Supporting Information

Henry et al. 10.1073/pnas.1408741111

SI Data

Further Analysis of Behavioral Performance. To test whether the nature of the observed behavioral modulation was consistent across participants, and therefore potentially attributable to differences in gap detectability on the basis of stimulus acoustics [e.g., whether gaps were potentially simply more detectable in the peak versus trough of the stimulus frequency modulation (FM)], we fit single-cycle cosine functions to the binned hit rate data as a function of FM stimulus phase, separately for the rising and falling amplitude modulation (AM) stimulus phases (Fig. S1*A*). From best-fitting cosine functions, we estimated three dependent measures (Fig. S1*B*): mean performance, performance range, and optimal FM phase.

First, mean performance was taken as the intercept of the fitted cosine function and reflects the overall performance level, ignoring modulation by FM stimulus phase. Second, performance range corresponded to the peak-to-trough distance of the fitted cosine function and is related to the depth of behavioral modulation by FM stimulus phase. Third, optimal FM stimulus phase was taken as the FM stimulus phase that corresponded to the peak of the fitted cosine function; that is, the peak predicted performance. Performance range and mean performance values were compared between the rising and falling AM phases, using separate paired-sample t tests. Optimal stimulus phases were tested for nonuniformity, using separate Rayleigh tests.

We did not observe significant differences between the rising and falling AM stimulus phases for either mean performance or performance range [mean performance: $t(16) = 0.90$, $P = 0.38$, $r = 0.22$; performance range: $t(16) = 1.22$, $P = 0.24$, $r = 0.29$]. With respect to optimal FM stimulus phase, separate Rayleigh tests were significant only for the falling AM stimulus phase ($z = 4.26$, $P = 0.01$, $r = 0.50$), not for the rising AM stimulus phase ($z = 0.64$, $P = 0.53$, $r = 0.19$).

Gap-Evoked Responses. We also analyzed the relationship of gap-evoked responses (ERPs) to behavioral performance and to prestimulus phase effects. First, we simply compared ERPs resulting from detected gaps (hits) versus undetected gaps (misses; Fig. S2*A*). ERPs for hits versus misses were compared using a paired-samples t test and a cluster-based correction for multiple comparisons, as implemented in the Fieldtrip framework (1). We observed higher ERP magnitudes for hits relative to misses, and in particular in the time windows of the first negative (N1; 0.13–0.28 s) and subsequent positive (P2; 0.43–0.64 s) deflections after gap onset. In our subsequent analyses, we focus on the N1 because the later positive deflection overlapped with the time window during which responses were made. The central topography of the N1 is consistent with auditory cortex generators.

Next, we conducted a more fine-grained analysis of the relation between gap detection and N1 magnitude. Single trials were binned on the basis of N1 magnitudes (into 10 percentile bins with 15% width), and hit rates were calculated for each percentile bin. For each participant, we fit a linear function to hit rates as a function of N1-magnitude percentile and tested the resulting slopes against zero, using a single-sample t test [$t(16) = -2.75$, $P = 0.01$, $r = 0.57$]. Higher hit rates were associated with larger N1 magnitudes (Fig. S2*B*).

Finally, we examined the effects of pre-gap phase on gap-evoked responses. The analysis pipeline was the same as for the analysis of pre-gap phase effects on hit rates. In brief, pre-gap phase was estimated from the Fourier output of a wavelet convolution applied to 4-s epochs centered on gap onset. Epochs were not high-pass filtered but were detrended before removing the ERP from the postgap time window by multiplication with half a Hann window (Fig. S2*A*). ERPs were separately estimated from the same epochs; for analysis of ERPs, the poststimulus time-window was left intact and the full epoch was bandpass-filtered between 1 and 15 Hz. N1 magnitudes were estimated for single trials from electrode Cz by averaging over the time window ranging between 0.18 and 0.22 s. N1 magnitudes were then binned simultaneously on the basis of pre-gap 3.1-Hz and 5.075-Hz neural phase and were plotted on a torus (Fig. S2*C*). N1 magnitudes were largest when the gap occurred in the rising phase of both entrained neural oscillations (3.1 Hz: -2.18 ± 0.47 rad, mean ± variance; 5.075 Hz: -1.00 ± 0.55 rad). Interestingly, the best phase for the behavioral effect and the best phase for N1 magnitude were separated by $\pi/2$ radians, which is consistent with what we observed in a previous study using only one entraining frequency (2).

