

# Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha

Stephanie J. Kayser,<sup>1</sup> Robin A.A. Ince,<sup>1</sup> Joachim Gross, and Christoph Kayser<sup>1</sup>

Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom

The entrainment of slow rhythmic auditory cortical activity to the temporal regularities in speech is considered to be a central mechanism underlying auditory perception. Previous work has shown that entrainment is reduced when the quality of the acoustic input is degraded, but has also linked rhythmic activity at similar time scales to the encoding of temporal expectations. To understand these bottom-up and top-down contributions to rhythmic entrainment, we manipulated the temporal predictive structure of speech by parametrically altering the distribution of pauses between syllables or words, thereby rendering the local speech rate irregular while preserving intelligibility and the envelope fluctuations of the acoustic signal. Recording EEG activity in human participants, we found that this manipulation did not alter neural processes reflecting the encoding of individual sound transients, such as evoked potentials. However, the manipulation significantly reduced the fidelity of auditory delta (but not theta) band entrainment to the speech envelope. It also reduced left frontal alpha power and this alpha reduction was predictive of the reduced delta entrainment across participants. Our results show that rhythmic auditory entrainment in delta and theta bands reflect functionally distinct processes. Furthermore, they reveal that delta entrainment is under top-down control and likely reflects prefrontal processes that are sensitive to acoustical regularities rather than the bottom-up encoding of acoustic features.

**Key words:** auditory cortex; delta band; information theory; rhythmic entrainment; speech

## Significance Statement

The entrainment of rhythmic auditory cortical activity to the speech envelope is considered to be critical for hearing. Previous work has proposed divergent views in which entrainment reflects either early evoked responses related to sound encoding or high-level processes related to expectation or cognitive selection. Using a manipulation of speech rate, we dissociated auditory entrainment at different time scales. Specifically, our results suggest that delta entrainment is controlled by frontal alpha mechanisms and thus support the notion that rhythmic auditory cortical entrainment is shaped by top-down mechanisms.

## Introduction

Natural sounds are characterized by statistical regularities at the scale of a few hundreds of milliseconds. For example, the pseudorhythmic structure imposed by syllables plays an important role for speech parsing and intelligibility (Elliott and Theunissen, 2009; Giraud and Poeppel, 2012; Leong and Goswami, 2014).

Recent work has shown that auditory cortical activity exhibits prominent fluctuations at similar time scales (Kayser et al., 2009; Szymanski et al., 2011; Ng et al., 2013). In particular, activity in the delta (~1 Hz) and theta (~4 Hz) bands systematically aligns to acoustic landmarks, a phenomenon known as cortical “entrainment” (Luo and Poeppel, 2007; Lakatos et al., 2009; Peelle and Davis, 2012). Given that rhythmic network activity indexes the gain of auditory cortex neurons (Lakatos et al., 2005; Kayser et al., 2015), it has been hypothesized that entrainment reflects a key mechanism underlying hearing, for example, by facilitating the parsing of individual syllables through adjusting the sensory gain relative to fluctuations in the acoustic energy (Giraud and Poeppel, 2012; Peelle and Davis, 2012; Ding and Simon, 2014).

Entrainment is observed for many types of nonspeech stimuli and is affected by manipulations of acoustic properties, suggesting that it is partly driven in a bottom-up manner by the auditory input (Henry and Obleser, 2012; Doelling et al., 2014; Millman et al., 2013; Ding and Simon, 2014). However, auditory activity at the same time scales has also been implicated in mediating mech-

Received June 11, 2015; revised Aug. 20, 2015; accepted Aug. 21, 2015.

Author contributions: S.J.K., J.G., and C.K. designed research; S.J.K. performed research; S.J.K., R.A.A.I., and J.G. contributed unpublished reagents/analytic tools; S.J.K. and C.K. analyzed data; S.J.K., R.A.A.I., J.G., and C.K. wrote the paper.

This work was supported by start-up funds from the University of Glasgow, the UK Biotechnology and Biological Sciences Research Council (Grant BB/L027534/1), and the Wellcome Trust (Grant 098433).

This article is freely available online through the *JNeurosci* Author Open Choice option.

Correspondence should be addressed to Christoph Kayser, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, United Kingdom. E-mail: christoph.kayser@glasgow.ac.uk.

DOI:10.1523/JNEUROSCI.2243-15.2015

Copyright © 2015 Kayser et al.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

anisms underlying active sensing, such as temporal expectations, rhythmic predictions, and attentional selection (Besle et al., 2011; Ding and Simon, 2012; Morillon et al., 2014; Hickok et al., 2015). For example, higher delta phase concentration is observed around expected sounds (Stefanics et al., 2010; Arnal et al., 2015; Wilsch et al., 2015b) and auditory cortex entrains more strongly to attended than unattended streams (Lakatos et al., 2008; Zion Golumbic et al., 2013). This suggests that entrainment is at least partly under top-down control by frontal and premotor cortices (Saur et al., 2008; Peelle and Davis, 2012; Park et al., 2015). As a result, current data suggest that entrained activity reflects both the feedforward tracking of sensory inputs and active mechanisms of sensory selection (Giraud and Poeppel, 2012; Hickok et al., 2015) and it remains difficult to disentangle these contributions (Doelling et al., 2014; Ding and Simon, 2014).

To better dissociate rhythmic auditory entrainment and neural activity reflecting the early encoding of acoustic inputs (e.g., evoked responses), we investigated whether and to what degree unpredictable changes in the temporal pattern of speech affect different neural indices of auditory function. To this end, we used an artificial manipulation of the local speech rate. We focused on the regularity emerging from the alternation of articulation and pauses (periods of relative silence) between words or syllables that is important for speech segmentation (Rosen, 1992; Zellner, 1994; Dilley and Pitt, 2010; Geiser and Shattuck-Hufnagel, 2012). By manipulating the statistical distribution of these pauses, we systematically rendered the local speech rate irregular while preserving the overall speech rate, the statistical structure of the overall sound envelope, and intelligibility. Quantifying different signatures of auditory function in human EEG data, we found: (1) a dissociation between rhythmic entrainment in the delta band and left frontal alpha power, which were reduced for manipulated speech rate; and (2) entrainment in other frequency bands, perceptual intelligibility, and auditory evoked responses that were preserved.

## Materials and Methods

**Study.** Nineteen healthy adult participants (age 18–37 years) took part in this study. All had self-reported normal hearing, were briefed about the nature and goal of this study, and received financial compensation for their participation. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee (College of Science and Engineering, University of Glasgow). Written informed consent was obtained from all participants.

**Stimulus material.** We presented 8 6-min-long speech samples. The samples were based on text transcripts taken from publicly available TED talks. Acoustic recordings (at 44.1 kHz sampling rate) of these texts were obtained while a trained male native English speaker narrated them. The root mean square (RMS) intensity of each recording was normalized using 10 s sliding windows to ensure a constant average intensity.

For the present study, we presented sections of the original samples and manipulations of these with altered speech rate. In brief, this manipulation was performed by detecting periods of relative silence (i.e., low amplitude) within the original speech (termed pauses), noting the mean and SD (jitter) of the length of these and subsequently creating manipulated speech by randomly shortening or lengthening pauses to preserve their overall mean duration but increase their jitter. We performed this manipulation using three different levels that increased the jitter by 30%, 60%, and 90%. The detection of pauses was performed using an algorithm based on acoustic properties agnostic to linguistic contents (Zellner, 1994; Loukina et al., 2011). The wideband amplitude envelope was computed following previous studies (Chandrasekaran et al., 2009; Gross et al., 2013) by band-pass filtering the original speech into 11 logarithmically spaced bands between 200 and 6000 Hz (third-order Butterworth filters), computing the amplitude envelope for each band using the Hil-

bert transform, down sampling to 1 ms resolution, and averaging across bands. The resulting envelope was smoothed using a 10 ms sliding Gaussian filter. Periods in which the normalized signal (relative to 1) was  $<0.1$  were considered as pauses and a clustering algorithm was used to identify continuous pauses of at least 30 ms duration, the beginning and end of which were at least 60 ms apart from neighboring pauses (see Fig. 1A). On average, across all eight text samples, we detected 6300 pauses. The length of each pause was then systematically manipulated to increase the jitter (i.e., the SD) of the distribution of pause durations (see Fig. 1A). This was done by extending (or shrinking) each pause randomly by adding (or subtracting) a normally distributed silent interval with zero mean and scaled SD (increasing the overall SD by 0%, 30%, 60%, or 90%), with the additional constraint that the resulting pause must remain longer than 20 ms and must not exceed 300% of its original length. For the 0% condition, we used the original duration of each pause. To reconstitute the speech material with manipulated pause, we assumed a zero amplitude during each pause and cosine ramped the onset and offset of the speech segments around each pause (5 ms ramp). A continuous white noise background with relative RMS level of 0.05 was added to the reconstituted speech material to mask minor acoustic artifacts introduced by the manipulation.

