Effects of Stimulus Rate and Periodicity on Auditory Cortical Entrainment to Continuous Sounds

Sara Momtaz[1,2] and Gavin M. Bidelman[3,4]

[1]School of Communication Sciences & Disorders, University of Memphis, Memphis, Tennessee 38152, [2]Boys Town National Research Hospital, Boys Town, Nebraska 68131, [3]Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, Indiana 47408, and [4]Program in Neuroscience, Indiana University, Bloomington, Indiana 47405

Abstract

The neural mechanisms underlying the exogenous coding and neural entrainment to repetitive auditory stimuli have seen a recent surge of interest. However, few studies have characterized how parametric changes in stimulus presentation alter entrained responses. We examined the degree to which the brain entrains to repeated speech (i.e., /ba/) and nonspeech (i.e., click) sounds using phase-locking value (PLV) analysis applied to multichannel human electroencephalogram (EEG) data. Passive cortico-acoustic tracking was investigated in $N = 24$ normal young adults utilizing EEG source analyses that isolated neural activity stemming from both auditory temporal cortices. We parametrically manipulated the rate and periodicity of repetitive, continuous speech and click stimuli to investigate how speed and jitter in ongoing sound streams affect oscillatory entrainment. Neuronal synchronization to speech was enhanced at 4.5 Hz (the putative universal rate of speech) and showed a differential pattern to that of clicks, particularly at higher rates. PLV to speech decreased with increasing jitter but remained superior to clicks. Surprisingly, PLV entrainment to clicks was invariant to periodicity manipulations. Our findings provide evidence that the brain's neural entrainment to complex sounds is enhanced and more sensitized when processing speech-like stimuli, even at the syllable level, relative to nonspeech sounds. The fact that this specialization is apparent even under passive listening suggests a priority of the auditory system for synchronizing to behaviorally relevant signals.

Key words: auditory evoked potentials; auditory neural oscillations; periodicity coding; rhythmicity; time-frequency processing

Significance Statement

To examine the effects of stimulus factors on auditory cortical entrainment, we compared cortico-acoustic tracking for speech versus click stimuli across various speeds (below, at, and above the nominal speech rate) and acoustic periodicity (jitter in 20% steps). Overall, our results demonstrate that brain responses are more sensitive to changes in rhythm and periodicity of speech even during passive listening. The prioritization of speech in the brain might be partially due to the increased neuronal entrainment it elicits.

Introduction

Temporal processing is a crucial component of audition (Picton, 2013) that influences all levels of auditory skills ranging from sensory levels to higher cognitive processing including attention and memory (Toplak et al., 2006; Allman and Meck, 2011). Indeed, the ability to time synchronize (i.e., entrain) to ongoing sound stimuli might represent a fundamental coding mechanism to support a variety of perceptual–cognitive processes including predictive coding and speech segmentation (Arnal and Giraud, 2012). In the broadest sense,

entrainment refers to the synchronization (i.e., phase coupling) between two signals (Cummins, 2009). Such rhythmic fluctuations in terms of brain-to-stimulus coupling are characterized by excitation–inhibition cycles of neuronal populations termed "neuronal oscillations" (Bishop, 1932; Buzsaki and Draguhn, 2004; Lakatos et al., 2005; Buzsaki, 2006). In the context of speech, oscillatory neural entrainment is the process by which auditory cortical activity precisely tracks and adjusts to modulations in the speech amplitude envelope via phase alignment between neural and acoustic signals (Pikovsky et al., 2002; Lakatos et al., 2019). Entrainment is one of several important functions in auditory processing that can make communication seem effortless and automatic to healthy listeners and conversely, difficult in individuals with language learning disorders (Momtaz et al., 2021, 2022).

Entrainment applies to a wide range of physiologically important behaviors including auditory–motor coupling (Fujii and Schlaug, 2013; Herbst and Landau, 2016; Nozaradan et al., 2016; Moumdjian et al., 2018; Rosso et al., 2021) as well as speech (Peelle and Davis, 2012) and music (Fiveash et al., 2021) perception. For example, listeners must segment the continuous input speech signal into proper discrete units (i.e., syllables), which are then used as input for future decoding stages by auditory and nonauditory brain regions. In speech, the amplitude envelope's rhythmic information reflects different aspects of sensory and motor processing such as segmentation, speech rate, and articulation place and manner (Peelle and Davis, 2012). Moreover, the perceptual system recovers these rhythmic structures in speech, which are important for spoken language comprehension (Poeppel and Assaneo, 2020). Sensory processing presumably benefits from neural entrainment as it could provide a temporal prediction mechanism to anticipate future auditory events before they arrive at the ear (Lakatos et al., 2013).

On the contrary, auditory signals like speech might challenge an entrained system given its quasi-rhythmic temporal structure (Levelt, 1993; Guenther, 2016) that fluctuates with a speaker's talking rate. Still, speech production imparts temporal regularity to the signal envelope that is, on average, remarkably consistent in speed across the world's languages (Poeppel and Assaneo, 2020). Indeed, the temporal syllabic rate of speech across various languages ranges from 2 to 8 Hz with a "characteristic" periodicity of 4.5 Hz (Poeppel and Assaneo, 2020). It has been hypothesized that this range of rhythmicity is prioritized for speech perception and production (Assaneo and Poeppel, 2018). Moreover, speech intelligibility is optimal when the rhythmic structure of the signal falls inside a syllabic rate of 4–8 Hz (Peelle et al., 2013; Doelling et al., 2014). Thus, in addition to various speeds of representation, the quasi-periodic nature of speech effectively produces jitter which presumably also impacts entrainment and subsequent auditory processing.

Prior work has not fully elucidated how and to what extent aperiodicity might affect auditory processing (Doelling and Poeppel, 2015; Breska and Deouell, 2017; Novembre and Iannetti, 2018). Some studies demonstrate comparable neuronal phase-locking patterns for both periodic and aperiodic nonspeech stimuli (Wilsch et al., 2015; Morillon et al., 2016; Breska and Deouell, 2017). Nonetheless, upcoming events in speech can be anticipated by nonperiodic cues, for example, based on syntactic or semantic features (Nguyen et al., 2015). The predictability that results from periodic stimuli can also facilitate auditory perception and learning (Falk et al., 2017; Rimmele et al., 2018). Therefore, periodicity might facilitate auditory processing by providing temporal *predictability* to the system (Hovsepyan et al., 2020). At the very least, the brain must remain flexible and continuously adjust to changes in signal (a)periodicity to maintain robust processing. Yet, how different types of stimuli and the degree to which their aperiodicity affects auditory neural coding and entrainment remains unclear.

In the present study, we aimed to characterize how (1) rate, (2) periodicity, and (3) stimulus domain (i.e., speech vs nonspeech) affect auditory neural entrainment. In passive listening paradigms, we recorded multichannel EEGs in young, normal-hearing adults to assess neural entrainment to rapid auditory stimuli that parametrically varied in their speed (rate) and periodicity (temporal jitter). We analyzed the data at the source level to assess possible differences in hemispheric lateralization for entrained neural responses. The pacing of our rate manipulation assessed changes in neural oscillation strength for sounds presented slower than, at, and faster than the nominal syllabic rate of typical speech (i.e., 4.5 Hz). We reasoned that characterizing phase-locking strength across rates may demonstrate a preferred entrainment frequency of the system relative to rates that are considered to have special importance for speech perception. As a second manipulation, we evaluated the effects of signal (a)periodicity on entrained brain activity. By adjusting the successive interstimulus interval between repeated tokens, we varied the stimulus delivery between fully aperiodic and periodic presentation. As a third aim, we assessed the domain specificity of auditory neural entrainment. Studies using unintelligible sounds (Howard and Poeppel, 2010) have raised questions about whether brain entrainment mechanisms reflect mere physical stimulus characteristics (Capilla et al., 2011) or higher-level functions unique to speech-language processing (Peelle and Davis, 2012). Thus, in addition to speech, we mapped rate and jitter functions for nonspeech (click) stimuli to test for possible domain specificity in entrainment strength. Neural responses were then compared with standard psychoacoustical assays of rate and periodicity sensitivity to assess the behavioral relevance of our EEG findings.

Materials and Methods

Participants

We recruited $N = 24$ young adults (aged 20–39 years; 12 female, 12 male) to participate in the study. All participants had no history of neuropsychiatric illness and had normal hearing (i.e., air conduction thresholds ≤25 dB HL, screened from 500 to 4,000 Hz; octave frequencies). History of music training, years of education, and handedness were documented.

We required participants to have <3 years of formal musical training since musicianship is known to enhance oscillatory EEG responses (Trainor et al., 2009; Bidelman, 2017). All participants were monolingual English speakers and were right-handed (mean score at the Edinburgh Handedness Inventory, 79; Oldfield, 1971). Participants gave written informed consent in compliance with a protocol approved by the Institutional Review Board at the University of Memphis (#2370) and were monetarily compensated for their time.