Reversing FM and AM Frequencies Yields Similar Behavioral Results. In an additional behavioral experiment, gap-detection data were obtained from an independent sample of 18 normal-hearing participants (mean age = 25.3 y, SD = 3.5 y; 7 women). Stimuli were identical to those used in the experiment proper, with the critical exception that the FM and AM rates were reversed; that is, the complex rhythm was composed of a 3.1-Hz AM and a 5.075-Hz FM. We did this to rule out the possibility that our main observed result (behavioral comodulation by two entraining stimulus rhythms) was specific to our choice of stimulation rates that mimicked characteristics of natural speech.

The results were very similar to the experiment proper (Fig. S3). Individual participants showed strong and significant modulation of behavior by FM-stimulus phase, as indexed by significant circular–linear correlations between FM phase and hit rate (rising AM phase: z = 3.67 $P < 0.001$, $r = 0.46$; falling AM phase: z = 3.71, $P < 0.001$, $r = 0.47$), which moreover did not differ between AM stimulus phases [$t(17) = 0.15$, $P = 0.88$, $r = 0.04$]. In addition, estimated periodicity in behavioral performance patterns reflected the FM rate (5.075 Hz) and its harmonic (10.15 Hz).

Finally, we confirmed that the behavioral modulation was not driven by consistent detectability differences attributable to stimulus acoustics. We observed a significant difference between AM phases for mean performance [$t(17) = 2.69$, $P = 0.02$, $r = 0.56$], but not for performance range [$t(17) = 0.45$, $P = 0.66$, $r = 0.11$]. Moreover, with respect to optimal FM phase, separate Rayleigh tests were nonsignificant for both the rising AM phase ($z = 0.71$, $P = 0.50$, $r = 0.20$) and the falling AM phase ($z = 2.13$, $P = 0.12$, $r = 0.35$).

Consequences of Independent Component Analysis on Phase Estimation. We also conducted an analysis to test whether application of independent component analysis (ICA) could have affected the phase of the electroencephalography (EEG) signal, and thereby our analysis of pre-gap phase. For each participant and for each single trial, we estimated pre-gap phase (in the 24 ms preceding gap onset). To explore the potential effects of ICA on EEG phase across frequencies, we estimated pre-gap phase for all frequencies between 0.5 and 15 Hz (in 0.5-Hz steps). We also estimated phase for data that had not been subjected to ICA but otherwise were preprocessed in the same manner as the ICA data. Then, for each trial, we calculated the difference between the estimated phases as the circular distance between the per trial phase estimates for the ICA data and the corresponding phase estimates for the non-ICA

data. If ICA did not introduce phase distortions into the data, we would expect the average circular distance between single-trial phase estimates to be zero.

Fig. S4 shows the distribution of single-trial phase estimates for a single participant for non-ICA data and ICA data for three frequency bands: two corresponding to the stimulation frequencies in the current study and 10 Hz (phase estimates for the same participant are shown for all three frequencies). We found no evidence that ICA distorted the phase of the signal. In fact, the distribution of phase differences (between ICA and non-ICA data) for each single participant had a mean of 0 radians and a resultant vector length equal to 1 for all tested frequencies (0.5–15 Hz). Moreover, taking all single trials for all participants together in a single calculation also yielded a mean phase difference of 0 radians and a resultant vector length equal to 1 for all frequency bands.