We ensured that this manipulation of the local speech rate did not alter the overall mean duration of pauses, only increased their jitter. To verify this, we compared the distributions of pause durations in the original material and manipulated versions of the very same text segments directly (see Fig. 1C). For statistical comparison, we used the data from the same sub-blocks for each condition that were also used in the main experiment (cf. Fig. 2A). We averaged the mean duration and jitter within each sub-block and compared their distribution between the 12 sub-blocks for each condition. We also ensured that the manipulation did not induce a significant difference in the overall envelope statistics across the different conditions (see Fig. 1D). To this end, we computed the power of the acoustic amplitude envelope in the same frequency bands used for the analysis of entrainment (below). We then compared the time-averaged power between the text segments of each experimental condition as presented in the experiment (i.e., between the 12 sub-blocks for each condition).

**Experimental design.** The experiment was based on a block design (see Fig. 2A). We presented each of the 8 original texts as continuous 6 min stimuli, but introduced the 4 experimental conditions (0%, 30%, 60%, and 90% increased jitter) in 1 min sub-blocks. Each 6 min text was divided into sub-blocks of  $\sim 1$  min (59.2–61.3 s) and the speech within this sub-block was manipulated according to the respective condition. The order of the conditions across texts and sub-blocks was pseudorandomized (see Fig. 2A). In total, we obtained 12 continuous 1 min blocks for each of the 4 conditions. Given that each original sample was used only once, each condition was based on distinct acoustic material with a clearly defined distribution of silent periods.

To obtain a behavioral assessment of speech intelligibility and to maintain the subject's attention during EEG recordings, we instructed participants to pay attention and to listen carefully to be able to complete a memory task after each block. At the end of each block, participants were presented with 12 words (nouns) on a computer monitor and had to indicate whether they had heard this word or not by pressing one of two buttons. Two words per block were taken from each sub-block, allowing us to compute performance separately for each condition (see Fig. 2B).

To judge the quality of auditory evoked responses in each participant, we also recorded responses to a brief acoustic localizer stimulus during passive listening (Ding and Simon, 2013). We presented 10 trials, each consisting of a sequence of 10 500 Hz tones (150 ms duration including a 30 ms on/off cosine ramp) spaced randomly between 400 and 700 ms apart.

**Recording procedures.** Experiments were performed in a dimly light and electrically shielded room. Acoustic stimuli were presented binaurally using Sennheiser headphones while stimulus presentation was controlled from MATLAB (The MathWorks) using routines from the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). Sound levels were calibrated using a sound level meter (Model 2250; Brüel & Kjær) to an average of 65 dB RMS level. EEG signals were continuously recorded

using an active 64 channel BioSemi system using Ag-AgCl electrodes mounted on an elastic cap (BioSemi) according to the 10/20 system. Four additional electrodes were placed at the outer canthi and below the eyes to obtain the electrooculogram. Electrode impedance was kept at <25 k $\Omega$ . Data were acquired at a sampling rate of 500 Hz using a low-pass filter of 208 Hz.

**General data analysis.** Data analysis was performed offline with MATLAB using the FieldTrip toolbox (Oostenveld et al., 2011) and custom-written routines. The EEG data from different recording blocks were preprocessed separately. The data were low-pass filtered at 70 Hz, resampled to 150 Hz, and subsequently denoised using independent component analysis. Usually, one or two components reflecting eye-movement-related artifacts were identified and removed following definitions provided by Debener et al. (2010). In addition, for some subjects, highly localized components reflecting muscular artifacts were detected and removed (O'Beirne and Patuzzi, 1999; Hipp and Siegel, 2013). To detect potential artifacts pertaining to remaining blinks or eye movements, we computed horizontal, vertical, and radial EOG signals following established procedures (Keren et al., 2010; Hipp and Siegel, 2013).

For the analysis of evoked potentials and oscillatory activity, the data were epoched ( $-0.8$  to  $+0.8$  s) around the end of each pause; that is, around the syllable onset after each pause. Potential artifacts in the epoched data were removed using an automatic procedure by excluding epochs if on any electrode the peak amplitude exceeded a level of  $\pm 110$   $\mu$ V. In addition, we removed epochs containing excessive EOG activity based on the vertical and radial EOGs. Specifically, we removed epochs in which potential eye movements were detected based on a threshold of 3 SDs above mean of the high-pass-filtered EOGs using the procedures suggested by Keren et al. (2010). Together, these criteria led to the rejection of  $12 \pm 8\%$  (mean  $\pm$  SD) of epochs across participants. Evoked responses for each condition were obtained by epoch averaging after low-pass (20 Hz, third-order Butterworth filters) and high-pass (1 Hz) filtering and baseline normalization ( $-0.8$  to  $0$  s). Complex valued time-frequency (TF) representations of the epoched data were obtained with Morlet wavelets using a frequency-dependent cycle widths to allow more smoothing at higher frequencies (Griffiths et al., 2010), ranging from 3.5 cycles at 2 Hz to 6 cycles at 30 Hz. TF representations were computed at 1 Hz frequency steps between 2 and 16 Hz and 2 Hz frequency steps between 16 and 30 Hz. The TF power spectrum was obtained by epoch averaging and Z-scoring the power time series in each band by the mean and SD over time within this band (Arnal et al., 2015; Wilsch et al., 2015a). Importantly, before computing the TF representation for power, we subtracted the trial averaged evoked responses for each condition from the single trial responses (Griffiths et al., 2010), thereby ensuring that the analysis of power specifically focused on so-called induced oscillations and therefore activity that is not strictly time locked to the epoch (Tallon-Baudry and Bertrand, 1999). The intertrial phase coherence (ITC) was obtained by computing the length of the epoch averaged complex representation of the instantaneous phase. To obtain a better separation of syllable onsets and the preceding pauses, we restricted this analysis to pauses with a minimal duration of 0.05 s. The analysis of evoked potentials and oscillatory activity was performed separately for epochs falling in each of the experimental conditions, ensuring that the same number of epochs was used per condition. As a control, we repeated these analyses after grouping pauses into four equi-populated groups defined by their duration (grouping by the 0–25<sup>th</sup>, 25–50<sup>th</sup>, 50–75<sup>th</sup>, and 75–100<sup>th</sup> length percentiles). This analysis served as control to verify that statistical approach was sufficiently sensitive to find potential changes across experimental conditions. We expected to find significant changes in evoked responses and ITC with longer pause duration.

For the auditory localizer paradigm, we analyzed the evoked responses as above using epochs ( $-0.5$  to  $+0.5$ ) around each individual tone. For most participants, this resulted in a strong and centrally located auditory evoked response. We used this to exclude participants not exhibiting a clear auditory evoked response. For the present study, this was the case for three of the 19 participants. Therefore, results reported in this study are from a sample of  $n = 16$  participants.

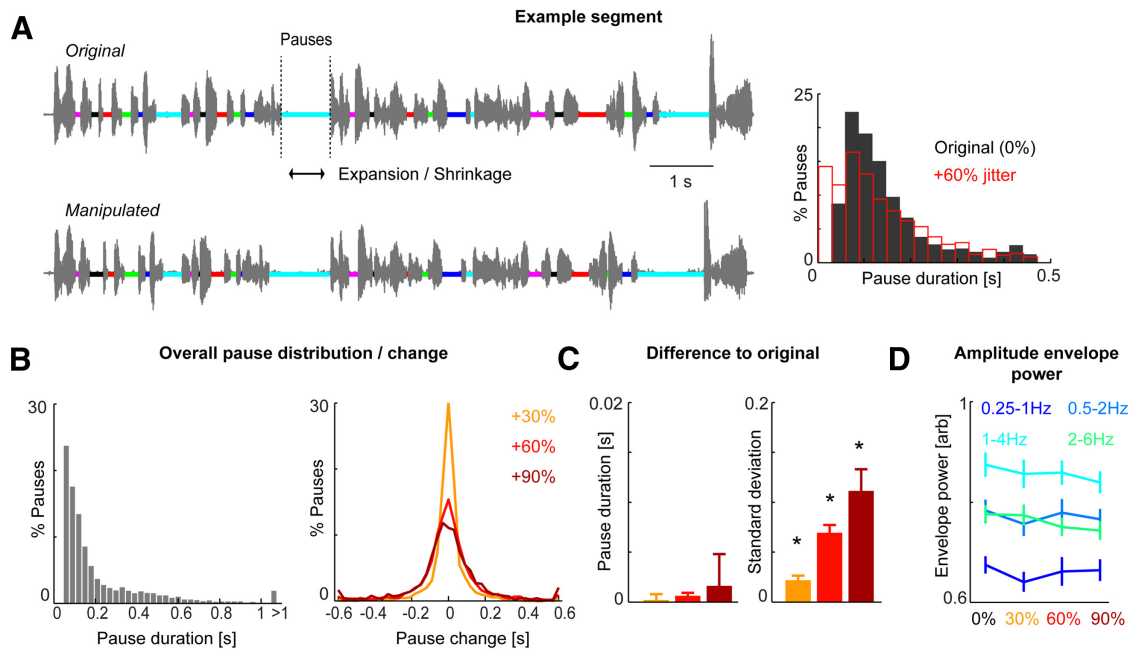
**Analysis of entrainment.** To compute the statistical relation between EEG activity and the acoustic stimulus, we used the framework of mutual