Behavioral tasks and procedure

We used TMTFs and the CA-BAT paradigm (Viemeister, 1979; Bidelman et al., 2015; Harrison and Müllensiefen, 2018) to assess listeners' perceptual rate and periodicity sensitivity and relate our neural findings to the behavior.

Temporal modulation transfer functions (TMTFs). TMTFs are generally performed by modulating a carrier signal (e.g., noise) with a sinusoid at various rates and measuring the threshold modulation amplitude. The TMTF is a psychoacoustic measure of listeners' sensitivity to track amplitude modulations. The TMTF function describes amplitude detection thresholds (i.e., absolute sensitivity) as a function of modulation frequency (Viemeister, 1973, 1979; Dau et al., 1997; Bidelman et al., 2015). TMTFs were measured using a forced-choice, adaptive tracking task. Three consecutive 500 ms bursts of wideband noise (100–10,000 Hz) with 300 ms interstimulus interval (ISI) and 25 ms rise/fall ramping were presented binaurally using circumaural headphones (Sennheiser HD 280 Pro). The noise was set at 74 dB SPL. The first and third noise bursts had no modulation; the second burst was modulated with a sinusoidal envelope at rates of 2.1, 3.3, 4.5, 8.5, and 14.9 Hz, identical to those used for the EEG recordings. Participants adjusted the degree of modulation imposed on the noise so that the fluctuation in the second noise burst was just detectable. Plotting the minimum detectable modulation depth across various carrier frequencies (rates) gives the TMTF. Participants were allowed to adjust the modulation depth (measured in dB) using a slider bar on the computer screen until the difference between the target (modulated) and reference (unmodulated) intervals were no longer audible. The threshold was taken as the smallest modulation depth needed to just detect amplitude fluctuations in the stimulus. More negative thresholds reflect better task performance. This was repeated across rates to measure thresholds as a function of frequency. TMTFs were measured using the Auditory Interactivities Software (Sensimetrics Corp., Gloucester, MA).

CA-BAT. The computerized adaptive beat alignment test is a version of the beat alignment test that assessed participants' behavioral sensitivity to periodicity, that is, jitter (Harrison and Müllensiefen, 2018). The test consisted of 27 items lasting ~10 min that were presented with an intensity of ~74 dB SPL. Each item consisted of a beep track superimposed on a musical clip. The beep track alignment (d_r) varied adaptively from trial to trial ($0.5 \leq d_r < 1$) such that it was displaced in the direction ahead or behind the music. Increasing d_r moved the beep track closer to the musical beat and made discrimination harder. Critically, d_r was varied adaptively based on the listener's trial-to-trial performance to converge onto their threshold for periodicity sensitivity. Participants were provided some sample music before the testing session as a training phase that includes instructions, audio demonstrations, and two practice questions. They were then given the 27 musical track test items in random order during the data collection phase, with no item-to-item feedback. The task was a two-alternative forced-choice (2-AFC) paradigm. On each trial, listeners heard two versions of the same musical track; they differed only in the overlaid metronome beep track. In one interval, the metronome and music were synchronized. In the other (lure) interval, they were displaced by a constant proportion of a beat. Participants were instructed to choose the one that was synchronized. The main output from the CA-BAT is an *ability score* (range, −4 to 4), corresponding to the listener's sensitivity to periodicity. A secondary output is an *ability_sem score*, corresponding to the standard error of measurement for the ability estimate. Both metrics are computed from the underlying item response theory model (Harrison and Müllensiefen, 2018). The paradigm was implemented in R (v.4.1.1; R Core Team, 2013).

EEG recording procedures

Stimuli. EEGs were elicited using trains of clicks (100 μs) and the synthesized speech (60 ms) token /ba/ (Fig. 1). Click and speech tokens were matched in overall level and bandwidth. Acoustic stimuli were presented at a sampling rate of 48,818 Hz to ensure maximal acoustic bandwidth during stimulus presentation. The speech token was selected as pilot testing determined it was the most identifiable token among several consonant–vowel options from previous neural oscillation studies (/ma/, /wa/, /va/, and /ba/; Assaneo and Poeppel, 2018). In the rate experiment, click and /ba/ tokens were presented at five different rates (2.1, 3.3, 4.5, 8.5, 14.9 Hz).

In the jitter experiment, this token was presented at the nominal syllabic rate of speech (4.5/s), but we varied the trains' periodicity by introducing random jitter in half the ISI between successive tokens. Jitter ranged from perfectly periodic (nominal ISI/2 ± 0% jitter) to aperiodic trains (nominal ISI/2 ± 80% jitter) in five equal steps from 0 to 80% (20% steps; Krumbholz et al., 2003). Importantly, ISIs were uniformly sampled around the nominal rate (222 ms = 1/4.5) which maintained the overall average rate of stimuli between periodic and aperiodic conditions, allowing only the degree of periodicity to vary (Fig. 2). Both click and speech stimuli were presented binaurally at 74.3 dB SPL via ER-2 insert earphones (Etymotic Research). Stimulus level was calibrated using a Larson–Davis SPL meter (Model LxT) measured in a 2 cc coupler (IEC 60126) (Stimulus level was equated using recommended procedures for calibrating transient stimuli
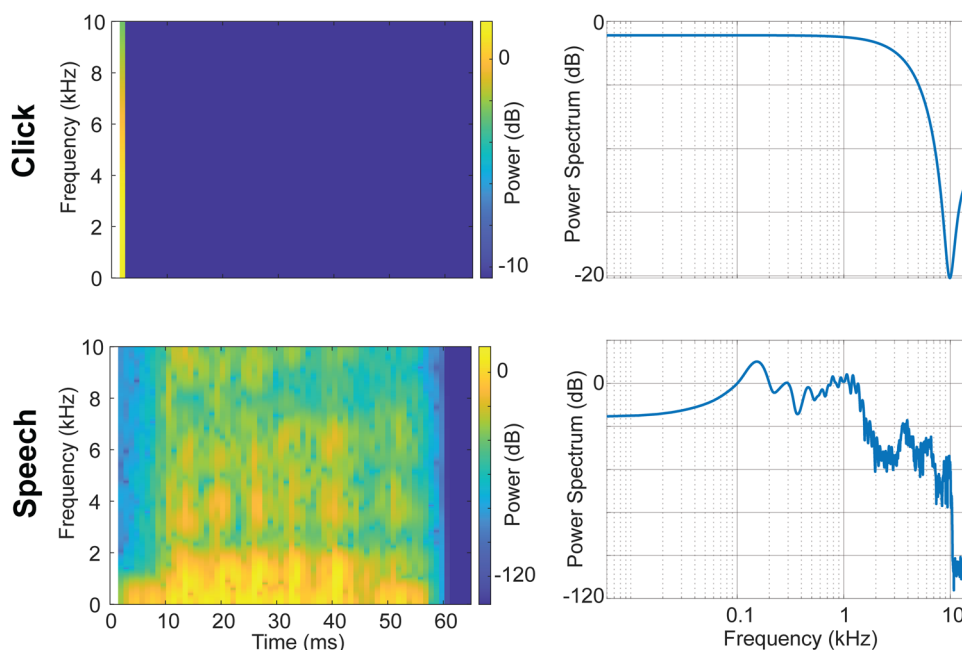
Figure 1. Spectral properties of the isolated click and speech /ba/ tokens. First column, Stimulus spectrograms and right column, power spectra. Note the similarity in overall bandwidth of the speech and nonspeech stimulus. Level is expressed in arbitrary units to show the relative changes in dB across the power spectrum of the isolated tokens (apparent overall level differences are arbitrary before calibration). For actual stimulus presentation, isolated tokens were strung together to produce trains of stimuli which were then equated in RMS amplitude. Thus, despite clear variation in their power spectra, overall perceptual loudness was similar between speech and click *trains* (see also footnotes #1 and #2).

(ISO389-6, 2007). Click/speech tokens were presented in a continuous train at the fastest rate (14.9 Hz). The resulting steady-state RMS was then adjusted to achieve 74.3 dB SPL for both speech and nonspeech stimuli to match overall level and thus perceptual loudness (see also footnote #2)). Left and right ear channels were calibrated separately. The study encompassed both speech and click conditions, involving a total of five rates (2.1, 3.3, 4.5, 8.5, 14.9 Hz) and five jitter conditions specifically applied at the 4.5 Hz rate (0, 20, 40, 60, 80%). In total, there were 18 distinct conditions (the 0% jitter was the same in the 4.5 Hz rate). Each condition consisted of 1,000 tokens. All conditions were randomized for each participant.