To confirm that this analysis would have been sensitive to single-trial phase differences, we also shuffled the non-ICA phases over trials separately for each participant and each frequency and then repeated the analysis on the basis of circular distances between phase estimates with and without ICA. When we compared ICA phase with shuffled non-ICA phase, single-trial circular distances were approximately uniformly distributed, and the average resultant vector length (across participants and frequencies was 0.046 ± 0.002 SEM).

Thus, we conclude that removal of components using ICA has no biasing effects on estimates of phase.

SI Materials and Methods

Behavioral Data. To test whether hit rate was significantly modulated by FM stimulus phase in each individual participant, circular–linear correlations were calculated between the binned FM stimulus phases and respective hit rates for each listener, separately for the rising and falling phases of the AM. To test the strength of these correlations across listeners, we conducted permutation tests. On each of 1,000 iterations (separately for AM phases), the correspondence between FM stimulus phase (after binning) and behavioral data was shuffled, and circular–linear correlations were calculated. On the basis of the resulting permutation distribution of circular–linear correlation coefficients, a z-score for the actual correlation between FM stimulus phase and hit rate was calculated by subtracting the distribution's mean from the observed correlation score and dividing this difference by the distribution's SD. Resulting z-scores were tested against 0, using single-sample t tests (two-tailed) separately for the rising and falling AM stimulus phases. Moreover, we compared correlation strengths (z-scores) between AM stimulus phases (rising versus falling), using a paired-samples t test (two-tailed).

Next, we estimated the periodicity present in the behavioral data to confirm it corresponded to the FM stimulation frequency. Separately for the rising and falling AM phases, six cycles of each participant's data were concatenated. Then a cosine function was fit to hit rates as a function of time. Critically, the frequency of the cosine fit took on values ranging between 1 and 10 Hz. For each frequency, the coefficient of determination, R^2, provided a goodness-of-fit measure.

We also investigated whether acoustic modulations consistently affected behavioral performance measures that were independent from the circular–linear correlation across participants. We fit single-cycle cosine functions to the binned hit rate data as a function of FM stimulus phase, separately for the rising and falling AM phases. From best-fitting cosine functions, we estimated three dependent measures. First, mean performance was taken as the intercept of the fitted cosine function and reflects the overall performance level, ignoring modulation by FM stimulus phase. Second, performance range corresponded to the peak-to-trough distance of the fitted cosine function and is related to the degree of behavioral modulation by stimulus phase. Third, optimal FM stimulus phase was taken as the FM phase that corresponded to the peak of the fitted cosine function; that is, the peak predicted performance. Performance range and mean performance values were compared between the rising and falling AM phases, using separate paired-samples t tests (two-tailed). Optimal stimulus phases were tested for nonuniformity, using separate Rayleigh tests.

Electroencephalography Data. Data were preprocessed twice; one pipeline was geared toward frequency-domain analysis of full-stimulus epochs, and the second was geared toward analysis of prestimulus phase in short epochs centered on targets (gaps). The former involved first high-pass filtering, at 0.9 Hz (zero-phase), the continuous EEG signal, and then defining full-stimulus epochs as 1.5 s before stimulus onset to 15.5 s after stimulus onset to capture the response to the full 14-s stimulus. Epoched data were then low-pass filtered at 100 Hz (zero-phase) and re-referenced to linked mastoids. Blinks, muscle activity, electrical heart activity, and noisy electrodes were removed from the signal with ICA, using the Fieldtrip-implemented runica method (3), which performs ICA decomposition using the logistic infomax algorithm (4) with principle component dimension reduction. Individual trials were subsequently removed if the amplitude range exceeded 120 μV; of the 200 presented trials, the median number of rejected trials was 10 (±12.5 semi-interquartile range). After artifact rejection, full-stimulus epochs were analyzed in the frequency domain to examine oscillatory brain responses entrained by the 3.1- and 5.075-Hz stimulation.

The latter pipeline omitted high-pass filtering, and thus first involved epoching (−1.5 to 15.5 s) and then low-pass filtering, re-referencing, and ICA artifact removal. We subsequently rejected the same set of trials identified by the artifact rejection routine from the previously described pipeline. After artifact rejection, shorter epochs were defined that ranged between −2 s and +2 s with respect to each gap onset. Further analysis of prestimulus phase effects is described here.

