuli (and stimulus rates). By analyzing phase differences between the neural signal and the stimulus envelope, we are able to compare phase differences across stimulus rates, which would not be possible with a typical ITPC analysis.

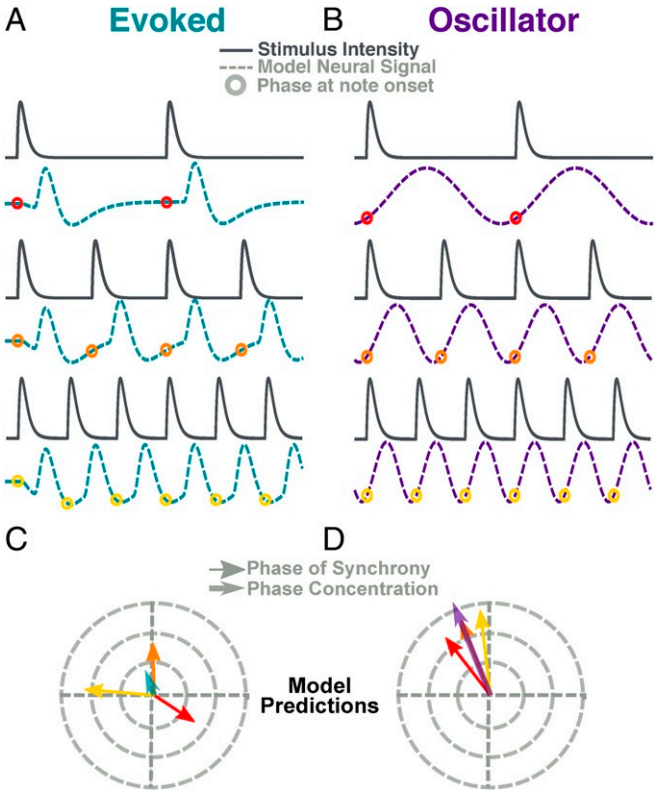

Fig. 1. Toy oscillatory and evoked models. Toy models demonstrate intuitions of PCM. (*A*) The evoked model (teal) convolves a response kernel to the stimulus envelope. As input rate increases, phase difference between stimulus and output shifts. (*B*) The oscillator model (violet) is a cosine function with a frequency that matches the stimulus note rate. Here the time lag shifts with frequency, maintaining a near-constant phase. (*C* and *D*). Phase lag calculated for each stimulus–response pair in evoked model (*C*) and oscillator (*D*). The angle of the arrow corresponds to the phase while the length corresponds to the strength of synchrony. The teal and violet arrows represent the PCM of each model.

The analysis shown in Fig. 1 uses idealized models to illustrate the key distinctions. The models are simplified for clarity. For example, the oscillator model is perfectly sinusoidal, an assumption unlikely to be replicated in the neural system (see ref. 25) and synchronized to a perfectly isochronous input, unlikely to occur in a natural environment. To further clarify the underlying mechanisms of activity observed using MEG, we need quantitative predictions of the PCM using more realistic models.

We, therefore, first performed a computational study to establish whether PCM can distinguish between more biologically plausible models that "listen" to ecologically valid stimuli: music clips of a wide range of note rates. The musical clips were drawn from piano pieces with one of six note rates ranging from 0.5 to 8 notes per second (nps). We then applied the PCM analysis to MEG responses of participants listening to the same clips and compared the results to our models' predictions. The responses are better predicted by the oscillator model than by the evoked model. However, evoked transient responses are clearly elicited and likely play a critical role. Consistent with this is that the oscillator model appears to overestimate the PCM of the MEG data. We conclude that the measured MEG responses consist of both components, an evoked, bottom-up, transient response as well as an internally generated oscillatory response, synchronizing to the input. By this reasoning, overestimation in the oscillator model prediction is due to the presence of an evoked response in the experimental data that is not in the model.

To demonstrate the coexistence of the two components and explain why the oscillator model overestimates the phase concentration, we aimed to manipulate each component independently. We designed a further study in which we manipulated the musical stimulus by smoothing the attack of each note. We hypothesized that such a manipulation will reduce the evoked component and thereby improve PCM. As predicted, modulating attack acoustics increased the phase concentration of the neural responses relative to the oscillator model. Taken together, our data demonstrate the coexistence of both evoked and oscillatory components that combine temporal prediction with bottom-up transient responses to efficiently process rhythmic or quasi-rhythmic stimuli such as speech and music.

Results

Model Analysis. Fig. 2 shows the design of the two models. The evoked model (Fig. 2*A*) convolves the stimulus envelope with an average kernel derived from the participants' MEG response to individual tones. The oscillator model (Fig. 2*B*) is an instantiation of the Wilson–Cowan model (26) with the stimulus envelope as a driving force to the excitatory population. See *Materials and Methods* for more information on model design.

Example model outputs and PCM analyses are shown in Fig. 3. PCM clearly distinguishes between the two models. Under the evoked model (Fig. 3*A*), the phase lag between model output and stimulus strongly depends on the musical note rate. As note rate increases, so does the phase difference between model output and stimulus. This is summarized in Fig. 3*A*, *Right*, where all note rates are plotted together. The results show a nearly full cycle shift of phase difference from 0.5 nps to 8 nps. For phase patterns based on single subject kernels see *SI Appendix*, Fig. S1; the pattern is remarkably consistent. To quantify this pattern, we compute the PCM. The resulting phase concentration vector (PCV) is plotted in teal. PCM = 0.17, calculated as the absolute value of the PCV. Small PCM is characteristic of a wide phase spread.

Note how highly coupled the model output is to the stimulus; each subject arrow shows a coupling value near 0.6. Such high synchrony in the evoked model clearly demonstrates how a model that does not actively synchronize can appear synchronized. In this way, our evoked model represents the alternative hypothesis well.

Fig. 3*B*, *Left* shows example outputs of the oscillator model (violet). The behavior of the oscillator changes depending on the frequency of the input. At the lowest frequency (0.5 nps), the model largely oscillates at its resting frequency (~4 Hz). As the stimulus rates gets closer to the natural resting-state frequency, the oscillator begins to synchronize more readily. The degree of synchrony changes depending on the stimulus frequency. This demonstrates the dynamic nature of such synchrony: It prioritizes certain timescales over others and as such is well matched to oscillatory entrainment theory.