EEG recording. Participants were seated in an electrically shielded, sound-attenuating booth during EEG testing. They were asked to keep their eyes open and watch a silent selected movie (i.e., passive listening task). Passive listening allows for the investigation of spontaneous neural responses without the influence of specific cognitive or attentional demands. We use a passive task here to explicitly test whether previously reported enhancements in neural speech tracking near the language-universal syllable rate (4–5 Hz) depend on attention (Assaneo and Poeppel, 2018; Assaneo et al., 2019a; He et al., 2023b) or instead reflect more automatic tuning of auditory entrainment processing. Continuous EEGs were recorded using a 64-channel electrode cap (Neuroscan Quik-Cap). Blink artifacts were monitored by placing electrodes on the outer canthi and superior and inferior orbit of the eyes. Electrode positions in the array followed the international 10–10 system (Oostenveld and Praamstra, 2001). Electrodes were maintained at <5 kΩ impedance during testing and were rehydrated halfway through the experiment as necessary. EEGs were recorded using Neuroscan SynAmps RT amplifiers at a sample rate of 500 Hz. Data were re-referenced to the common average offline for subsequent analysis.

EEG analysis
We used BESA Research 7.1 (BESA) to transform each listener's single-trial scalp data into source space using BESA's auditory evoked potential (AEP) virtual source montage (Scherg et al., 2002; Bidelman, 2017). This applied a spatial filter to all electrodes that calculate their weighted contribution to the scalp recordings. We used a four-shell spherical volume conductor head model (Sarvas, 1987; Berg and Scherg, 1994) with relative conductivities (1/Ωm) of 0.33, 0.33, 0.0042, and 1 for the head, scalp, skull, and cerebrospinal fluid, respectively, and compartment sizes of 85 mm (radius), 6 mm (thickness), 7 mm (thickness), and 1 mm (thickness; Picton et al., 1999; Herdman et al., 2002). The AEP model includes 11 regional dipoles distributed across the brain including bilateral auditory cortex [AC; Talairach coordinates (*x*, *y*, *z*; in mm): left (−37, −18, 17) and right (37, −18, 17)]. Regional sources consist of dipoles describing current flow (units nAm) in tangential planes. We extracted the time courses of the tangential components for left and right AC sources as this orientation captures the majority of variance describing the auditory cortical ERPs (Picton et al., 1999). This approach
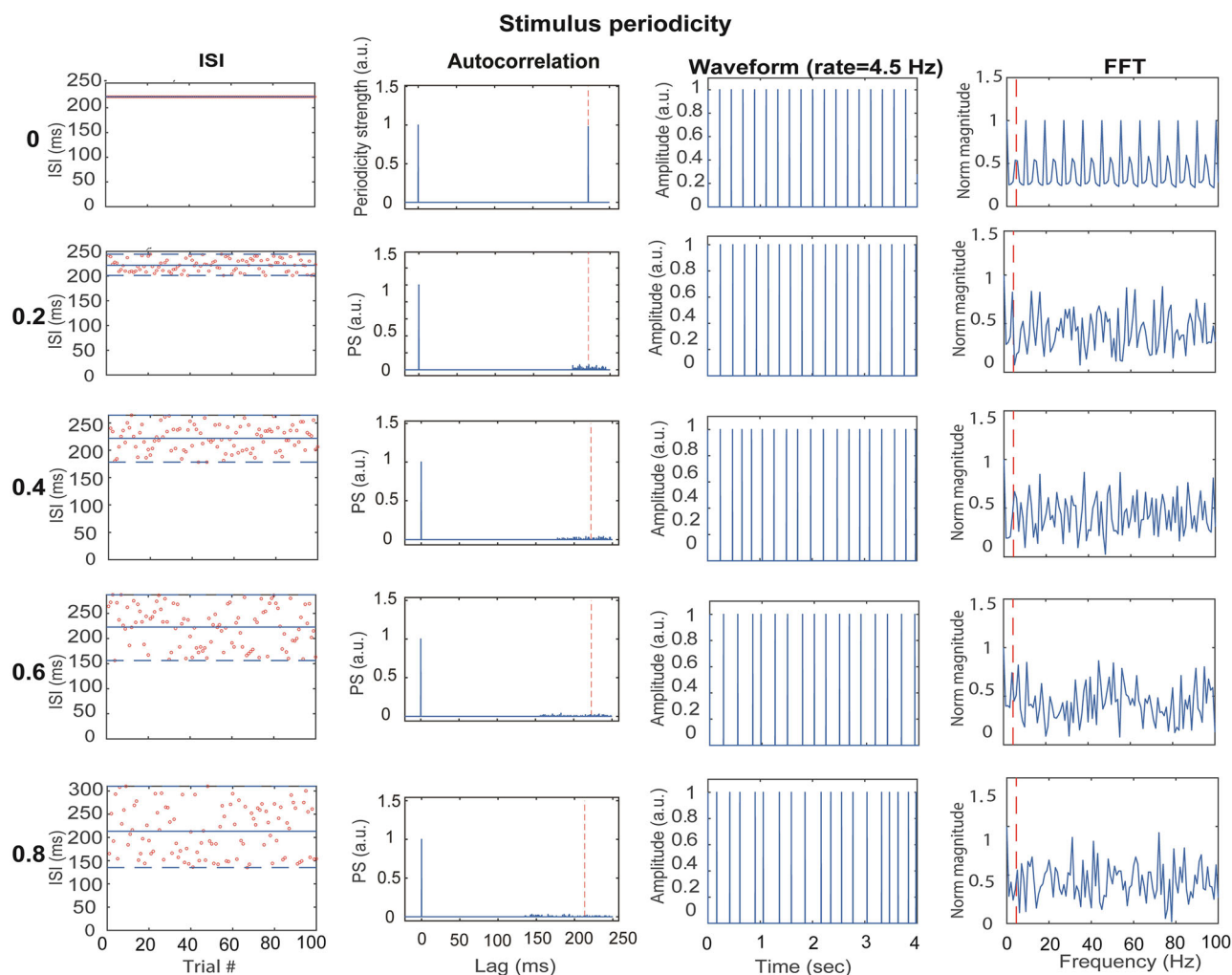
Figure 2. Acoustic properties illustrating the effects of (a)periodicity jitter (all clicks at 4.5 Hz rate; nominal ISI = 1/4.5 Hz = 222 ms). First column, Distribution of ISI around the nominal ISI as a function of jitter from 0 to 80%. Second column, Autocorrelation functions (ACFs) show the degree of periodicity in the stimuli. Note the periodicities at 222 ms which becomes blurred around the nominal period (dotted lines) with increasing jitter. Third column, Time waveforms. Fourth column, Fast Fourier transforms (FFTs) as a function of jitter. As with the ACFs, note the decreased energy at 4.5 Hz (dotted lines) with increasing stimulus jitter/aperiodicity.

allowed us to reduce each listener's 64-channel data to two source channels describing neuronal currents localized to the left and right AC (Price et al., 2019; Momtaz et al., 2021).

Phase-locking value (PLV). We computed phase-locking value (PLV; Lachaux et al., 1999) between brain and stimulus waveforms to evaluate how neural oscillatory responses track speech and nonspeech acoustic signals across different presentation speeds (rates) and periodicities (jitter). This allowed us to assess neuro-acoustic synchronization across tokens in the ongoing sound stream. We first transformed the continuous EEGs into source waveforms (SWFs) via matrix multiplication of the sensor data (EEG) with the AEP source montage's dipole leadfield (L) matrix (i.e., SWF = $L^{-1} \times$ EEG; Scherg et al., 2002; Bidelman, 2018). The leadfield was based on the dipole configuration detailed above. This resulted in two time series representing current waveforms in source space projected from left and right AC. For submission to PLV analysis, 30 s of continuous data was then extracted. Importantly, this yielded equal-length neural data per stimulus condition and listener. Identical processing was then applied to all rate and jitter conditions per participant.

We measured brain-to-stimulus synchronization as a function of frequency via PLV (Lachaux et al., 1999). First, we computed the full-band envelope from the stimulus via the Hilbert transform. We then down-sampled the stimulus envelope to match the sampling rate of the EEG (i.e., 500 Hz). To be reasonably interpreted, PLV requires bandpass filtering the signals to assess how entrainment changes in a frequency-dependent manner. Following approaches by Assaneo and Poeppel (2018) and He et al. (2023a,b), neural and acoustic stimulus signals were bandpass filtered (±0.5 Hz) around each nominal

frequency, and PLV was then computed according to Eq. 1:

$$PLV = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i[\theta_1(t)-\theta_2(t)]}\right|, \qquad (1)$$

where $\theta_1(t)$ and $\theta_2(t)$ are the Hilbert phases of the EEG and corresponding evoking stimulus signal, respectively. Intuitively, PLV describes the average phase difference (and by reciprocal, the correspondence) between the two signals. PLV ranges from 0 to 1, where 0 represents no (random) phase synchrony and 1 reflects perfect phase synchrony between signals. We then repeated this procedure—that is, isolating a 1 Hz band and computing PLV—for center frequencies between 1.1 and 30 Hz (0.3 Hz steps). This resulted in a continuous function of PLV describing the degree of brain-to-stimulus synchronization across the bandwidth of interest (Assaneo et al., 2019b; Fig. 4). We then measured PLV magnitude for each rate/jitter, stimulus type (speech vs click), and participant. The magnitude was taken as the peak of each individual frequency-dependent PLV function within ±0.5 Hz of the nominal stimulus rate (Fig. 4, ▼s; He et al., 2023a,b). Comparing PLV magnitude across increasing rates/jitters allowed us to characterize how brain-to-stimulus synchronization varied for speech versus nonspeech stimuli and between cerebral hemispheres. However, the omnibus ANOVA on PLV measures failed to reveal main or interaction effects with hemisphere (results reported below). Consequently, we collapsed LH and RH responses to focus on variations in peak PLV across stimulus manipulations (i.e., rates and jitters).