Frequency-domain analysis. The initial and final seconds of stimulation were first removed to eliminate onset- and offset-evoked responses. Then time-domain data were multiplied with a Hann window before analysis to eliminate artifacts related to the assumption of periodic data that is inbuilt in the fast Fourier transform (FFT). For the current data, we performed the FFT in two different ways. First, "total" amplitude was calculated by averaging frequency-domain representations of single-trial data. Second, "evoked" amplitude was calculated in two FFTs on trial-averaged time-domain data that were first realigned either with respect to FM stimulus phase ("FM evoked") or AM stimulus phase ("AM evoked"). That is, because the starting phase of both the AM and FM stimulation was randomized from trial to trial, the FFT analysis was performed twice: once with attention to each modulation type. Single-trial brain responses were shifted in time so that either the FM (3.1 Hz) or the AM (5.075 Hz) stimulation would have been perfectly phase-locked across trials, and FFTs were performed on trial-averaged time-domain data.

To test for entrained neural responses, we performed hypothesis-directed nonparametric Wilcoxon signed-rank tests for spectral amplitudes at the stimulation frequencies and for their second harmonics against the median spectral amplitude over the 16 neighboring frequency bins centered on the stimulation frequencies (8 on either side of the target frequency; two-tailed tests, ref. 5). For total amplitude representations, we tested both FM and AM frequencies as well as harmonics. For FM- and AM-evoked amplitude representations, we tested only the FM or AM stimulation frequencies, respectively, as well as their harmonics. For visualization, we normalized amplitudes by subtracting from every frequency bin the median over the 16 neighboring bins in the same way that we performed statistics. Both FFT plots and topographies show normalized amplitudes.

Prestimulus phase effects. Prestimulus phase analyses were conducted only for electrode Cz. First, the single-trial time-domain signal was

detrended (using linear regression), and then gap-evoked responses were removed from the postgap onset time window by multiplication with half of a Hann window that ranged between 0 and 50 ms postgap onset and was zero thereafter ("ERP-free"; Fig. S2). This was done to eliminate the possibility that "smearing" of the evoked response back into the prestimulus period by the wavelet convolution could produce spurious prestimulus phase effects (for a similar approach, see ref. 6). Next, the time-domain data were submitted to a wavelet convolution (wavelet width = 3 cycles) that yielded complex output in two frequency bins centered on 3.1 and 5.075 Hz (± 0.25 Hz) and with 2-ms temporal resolution. Complex output was then converted to phase-angle time series.

Prestimulus phase in each frequency band was estimated for each trial as the circular mean of phase angle values in the 24-ms time window preceding gap onset. Trials were then sorted into an 18 × 18 grid of overlapping bins (bin width = 0.6π), according to pre-gap neural phase in the frequency bands of interest (i.e., 3.1 × 5.075 Hz). We calculated the same three dependent measures as for the behavioral data, except that cosine fits were conducted for hit rate as a function of 3.1-Hz neural phase separately for each of the 18 values of 5.075-Hz neural phase (Fig. 4*C*). That is, for each listener, we estimated 18 values of 3.1-Hz-driven mean performance, performance range, and optimal 3.1-Hz-driven neural phase, respectively. We then tested whether any of these behavioral measures (estimated as a function of 3.1-Hz neural phase) depended on the specific phase in the 5.075-Hz frequency band. For optimal neural phase, we did this using a Watson-Williams circular one-way ANOVA. For performance range and mean performance, we used separate one-way repeated-measures ANOVAs.