information (MI) (Quian Quiroga and Panzeri, 2009; Panzeri et al., 2010). Following previous studies, we computed the MI between band-limited components extracted from EEG data and from the sound amplitude envelope in the same frequency bands (Kayser et al., 2010; Cogan and Poeppel, 2011; Gross et al., 2013; Ng et al., 2013). The wide-band speech amplitude envelope was computed using nine equi-spaced frequency bands (100 Hz–10 kHz) at a temporal resolution of 150 Hz following the methods of Gross et al. (2013). Both the wide-band envelope and the EEG data were then filtered into 10 partly overlapping frequency bands using Kaiser filters (1 Hz transition bandwidth, 0.01 dB pass-band ripple and 50 dB stop-band attenuation; forward and backward filtering was used to prevent phase shifts). The precise bands were 0.25–1, 0.5–2, 1–4, 2–6 Hz, 4–8, 8–12, 12–18, 18–24, 24–36, and 30–48 Hz. The MI was then computed between the Hilbert representation (or the power, or phase) of the band-limited EEG data on each channel and the Hilbert representation (or the power or phase) of the band-limited amplitude envelope (Gross et al., 2013). The calculation was performed twice, once using the entire data from each block to yield the overall acoustic information and once separately for each condition.

The MI was calculated using a binless, rank-based approach based on statistical copulas. This method greatly reduces the statistical bias that is inherent to direct information estimates (Panzeri et al., 2007) and is highly robust to outliers in the EEG data because it relies on ranked rather than raw data values. The theoretical background is provided by the observation that a copula expresses the relationship between two random variables and that the negative entropy of a copula between two variables is equal to their mutual information (Ma and Sun, 2008; Kumar, 2012). On this basis, we estimated MI via the entropy of a Gaussian copula fit to the empirical copula obtained from the observed data. Whereas the use of a Gaussian copula does impose a parametric assumption on the form of the interaction between the variables, it does not impose any assumptions on the marginal distributions of each variable. Further, because the Gaussian distribution has maximum entropy for a given mean and covariance, this method provides a lower bound of the true information value. In practice, for a given data vector, we calculated its empirical cumulative distribution by ranking the data points, scaling the ranks between 0 and 1, and then obtaining the corresponding standardized value from the inverse of a standard normal distribution. We then computed the MI between the two time series consisting of standardized variables using the analytic expressions for the entropy of Gaussian variables (Cover and Thomas, 1991). Note that this procedure is conceptually the same as for other approaches to mutual information using a binning procedure (Panzeri et al., 2007), but rather than associating each point in a time series with a bin index (indicating the binned amplitude of the respective value), we used the standardized rank of each value computed within the entire time series. Conceptually, this can be imagined as computing correlations between two time series based on a rank correlation rather than a Pearson correlation. For the present analysis, we used the real and imaginary parts of the Hilbert representation of the data as a 2D feature vector to compute the MI and each component was standardized separately. We also repeated the analysis by extracting the phase and power of the Hilbert representation and using these as data representations (see also Gross et al., 2013). As in previous work (Luo and Poeppel, 2007; Belitski et al., 2010; Gross et al., 2013; Ng et al., 2013), we found that the majority of the MI was carried by the phase of the band-passed signals rather than their power (cf. Fig. 3A, right). Unless stated otherwise, all results are based on the full Hilbert representation.

We computed the MI separately for each sub-block and averaged the resulting values across sub-blocks within each condition. To derive an estimate of the information value observed due to random variations in the data, we used a randomization procedure (Montemurro et al., 2007; Kayser et al., 2009). We estimated the MI after time shifting the acoustic envelope by a random lag, which destroys the specific relation between acoustic input and EEG activity but preserves the statistical structure of each individual signal. Based on a distribution of 1000 randomized values for each participant, we derived the group-level probability that the subject-averaged MI values at each electrode exceeded the 99% confidence interval of the null distribution. We corrected for multiple tests





**Figure 1.** Manipulation of speech rate. **A**, The rhythmic structure of speech was characterized by detecting periods of relative silence (termed “pauses”), defined by continuous segments of at least 30 ms of low ( $<0.1$ , normalized) amplitude. These are indicated by colored bars on the sound wave. The duration of individual pauses was manipulated by expanding or shrinking each pause by a random duration drawn from a Gaussian distribution with zero mean and defined variance. The resulting manipulation is shown in the bottom panel (matching colors). Right, Distribution of pause duration in one example text (6 min) before (black) and after applying the manipulation (60% jitter; red). **B**, Distribution of pause duration (left) and change in duration (right) across all 8 6-min texts. **C**, Changes in mean duration and SD imposed by each level of manipulation expressed as a difference from the original material. For this comparison, each manipulation was applied to the entire text corpus. Bars show the difference across sub-blocks, as used in the experiment (cf. Fig. 2;  $n = 12$ ; mean and SEM). **D**, Band-limited power of the speech amplitude envelope for each condition in different frequency bands. Bars show the power across sub-blocks ( $n = 12$ ; mean and SEM).

across electrodes and bands using the maximum statistics (Nichols and Holmes, 2002; Maris and Oostenveld, 2007).

We also repeated the MI analysis by restricting the calculation to those time epochs used for the analysis of evoked responses. Specifically, we used the interval of  $-0.1$  s before and  $0.4$  s after the detected syllable onset following each pause to calculate MI for each condition. Finally, to quantify the overall signal power within each of these frequency bands (see Fig. 1D), we averaged the power of the Hilbert signals over all time points within each sub-block.

**Statistical analysis.** Our main hypotheses concern changes in MI, evoked potentials, or induced oscillatory power across the experimental conditions. Therefore, our statistical analysis focused on systematic increases or decreases across conditions. We implemented a two-level statistical approach using a cluster-based permutation procedure controlling for multiple comparisons for all regression analyses (Maris and Oostenveld, 2007; Strauss et al., 2015). First-level contrasts reflecting systematic increases or decreases with conditions were derived using single-subject data based on rank-ordered regression of the observed data with the condition label (1–4). We used rank-regression rather than linear regression because the latter carries the implicit assumption of equal differences between conditions, which may not be justified. In practice, however, we found little difference between the tests. Beta values for the regression were obtained for each electrode and TF bin (where applicable). The second-level group statistical analysis used a cluster-based permutation procedure implemented in Fieldtrip (Maris and Oostenveld, 2007). This procedure tests the first-level statistics against zero controlling for multiple comparisons (detailed parameters: 1000 iterations, including only bins with  $t$  values exceeding a two-sided  $p < 0.05$  in the clustering procedure, requiring a cluster size of at least 2 significant neighbors, performing a two-sided  $t$  test at  $p < 0.05$  on the clustered data). For MI, this test was performed across all electrodes and frequency bins and, for evoked potentials (ITC or power), across all time (TF) points but restricting the analysis to frontocentral electrodes exhibiting significant overall acoustic information (see Fig. 4, inset).

Effect sizes for cluster-based  $t$ -statistics are reported as the summed  $t$  value across all bins (electrode, time, frequency) within a cluster ( $T_{\text{sum}}$ ) and by providing the equivalent  $r$  value that is bounded between 0 and 1 (Rosenthal and Rubin, 2003; Strauss et al., 2015). The equivalent  $r$  was averaged across

all bins and denoted  $R$ . We provide exact  $p$  values where possible (for parametric tests), but values  $<10^{-5}$  are abbreviated as such.