Statistical analysis

Unless otherwise noted, we used mixed-model ANOVAs implemented in R (lme4 package;(Bates et al., 2014) to assess all dependent variables of interest. Fixed factors were rate (2.1, 3.3, 4.5, 8.5, 14.9 Hz), periodicity (0, 20, 40, 60, 80% jitter), and stimulus domain (click, /ba/). Subjects served as a random effect. Based on the distribution of the data and initial diagnostics, we transformed the data using a square root transformation. The significance level was set at $\alpha = 0.05$. Tukey–Kramer adjustments were used for post hoc contrasts. Correlations (Pearson's r) were used to evaluate relationships between neural oscillations and behavior.

Results

Behavioral data

TMTFs (rate sensitivity)

Figure 3 shows the average TMTFs of our participants and the data of Viemeister (1979) for comparison. TMTFs show sensitivity (threshold) to amplitude modulation in wide-band noise measured at five different rates. With increasing rates, participants showed better (i.e., more negative) detection thresholds corresponding to better sensitivity (i.e., temporal resolution). An ANOVA on TMTF thresholds revealed a rate effect on TMTF thresholds ($F_{(4,115)} = 10.44$; $p < 0.001$). TMTF thresholds typically worsen with increasing rates up to ~100 Hz. However, at the low modulation rates used in this study—and consistent with prior psychoacoustic studies (Viemeister, 1979)—we find that rate sensitivity increases slightly between 2.1 and 14.9 Hz.

CA-BAT (aperiodicity sensitivity)

The CA-BAT produced two different scores of periodicity sensitivity for each participant related to the absolute threshold (ability score) and its variance (ability_sem; Harrison and Müllensiefen, 2018). Ability scores averaged 0.17 ± 0.92 across participants, consistent with prior psychoacoustic studies on jitter sensitivity (Harrison and Müllensiefen, 2018).

EEG oscillations across token (PLV)

We used PLV to quantify neural phase locking across tokens and how the brain entrains the ongoing stream of acoustic stimuli. Raw PLV response functions illustrating changes in phase-locking strength as a function of frequency, stimulus manipulations (rate, jitter), hemispheres (RH, LH), and stimulus type (click, /ba/) are shown in Figure 4. Peak quantification of the PLV functions is shown in Figure 5. In general, neural responses closely followed the speed of the auditory stimuli, showing stark increases in PLV strength that closely followed the fundamental rate of presentation (i.e., $F0 = 2.1–14.9$ Hz). Harmonics were also observed, which are common in the spectra of sustained auditory potentials and are due to nonlinearities of the EEG that result in phase locking at the $F0$ and its integer multiples ($2F0$, $3F0$, …, $nF0$; Lins et al., 1995; Bidelman and Bhagat, 2016). PLV strength also varied for the 4.5 Hz stream with changes in jitter; more aperiodic sounds produced weaker neural entrainment.

An ANOVA assessing rate effects on peak PLV showed a rate × stimulus domain interaction ($F_{(4,437)} = 13.64$; $p < 0.0001$; Fig. 5A). Multiple comparisons showed stronger PLV for speech versus clicks at 4.5 Hz ($p < 0.0001$). This pattern reversed at higher rates with speech eliciting lower PLV than clicks at 8.5 ($p = 0.0095$) and 14.9 Hz ($p = 0.0002$). These results suggest neural entrainment for speech stimuli is enhanced relative to nonspeech signals specifically at 4.5 Hz and weakens precipitously for higher rates beyond what typically occurs in normal speech production.
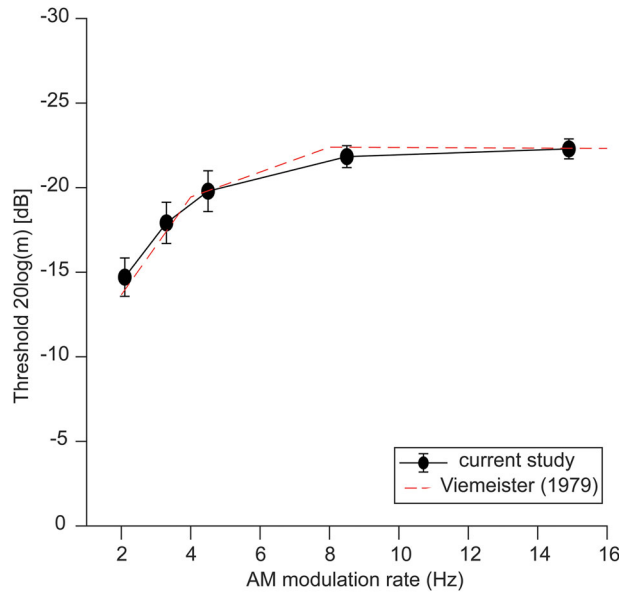
Figure 3. TMTFs show behavioral rate sensitivity. TMTFs demonstrate temporal acuity for detecting AM fluctuations in sounds as a function of modulation frequency. For very low frequencies <20 Hz, and consistent with Viemeister (1979; see 500 ms gated carrier condition, their Fig. 6), TMTFs demonstrate a high-pass filter shape, indicating slight improvements in behavioral rate sensitivity from 2 to 20 Hz. Error bars = ±1 SEM.
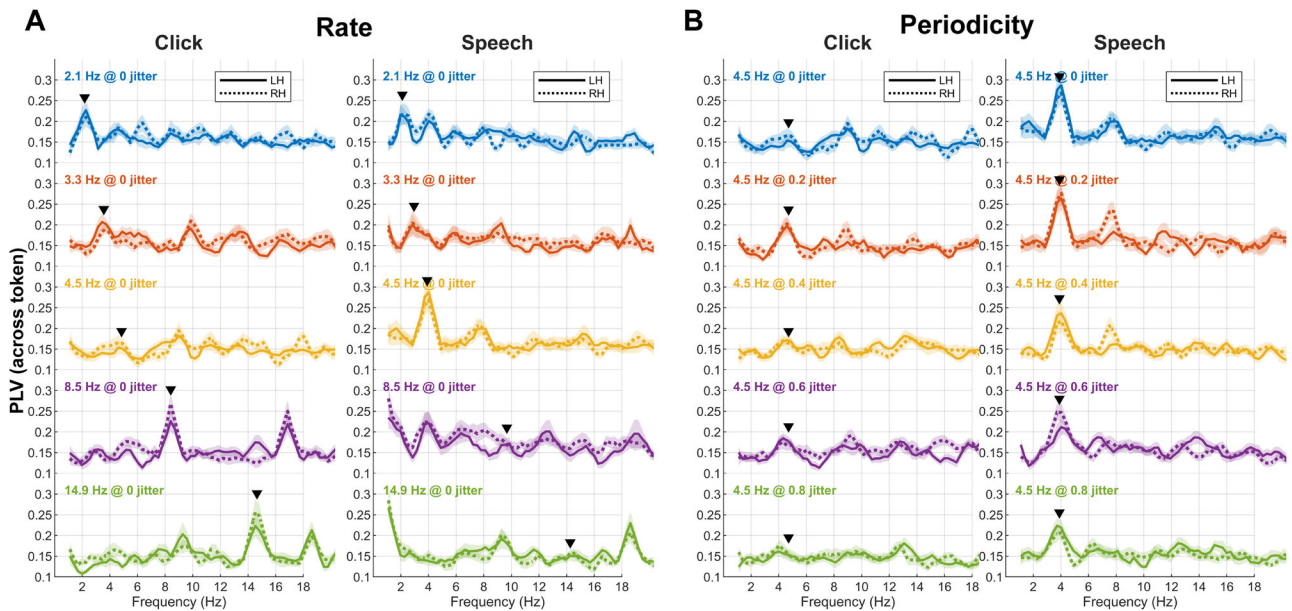

Figure 4. Grand average PLV across subjects as a function of rate, jitter, stimulus, and hemisphere. **A**, Rate results. **B**, Periodicity results. PLV showed enhanced activity at each fundamental frequency (▼) and integer-related harmonics. The right and left hemispheres exhibited comparable responses. Shading = ±1 SEM.