To test whether neural phase effects on gap-detection performance were interactive, we estimated the degree to which hit rates were modulated by neural phase in the 3.1-Hz frequency band alone, in the 5.075-Hz frequency band alone, and in the two frequency bands taken together. To do so, we sorted single-trial accuracies into 18 bins on the basis of neural phase only in the 3.1- or 5.075-Hz frequency bands. For each frequency band, we calculated a behavioral modulation index, which was simply the maximum minus the minimum hit rate after sorting and averaging in each bin. For the "interaction" effect, we sorted single-trial accuracies simultaneously by 3.1- and 5.075-Hz pre-gap neural phase, and we again calculated a behavioral modulation index by subtracting the minimum hit rate from the maximum hit rate for each participant. Finally, we conducted pairwise tests of the behavioral modulation index in each single frequency band relative to the combined frequency bands (and for the two single frequency bands against each other), using paired-samples *t* tests (two-tailed).

Finally, we tested whether the observed behavioral comodulation was specific to the neural frequency bands entrained by our stimulation. From the complex output of a wavelet convolution yielding prestimulus complex values between 1 and 10 Hz (in 0.25-Hz steps), we estimated prestimulus neural phase in a time window corresponding to 10% of a cycle. We then calculated a behavioral modulation index for every frequency band between 1 and 10 Hz separately, as well as for every pairwise combination of frequencies between 1 and 10 Hz. We subsequently calculated an interaction strength metric by taking the difference between the "interaction" modulation index and the mean of the two individual-frequency modulation indices. We tested the interaction strength for the combination of the two entrained frequency bands (3.1 and 5.075 Hz) against all other frequency combinations, using a permutation test. That is, for each listener, we formed a permutation distribution by calculating interaction strength for 1,000 random pairwise frequency combinations. Then, we compared interaction strength in the 3.1-Hz × 5.075-Hz bin to the permutation distribution, which yielded a z-score for each participant that reflected the frequency-specificity of the behavioral comodulation effect. Z-scores were tested against 0, using a single-sample *t* test at the second-level (two-tailed).

1. Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164(1):177–190.
2. Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci USA* 109(49):20095–20100.
3. Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1996) *Independent Component Analysis of Electroencephalographic Data*, eds Touretzky D, Mozer M, Hasselmo M, Advances in Neural Information Processing Systems (MIT Press, Cambridge, MA), Vol 8.
4. Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 7(6):1129–1159.
5. Picton TW, John MS, Dimitrijevic A, Purcell D (2003) Human auditory steady-state responses. *Int J Audiol* 42(4):177–219.
6. Lakatos P, et al. (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77(4):750–761.