Critically, the phase lag remains concentrated (Fig. 3*B*, *Right*) so that the range of synchronous phase is smaller for the oscillatory model than for the evoked model. The PCV (plotted in violet) confirms this quantitatively to show a longer vector (PCM = 0.66) than seen in the evoked model. Interestingly, for this Wilson–Cowan oscillator, the responses to 5- and 8-nps inputs are grouped in at a slightly different phase regime than the other frequencies, which we did not predict. Future work is needed to understand how this phase lag depends on the underlying oscillator mechanism. Using a permutation test across groups, we found that the difference in PCM is significant (ΔPCM = 0.49, $P = 0.017$).

Exp. 1. We next turned to investigate MEG responses to the same rhythmic stimuli. Participants listened to music clips with varying note rates and made judgments about their pitch. Previous analyses of these data used ITPC, a measure of phase consistency across trials but not directly to the stimulus, to reveal entrainment at frequencies 1 to 8 Hz (4). Here, we reanalyzed these data to more directly relate the MEG response to the stimulus. We selected the best auditory channels based on responses to single tones (Fig. 4*A*) and analyzed the direct synchrony from stimulus to brain using cerebro-acoustic coherence (CACoh; ref. 7). An example of high synchrony is shown in Fig. 4*A*, *Right*, where the acoustic signal and an average signal from auditory channels are plotted; notice the alignment of signal peak and note

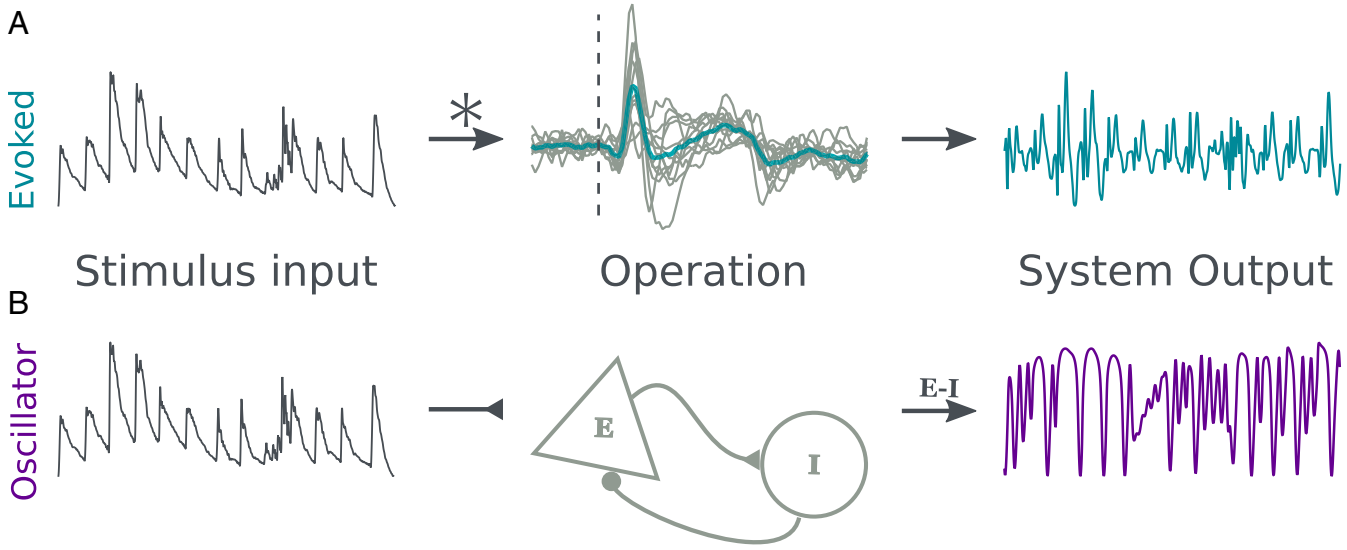

Fig. 2. Model design. The figure shows the process that generated model outputs for the evoked model (*A*) and the oscillator (*B*). (*A*) The stimulus envelope (dark gray, *Left*) is fed into the evoked model through a convolution with an evoked response kernel. A kernel was created for each participant based on their average response to a single tone (light gray lines, *Middle*). The kernel used is the average across subjects (teal line, *Middle*). (*B*) The stimulus envelope is an added drive to the excitatory population of the oscillator model. The output used for analysis is the difference between the activity of the excitatory and inhibitory populations.

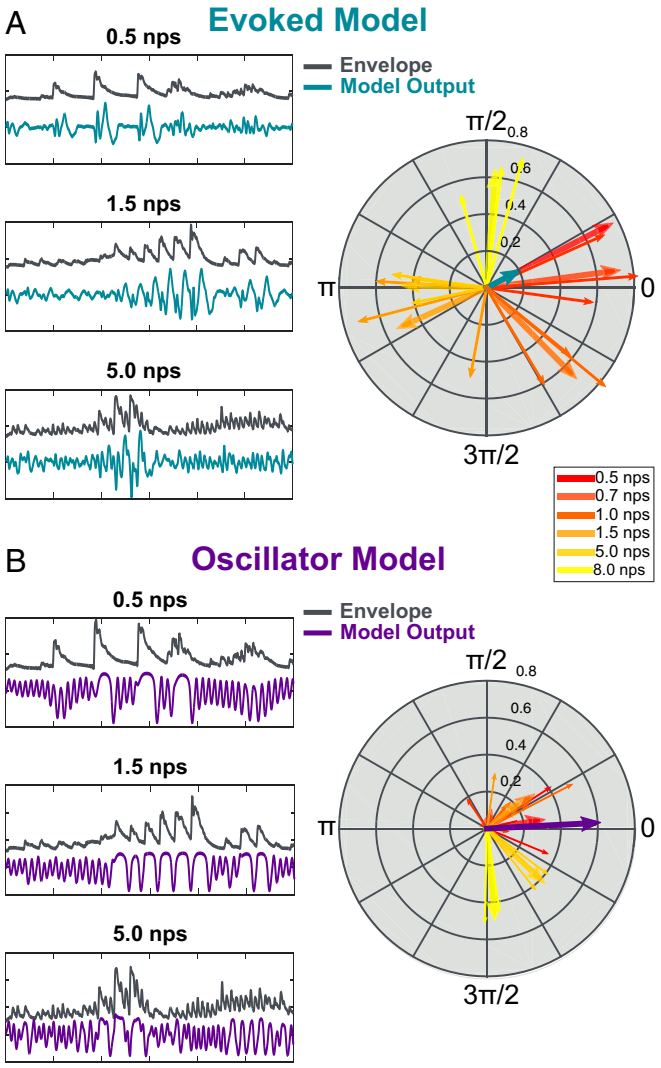

Fig. 3. Model outputs. (*A*) Phase lag in the evoked model is frequency-dependent. (*Left*) Example inputs (dark gray) and outputs (teal) of the evoked model at a 0.5, 1.5, and 5 nps. (*Right*) Averages across subjects are plotted for each clip (thin arrows) and an average across clips for each note rate (thick arrows). Average phase concentration is plotted as the teal arrow. (*B*) Phase lag in the oscillator model is highly concentrated. (*Left*) Stimulus (dark gray) and output (violet) of the oscillator model for examples at the same clips as in *A*. (*Right*) Phase lag is plotted for each clip (thin line) and for the average across clips for each note rate (thick lines). PCV is plotted in violet.

onset (shown above the figure). This analysis technique replicated the results from the previous study. Fig. 4*B* shows the CACoh values at each frequency across stimulus rates for the previously recorded data. The data for all stimulus rates was collected in two studies (study 1: 0.5, 5, and 8 nps; study 2: 0.7, 1, and 1.5 nps) and are compared separately. Only values at 1 nps and above show significant results. As such, from here on, we only consider frequencies from 1 to 8 nps where synchrony was successful.