For periodicity, we found a main effect of stimulus domain ($F_{(1,437)} = 67.75$; $p < 0.0001$), but more importantly, a periodicity × stimulus domain interaction ($F_{(4,437)} = 3.18$; $p = 0.013$; Fig. 5B). The main effect of stimulus was due to speech (all 4.5 Hz rate) eliciting stronger PLV than clicks across the board. Indeed, individual click versus speech contrasts at each jitter revealed higher PLV for speech than clicks that diminished with increasing aperiodicity: 0% ($p < 0.0001$), 20% ($p = 0.001$), 40% ($p = 0.0021$), 60% ($p = 0.0042$), and 80% ($p = 0.058$). In other words, the overall pattern for speech exhibited a linear decline in PLV (linear effect; $t_{437} = -4.01$; $p < 0.0001$) where the strength decreased with increasing jitter. In contrast, entrainment to clicks remained constant with increasing jitter (linear effect; $t_{(437)} = 0.66$; $p = 0.51$). Indeed, when compared directly, the linear trend effect was stronger for speech than clicks ($t_{(437)} = -3.299$; $p = 0.0011$). This interaction suggests that speech produces stronger neural entrainment across tokens than nonspeech sounds and is also more impervious to disruptions in periodicity (i.e., signal jitter).
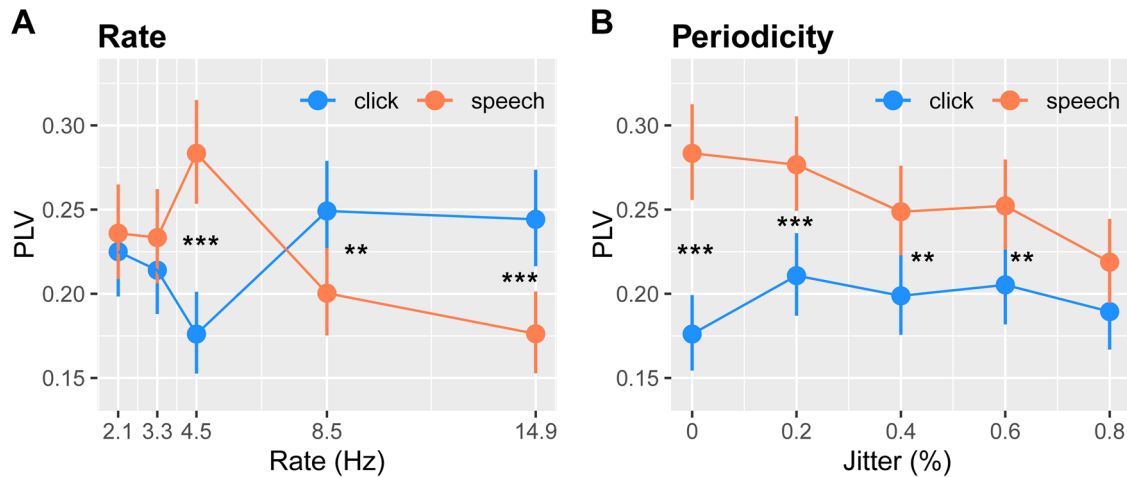
Figure 5. Grand average peak PLV across rate and jitter (aperiodicity). **A**, Rate effects show a rate × stimulus domain interaction. The speech token /ba/ elicited increased neural entrainment at 4.5 Hz. **B**, Periodicity effects show a stimulus domain × jitter interaction. Neural entrainment to speech is initially enhanced compared with clicks at low jitters but declines within increasing aperiodicity. PLV strength is invariant to increasing jitter for clicks. PLV, phase-locking value. Responses are collapsed across hemispheres. Error bars = ±95% CI. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. See also Extended Data Fig. 5-1.
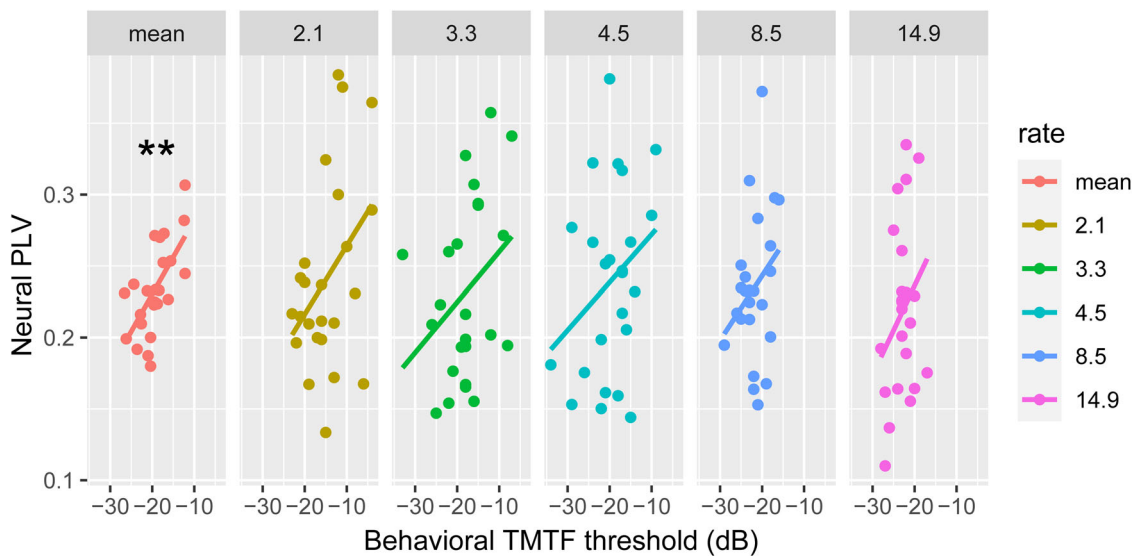

Figure 6. Brain–behavioral correlations for rate sensitivity. Behavioral TMTF thresholds show an association with the mean PLV averaged across rates (**$p < 0.01$). Values are pooled across speech/click conditions and hemispheres. Scatters show brain–behavior correlations between TMTF thresholds and the five individual rates (all n.s. by themselves).

One concern is that the increased jitter of a signal spreads the power across a wider frequency range and conversely, lessens the power in a narrow frequency band near the $F0$ rate. This might artificially reduce PLV with increasing aperiodicity. To test this possibility, we computed cross-correlations (MATLAB xcorr function) between the stimulus acoustic envelope waveforms and the EEG (Extended Data Fig. 5-1). In general, cross-correlation effects mirrored but were less salient than PLV effects. An ANOVA conducted on cross-correlations revealed only main effects of jitter ($F_{(4,437)} = 4.64$; $p = 0.0011$) and stimulus domain ($F_{(1,437)} = 15.32$; $p < 0.0001$). Critically, however, we did not find a jitter × stimulus interaction ($F_{(4,437)} = 0.76$; $p = 0.55$) as observed for PLV. Similarly, cross-correlations across rate showed a different pattern than for PLV. While there was a rate ($F_{(4,437)} = 5.18$; $p = 0.0004$) and stimulus effect ($F_{(1,437)} = 12.19$; $p = 0.0005$), there again was no rate × stimulus interaction ($F_{(4,437)} = 1.21$; $p = 0.31$) as observed for PLV. These findings suggest that while the gross pattern of PLV effects observed in the data (Fig. 5) might be partially accounted for changes in stimulus-to-response correlation with increasing jitter, PLV reveals an additional *differential* sensitivity in neural phase-locked entrainment between speech and nonspeech stimuli.

Brain–behavior relationships

We used correlations to explore the correspondence between neural responses and behavioral measures. The average PLV across rates was highly correlated with TMTF thresholds ($r = 0.65$; $p = 0.0006$); larger neural PLV was associated with poorer (less negative) behavioral thresholds. The correlations for each individual rate were not significant at 2.1 Hz ($r = 0.39$; $p = 0.06$), 3.3 Hz ($r = 0.33$; $p = 0.11$), 4.5 Hz ($r = 0.29$; $p = 0.16$), 8.5 Hz ($r = 0.28$; $p = 0.19$), or 14.9 Hz ($r = 0.28$; $p = 0.19$). For the CA-BAT test, the degree of periodicity sensitivity at 0.6 jitter was negatively correlated with PLV ($r = -0.41$; $p = 0.044$; data not shown). That is, better periodicity detection thresholds were associated with improved entrainment for ~60% jittered stimuli. However, we note this correlation was weak and occurred only after poling across hemispheres and stimulus type. Still, these findings indicate participants' behavioral sensitivity to periodicity was loosely associated with their entrained brain responses.