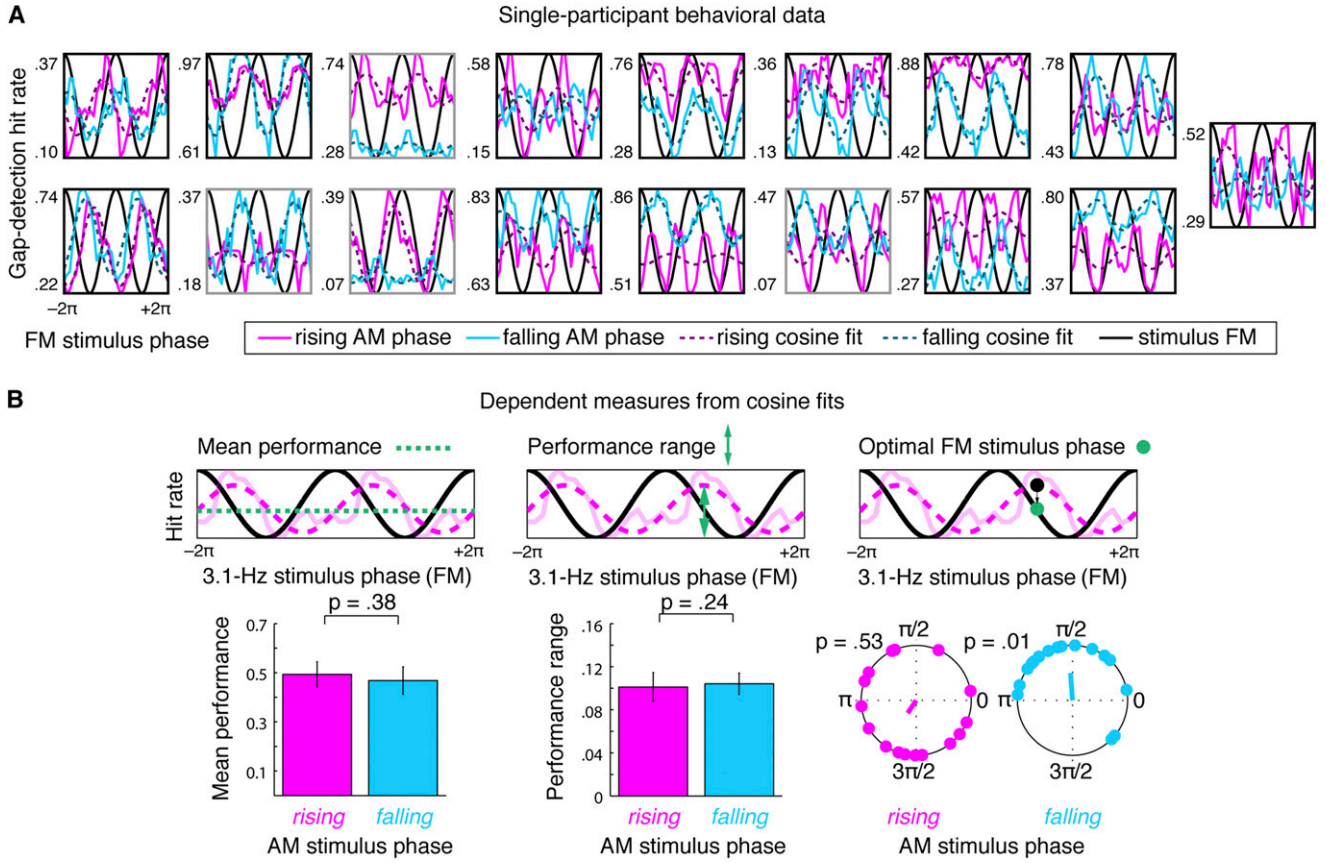

Fig. S1. Dependent measures from cosine fits to single-participant behavioral data. (*A*) Single-participant hit rate data for all participants as a function of FM stimulus phase, separately for the rising (magenta) and falling (cyan) AM phases. Best-fit cosine functions are overlaid (dotted lines). (*B*) Dependent measures from cosine fits: mean performance (*Left*) corresponds to the intercept of the fitted cosine function, performance range (*Center*) corresponds to the peak-to-trough distance of the fitted function, and optimal FM stimulus phase (*Right*) corresponds to the FM stimulus phase yielding peak predicted performance. Neither mean performance nor performance range differed significantly between the rising and falling AM phases. For optimal FM phase, a test to compare rising versus falling AM phases was not performed because the mean resultant vectors were not sufficiently long.