Fig. 4*C* shows the average phase lag across subjects for each frequency using the same analysis pipeline as in the model outputs. The pattern is similar to the phase pattern of the oscillator model, specifically a narrow range for most frequencies. In contrast, the evoked model's phase is monotonically dependent on stimulation frequency and is spread to a wider phase range. In gray, we plot the PCV for the average data, which has a length more similar to the oscillatory model than the evoked model. While the shift in phase at 8 Hz is surprising based on the intuitions from our toy models, it does follow from the predictions of the Wilson–Cowan model.

While Fig. 4*C* shows the average PCV pattern per hemisphere across all subjects, and Fig. 4*D* shows the PCM for each subject. We compare the resulting values with single predictions from each model using again only the stimulus rates from 1 to 8 nps. We established CIs for the left ($CI_L = (0.30, 0.57)$) and right ($CI_R = (0.35, 0.59)$) hemisphere responses and compared these to the model predictions. The prediction of the evoked model is significantly outside the CI for the mean PCM in either hemisphere ($PCM_E = 0.245$; left: $t(14) = 3.08$, $P = 0.0082$; right: $t(14) = 4.11$, $P = 0.0011$). The prediction of the oscillator model is on the border, just outside the left hemisphere's CI and just inside the right's ($PCM_O = 0.58$; left: $t(14) = -2.32$, $P = 0.037$; right: $t(14) = -2.01$, $P = 0.065$).

To compare the models' performance, we fit Gaussian distributions to the responses with means set by the two models' predictions. We then compared the fit of the distributions using the Akaike information criteria (AIC) (*Materials and Methods*). In both hemispheres, the oscillator model showed a better fit to the data (left: $AIC_E = 8.40$, $AIC_O = 5.51$, $\Delta = 2.89$; right: $AIC_E = 8.73$, $AIC_O = 0.63$, $\Delta = 8.1$).

Exp. 2. Exp. 1 provides evidence that the oscillatory model is a better predictor of the MEG activity than the evoked. However, the results do not conclusively validate the oscillatory model since its prediction is statistically ambiguous: In both hemispheres, the oscillator model's estimate of mean PCM hovers around the threshold for statistical difference. Given that evidence for an evoked response is well documented in the literature, we hypothesize an interplay between evoked and oscillatory mechanisms is present in the data. Such an interplay could explain the oscillator model's overestimation of the MEG PCM.

We conducted a second experiment in which the stimuli were designed to test the relationship between the evoked response and the accuracy of our models' predictions—hypothesizing that the evoked response would be reduced by smoothing the attack of each note. To carefully control the attack of each note, we rebuilt artificial versions of the stimuli note by note, rendering the clips perfectly rhythmic. To avoid this potential confound, we had the participants listen to two stimulus types: sharp attack and smooth attack. Sharp attack differs from the original stimuli in that the notes are now perfectly rhythmic (rather than natural recordings), while the smooth attack differs from the original both in its perfect rhythmicity and in the smoothed attack of note onsets (*Materials and Methods*). An example of a sharp and smoothed note is shown in *SI Appendix*, Fig. S2*A* for comparison. We then acquired new MEG recordings, this time using clips from all six note rates in each participant, and we ran the same analysis as in the previous experiment for the two stimulus types.

An important first question is whether smoothing the attack reduced the evoked response as expected. We segmented the data to align trials to each individual note and compared the event-related potential (ERP) response to the smooth and sharp stimuli. The results of this analysis are shown in *SI Appendix*, Fig. S2*B*. We found a significant reduction in the amplitude of the response in two clusters from 100 to 130 ms and from 160 to 240 ms consistent with M100 and M200 responses. This suggests our note transformation did indeed reduce the evoked response as predicted.

Fig. 5*A* shows the phase pattern of left and right hemispheric responses for perfectly rhythmic stimuli with a sharp note attack. The phase patterns are consistent with results from Exp. 1. Fig. 5*B* shows the comparison of the subject data to the model