Discussion

We compared cortico-acoustic tracking for repeated speech versus click train stimuli parameterized across various speeds (below, at, and above the nominal speech rate) and degrees of acoustic periodicity (jitter in 20% steps) to probe the effects of these stimulus factors and map the profiles of oscillatory activity for auditory cortical entrainment. For *rate*, we found that phase locking to the repetitive speech token /ba/ showed a surprising improvement in neural entrainment at 4.5 Hz but deteriorated at higher rates. For clicks, however, phase-locking strength dropped at 4.5 Hz and then rebounded sharply at 8.5 Hz until peaking at 14.9 Hz. For *periodicity*, we found that while phase locking to speech declined with increasing jitter, entrainment to speech was still superior to that of clicks. In contrast to /ba/ sounds, click responses were largely resistant to disruptions in periodicity. Collectively, our findings show that even under passive listening, auditory neural entrainment for speech relative to nonspeech sounds is (1) more sensitive to changes in rate and periodicity and (2) enhanced in strength, perhaps reflecting a prioritization for synchronizing to behaviorally relevant sounds.

Rate effects

Continuously changing speech rate/context requires ongoing adjustment by the listener (Casas et al., 2021). Our stimulus design included five rates intending to examine the auditory system's temporal processing of speech and nonspeech sounds at speeds below, at, and above typical syllabic rates for speech (Poeppel and Assaneo, 2020). We found an interaction between rate and stimulus type for PLV measures. Slow theta waves easily follow the envelope of ongoing speech and are thus thought to integrate syllable representations across tokens—a larger level of the speech-analysis hierarchy (Casas et al., 2021). Our findings indicate that PLV exhibited a strong differentiation between repeated click and speech-syllable trains. PLV reflects the phase locking between the acoustic stimulus and EEG response across tokens and over the time course of the stimulus stream. Surprisingly, across-token PLV to speech revealed strong enhancements in phase locking at 4.5 Hz, validating the concept of a universal speech syllabic rate across global languages (Assaneo and Poeppel, 2018; He et al., 2023b). Our results here extend these prior studies, however, by demonstrating such enhanced synchronization at 4.5 Hz during passive listening, in the absence of an active speech perception task. Critically, our data also show that phase-locking enhancement at ~4.5 Hz was specific to speech and did not occur for clicks (i.e., domain-specific effect). Moreover, we found phase locking to speech plummeted at higher frequencies, suggesting these speech-specific enhancements are limited to stimulus rates that occur naturally for speech (Assaneo and Poeppel, 2018).

It is difficult to see how duration effects alone could account for the differential (interaction) pattern observed in our data between speech and click trains. Indeed, there is some evidence that auditory cortical entrainment is enhanced for intelligible signals which might explain the larger PLV we see for the syllable train relative to click stimuli (Xu et al., 2023). Moreover, token duration seems to play a less prominent role in the strength of cortical entrainment since tracking is dominated by locking to an acoustic signal's edge landmarks rather than its nucleus, per se (Oganian and Chang, 2019; Oganian et al., 2023).

At higher rates, and in contrast to speech, click synchronization actually improved slightly. The counterintuitive enhancement of click responses at the higher rates is however consistent with psychoacoustical findings (including those here), explained as increased summation (i.e., temporal integration) of postsynaptic potentials as a result of activating a broader neural population (Forss et al., 1993; Brugge et al., 2009). Thus, while the auditory system is certainly capable of synchronizing to higher stimulus rates (as evidenced by our click data), it appears as though the sensitivity to modulations in speech synchronization is more restricted.

Periodicity effects

Various temporal scales of the speech signal and linguistic hierarchy are correlated with neural entrainment in different frequency bands of brain activity (Ghitza and Greenberg, 2009). Given that speech is not perfectly periodic and shorter syllables frequently follow or precede longer ones, it is intriguing to examine how the (a) periodicity of speech influences brain entrainment (Ghitza and Greenberg, 2009). By parametrically varying the jitter of otherwise periodic signals, we aimed to perturb the input integrity and examine how brain rhythms to the optimal speech rate (at 4.5 Hz; Assaneo and Poeppel, 2018) are disrupted by aperiodicity. In classic evoked response paradigms, which only consider token-wise responses, the temporal information between successive stimulus events does not directly affect the neural representation

of the acoustic signal, at least at lower rates where adaptation would be at play. However, other studies showed that "cross-token comparisons" could indeed be a cue for temporal coding that has a significant impact on intelligibility (Miller and Licklider, 1950; Huggins, 1975). Thus, temporal perturbation can affect how sounds are organized and subsequently perceived. Our PLV measurements robustly detect these jitter effects especially for speech events.

Periodicity in speech has been shown to facilitate perception and intelligibility (Benesty et al., 2008). Ghitza and Greenburg (2009) altered the interstimulus intervals between syllables using periodic and aperiodic interruption to affect intelligibility. Their results ascribed neural entrainment to internal processing rather than acoustic aspects of the sounds (Ghitza and Greenberg, 2009). They showed the speech intelligibility of the compressed signal is poorer than that of the original signal. In their study, inserting 20–120-ms-long gaps of silence into a speech significantly increases its intelligibility. Our study design excluded an intelligibility component as it was passive and focused arguably on only the acoustic-phonetic characteristics of speech. Still, our results reveal that neural entrainment is improved for rhythmic speech as opposed to click or even aperiodic speech. It is possible such neural effects account for the perceptual facilitation observed for periodic signals in previous behavioral studies (Ghitza and Greenburg, 2009).

Only the 4.5 Hz presentation rate was chosen in the jitter experiment; we anticipated optimal phase locking at this pace due to its alignment with the nominal syllable rate of most languages (Assaneo and Poeppel, 2018). PLV demonstrated that at this specialized rate, periodicity did not affect cortical entrainment to clicks. Despite the fact that clicks are more perceptually salient than speech due to their rapid onset (Vigil and Pinto, 2020), jitter did not affect PLV entrainment sensitivity for clicks. With increasing the jitter of speech stimuli, however, phase locking deteriorated. It is possible that as speech becomes more aperiodic, the brain treats the signal more like a nonspeech stimulus, resulting in similarly low PLV as we observe for click stimuli. This could explain why stimulus type and jitter interacted in our PLV analysis. We can rule out explanations of these effects due to loudness differences as all tokens were similar in overall SPL as well as perceptual loudness (96.9 ± 2.7 phons; Moore et al., 1997) (Loudness was computed using the MATLAB function acousticLoudness() according to standard ISO532-2 (2017). The similarity in (estimated) loudness across stimuli is perhaps unsurprising given that the RMS amplitude was equated across tokens (see Materials and Methods and footnote #1)). In fact, speech stimuli, which elicited more robust PLV, were actually ~5 phons *weaker* in perceptual loudness than clicks, contrary to a temporal integration explanation of the data. On the contrary, periodicity in speech may boost neural responses via predictive coding, which could account for the stronger entrainment to speech we find at low relative to high degrees of jitter (Thut et al., 2011; Peelle and Davis, 2012). Previous research also suggests specific populations of neurons that respond to aperiodic but not periodic stimuli (Yrttiaho et al., 2008). Also, imaging studies demonstrate the location of periodicity-sensitive cortical areas can be more anterior than aperiodic stimuli, depending on the stimulus type (Hall and Plack, 2009). Our technique for EEG source analysis only localized responses in the auditory cortex, which may not elicit equivalent responses for speech and click stimuli (Steinschneider et al., 1998).

Brain–behavior relations between entrainment and rate/periodicity sensitivity

We demonstrate an association between behavioral sensitivity and brain synchronization to stimuli varying in rates. Individuals who showed greater sensitivity in our psychoacoustic measures exhibited weaker phase locking in cortical entrainment. A recent study by Casas et al. (2021) also showed that participants that are more behaviorally responsive to temporal changes in ongoing sounds exhibit weaker phase-locked responses. They attributed their counterintuitive findings to the sample size and other methodological concerns including the perceptual difficulty of their stimulus set (Casas et al., 2021). However, one possibility is that their neural recordings included several entrained responses from multiple brain areas (not exclusively auditory regions). Indeed, conventional EEG suffers from volume conduction resulting in frontal generators contributing to auditory evoked responses (Knight et al., 1989; Picton et al., 1999). Our use of passively presented stimuli and source analysis helps exclude attentional or task confounds (as suggested by Casas et al., 2021) that are likely modulated by frontal cortical areas. Having the same result, however, it is possible that individuals who show better brain-to-acoustic coupling relegate temporal processing to lower levels of the auditory system (e.g., brainstem, thalamus), more peripheral to the behavior and cortical responses assessed here. Whereas others who perform worse or more laboriously in temporal processing tasks might require high levels of auditory processing at the cortical level as a form of compensatory mechanism (Momtaz et al., 2021, 2022). This might account for the counterintuitive negative correlation we find between cortical phase-locking strength and behavior for the rate manipulation, that is, more effortful encoding (higher PLV) in less perceptually sensitive individuals. Other electrophysiological TMTF studies reveal a decrease in neural sensitivity with increasing modulation rate (Wang et al., 2012). These results run counter to what we observe here and suggest differences in stimulus acoustics (e.g., bandwidth) and analysis techniques might account for discrepancies across studies. However, we did find that CA-BAT periodicity detection thresholds weakly predicted PLV strength for jitter manipulations (at least the ~60% jitter). We note that a significant distinction between our brain and behavior data is that although TMTFs and CA-BAT measure rate modulation and jitter detection thresholds in active tasks, our neural recordings were conducted under strictly passive listening. It is highly likely the nature of the correlation would change had we conducted our behavioral tasks during the EEG recordings. Consequently, given our strictly passive listening paradigm, interpretations that our brain–behavioral correlations index listening effort remains speculative and should be confirmed in future studies.