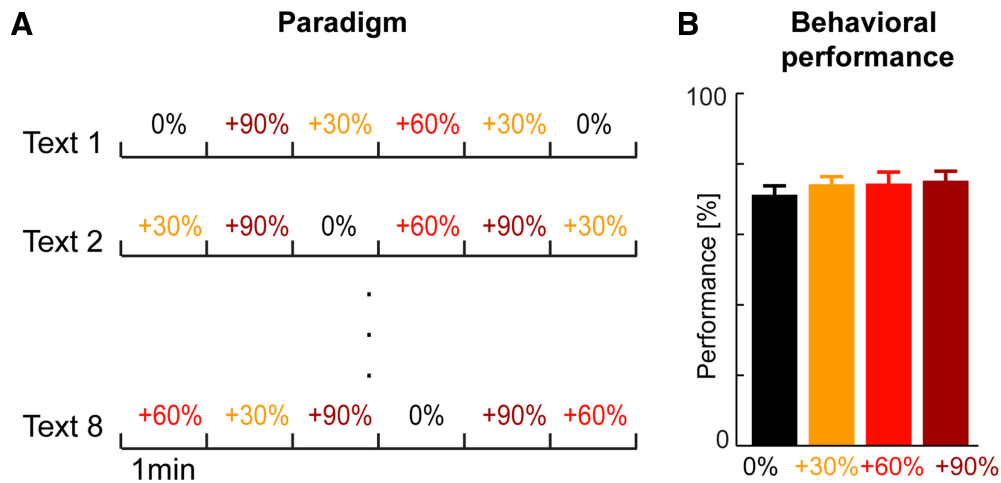
Given that potential effects of our experimental conditions may be more prevalent after a longer compared with a shorter pause, we performed a secondary analysis of interaction. Having identified time or TF regions of interest based on group-level statistics, we subjected the time (TF) averaged data to a  $2 \times 2$  ANOVA to test for an interaction of condition and pause duration. To reduce the number of effective conditions, we only considered two levels of manipulation (grouping 0% and 30% jitter and grouping 60% and 90% jitter, each one category each) and only two levels of duration (defined by the lowest and highest 30<sup>th</sup> percentiles of the distribution of all durations).

## Results

### Manipulation of speech rate

We systematically manipulated the rhythmic structure of speech arising from the alternation of periods of articulation and relative silence; that is, the speech rate (Tauroza and Allison, 1990; Zellner, 1994). Based on a corpus of 8 6-min-long texts, we first segmented the speech amplitude envelope into periods of acoustic signal and pauses by using a thresholding procedure (Fig. 1A, left). We then manipulated the statistical distribution of these pauses by randomly shortening or extending their duration in a manner that preserved their overall mean duration but increased the jitter (SD). This is illustrated in Figure 1A for one example segment, showing the pauses in the original segment (top) and after increasing the jitter by 60% (bottom). Directly comparing matching pauses (color code) across the two samples illustrates the local changes in speech rate while the overall rate and text duration are maintained.

Across the 8 texts, our algorithm recovered 6300 pauses, with an average duration of 0.233 s and an SD of 0.289 s. Based on the broad distribution of pause durations (Fig. 1B), these



**Figure 2.** Experimental design and behavioral results. **A**, Design of the EEG study. Each of the 8 6-min text samples was presented once and the 4 conditions were introduced in 1 min sub-blocks. The order of conditions was randomized across (sub-) blocks to obtain  $n = 12$  unique 1 min text segments for each condition. **B**, Behavioral recognition performance across participants ( $n = 16$ ; mean and SEM).

include both intrasegmental and interlexical pauses; that is, periods between syllables within a word as well as periods between words or sentences (Zellner, 1994; Loukina et al., 2011). However, the average duration of ~230 ms and the peak at even shorter durations is consistent with syllable rate segmentation rather than a word-based segmentation (Taurroza and Allison, 1990). Our experimental manipulation thus altered the regularity of the speech rate largely on the basis of local syllable-scale manipulations.

For this study, we used four conditions consisting of the original rate (0% manipulation) and three conditions with systematically increased jitter (30%, 60%, and 90%). We verified that our manipulation increased the jitter without significantly affecting the mean duration of pauses (i.e., global speech rate). To this end, we compared directly the distributions of pauses in the original material and the same text segments after introducing the manipulation (Fig. 1C). Changes in mean duration were <5 ms and did not differ significantly between conditions (one-way ANOVA  $F_{(2,22)} = 0.37, p = 0.7$ ). In contrast, changes in jitter differed significantly between conditions ( $F_{(2,22)} = 9.9, p = 0.0008$ ) and a regression of the mean change revealed a significant increase in jitter with condition ( $r^2 = 0.95, F = 651, p = 0.024$ ). For completeness, Figure 1B also shows the distribution of the introduced changes in pause duration across the full corpus.

The sound amplitude envelope is a critical determinant for the entrainment of auditory cortex activity (Peelle and Davis, 2012; Doelling et al., 2014; Ding and Simon, 2014) and we verified that the statistical properties of the amplitude envelope of the manipulated material were comparable across conditions (Fig. 1D). We computed the power of the speech envelope of each text segment as it was presented during the experiment and then compared the power of envelope fluctuations between conditions using the same frequency bands as for the analysis of cortical entrainment below. This revealed no significant effect of condition on band-limited power for any of the frequency bands (one-way ANOVA; e.g., 0.25 Hz:  $F_{(3,33)} = 0.71, p = 0.55$ ; 0.5 Hz:  $F = 0.37, p = 0.77$ ; 1 Hz:  $F = 0.79, p = 0.5$ ; 2 Hz:  $F = 2.0, p = 0.12$ ; Fig. 1D).

During the experiment, we presented the different levels of manipulation in a block design (Fig. 2A) in which each level of jitter was present for 1 min and followed by another level in a pseudorandom sequence. Given the statistical nature of the manipulation, the transition between conditions was perceptually

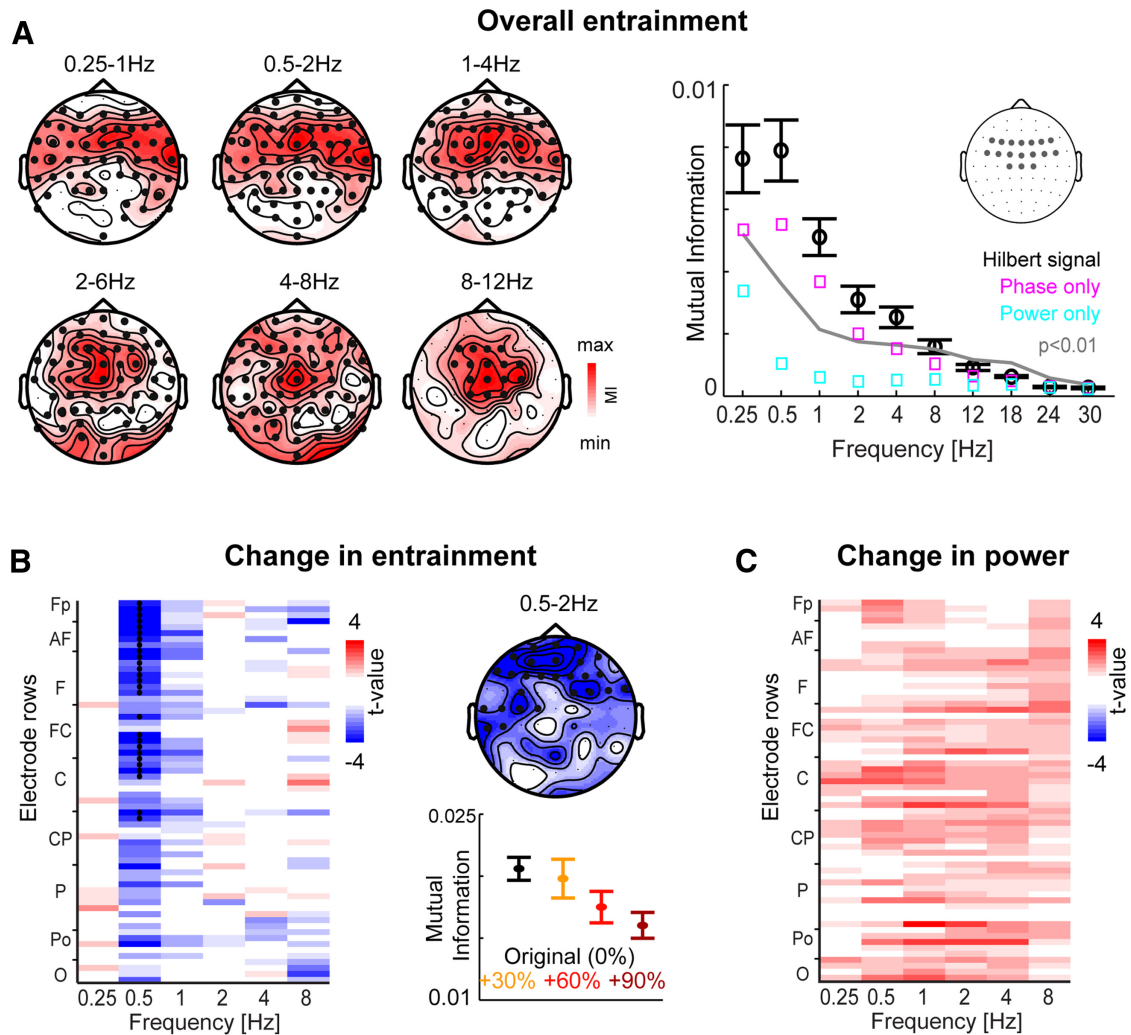
continuous rather than discrete. However, because we were not interested in the perceived rhythmicity of the speech, but rather the impact of the statistical regularity on brain activity, we pooled data from the full 1 min segments for analysis.