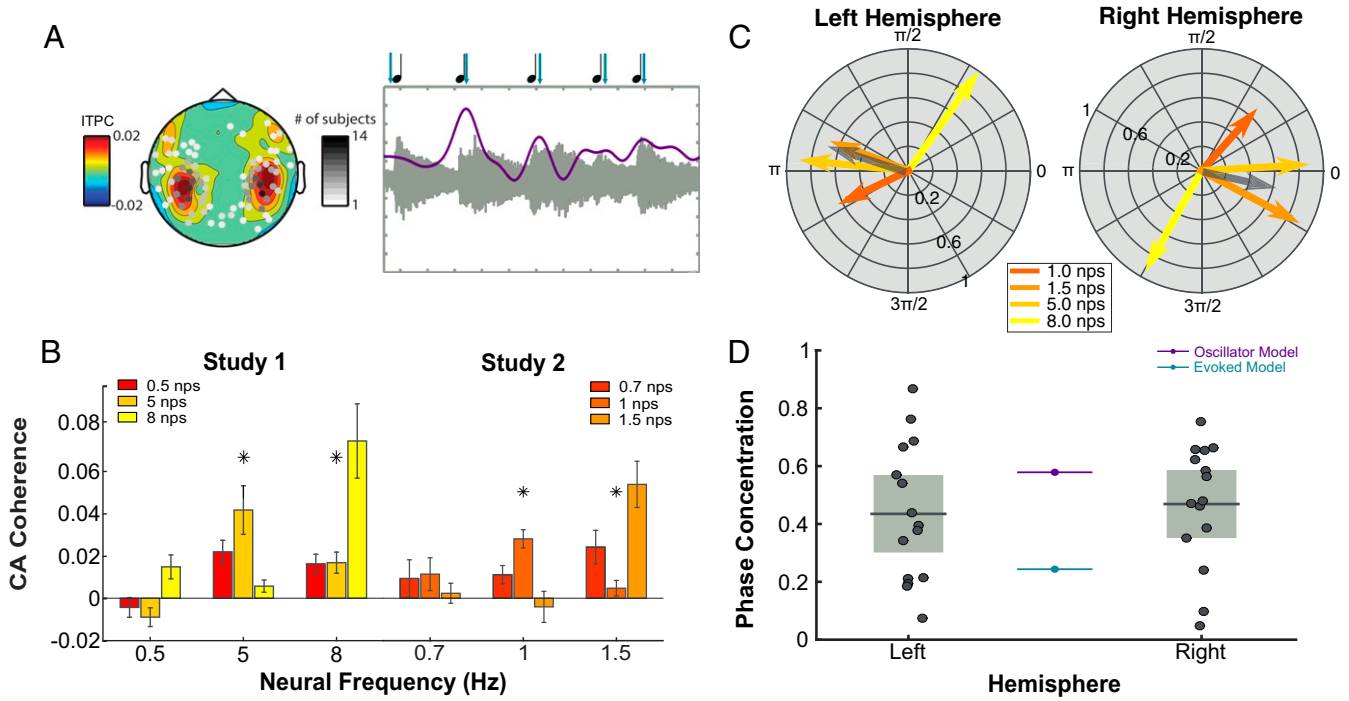

Fig. 4. Phase lag to various note rates. (*A*, *Left*) The topography of ITPC as analyzed in a previous paper (reprinted from ref. 4). Sensors shown are those selected based on response to single tones. Grayscale represents the number of participants for which that channel was picked. (*A*, *Right*) An example from single subject averaged over 20 trials of high synchrony. Note how the peak of the neural response (violet) is well aligned to the onset of acoustics (light gray). Markers for peak and onset are presented above. (*B*) CACoh analysis shows significant synchrony at frequencies from 1 to 8 Hz. Colors refer to note rate of each stimulus condition. They are analyzed separately by study. Asterisks denote significant main effects of note rate at the specific neural frequency in a one-way ANOVA. (*C*) Phase lag in left and right hemispheres on the unit circle for the four note rates that showed successful synchrony. PCV is plotted in dark gray. (*D*) Phase concentration for both hemispheres. Box plots show mean and CI of the MEG data (participant data in gray dots). The oscillator and evoked model predictions are represented by lines in violet and teal, respectively.

outputs with new stimulus input. The sharp attack stimulus is essentially a replication of Exp. 1; we find similar results. We again compared the CIs of the left ($CI_L = (0.49, 0.67)$) and right ($CI_R = (0.46, 0.69)$) hemispheres data with model predictions. While the prediction of the evoked model is significantly outside of the CIs of the MEG data ($PCM_E = .278$; left: $t(11) = 7.27$, $P = 0.00002$; right: $t(11) = 5.72$, $P = 0.00013$), the oscillator model prediction is within their upper border

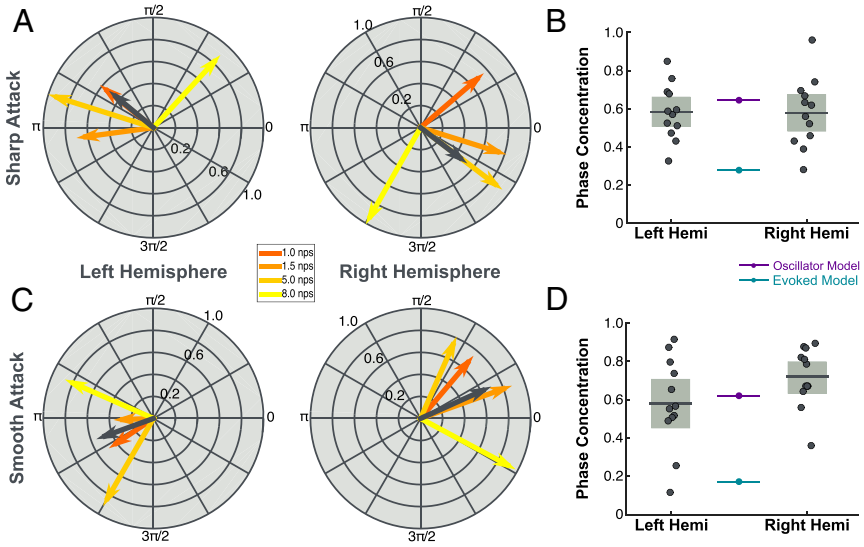

Fig. 5. Smooth attacks place MEG data in line with oscillator model. (*A*) Phase lag for the four note rates that showed significance in the previous study in response to new stimuli with sharp attack. The phase concentration angle is shown in dark gray. (*B*). PCM for models and data with the new stimulus input. Box plots (light gray) show 95% CIs and mean (dark gray line) of the sample data (gray dots). Colored lines (teal and violet) reflect predictions of the models (evoked and oscillator, respectively). (*C* and *D*) A plot of phase synchrony and phase concentration in response to new stimuli with smoothed attacks.

($PCM_O = .644$; left: $t(11) = -1.46, P = 0.17$; right: $t(11) = -1.2, P = 0.24$). In this case, the oscillator model prediction is always inside the CIs while the evoked model prediction is always outside the CIs. This result is confirmed by our direct model comparison which fit Gaussian distributions with means set either by the evoked or oscillator model. Again, the oscillator model is the better fit in both hemispheres (left: $AIC_E = 9.81, AIC_O = -9.19, \Delta = 19.00$; right: $AIC_E = 10.71, AIC_O = -4.42, \Delta = 15.12$).

Fig. 5*C* shows the phase lag to smoothed stimuli. Here the phase pattern has narrowed relative both to the sharp attack and the pattern in the previous experiment. This is made clear in Fig. 5*D*, where the oscillatory model prediction is much more in line with the MEG data. We compared the CIs of the PCM for left ($CI_L = (0.43, 0.73)$) and right ($CI_R = (0.62, 0.82)$) hemispheres with model predictions in the smooth condition. The prediction of the evoked model is again significantly lower than the CIs of the data ($PCM_E = .172$; left: $t(11) = 6.03, P = 0.00009$; right: $t(11) = 12.17, P < 0.00001$). The prediction of the oscillator model ($PCM_O = 0.619$), however, is well within the CIs of the left hemisphere ($t(11) = -0.54, P = 0.60$) and even slightly underestimates the right hemisphere data ($t(11) = 2.23, P = 0.047$). This suggests that the smooth attack has brought the MEG data more in line with expected results of an oscillatory mechanism and, in the case of the right hemisphere, even more oscillatory than the parameters set for our oscillator model. Further, a comparison of the right hemisphere across conditions shows a significant increase in PCM in the smooth condition compared with sharp ($t(11) = 2.26, P = 0.045$). No such effect was found in the left hemisphere ($t(11) = -.005, P = 0.996$).