In this vein, while our results here demonstrate periodic speech-like stimuli differentially affects neuronal entrainment in passive listening conditions, how these mechanisms might change in active attentional states remains unknown.

Presumably, active tasks during entraining stimuli might also recruit additional (nonauditory) brain regions other than the auditory cortex, which was the focus of this investigation. As such, functional connectivity approaches (Rimmele et al., 2018) might be used to further tease out the dynamics of bottom-up (sensory-driven) versus top-down (cognition-driven) processing and interhemispheric connections that affect the brain's ability to entrain to rapid auditory stimuli.

Conclusions

Overall, this study aimed to address questions regarding rate, periodicity, and stimulus differences in EEG neural entrainment and their association with behavioral responses. By examining the same parameters, our data reveal unique distinctions in how each of these factors impacts the neural encoding of ongoing complex sounds.

Our findings show that the brain's entrainment to repeated speech tokens (phonemic level processing) is rate and periodicity sensitive and more so than for nonspeech clicks. These findings might inform broader work examining temporal processing issues in patient populations such as those with certain auditory processing disorders (Momtaz et al., 2021, 2022) or dyslexia (Tallal et al., 1993; Ben-Yehudah and Ahissar, 2004) which impact auditory temporal processing. The data here helps characterize the constraints on temporal capabilities of auditory cortex and neural entrainment at early stages of speech perception. It would be interesting to extend the current paradigm to future studies in these clinical populations. Auditory plasticity induced by training and rehabilitative programs that aim to enhance temporal processing (Chermak and Musiek, 2002; Anderson and Kraus, 2013; Moreno and Bidelman, 2014) could be used to enhance cortical phase locking and, subsequently, speech understanding in challenging listening environments. The stimulus specificity we observe in entrainment patterns also speaks to the need to incorporate the correct stimuli in training plans if the goal is to maximize neural and behavioral outcomes. Indeed, our data suggest periodic speech at or near nominal syllabic rates (4.5 Hz) might have the largest impact on perception and cognition following rehabilitation. Future studies could test these possibilities.

PLV is by far the most common metric in the neuroimaging literature to compute neuroacoustic tracking of brain signals (Assaneo and Poeppel, 2018; Assaneo et al., 2019b; He et al., 2023a,b). This motivated its adoption here. In principle, both PLV and cross-correlation are methods to assess signal similarity and are in fact equivalent under a variety of circumstances (Aydore et al., 2013). One advantage of PLV is that by definition (Lachaux et al., 1999), the metric is largely invariant to amplitude scaling and depends only on phase consistency. Thus, PLV is largely impervious to fluctuations in EEG amplitude that might artificially reduce phase locking. However, PLV has been shown to be a potential biased measure under some circumstances (Aydore et al., 2013). Our data here partially confirm these concerns. We found that in the periodicity/jitter manipulation, cross-correlation largely mirrored (but was less salient than) the pattern observed for PLV (compare Fig. 5 and Extended Data Fig. 5-1). This suggests, prima facie, that mere signal correspondence may have partially driven our PLV jitter results. However, we also found cross-correlation effects were much more muted and failed to yield a differential effect (i.e., jitter × stimulus interaction) across speech/nonspeech stimuli as in PLV. An exhaustive comparison between metrics is beyond the scope of this investigation and is addressed elsewhere (Aydore et al., 2013; Soltanzadeh and Daliri, 2014). Still, the higher sensitivity and differential entrainment strength observed for PLV (but not cross-correlation) suggests the PLV metric perhaps overinflates the degree of neural phase-locked entrainment. Future studies are needed to directly assess such potential weaknesses in brain entrainment measures.

Lastly, we acknowledge the simplicity of our single-syllable tokens and the limitations of using repeated syllables to describe natural or canonical "speech" processing. While there is precedent in the literature for describing such periodic syllable trains as "speech" (Assaneo and Poeppel, 2018; Poeppel and Assaneo, 2020; He et al., 2023b), whether the differential patterns we observe for our /ba/ versus click stimuli would hold for more naturalistic (e.g., continuous and mixed-syllabic) speech remains unknown. Thus, interpretation of our findings should be limited to understanding the encoding and processing of syllable features in speech. Still, syllable approaches offer controlled and targeted investigations of specific speech features, while continuous speech approaches capture the complexity and naturalistic aspects of speech processing. Comparing these two approaches would be an interesting future direction of study.

References

Allman MJ, Meck WH (2011) Pathophysiological distortions in time perception and timed performance. Brain 135:656–677.

Anderson S, Kraus N (2013) Auditory training: evidence for neural plasticity in older adults. Perspect Hear Hear Disord Res Res Diagn 17:37–57.

Arnal LH, Giraud A-L (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci 16:390–398.

Assaneo MF, Poeppel D (2018) The coupling between auditory and motor cortices is rate-restricted: evidence for an intrinsic speech–motor rhythm. Sci Adv 4:eaao3842.

Assaneo MF, Rimmele JM, Orpella J, Ripollés P, de Diego-Balaguer R, Poeppel D (2019a) The lateralization of speech–brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. Front Integr Neurosci 13:28.

Assaneo MF, Ripollés P, Orpella J, Lin WM, de Diego-Balaguer R, Poeppel D (2019b) Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nat Neurosci 22:627–632.

Aydore S, Pantazis D, Leahy RM (2013) A note on the phase locking value and its properties. Neuroimage 74:231–244.

Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Ben-Yehudah G, Ahissar M (2004) Sequential spatial frequency discrimination is consistently impaired among adult dyslexics. Vis Res 44:1047–1063.

Benesty J, Sondhi MM, Huang Y (2008) *Springer handbook of speech processing*. Vol. 1. Berlin: Springer.

Berg P, Scherg M (1994) A fast method for forward computation of multiple-shell spherical head models. Electroencephalogr Clin Neurophysiol 90:58–64.

Bidelman GM (2017) Amplified induced neural oscillatory activity predicts musicians' benefits in categorical speech perception. Neuroscience 348:107–113.

Bidelman GM (2018) Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. Neuroimage 175:56–69.

Bidelman GM, Bhagat SP (2016) Objective detection of auditory steady-state evoked potentials based on mutual information. Int J Audiol 55:313–319.

Bidelman GM, Jennings SG, Strickland EA (2015) PsyAcoustX: a flexible MATLAB® package for psychoacoustics research. Front Psychol 6:1498.

Bishop GH (1932) Cyclic changes in excitability of the optic pathway of the rabbit. Am J Physiol 103:213–224.

Breska A, Deouell LY (2017) Neural mechanisms of rhythm-based temporal prediction: delta phase-locking reflects temporal predictability but not rhythmic entrainment. PLoS Biol 15:e2001665.

Brugge JF, Nourski KV, Oya H, Reale RA, Kawasaki H, Steinschneider M, Howard MA 3rd (2009) Coding of repetitive transients by auditory cortex on Heschl's gyrus. J Neurophysiol 102:2358–2374.

Buzsaki G (2006) *Rhythms of the brain*. New York: Oxford University Press.

Buzsaki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929.

Capilla A, Pazo-Alvarez P, Darriba A, Campo P, Gross J (2011) Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. PLoS One 6:e14543.

Casas ASH, Lajnef T, Pascarella A, Guiraud-Vinatea H, Laaksonen H, Bayle D, Jerbi K, Boulenger V (2021) Neural oscillations track natural but not artificial fast speech: novel insights from speech-brain coupling using MEG. Neuroimage 244:118577.

Chermak GD, Musiek FE (2002) Auditory training: principles and approaches for remediating and managing auditory processing disorders. Semin Hear 23:297–308.

Cummins F (2009) Rhythm as an affordance for the entrainment of movement. Phonetica 66:15–28.

Dau T, Kollmeier B, Kohlrausch A (1997) Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. J Acoust Soc Am 102:2906–2919.

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.

Doelling KB, Poeppel D (2015) Cortical entrainment to music and its modulation by expertise. Proc Natl Acad Sci U S A 112:E6233–E6242.