**Behavioral results**

The manipulation imposed on the local speech rate affected the timing of individual words or syllables, but did not distort the acoustic structure of these. As a result, it did not affect speech intelligibility. The behavioral reports obtained after each block confirmed that, across participants ( $n = 16$ ) words were identified equally well across conditions: there was no significant effect of condition on recognition rates (one-way ANOVA,  $F_{(3,45)} = 0.56, p = 0.64$ ; Fig. 2B).

**Cortical signatures of auditory entrainment**

The entrainment of rhythmic auditory activity to the speech envelope can be measured by quantifying the consistency of the relative timing between brain activity and the envelope, for example, by measures of the relative phase locking between changes in both signals (Luo and Poeppel, 2007; Peelle and Davis, 2012; Gross, 2014). One approach that has proven to be versatile with respect to the neural signals of interest and that is robust to data outliers is the MI between brain activity and sound envelope (Belitski et al., 2010; Cogan and Poeppel, 2011; Gross et al., 2013). Following this approach, we separated the EEG data into band-limited signals between 0.25 and 48 Hz and calculated the MI between the Hilbert representations of the speech signal and of the EEG activity separately for each band. We first performed this analysis across the full 6 min text samples to quantify the overall acoustic information carried by different electrodes and bands. Based on group-level randomization statistics controlling for multiple comparisons, we found significant ( $p < 0.01$ ) MI over central and frontal electrodes at frequencies <12 Hz. No significant MI was found for any of the bands starting at 12 Hz or above (Fig. 3A). The highest MI occurred in the two delta bands: 0.25–1 and 0.5–2 Hz. The topographies for these bands reveal a slight dominance of the right hemisphere, but MI values were significant for a large cluster of central and frontal electrodes. Additional analysis revealed that the MI between EEG activity and the speech envelope was carried mostly by the phase and not the power of both signals (Fig. 3A, right); computing MI for power



**Figure 3.** Entrainment of EEG activity to the speech envelope. **A**, Overall entrainment. For each frequency band, we quantified the MI between band-limited EEG activity and the speech amplitude envelope in the same band across the full eight text samples. Left, Topographies displaying the MI values averaged across participants ( $n = 16$ ) for all bands containing significant clusters (MI calculated using the Hilbert representation of each signal;  $p < 0.01$  permutation statistics; dark dots). Right, MI within a region of interest (frontocentral electrodes; see inset) across participants together with the 99% confidence interval (permutation statistics; red). Black, MI obtained from the Hilbert representation of each signal; cyan, MI obtained using power; magenta, MI obtained using phase (see also Materials and Methods for details). **B**, Reduction of MI with manipulated speech rate. Left, Group-level statistics for systematic increases or decreases in MI (based on the Hilbert representation) across conditions, frequency bands, and electrodes (sorted according to their label). Right top, Delta band topography (0.5–2 Hz). Dark dots are cluster of electrodes with a significant decrease in MI with increasing jitter in speech rate ( $p < 0.01$ ; permutation statistics). Right bottom, MI values for the delta cluster across participants (mean and SEM). **C**, Group-level statistic for changes in band limited power of EEG activity across conditions. There was no significant effect (at  $p < 0.05$ ).

only revealed no significant clusters, whereas the MI for phase was highly significant, an observation that is highly consistent with previous studies (Gross et al., 2013; Ng et al., 2013).

### Speech rate manipulation reduces entrainment

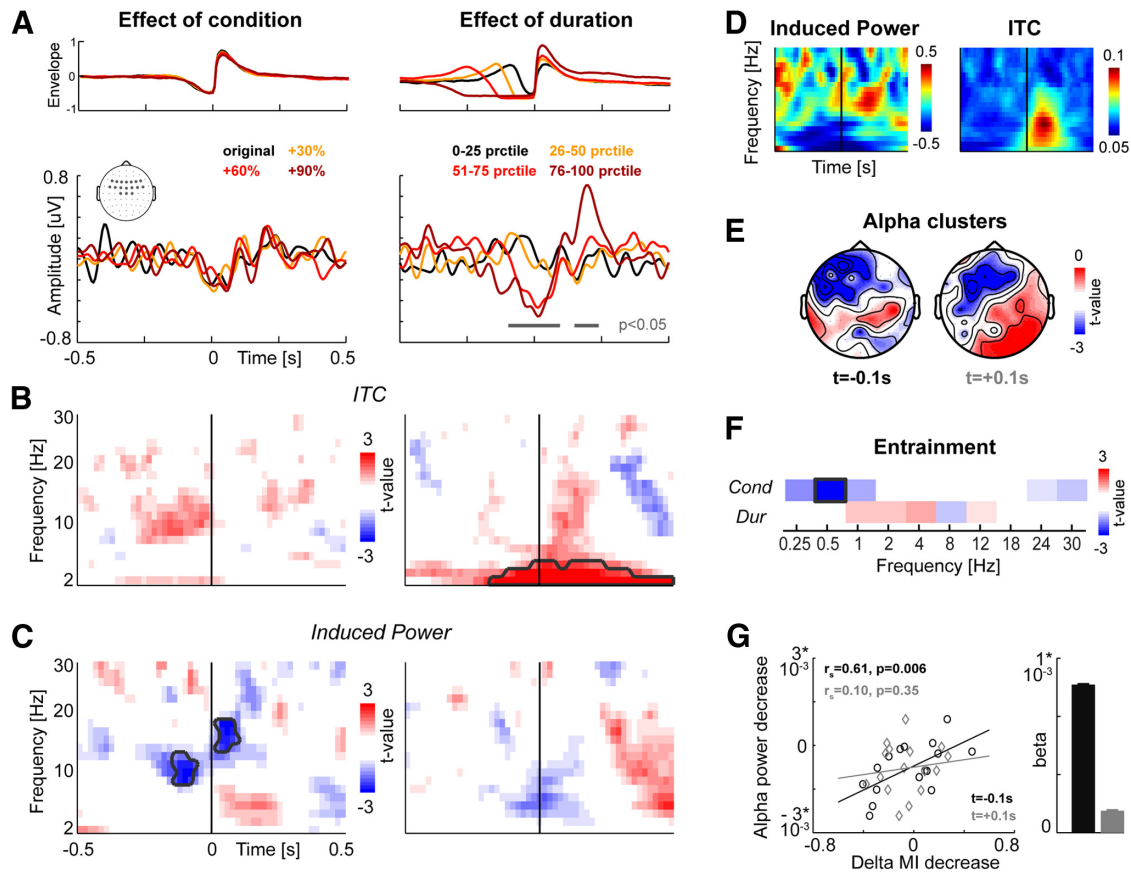
We then calculated MI separately for each sub-block (Fig. 3B). Using regression statistics on single-subject data, we tested whether and for which electrodes or frequency bands entrainment significantly and systematically changed across conditions. Group-level statistics revealed a significant cluster of electrodes for which MI decreased with increasing jitter in the delta band (0.5–2 Hz;  $p = 0.0025$ ;  $T_{\text{sum}} = 120$ ,  $r = 0.67$ ), but not in any other band. The delta cluster was concentrated over frontal and temporal electrodes (Fig. 3B).

We ruled out that this decrease in entrainment was simply the result of an overall decrease in the power of oscillatory EEG activity (but see below for local changes in power). For each condition, we quantified the time-averaged power for those frequency

bands with significant entrainment (i.e., between 0.25 and 8 Hz). Group-level regression statistics revealed no cluster in which power changed significantly across conditions ( $p < 0.05$ ; see Fig. 3C for  $t$ -maps).