Our direct model comparison—fitting Gaussian distributions with means set by the model predictions—confirms these results. Again, for the smooth condition in both hemispheres the oscillator model is the better fit (left: $AIC_E = 17.87, AIC_O = 0.67, \Delta = 17.20$; right: $AIC_E = 22.47, AIC_O = -5.11, \Delta = 27.59$). The summary results for AIC model comparison across all experiments are shown in Table 1.

Discussion

We present the PCM as a method for direct analysis of the oscillatory nature of cortical activity during processing of naturalistic stimuli. We validated the metric on two models: an oscillatory, Wilson–Cowan model and an evoked convolution-based model. The inputs for both models were musical stimuli of varying rates. PCM clearly distinguished the two models. We then used PCM to analyze previously recorded data of participants listening to the same clips. We found that the mean PCM of participants matched that of the oscillatory model better than that of the evoked model. The results show clear evidence for an oscillatory mechanism in auditory cortex, likely coordinated with a bottom-up evoked response. We then collected new data on new participants listening to altered versions of the music clips with sharp and smoothed attacks to reduce the evoked response. Our results represent a three-time replication showing across different participants and different stimulus types that the oscillator model is a better predictor of the PCM of MEG data than a purely evoked model. We conclude that the MEG signal contains both oscillatory and evoked responses, with their relative weights determined, in part, by the sharpness of the note onsets.

Table 1. Summary of AIC results for evoked and oscillator models across all experiments

Model	df	AIC	ΔAIC	w_i	ER_i	$\log_{10} ER_i$
Exp. 1, left						
Evoked	1	8.40	2.89	0.191	4.24	0.63
Oscillator*	1	5.51	0	0.81	1	0
Exp. 1, right						
Evoked	1	8.73	8.10	0.017	57.4	1.76
Oscillator*	1	0.63	0	0.983	1	0
Exp. 2, sharp, left						
Evoked	1	9.81	19.00	1.00×10^{-4}	$1.34 \times 10^{+4}$	4.13
Oscillator*	1	−9.19	0	0.9999	1	0
Exp. 2, sharp, right						
Evoked	1	10.71	15.12	5.00×10^{-4}	$1.93 \times 10^{+3}$	3.29
Oscillator*	1	−4.42	0	0.9995	1	0
Exp. 2, smooth, left						
Evoked	1	17.87	17.2	2.00×10^{-4}	$5.43 \times 10^{+3}$	3.73
Oscillator*	1	0.67	0	0.9998	1	0
Exp. 2, smooth, right						
Evoked	1	22.47	27.59	0	$9.75 \times 10^{+5}$	5.99
Oscillator*	1	−5.11	0	1	1	0

See *Materials and Methods* for a description of the statistics reported.
*Denotes the model that best fits the PCM of MEG data as defined by AIC.

Models Demonstrate a Clear Prediction. To tease apart the competing hypotheses (evoked vs. oscillator), we contrasted the phase lag of two computational models as a function of stimulation rate. We hypothesized that the phase lag of the oscillator would remain stable across rates. However, the evoked model, based on real M100 recordings, would not constrain the phase lag. These predictions are borne out in the model outputs (Fig. 3).

The evoked model was designed to react to an input with a stereotyped response. To identify the shape of that response, we recorded our participants in the MEG as they listened to tones. We then used the average response across participants to a tone as a kernel that was convolved with the stimulus envelope to generate the model output. The kernel, while fixed, does have its own timescale, which depends on its length and shape. It, therefore, can have the ability to prioritize certain timescales over others and as such could generate meaningful temporal predictions in limited circumstances. A common critique of the work studying oscillatory mechanisms in perception is that a model such as this—with no oscillatory mechanism—can generate an oscillatory output. Our model demonstrates this very well. The model yields high synchrony values (measured by the length of the vectors in Fig. 3*A*) to all of the presented clips. Still, while the data are rhythmic, the underlying mechanism has no oscillatory properties, and by investigating the relative phase alignment across stimulation rates we can discern that there is no active synchronization to a specific phase.

In contrast, the oscillator model does constrain phase lag to a narrower range. As the oscillatory behavior arises from balanced activation of its excitatory and inhibitory populations, the temporal constant of their interaction imposes its own timescale onto the stimulus input and drives the alignment between input and output. The exact phase of alignment is not something we expect to match the recorded neural data, particularly from the perspective of MEG, where the signal of interest is altered in each participant by changes in source orientation due to neural anatomy and head position. Instead, we emphasize the relative phase across stimulation rates.

The oscillatory model is more selective in terms of the rates it will synchronize. While the evoked model showed high synchrony to all stimulus rates, the oscillatory model prioritizes narrower timescales. This specific model, implemented using the Wilson–Cowan model, is not meant to fit the features and characteristics of our recorded MEG data. Still, the model demonstrates the kinds of features we should expect to see from an oscillatory mechanism, specifically a constrained phase regime across stimulation rates.

The evoked model is implemented here using convolution between the stimulus envelope and a response kernel. By the convolution theorem, the phase transfer function in this linear system is entirely dependent on the Fourier transform of the kernel that is convolved with the input. There do exist kernels, therefore, that maintain a constant phase lag within some frequency range, which could be confused with an oscillator model by PCM. Some of these kernels may even be biologically plausible in various neural contexts. Our results depend critically on the empirical data that allowed us to characterize the kernel shape in auditory cortex and its spectral content by averaging the response to individual tones. Further use of the PCM as a distinguishing feature between evoked and oscillator models in other domains will similarly depend on the spectral content of the kernel in the probed region.

Our evoked model is in part based on the assumption that the evoked response should not change much depending on the stimulus rate. However, some classic studies (27–29) have shown that the amplitude of the evoked response changes with stimulus rates, decreasing as rates reach 2 Hz and then increasing as they get faster. Of critical interest to us is the peak latency, as we expected the Δt between stimulus input and output to heavily affect the $\Delta\phi$. In Exp. 2, we were able to address this question directly by looking at how peak latency was affected by note rate. If the evoked model's underestimation of PCM were due to a change in peak latency across stimulus rates, we would expect that the latency of the peak response should decrease with increasing note rate. *SI Appendix*, Fig. S2 *C* and *D* show this not to be true. While the peak amplitude of the response decreases with increasing note rate, the peak lag of the M100 increases. Given the logic of Fig. 1 and Eq. **1**, we would expect this effect—a longer lag for smaller cycle lengths—to decrease the PCM of the MEG data further than predicted by the evoked model. Therefore, that the true PCM is significantly higher than the evoked model prediction further refutes the model's viability.