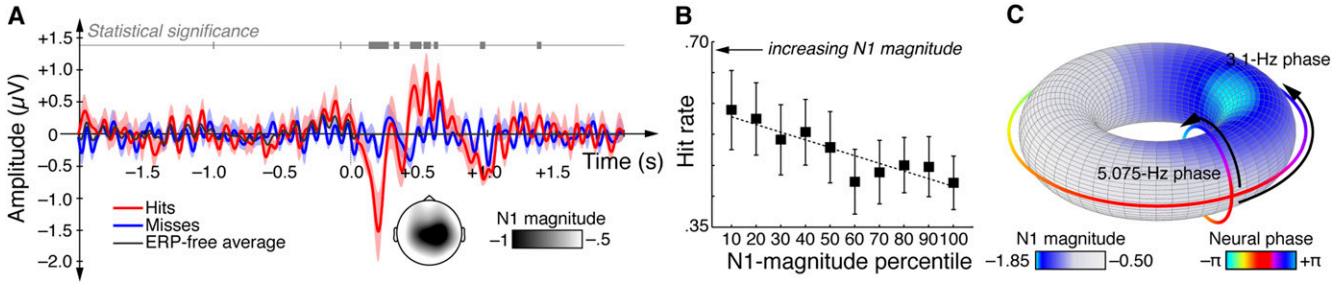

Fig. S2. (*A*) ERPs for detected gaps (hits, red) and undetected gaps (misses, blue). Shading shows SEM, and the gray bars at the top of the figure mark time windows within which the ERP magnitudes differed significantly ($P < 0.05$ with cluster correction) for hits versus misses. The topography corresponds with the grand average magnitude (microvolts) in the time window ranging between 0.18 and 0.22 s after gap-onset (N1 time window). The thin black line in the time-domain representation ("ERP-free average") shows the grand average ERP where the poststimulus gap-evoked response has been muted by multiplication with half a Hann window. The ERP-free signal was used for estimating prestimulus phase. (*B*) Hit rates plotted as a function of N1-magnitude percentile bins. Small percentile values correspond to stronger (increasingly negative) N1s. Larger hit rates were associated with larger N1s. (*C*) N1 magnitudes plotted on a torus as a function of prestimulus 3.1-Hz neural phase (larger, outside circle) and 5.075-Hz neural phase (smaller, inner circle). N1s were maximal (most negative) when the gap fell simultaneously into the rising phase of both entrained neural oscillations.