### Speech rate manipulation does not alter evoked responses

The reduction in delta entrainment with increasingly less predictable speech rate could reflect two processes. It could indicate a reduced fidelity with which slow rhythmic activity tracks the regularity of the acoustic input in the absence of changes in the encoding of individual sound tokens by time-localized brain activity. Conversely, it could primarily reflect such changes in auditory evoked responses to individual sound tokens, which are then reflected in reduced entrainment (Ding and Simon, 2014). To disentangle these possibilities, we quantified transient changes in brain activity around the pauses and the subsequent syllable onset. We first focus on responses time-locked to syllable onsets, quantified by evoked potentials and the intertrial coher-



**Figure 4.** Evoked potentials and changes in oscillatory activity around pauses and syllable onsets. **A**, Evoked potentials for each condition (left) and for pauses grouped by duration (right). The top panel displays the epoch-averaged speech amplitude envelope (in peak-normalized units); the bottom panel the epoch- and participant-averaged evoked potentials for frontocentral electrodes (see inset). Gray lines indicate epochs with a significant group-level effect ( $p < 0.05$ ). Time 0 corresponds to the detected syllable onset following a pause. **B**, TF maps showing the group-level statistics for a systematic effect of condition (or duration) on ITC within the electrode region of interest. Gray lines indicate TF clusters with a significant group-level effect ( $p < 0.05$ ). **C**, Same as in **B** but for induced oscillatory power. **D**, Average TF representations for induced power and ITC within the region of interest averaged across all epochs and participants. Power was z-scored within each frequency band and participant. **E**, Topographies of the group-level effect for a systematic effect of condition of induced alpha power separately for each of the two alpha clusters (at  $-0.1$  and  $+0.1$  s) revealed in **C**. **F**, Group-level statistics for a systematic effect of condition (duration) on entrainment for each frequency band; MI was calculated using only time epochs around articulation onset ( $-0.1$  to  $0.4$  s). **G**, Regression of delta band MI on alpha power. Left, Correlation between changes in alpha power (regression betas vs jitter level) versus changes in delta MI across participants separately for each cluster. Results of separate Spearman’s rank correlations are indicated. Right, Beta values (mean, SEM across participants) for each alpha cluster obtained from a joined regression model of delta MI on both alpha clusters. The contribution of the early ( $-0.1$  s) cluster was significant, but that of the later ( $+0.1$  s) cluster was not (see main text).

ence of oscillatory activity. We restricted these analyses to a region of interest of frontocentral electrodes where we observed the strongest entrainment (cf. Fig. 4A, inset).

If changes in the regularity of speech rate were to affect the encoding of subsequent syllables, then we would expect to find changes in evoked responses at syllable onset. As a control analysis, we compared evoked activity between onsets selected based on the length of the preceding pause. Based on the adaptive mechanisms of auditory cortex, one would expect to find significantly reduced activity during longer pauses and significantly stronger responses during articulation after longer pauses (Fishman, 2014; Pérez-González and Malmierca, 2014).

Group-level regression statistics for an effect of condition revealed no cluster with a significant change in evoked potentials across conditions ( $p < 0.05$ ; Fig. 4A, left). In contrast, we found significant changes in evoked potentials with pause duration, consisting of a reduction of evoked activity during longer pauses ( $-0.1$  to  $0.08$  s;  $p = 0.0025$ ,  $T_{\text{sum}} = 2700$ ,  $r = 0.68$ ) and an increase in response during stimulation after longer pauses ( $0.16$  to  $0.24$  s;  $p = 0.005$ ,  $T_{\text{sum}} = 809$ ,  $r = 0.71$ ). To ensure that the null result for condition was not obscured by a potential interaction

between an effect of condition and pause duration, we subjected the evoked responses in these 2 time windows to a  $2 \times 2$  ANOVA. This replicated a main effect of duration ( $-0.1$  to  $0.08$  s:  $F_{(1,60)} = 8.91$ ,  $p = 0.004$ ;  $0.16$  to  $0.24$  s:  $F = 7.68$ ,  $p = 0.007$ ) and a null effect of condition ( $F = 0.2$ ,  $p = 0.64$  and  $F = 0.03$ ,  $p = 0.86$ ), but did not reveal an interaction ( $F = 0.03$ ,  $p = 0.87$  and  $F = 0.12$ ,  $p = 0.73$ ).

Group-level regression statistics on the ITC of oscillatory activity revealed no cluster with a significant change in ITC across conditions ( $p < 0.05$ ; Fig. 4B, left). Again, there was a significant effect of duration (Fig. 4B, right) consisting of an increase in ITC at frequencies  $< 4$  Hz with increasing pause duration ( $-0.1$  to  $0.5$  s;  $2-4$  Hz;  $p = 0.002$ ,  $T_{\text{sum}} = 4566$ ,  $r = 0.75$ ). Again, there was no interaction of condition and duration (effect of duration:  $F_{(1,60)} = 12.0$ ,  $p = 0.001$ ; condition  $F = 0.1$ ,  $p = 0.75$ ; interaction:  $F = 0.42$ ,  $p = 0.52$ ). For illustration, Figure 4D displays the subject- and epoch-averaged ITC TF distribution.

**Speech rate manipulation reduces frontal alpha power**

We then quantified changes in induced oscillatory power. We subtracted the trial-averaged response before computing TF rep-



representations of power (Griffiths et al., 2010). For illustration, Figure 4D displays the subject- and epoch-averaged power for the frontocentral region of interest. Group-level regression statistics revealed significant effects of condition on induced power: we found two TF clusters in the alpha band exhibiting a reduction in power with increasing jitter (Fig. 4C, left). One cluster was found during the pause preceding syllable onset ( $-0.16$  to  $-0.06$  s,  $8$ – $12$  Hz,  $p = 0.002$ ,  $T_{\text{sum}} = 415$ ,  $r = 0.67$ ) and one cluster during syllable onset ( $0.02$  to  $0.12$  s,  $12$ – $15$  Hz,  $p = 0.004$ ,  $T_{\text{sum}} = 303$ ,  $r = 0.68$ ). In contrast, there was no significant effect of duration on induced oscillatory power (Fig. 4C, right). There was also no interaction between condition and duration (tested for the combined alpha clusters;  $F_{(1,60)} = 0.67$ ,  $p = 0.40$ ). To illustrate the scalp distribution of the changes in alpha power, Figure 4E displays the topographies of the group-level statistics for each alpha cluster. Both clusters are centered over left frontal regions.

### Alpha power reduction correlates with reduced delta entrainment

We then investigated whether the reduction in alpha power with increasing jitter in speech rate correlates with the reduction of auditory delta entrainment. We first verified that the decrease in entrainment reported above for the entire text epochs was also present locally within the epochs around syllable onset. Restricting the MI analysis to epochs around syllable onsets ( $-0.1$  to  $0.4$  s), group-level regression statistics confirmed a significant reduction in MI for the delta band ( $0.5$ – $2$  Hz;  $p = 0.002$ ,  $T_{\text{sum}} = 45$ ,  $r = 0.67$ ; Fig. 4F) and revealed no significant effect for any other band ( $p < 0.05$ ). In addition, there was no significant change in MI with duration (group-level statistics at  $p < 0.05$ ). We then quantified the predictive value of the power in both alpha clusters on delta entrainment across participants. We first compared the slope (regression beta vs condition label) of alpha changes to the slope of entrainment changes across participants (Fig. 4F, left). This revealed a significant Spearman correlation for the early alpha cluster ( $-0.16$  to  $-0.06$  s;  $r_s = 0.61$ ,  $p = 0.006$ ,  $p = 0.012$  when Bonferroni corrected) but not for the later cluster ( $0.02$  to  $0.12$  s;  $r_s = 0.10$ ,  $p = 0.35$ ). Further, entering both clusters into a single joined regression of delta MI on alpha power revealed a significant contribution of the early cluster ( $t_{(15)} = 2.4$ ,  $p = 0.03$ ; Fig. 4F, right), but not of the later cluster ( $t_{(15)} = 0.38$ ,  $p = 0.70$ ). Therefore, the reduction in frontal alpha power during pauses is significantly related to the reduction in delta entrainment to the speech envelope during the following syllable onset.

## Discussion

We manipulated the rhythmic pattern in speech imposed by the alternation of pauses and syllables. This pattern sets the time scale of sound envelope fluctuations, provides a temporal reference for expectation, and is critical for comprehension (Ghitza and Greenberg, 2009; Giraud and Poeppel, 2012; Peelle and Davis, 2012; Hickok et al., 2015). We reduced the predictiveness of this pattern by manipulating the local speech rate while maintaining the overall rate, the power of the sound envelope, and intelligibility. This provided several results: (1) a dissociation between delta band ( $0.5$ – $2$  Hz) entrainment to the sound envelope, which were reduced, and the encoding of sound transients by evoked responses, which was preserved; (2) a dissociation between entrainment at delta (reduced) and higher frequencies (preserved); and (3) a correlation between left frontal alpha power and subsequent delta entrainment. These results foster the notion that delta

entrainment reflects processes that are under top-down control rather than reflecting the early encoding of acoustic features.

### Auditory entrainment as bottom-up reflection of acoustic features

Rhythmic sounds induce a series of transient auditory cortical activity that is time locked with the relevant acoustic features. *In vivo* recordings demonstrated correlations between different neural signatures of auditory encoding, such as population spiking and rhythmic network activity within auditory cortex or between auditory spiking and human EEG (Kayser et al., 2009; Szymanski et al., 2011; Ng et al., 2013; Kayser et al., 2015). It has been suggested that auditory entrainment may to a large extent reflect the recurring series of transient evoked responses in auditory cortex (Howard and Poeppel, 2010; Szymanski et al., 2011; Doelling et al., 2014). This hypothesis is also supported by the observation that entrainment is strongest around sound envelope transients (Gross et al., 2013) and is reduced when the speech envelope is artificially flattened (Ghitza, 2011; Doelling et al., 2014).