While there may be other added features that one may wish to apply to the evoked model and improve its performance, we have shown that the purest form of an evoked model is not sufficient to explain auditory cortical synchrony, as is often claimed (21). Indeed, the addition of complexity would have the effect of rendering the evoked and oscillatory models more and more indistinguishable. If a more complex evoked model naturally fluctuates at a certain timescale, adapts its shape to match a range of stimulation rates, and rebounds to align with stimulus onset, then it may also be described as an oscillation that synchronizes and aligns its phase to the note. To us, either description would be acceptable.

Weighing Evoked and Oscillatory Components. The models allowed us to generate quantitative predictions of the phase concentration we should expect to see in the MEG data. By comparing the measured results with our models, we were able to get a sense of how oscillatory the underlying neural mechanism may be. In the initial experiment, the oscillatory model was clearly a better predictor of the PCM. Still, the oscillatory model overestimated the phase concentration numerically, particularly in the left hemisphere. This may reflect the interplay between oscillatory and evoked components within the auditory cortex. That a new and unpredicted input generates an evoked response is uncontroversial, and the basis for a large field of research in ERP studies (30). What our data may point to is a system in which these bottom-up input responses are fed into neural circuitry that attempts to predict the timing of new inputs through oscillatory dynamics. Thus, both evoked and oscillatory components exist and are more or less weighted depending on the predictability of the stimulus.

If this hypothesis of cortical activity is reasonable, then we should be able to increase the PCM of the MEG data by reducing the evoked response to each note of the stimulus. We tested this hypothesis by designing the stimulus to have a smoother attack for each note (*SI Appendix*, Fig. S2*A*). We expected this to result in a lower magnitude of the evoked response. This should in turn increase the PCM. Our predictions were confirmed in the right hemisphere. We compared stimuli with a sharp attack to those with a smooth attack. The sharp stimuli generated a phase concentration similar to what we saw in Exp. 1, with the oscillatory model as a better predictor but slightly overestimating the MEG data. However, the smooth stimuli elicited a higher PCM in the right hemisphere that was well underestimated by the oscillatory model. This right hemisphere effect fits well with the asymmetric sampling in time hypothesis (31), which predicts that the right hemisphere is biased toward oscillatory temporal processing in the theta (1 to 8 Hz) range. To date, many studies have demonstrated stronger low-frequency oscillatory behavior in the right hemisphere compared with the left (2, 32, 33). Our study provides further evidence for this by suggesting that the processing at this temporal scale is more oscillatory in the right hemisphere.

This study focused on testing an oscillatory mechanism in auditory cortical regions and thus used smooth stimuli to reduce the evoked component and boost the oscillatory. The reverse should also be possible. For example, by presenting participants with extremely sharp and highly unpredictable notes, the oscillatory component may be disrupted or reduced and less useful. In this case, a bottom-up evoked response may be the only useful processing method and the PCM should theoretically decrease. In practice, however, the PCM may not be so useful in this case, as the phase loses its meaning in the context of an arrhythmic stimulus.

The analysis of model data has illustrated some interesting points that are crucial to understanding the oscillatory entrainment theory. First, the evoked model actually showed higher overall synchrony and to a wider range of stimulation frequencies than did the oscillator model. This may at first seem counterintuitive. However, the evoked model, in essence, mimics the input with some delay. This will naturally give rise to high synchrony values. However, because the model is always reactive, it has limited capacity to predict note onset and provides few theoretical benefits in this regard. The oscillator model, however, imposes its own temporal structure on to the incoming stimulus. By synchronizing in this way, it is able to predict note onsets, often rising just before new inputs (as the phase patterns in Fig. 3*B* show). The oscillator also prioritizes certain timescales near its natural frequency over others. This is useful only if relevant information exists at this timescale. In line with this, both syllable and note rates have been shown to consistently fall in this specific timescale (34–37). Therefore, while the overall synchrony values are lower in the oscillator model, the dynamic nature of the model allows for the many benefits described in previous research on auditory entrainment (7, 11, 38).

Conclusion. Taken together, this model comparison and the human MEG data argue in favor of an oscillatory model of auditory entrainment. That is to say, the auditory cortex actively synchronizes a low-frequency oscillator with the rhythms present in sound from 1 to 8 Hz. To our knowledge, this represents the clearest evidence to date of an oscillator mechanism in humans (see ref. 39 for foundational work in macaque monkeys) for processing auditory inputs with temporal regularity, in this case music. We propose that the method be extended to other cognitive domains, including speech, visual, and somatosensory perception. In so doing, we hope to take a critical step forward in understanding the role of oscillatory activity in the brain.

Materials and Methods

Model Simulation.

Evoked model. The evoked model is designed to simulate a system that has a clear impulse response and responds in the same way to all stimuli. It is designed using a separate dataset in which participants listened to a single tone for 200 trials. These trials are averaged and the 20 channels with the largest response (10 on the left hemisphere and 10 on the right) are picked. The average trace of these channels (adjusted to have the same sign at the peak response) is used as a response kernel for each subject (gray traces in Fig. 2*A*, *Middle*), which is then averaged across subjects to generate the average response kernel (teal trace). The stimulus envelope (dark gray in Fig. 2*A*, *Left*) is convolved with this average response kernel to generate model outputs (light gray in Fig. 2*A*, *Right*).

Oscillator model. The oscillator model is based on a model of excitatory and inhibitory neural populations first designed by Wilson and Cowan (26). Our design is inspired by a recent paper (40) that modeled inputs from auditory regions into motor cortex to explain selective coupling between motor and auditory regions at specific rates. We have used a similar to design to model coupling between auditory regions and the auditory stimulus. Here, we model auditory cortex as an interaction between inhibitory and excitatory populations, where the excitatory one receives the stimulus envelope as input. Fig. 2*B* shows a diagram of the model overall. The signal generated for further analysis is the difference between the excitatory and inhibitory populations. The dynamics of these populations are governed by Eqs. **2** and **3**:

$$\tau \frac{dE}{dt} = -E + S(\rho_E + cE - aI + \kappa A(t)) \quad [2]$$

$$\tau \frac{dI}{dt} = -I + S(\rho_I + bE - dI), \quad [3]$$

where $S(z) = \frac{1}{1+e^{-z}}$ is a sigmoid function whose argument represents the input activity of each neural population, E and I represent the activity of the excitatory and inhibitory populations, respectively, and τ represents the membrane time constant. a and b represent synaptic coefficients, and c and d represent feedback connections. ρ represents a constant base input from other brain regions. $A(t)$ represents the acoustic input—the stimulus envelope—and κ represents the coupling value. a, b, c, d, ρ_E, and ρ_I were fixed based on established literature (41), such that $a = b = c = 10$, $d = -2$, $\rho_E = 2.3$, and $\rho_I = -3.2$. Previous work (ref. 41, p. 46) has shown that these parameters are consistent with an Andronov–Hopf bifurcation which is characteristic of the onset of spontaneous periodic activity. The remaining parameters were set such that $\tau = 66$ ms and $\kappa = 1.5$. This is consistent with a resting state spontaneous frequency of ∼4 Hz.