Falk S, Lanzilotti C, Schön D (2017) Tuning neural phase entrainment to speech. J Cogn Neurosci 29:1378–1389.

Fiveash A, Bedoin N, Gordon RL, Tillmann B (2021) Processing rhythm in speech and music: shared mechanisms and implications for developmental speech and language disorders. Neuropsychology 35:771.

Forss N, Mäkelä JP, McEvoy L, Hari R (1993) Temporal integration and oscillatory responses of the human auditory cortex revealed by evoked magnetic fields to click trains. Hear Res 68:89–96.

Fujii S, Schlaug G (2013) The Harvard beat assessment test (H-BAT): a battery for assessing beat perception and production and their dissociation. Front Hum Neurosci 7:771.

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.

Guenther FH (2016) *Neural control of speech*. Cambridge, Massachusetts: MIT Press.

Hall DA, Plack CJ (2009) Pitch processing sites in the human auditory brain. Cereb Cortex 19:576–585.

Harrison PM, Müllensiefen D (2018) Development and validation of the computerised adaptive beat alignment test (CA-BAT). Sci Rep 8:12395.

He D, Buder EH, Bidelman GM (2023a) Cross-linguistic and acoustic-driven effects on multiscale neural synchrony to stress rhythms. bioRxiv [preprint].

He D, Buder EH, Bidelman GM (2023b) Effects of syllable rate on neuro-behavioral synchronization across modalities: brain oscillations and speech productions. Neurobiol Lang 4:344–360.

Herbst SK, Landau AN (2016) Rhythms for cognition: the case of temporal processing. Curr Opin Behav Sci 8:85–93.

Herdman AT, Lins O, van Roon P, Stapells DR, Scherg M, Picton T (2002) Intracerebral sources of human auditory steady-state responses. Brain Topogr 15:69–86.

Hovsepyan S, Olasagasti I, Giraud A-L (2020) Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. Nat Commun 11:3117.

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.

Huggins A (1975) Temporally segmented speech. Percept Psychophys 18:149–157.

ISO532-2 (2017). Acoustics — methods for calculating loudness — Part 2: Moore–Glasberg method.

ISO389-6 (2007) Acoustics – reference zero for the calibration of audiometric equipment – Part 6: reference hearing threshold levels for test signals of short duration. Geneva, Switzerland: International Organization for Standardization.

Knight RT, Scabini D, Woods DL (1989) Prefrontal cortex gating of auditory transmission in humans. Brain Res 504:338–342.

Krumbholz K, Patterson R, Seither-Preisler A, Lammertmann C, Lütkenhöner B (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb Cortex 13:765–772.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194–208.

Lakatos P, Gross J, Thut G (2019) A new unifying account of the roles of neuronal entrainment. Curr Biol 29:R890–R905.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

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.

Levelt WJ (1993) *Speaking: from intention to articulation*. Cambridge, Massachusetts: MIT Press.

Lins OG, Picton PE, Picton TW, Champagne SC, Durieux-Smith A (1995) Auditory steady-state responses to tones amplitude-modulated at 80–110 Hz. J Acoust Soc Am 97:3051–3063.

Miller GA, Licklider JC (1950) The intelligibility of interrupted speech. J Acoust Soc Am 22:167–173.

Momtaz S, Moncrieff D, Bidelman GM (2021) Dichotic listening deficits in amblyaudia are characterized by aberrant neural oscillations in auditory cortex. Clin Neurophysiol 132:2152–2162.

Momtaz S, Moncrieff D, Ray MA, Bidelman GM (2022) Children with amblyaudia show less flexibility in auditory cortical entrainment to periodic non-speech sounds. Int J Audiol 62:920–926.

Moore BCJ, Glasberg BR, Baer T (1997) A model for the prediction of thresholds loudness and partial loudness. J Audio Eng Soc 45:224–240.

Moreno S, Bidelman GM (2014) Examining neural plasticity and cognitive benefit through the unique lens of musical training. Hear Res 308:84–97.

Morillon B, Schroeder CE, Wyart V, Arnal LH (2016) Temporal prediction in lieu of periodic stimulation. J Neurosci 36:2342–2347.

Moumdjian L, Buhmann J, Willems I, Feys P, Leman M (2018) Entrainment and synchronization to auditory stimuli during walking in healthy and neurological populations: a methodological systematic review. Front Hum Neurosci 12:263.

Nguyen TT, Neubig G, Shindo H, Sakti S, Toda T, Nakamura S (2015) A latent variable model for joint pause prediction and dependency parsing. In Sixteenth Annual Conference of the International Speech Communication Association.

Novembre G, Iannetti GD (2018) Tagging the musical beat: neural entrainment or event-related potentials? Proc Natl Acad Sci U S A 115:E11002–E11003.

Nozaradan S, Peretz I, Keller PE (2016) Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. Sci Rep 6:20612.

Oganian Y, Chang EF (2019) A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci Adv 5:eaay6279.

Oganian Y, Kojima K, Breska A, Cai C, Findlay A, Chang E, Nagarajan SS (2023) Phase alignment of low-frequency neural activity to the amplitude envelope of speech reflects evoked responses to acoustic edges, not oscillatory entrainment. J Neurosci 43:3909–3921.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Oostenveld R, Praamstra P (2001) The five percent electrode system for high-resolution EEG and ERP measurements. Clin Neurophysiol 112:713–719.

Peelle JE, Davis MH (2012) Neural oscillations carry speech rhythm through to comprehension. Front Psychol 3:320.

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.

Picton T (2013) Hearing in time: evoked potential studies of temporal processing. Ear Hear 34:385–401.

Picton TW, Alain C, Woods DL, John MS, Scherg M, Valdes-Sosa P, Bosch-Bayard J, Trujillo NJ (1999) Intracerebral sources of human auditory-evoked potentials. Audiol Neurootol 4:64–79.

Pikovsky A, Rosenblum M, Kurths J (2002) Synchronization: a universal concept in nonlinear science. Cambridge: The press syndicate of the University of Cambridge.

Poeppel D, Assaneo MF (2020) Speech rhythms and their neural foundations. Nat Rev Neurosci 21:322–334.

Price CN, Alain C, Bidelman GM (2019) Auditory-frontal channeling in α and β bands is altered by age-related hearing loss and relates to speech perception in noise. Neuroscience 423:18–28.

R Core Team (2013) R: a language and environment for statistical computing.

Rimmele JM, Morillon B, Poeppel D, Arnal LH (2018) Proactive sensing of periodic and aperiodic auditory patterns. Trends Cogn Sci 22:870–882.

Rosso M, Leman M, Moumdjian L (2021) Neural entrainment meets behavior: the stability index as a neural outcome measure of auditory–motor coupling. Front Hum Neurosci 15:668918.

Sarvas J (1987) Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys Med Biol 32:11–22.

Scherg M, Ille N, Bornfleth H, Berg P (2002) Advanced tools for digital EEG review: virtual source montages, whole-head mapping, correlation, and phase analysis. J Clin Neurophysiol 19:91–112.

Soltanzadeh MJ, Daliri MR (2014) Evaluation of phase locking and cross correlation methods for estimating the time lag between brain sites: a simulation approach. Basic Clin Neurosci 5:205–211.

Steinschneider M, Reser DH, Fishman YI, Schroeder CE, Arezzo JC (1998) Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. J Acoust Soc Am 104:2935–2955.

Tallal P, Miller S, Fitch RH (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann N Y Acad Sci 682:27.

Thut G, Schyns PG, Gross J (2011) Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. Front Psychol 2:170.

Toplak ME, Dockstader C, Tannock R (2006) Temporal information processing in ADHD: findings to date and new methods. J Neurosci Methods 151:15–29.

Trainor LJ, Shahin AJ, Roberts LE (2009) Understanding the benefits of musical training: effects on oscillatory brain activity. Ann N Y Acad Sci 1169:133–142.

Viemeister NF (1973) Temporal modulation transfer functions for audition. J Acoust Soc Am 53:312.

Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am 66:1364–1380.

Vigil D, Pinto D (2020) An experimental study of the detection of clicks in English. Pragmat Cogn 27:457–473.

Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. J Neurophysiol 107:2033–2041.

Wilsch A, Henry MJ, Herrmann B, Maess B, Obleser J (2015) Slow-delta phase concentration marks improved temporal expectations based on the passage of time. Psychophysiology 52:910–918.

Xu N, Zhao B, Luo L, Zhang K, Shao X, Luan G, Wang Q, Hu W, Wang Q (2023) Two stages of speech envelope tracking in human auditory cortex modulated by speech intelligibility. Cereb Cortex 33:2215–2228.

Yrttiaho S, Tiitinen H, May PJ, Leino S, Alku P (2008) Cortical sensitivity to periodicity of speech sounds. J Acoust Soc Am 123:2191–2199.