However, the notion of entrainment being a bottom-up-driven process has been challenged based on changes in entrainment with expectations, task relevance, or attention (Lakatos et al., 2008; Peelle and Davis, 2012; Zion Golumbic et al., 2013; Arnal et al., 2015; Hickok et al., 2015). Consistent with a view that top-down mechanisms control entrainment, we demonstrate a direct dissociation of early evoked responses reflecting the encoding of sound transients in auditory cortices and delta entrainment. We observed reduced entrainment in the absence of significant changes in delta power or the delta fluctuations in the speech envelope. Therefore, our results are best explained by a reduced temporal fidelity with which high-level processes track the speech envelope in the absence of changes in early auditory responses. Although this interpretation resonates well with other data favoring a top-down interpretation (see below), we note that we cannot rule out bottom-up contributions of speech rhythm to the observed changes in entrainment because the time scales of entrained activity and our experimental manipulation partly overlap.

The observed dissociation of delta entrainment and evoked responses is also consistent with recent results disentangling the functional roles of auditory network activity at different time scales. By modeling the sensory transfer function of auditory cortex neurons relative to the state of rhythmic field potentials, we suggested that the sensory gain of auditory neurons is linked more to frequencies  $>6$  Hz, whereas the delta rhythms index changes in stimulus-unrelated spiking (Kayser et al., 2015). Assuming that auditory evoked potential reflects activity within auditory cortex (Verkindt et al., 1995), these previous results directly predict a dissociation of delta entrainment and evoked potentials as observed here.

The absence of changes in evoked potentials with increasingly irregular speech rate agrees with other work on the impact of sentence structure on evoked potentials. Changes in evoked activity with rhythmic primes or changes in speech accent were found mostly later than 300 ms after syllable onset (Cason and Schön, 2012; Goslin et al., 2012; Roncaglia-Denissen et al., 2013), consistent with higher level processes relating to lexical integration (Haupt et al., 2008; Chennu et al., 2013). Our finding of reduced left frontal alpha power with decreasing speech regularity is consistent with such a hypothesis. Further, the absence of changes in evoked potentials with condition also unlikely results from a lack of statistical sensitivity because we observed a signifi-



icant effect of pause duration. The latter effect may reflect signs of expectancy or adaptation of auditory processes, contributions that are difficult to dissociate with the present data (Todorovic and de Lange, 2012; Fishman, 2014; Pérez-González and Malmierca, 2014).

### Multiple time scales of auditory entrainment

The rhythmic syllable pattern is important for speech segmentation (Rosen, 1992; Ghitza and Greenberg, 2009; Geiser and Shattuck-Hufnagel, 2012). For example, phonemes placed at expected times or presented in concordance with a prominent beat are detected more efficiently (Meltzer et al., 1976; Cason and Schön, 2012). Our manipulation mostly concerned pauses of about 250 ms or longer and thus affected speech regularity at the scale corresponding to delta and theta frequencies. That speech intelligibility and theta entrainment were preserved while delta entrainment was reduced suggests functional differences between the entrainment at different time scales. Although auditory entrainment per se has been reported for essentially all frequencies between 0.5 and 10 Hz (Gross et al., 2013; Ng et al., 2013; Ding and Simon, 2014), there is good evidence to support a dissociation between individual frequencies. For example, acoustic manipulations such as background noise or noise vocoding affect theta entrainment and speech intelligibility in similar ways (Ding and Simon, 2013; Ding et al., 2013; Peelle et al., 2013), whereas intelligibility across participants tends to correlate with delta entrainment (Doelling et al., 2014; Ding and Simon, 2014). One other previous study also found a dissociation of delta and theta entrainment (Ding et al., 2013). The previous evidence thus suggests that theta entrainment may more directly reflect the encoding or parsing of acoustic features to guide speech segmentation, whereas delta entrainment reflects perceived qualities of speech such as irregular speech rate or top-down control over auditory cortex.

### Top-down control of entrainment

We suggest that the reduction of delta entrainment is induced by top-down processes that are sensitive to acoustic regularities and align rhythmic auditory activity to specific points in time (Schroeder and Lakatos, 2009; Lakatos et al., 2013; Hickok et al., 2015). Our results pinpoint the left frontal alpha activity as one key player in this top-down control over auditory entrainment.

Consistent with our hypothesis, a recent study on functional connectivity reported direct top-down influences on auditory entrainment during speech processing that were stronger for delta compared with theta entrainment (Park et al., 2015). This study suggested that left inferior frontal regions modulate entrainment over auditory cortex, which is consistent with the anatomical connectivity between frontal and temporal cortices (Hackett et al., 1999; Binder et al., 2004; Saur et al., 2008) and increases in frontal activation during the processing of degraded speech (Davis and Johnsrude, 2007; Hervais-Adelman et al., 2012). Our results further show that delta entrainment correlates directly with left frontal alpha activity, in particular with changes in alpha power before the reduction of delta entrainment. Although this correlation does not imply a causal relation, the fact that alpha before syllable onset had a stronger correlation with entrainment than alpha during articulation is at least positive evidence. In addition, a recent study demonstrated the entrainment of perception to speech rhythm in the absence of fluctuations in sound amplitude or spectral content, suggesting a linguistic driver of entrainment (Zoefel and VanRullen, 2015).

Frontal alpha activity has been implied in inhibitory processes

and the disengagement of task-relevant regions (Klimesch, 1999; Jensen and Mazaheri, 2010). Decreases in alpha power occur in response to increased attention, memory retrieval, and other top-down regulatory processes (Dockree et al., 2004; Hwang et al., 2005) and index increased engagement of the respective regions. Left prefrontal regions such as the inferior frontal gyrus (IFG) are implicated in verbal tasks such as semantic selection and interference resolution during memory (D'Esposito et al., 2000; Thompson-Schill et al., 2002; Swick et al., 2008) and their activity has been shown directly to correlate negatively with frontal alpha (Goldman et al., 2002). In addition, activity in the alpha band may be directly involved in top-down functional connectivity, as shown in the visual (Bastos et al., 2015) and auditory systems (Fontolan et al., 2014). Therefore, the finding of reduced left frontal alpha power with increasingly irregular speech rate is consistent with an increasing activation of the left IFG, which then influences auditory delta entrainment directly via top-down connectivity (Park et al., 2015).

### References

- Arnal LH, Doelling KB, Poeppel D (2015) Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb Cortex* 25:3077–3085. CrossRef Medline
- Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85:390–401. CrossRef Medline
- Belitski A, Panzeri S, Magri C, Logothetis NK, Kayser C (2010) Sensory information in local field potentials and spikes from visual and auditory cortices: time scales and frequency bands. *J Comput Neurosci* 29:533–545. CrossRef Medline
- Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE (2011) Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci* 31:3176–3185. CrossRef Medline
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7:295–301. CrossRef Medline
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436. CrossRef Medline
- Cason N, Schön D (2012) Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* 50:2652–2658. CrossRef Medline
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5:e1000436. CrossRef Medline
- Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. *J Neurosci* 33:11194–11205. CrossRef Medline
- Cogan GB, Poeppel D (2011) A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *J Neurophysiol* 106:554–563. CrossRef Medline
- Cover TM, Thomas JA (1991) Elements of information theory. New York: Wiley-Interscience.
- Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear Res* 229:132–147. CrossRef Medline
- Debener S, Thorne JD, Schneider TR, Viola FC (2010) Using ICA for the analysis of multi-channel EEG data. In: Simultaneous EEG and fMRI: recording, analysis, and application (Ullsperger M, Debener S, eds), pp 121–134. Oxford: OUP.
- D'Esposito M, Postle BR, Rypma B (2000) Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Exp Brain Res* 133:3–11. CrossRef Medline
- Dilley LC, Pitt MA (2010) Altering context speech rate can cause words to appear or disappear. *Psychol Sci* 21:1664–1670. CrossRef Medline
- Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci U S A* 109:11854–11859. CrossRef Medline
- Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a