Experimental Design.

Participants. Data from Exp. 1 were previously collected and analyzed from studies 1 and 2 of a previous paper (4). In this previous experiment, 27 participants were collected undergoing MEG recordings across study 1 (listening to note rates of 0.5, 5, and 8 nps) and study 2 (listening to note rates of 0.7, 1, and 1.5 nps) after providing informed consent. Further details about these participants can be found in ref. 4. In Exp. 2 of this paper, we collected new data with 12 new participants (8 female; median age 28 y; age range 22 to 51 y; average musical experience 1.53 y) undergoing MEG recording after providing informed consent. Participants received either payment of $40 or course credit for their time. Participants reported normal hearing and no neurological deficits. Both experiments were approved by the New York University Internal Review Board and the University Committee on Activities Involving Human Subjects.

Stimuli. The stimuli in Exp. 1 were described in our previous study (4). They consist of three clips from six piano pieces played by Murray Perahia. The pieces were chosen for their nps rate and were meant to span the range of possible speeds of music. The rates chosen were at 0.5, 0.7, 1, 1.5, 5, and 8 nps. The clip length ranged from 11 to 17 s.

Stimuli for Exp. 2 were designed from the same clips of music from Exp. 1 with two important differences: (*i*) the timing of each note is precisely defined by the written music and (*ii*) the onset (or attack) of the note could be smoothed so that we could determine the effect of sharp onsets. To accomplish this, we copied the written form of each music clip into the music notation software Sibelius 7 (https://www.avid.com/sibelius), set the metronome mark of each to correspond exactly with the note rates specified in our previous experiment, and exported the written music form into MIDI files. MIDI files store musical playback information and contain a list of note identities and their amplitude, duration, and start times. We then were able to read these files into MATLAB using the matlab-midi toolbox (https://github.com/kts/matlab-midi) to extract note identities and start times and recreate the clips note by note. For the notes, we used a database of individual piano notes at the University of Iowa Electronic Music Studios (theremin.music.uiowa.edu/MISpiano.html). Each note in the database lasts as long as the piano will ring out, ∼5 to 30 s depending on note frequency. We shorten the duration of each note by adding a cosine off-ramp to match the duration specified in the MIDI file. *SI Appendix*, Fig. S2*A* shows an example note for both sharp and smooth conditions shortened to 300-ms duration.

By designing the stimuli in this way, we are able to directly control the strength of each attack. We developed these stimuli in two conditions: with sharp attack and with smooth attack. In the sharp attack condition, we used the note database as described above. In the case of the smooth attack, before placing each note we multiply the note by a sigmoid function moving from 0 to 1 in this first 150 ms. This effectively softened or smoothed the attack of the note. After the full music piece is created, we normalize each clip to have the same overall amplitude as its hard attack counterpart.

Task. Participants in Exp. 2 performed a modified version of the task from the previous experiment (4). Each participant listened to 18 repetitions of each clip, one clip from each of the six note rates in both sharp and smooth attack conditions for a total of 6 note rates * 2 attacks * 18 repetitions = 216 trials. In 3 of the 18 repetitions, there was a short pitch distortion which the participants were asked to detect to keep them focused on the stimuli. The distortion was randomly placed using a uniform distribution from 1 s after onset to 1 s before offset. After the trial, participants were asked to identify whether (*i*) a pitch distortion shifted the music down, (*ii*) there was no pitch distortion, or (*iii*) the pitch shifted the music up. Their accuracy was not analyzed. Clips with pitch distortion were included in the analysis shown here.

Analysis and Data.

MEG recording. Neuromagnetic signals were measured using a 157-channel whole-head axial gradiometer system (Kanazawa Institute of Technology). The MEG data were acquired with a sampling rate of 1,000 Hz and filtered online with a low-pass filter of 200 Hz, with a notch filter at 60 Hz. The data were high-pass-filtered after acquisition at 0.1 Hz using a sixth-order Butterworth filter.

Channel selection. As we focused on auditory cortical responses, we used a functional auditory localizer to select channels for each subject. Channels were selected for further analysis on the basis of the magnitude of their recorded M100 evoked response elicited by a 400-ms, 1,000-Hz sinusoidal tone recorded in a pretest and averaged over 200 trials. In each hemisphere, the 10 channels with largest M100 response were selected for analysis. This method of channel selection allowed us to select channels recording signals generated in auditory cortex and surrounding areas while avoiding "double-dipping" bias.

Phase analysis. After generating the model outputs (and MEG data) we run both the output and the stimulus envelope through a Gaussian filter in the frequency domain with peak at the relevant frequency and standard deviation at half that frequency (e.g., for 8 nps condition $\mu = 8$ Hz, $\sigma = 4$ Hz). We then run the filtered signal through a Hilbert analysis to extract the instantaneous phase of both output and input. We calculate the phase difference between the two signals at each time point and convert to complex format. We then average this value across time, trials yielding a complex value for each clip. We confirm that the absolute value (equivalent to the phase locking value) is significantly greater for each frequency at the preferred note rate compared with others using permutation testing. Then, we average the clip values across to yield an average value for each note rate.

CACoh. To analyze the phase meaningfully, we must first confirm that the brain successfully synchronized to the music. To do so we used CACoh, which measures the coherence between neural signal and stimulus response normalized by the power of each signal. Eq. **4** shows how the value is calculated:

$$CA_f = \frac{\left|\sum_t \left(e^{i\theta_t}\sqrt{P_{a,t}\cdot P_{c,t}}\right)\right|}{\sum_t \left(\sqrt{P_{a,t}\cdot P_{c,t}}\right)}, \quad [4]$$

where θ, P_a, and P_c are the phase difference between neural signal and stimulus envelope, the power of the acoustic signal, and the power of the neural signal, respectively, at each time point and frequency. Phase angle difference is calculated as the angle of the cross-spectral density between the two signals.