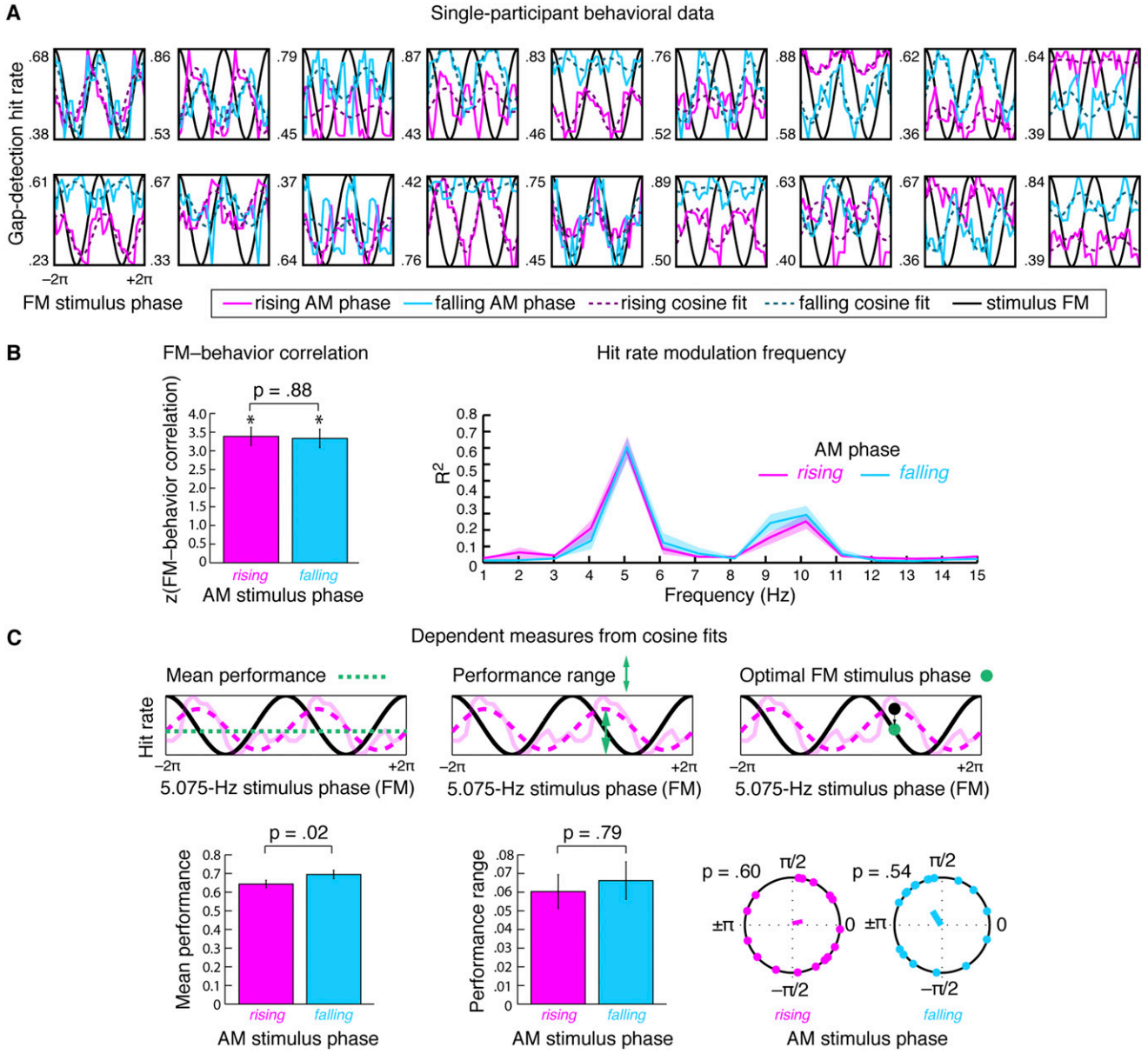

Fig. S3. (*A*) Single-participant behavioral data (gap-detection hit rates as a function of FM stimulus phase); best-fitting cosine functions are overlaid. (*B*) Mean (±SEM) z-values corresponding to circular–linear correlation strengths between hit rates and FM stimulus phase (*Left*). Across participants, behavior was significantly modulated by FM stimulus phase for both the rising and falling AM phases. Asterisks indicate significance at $P < 0.001$. Moreover, estimated hit-rate modulation frequencies (*Right*) matched the FM stimulation frequency and its harmonic. (*C*) Mean performance (*Left*), performance range (*Center*), and optimal FM phase (*Right*) shown separately for the rising (magenta) and falling (cyan) AM stimulus phases. Mean performance differed somewhat between rising and falling AM stimulus phases ($P = 0.02$), but performance range did not ($P = 0.79$). Distributions of optimal FM stimulus phases were not significantly different from uniform for either the rising or the falling AM phase. A test to compare rising versus falling AM stimulus phases was not performed because the mean resultant vectors were not sufficiently long.

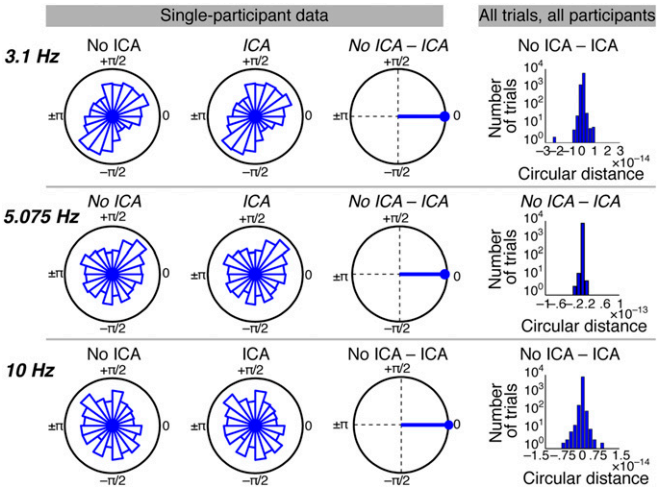

Fig. S4. Results of a comparison of phase estimates for data subjected to ICA for artifact removal compared with non-ICA data. Rose plots show the distribution of single-trial phase estimates for a single participant from non-ICA data ("No ICA") and data subjected to ICA ("ICA"). Calculating the circular distance between single-trial phase estimates for non-ICA versus ICA data yielded a mean phase difference of 0 radians and a resultant vector length of 1 for all participants and all frequencies ("No ICA – ICA"). Finally, taking all trials from all participants together in one analysis (histogram, "No ICA – ICA") also yielded a mean phase difference of 0 and resultant vector length of 1.