- background-insensitive cortical representation of speech. *J Neurosci* 33: 5728–5735. [CrossRef Medline](#)
- Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. *Front Hum Neurosci* 8:311. [Medline](#)
- Ding N, Chatterjee M, Simon JZ (2013) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88C:41–46. [Medline](#)
- Dockree PM, Kelly SP, Roche RA, Hogan MJ, Reilly RB, Robertson IH (2004) Behavioural and physiological impairments of sustained attention after traumatic brain injury. *Brain Res Cogn Brain Res* 20:403–414. [CrossRef Medline](#)
- 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. [Medline](#)
- Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. *PLoS Comput Biol* 5:e1000302. [CrossRef Medline](#)
- Fishman YI (2014) The mechanisms and meaning of the mismatch negativity. *Brain Topogr* 27:500–526. [CrossRef Medline](#)
- Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud AL (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat Commun* 5:4694. [CrossRef Medline](#)
- Geiser E, Shattuck-Hufnagel S (2012) Temporal regularity in speech perception: Is regularity beneficial or deleterious? *Proc Meet Acoust* 14:pii: 060004. [Medline](#)
- Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front Psychol* 2:130. [Medline](#)
- Ghitza O, Greenberg S (2009) On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66:113–126. [CrossRef Medline](#)
- Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15: 511–517. [CrossRef Medline](#)
- Goldman RI, Stern JM, Engel J Jr, Cohen MS (2002) Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport* 13:2487–2492. [CrossRef Medline](#)
- Goslin J, Duffy H, Floccia C (2012) An ERP investigation of regional and foreign accent processing. *Brain Lang* 122:92–102. [CrossRef Medline](#)
- Griffiths TD, Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Brugge JF, Howard MA (2010) Direct recordings of pitch responses from human auditory cortex. *Curr Biol* 20:1128–1132. [CrossRef Medline](#)
- Gross J (2014) Analytical methods and experimental approaches for electrophysiological studies of brain oscillations. *J Neurosci Methods* 228: 57–66. [CrossRef Medline](#)
- Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol* 11:e1001752. [CrossRef Medline](#)
- Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817:45–58. [CrossRef Medline](#)
- Haupt FS, Schlesewsky M, Roehm D, Friederici AD, Bornkessel-Schlesewsky I (2008) The status of subject-object reanalyses in the language comprehension architecture. *Journal of Memory and Language* 59:54–96. [CrossRef](#)
- Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci U S A* 109:20095–20100. [CrossRef Medline](#)
- Hervais-Adelman AG, Carlyon RP, Johnsrude IS, Davis MH (2012) Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes* 27:1145–1166. [CrossRef](#)
- Hickok G, Farahbod H, Saberi K (2015) The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychol Sci* 26:1006–1013. [CrossRef Medline](#)
- Hipp JF, Siegel M (2013) Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Front Hum Neurosci* 7:338. [Medline](#)
- Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophysiol* 104:2500–2511. [CrossRef Medline](#)
- Hwang G, Jacobs J, Geller A, Danker J, Sekuler R, Kahana MJ (2005) EEG correlates of verbal and nonverbal working memory. *Behav Brain Funct* 1:20. [CrossRef Medline](#)
- Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Hum Neurosci* 4:186. [Medline](#)
- Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61:597–608. [CrossRef Medline](#)
- Kayser C, Logothetis NK, Panzeri S (2010) Visual enhancement of the information representation in auditory cortex. *Curr Biol* 20:19–24. [CrossRef Medline](#)
- Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S (2015) Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. *J Neurosci* 35:7750–7762. [CrossRef Medline](#)
- Keren AS, Yuval-Greenberg S, Deouell LY (2010) Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. *Neuroimage* 49:2248–2263. [CrossRef Medline](#)
- Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* 29:169–195. [CrossRef Medline](#)
- Kumar P (2012) Statistical dependence: copula functions and mutual information based measures. *J Stat Appl Pro* 1:1–14. [CrossRef](#)
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94:1904–1911. [CrossRef Medline](#)
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320:110–113. [CrossRef Medline](#)
- Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64:419–430. [CrossRef Medline](#)
- Lakatos P, Musacchia G, O'Connell MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77:750–761. [CrossRef Medline](#)
- Leong V, Goswami U (2014) Assessment of rhythmic entrainment at multiple timescales in dyslexia: evidence for disruption to syllable timing. *Hear Res* 308:141–161. [CrossRef Medline](#)
- Loukina A, Kochanski G, Rosner B, Keane E, Shih C (2011) Rhythm measures and dimensions of durational variation in speech. *J Acoust Soc Am* 129:3258–3270. [CrossRef Medline](#)
- Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010. [CrossRef Medline](#)
- Ma J, Sun Z (2008) Mutual information is copula entropy. Available at: <http://arxiv.org/abs/0808.0845>
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190. [CrossRef Medline](#)
- Meltzer RH, Martin JG, Mills CB, Imhoff DL, Zohar D (1976) Reaction time to temporally-displaced phoneme targets in continuous speech. *J Exp Psychol Hum Percept Perform* 2:277–290. [CrossRef Medline](#)
- Millman RE, Prendergast G, Hymers M, Green GG (2013) Representations of the temporal envelope of sounds in human auditory cortex: can the results from invasive intracortical “depth” electrode recordings be replicated using non-invasive MEG “virtual electrodes”? *Neuroimage* 64:185–196. [CrossRef Medline](#)
- Montemurro MA, Senatore R, Panzeri S (2007) Tight data-robust bounds to mutual information combining shuffling and model selection techniques. *Neural Comput* 19:2913–2957. [CrossRef Medline](#)
- Morillon B, Schroeder CE, Wyart V (2014) Motor contributions to the temporal precision of auditory attention. *Nat Commun* 5:5255. [CrossRef Medline](#)
- Ng BS, Logothetis NK, Kayser C (2013) EEG phase patterns reflect the selectivity of neural firing. *Cereb Cortex* 23:389–398. [CrossRef Medline](#)
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15:1–25. [CrossRef Medline](#)
- O'Beirne GA, Patuzzi RB (1999) Basic properties of the sound-evoked post-auricular muscle response (PAMR). *Hear Res* 138:115–132. [CrossRef Medline](#)
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869. [Medline](#)
- Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for

- the sampling bias problem in spike train information measures. *J Neurophysiol* 98:1064–1072. [CrossRef Medline](#)
- Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. *Trends Neurosci* 33:111–120. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Peelle JE, Davis MH (2012) Neural oscillations carry speech rhythm through to comprehension. *Front Psychol* 3:320. [Medline](#)
- 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. [CrossRef Medline](#)
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442. [CrossRef Medline](#)
- Pérez-González D, Malmierca MS (2014) Adaptation in the auditory system: an overview. *Front Integr Neurosci* 8:19. [Medline](#)
- Quiñero Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci* 10:173–185. [Medline](#)
- Roncaglia-Denissen MP, Schmidt-Kassow M, Kotz SA (2013) Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS One* 8:e56000. [CrossRef Medline](#)
- Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* 336:367–373. [CrossRef Medline](#)
- Rosenthal R, Rubin DB (2003) *r*(equivalent): a simple effect size indicator. *Psychol Methods* 8:492–496. [CrossRef Medline](#)
- Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry MS, Umarova R, Musso M, Glauche V, Abel S, Huber W, Rijntjes M, Hennig J, Weiller C (2008) Ventral and dorsal pathways for language. *Proc Natl Acad Sci U S A* 105:18035–18040. [CrossRef Medline](#)
- Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32:9–18. [CrossRef Medline](#)
- Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I (2010) Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J Neurosci* 30:13578–13585. [CrossRef Medline](#)
- Strauß A, Henry MJ, Scharinger M, Obleser J (2015) Alpha phase determines successful lexical decision in noise. *J Neurosci* 35:3256–3262. [CrossRef Medline](#)
- Swick D, Ashley V, Turken AU (2008) Left inferior frontal gyrus is critical for response inhibition. *BMC Neurosci* 9:102. [CrossRef Medline](#)
- Szymanski FD, Rabinowitz NC, Magri C, Panzeri S, Schnupp JW (2011) The laminar and temporal structure of stimulus information in the phase of field potentials of auditory cortex. *J Neurosci* 31:15787–15801. [CrossRef Medline](#)
- Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3:151–162. [CrossRef Medline](#)
- Tauroza S, Allison D (1990) Speech rates in British English. *Applied Linguistics* 11:90–105. [CrossRef](#)
- Thompson-Schill SL, Jonides J, Marshuetz C, Smith EE, D'Esposito M, Kan IP, Knight RT, Swick D (2002) Effects of frontal lobe damage on interference effects in working memory. *Cogn Affect Behav Neurosci* 2:109–120. [CrossRef Medline](#)
- Todorovic A, de Lange FP (2012) Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci* 32:13389–13395. [CrossRef Medline](#)
- Verkindt C, Bertrand O, Perrin F, Echallier JF, Pernier J (1995) Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalogr Clin Neurophysiol* 96:143–156. [CrossRef Medline](#)
- Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J (2015a) Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cereb Cortex* 25:1938–1946. [Medline](#)
- Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J (2015b) Slow-delta phase concentration marks improved temporal expectations based on the passage of time. *Psychophysiology* 52:910–918. [CrossRef Medline](#)
- Zellner B (1994) Pauses and the Temporal Structure of Speech. In: *Fundamentals of speech synthesis and speech recognition* (Keller E, ed), pp 41–62. New York: Wiley.
- Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron* 77:980–991. [CrossRef Medline](#)
- Zoefel B, VanRullen R (2015) Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J Neurosci* 35:1954–1964. [CrossRef Medline](#)