We then compare the CACoh to a randomized CACoh in which the neural signal and acoustic clip were not matched as a control.

Statistics are analyzed comparing neural frequency in the corresponding note rate (e.g., 5 Hz CACoh in 5-nps stimulus) to an average of the other note rates (e.g., 5-Hz CACoh at all other stimulus rates).

Model comparison. In each experiment, the two models each generate a single prediction for mean PCM across frequencies. To assess the accuracy of predictions relative the data we use two methods: (*i*) CIs and (*ii*) Gaussian fitting. First, we use the Student's *t* distribution to identify the 95% CIs of the PCM across our subjects. We then assess which of the predictions exist inside these CIs for both left and right hemispheres, establishing significance using a *t* test. Next, we assess the likelihood of each model prediction given the PCM data for each subject. We first do a maximum likelihood fit of the SD for each model with the mean set to the prediction of each model. We then use AIC (42) to compare the model fit performance for the evoked and oscillator prediction. This affords the opportunity to compare the odds of each model using the evidence ratio (43). Included with AIC we report the following statistics for model comparison:

$$\begin{aligned}
&\Delta AIC_i = AIC_i - AIC_{min}\\
&Akaike\ Weight: w_i = \frac{\exp(-\frac{1}{2}\Delta AIC_i)}{\sum_{m=1}^{M}\exp(-\frac{1}{2}\Delta AIC_i)}\\
&Evidence\ Ratio: ER_i = \frac{w_{best}}{w_i}\\
&Log_{10}\ Evidence\ Ratio: LER_i = \log_{10} ER_i
\end{aligned} \quad [5]$$

w_i represents the weight that should be given to each model. The values should sum to 1 across models and the difference in weights between models can be used as a metric of certainty of model selection. ER_i represents the strength of evidence for the best model over the current model and LER_i is the $\log_{10}$ of the ER_i

ACKNOWLEDGMENTS. We thank Jeff Walker for his technical support in the collection of MEG data and Jon Winawer, Michael Landy, and Jonathan Simon for their comments and advice. This work was supported by National Institutes of Health Grant 2R01DC05660 (to D.P.), National Science Foundation Graduate Research Fellowship Grant DGE1342536 (to K.B.D.), and Army Research Office Grant W911NF-16-1-0388 (to B.P.).

1. Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophys* 104:2500–2511.
2. Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010.
3. Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex* 23:1378–1387.
4. Doelling KB, Poeppel D (2015) Cortical entrainment to music and its modulation by expertise. *Proc Natl Acad Sci USA* 112:E6233–E6242.
5. Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci USA* 109:20095–20100.
6. Arnal LH, Doelling KB, Poeppel D (2015) Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb Cortex* 25:3077–3085.
7. Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85:761–768.
8. Zoefel B, van Rullen R (2015) The role of high-level processes for oscillatory phase entrainment to speech sound. *Front Hum Neurosci* 9:651.
9. Park H, Ince RA, Schyns PG, Thut G, Gross J (2015) Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr Biol* 25:1649–1653.
10. Zion Golumbic EM, et al. (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77:980–991.
11. Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat Neurosci* 15:511–517.
12. Pikovsky A, Rosenblum M, Kurths J (2003) *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge Univ Press, Cambridge, UK), Vol 12.
13. 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:e14543.
14. Lakatos P, et al. (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophys* 94:1904–1911.
15. Keitel A, Gross J (2016) Individual human brain areas can be identified from their characteristic spectral activation fingerprints. *PLoS Biol* 14:e1002498.
16. Giraud A-L, et al. (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56:1127–1134.
17. Zoefel B, Oever S, Sack AT (2018) The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. *Front Neurosci* 12:95.
18. Teng X, Tian X, Doelling K, Poeppel D (2017) Theta band oscillations reflect more than entrainment: Behavioral and neural evidence demonstrates an active chunking process. *Eur J Neurosci* 48:2770–2782.
19. Teng X, Tian X, Rowland J, Poeppel D (2017) Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLOS Biol* 15:e2000812.
20. Lenc T, Keller PE, Varlet M, Nozaradan S (2018) Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proc Natl Acad Sci USA* 115:8221–8226.
21. Novembre G, Domenico Iannetti G (2018) Tagging the musical beat: Neural entrainment or event-related potentials? *Proc Natl Acad Sci USA* 115:E11002–E11003.
22. Lenc T, Keller PE, Varlet M, Nozaradan S (2018) Reply to Novembre and Iannetti: Conceptual and methodological issues. *Proc Natl Acad Sci USA* 115:E11004–E11004.
23. Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus specificity of phase-locked and non-phase-locked 40 hz visual responses in human. *J Neurosci* 16:4240–4249.
24. Jervis B, Nichols M, Johnson E, Allen E, Hudson NR (1983) A fundamental investigation of the composition of auditory evoked potentials. *IEEE Trans Biomed Eng* 1:43–50.
25. ScottCole R, Voytek B (2017) Brain oscillations and the importance of waveform shape. *Trends Cogn Sci* 21:137–149.
26. Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* 12:1–24.
27. 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 Neurophys* 54:561–569.
28. Michie PT, et al. (2000) Duration and frequency mismatch negativity in schizophrenia. *Clin Neurophys* 111:1054–1065.
29. Li Wang A, Mouraux A, Liang M, Domenico Iannetti G (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:e3929.
30. Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (mmn) in basic research of central auditory processing: A review. *Clin Neurophys* 118:2544–2590.
31. Poeppel D (2003) The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun* 41:245–255.
32. Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
33. Abrams DA, Nicol T, Zecker S, Kraus N (2008) Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci* 28:3958–3965.
34. Ding N, et al. (2017) Temporal modulations in speech and music. *Neurosci Biobehav Rev* 81:181–187.
35. Greenberg S, Arai T (2004) What are the essential cues for understanding spoken language? *Ieice Trans Inform Syst* E87d:1059–1070.
36. Varnet L, Clemencia Ortiz-Barajas M, Guevara Erra R, Gervain J, Lorenzi C (Oct 2017) A cross-linguistic study of speech modulation spectra. *J Acoust Soc Am* 142:1976–1989.
37. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5:e1000436.
38. Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Front Psychol* 3:238.
39. Lakatos P, et al. (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77:750–761.
40. Assaneo MF, Poeppel D (2018) The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Sci Adv* 4:eaao3842.
41. Hoppensteadt FC, Izhikevich EM (2012) *Weakly Connected Neural Networks* (Springer, New York), Vol 126.
42. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723.
43. Burnham KP, Anderson DR (2003) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York).