eNeuro

Research Article: Confirmation | Cognition and Behavior

Acoustically driven cortical delta oscillations underpin prosodic chunking

https://doi.org/10.1523/ENEURO.0562-20.2021

Cite as: eNeuro 2021; 10.1523/ENEURO.0562-20.2021

Received: 27 December 2020
Revised: 5 May 2021
Accepted: 9 May 2021

1. **Title:** Acoustically driven cortical delta oscillations underpin prosodic chunking
2. **Abbreviated Title:** Delta oscillations underpin prosodic chunking
3. Rimmele, J.M.[1,2*], Poeppel, D. [1,2,3] Ghitza, O.[1,4]

[1]Department of Neuroscience

Max-Planck-Institute for Empirical Aesthetics

Grüneburgweg 14

60322 Frankfurt am Main, Germany

[2]Max Planck NYU Center for Language, Music and Emotion
Frankfurt am Main, Germany, New York, USA

[3]Department of Psychology and Center for Neural Science

New York University

6 Washington Place New York , NY 10003, USA

[4]Department of Biomedical Engineering & Hearing Research Center

Boston University

44 Cummington Mall,

Boston, MA, 02215, USA

4. **Author Contributions:** Each author must be identified with at least one of the following: Designed research, Performed research, Contributed unpublished reagents/ analytic tools, Analyzed data, Wrote the paper. Example: CS and JS Designed Research; MG and GT Performed Research; JS Wrote the paper

JR: designed research, performed research, analyzed data, wrote the paper

DP: designed research, wrote the paper

OG: designed research, analyzed data, contributed analytic tools, wrote the paper

5. **Correspondence should be addressed to (include email address)**

johanna.rimmele@ae.mpg.de (J.M. Rimmele), Department of Neuroscience, Max, Planck-Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany

6. **Number of Figures:** 9
7. **Number of Tables:** 0
8. **Number of Multimedia**: 0

13. **Acknowledgements:** This work was funded by the Max-Planck-Institute for Empirical Aesthetics, supported by CLaME Max Planck NYU Center for Language Music and Emotion, and by a research grant from the US Air Force Office of Scientific Research [grant number FA9550-18-1-0055]. We thank Simone Franz and Daniela van Hinsberg for help with the data recording.
14. **Conflict of Interest:** The authors declare no competing financial interests.
15. **Funding sources:** This work was funded by the Max-Planck-Institute for Empirical Aesthetics, supported by CLaME Max Planck NYU Center for Language Music and Emotion, and by a research grant from the US Air Force Office of Scientific Research [grant number FA9550-18-1-0055].
16. **Number of words for Abstract:** 155
17. **Number of words for Significance Statement:** 118
18. **Number of words for Introduction:** 1001
19. **12. Number of words for Discussion:** 2317

Abstract

Oscillation-based models of speech perception postulate a cortical computational principle by which decoding is performed within a window structure derived by a segmentation process. Segmentation of syllable-size chunks is realized by a theta oscillator. We provide evidence for an analogous role of a delta oscillator in the segmentation of phrase-sized chunks. We recorded Magnetoencephalography (MEG) in humans, while participants performed a target identification task. Random-digit strings, with phrase-long chunks of two digits, were presented at chunk rates of 1.8 Hz or 2.6 Hz, inside or outside the delta frequency band (defined here to be 0.5 - 2 Hz). Strong periodicities were elicited by chunk rates inside of delta in superior, middle temporal areas and speech-motor integration areas. Periodicities were diminished or absent for chunk rates outside delta, in line with behavioral performance. Our findings show that prosodic chunking of phrase-sized acoustic segments is correlated with acoustic-driven delta oscillations, expressing anatomically specific patterns of neuronal periodicities.

Significance Statement

Oscillation-based models of speech perception postulate a cortical computational principle by which decoding is performed within a time-varying window structure, synchronized with the input on multiple time scales. At pre-lexical level, cycles of a *flexible theta oscillator* – locked to the input syllabic rhythm – constitute the syllabic windows. We find that the presence of cortical delta oscillations correlates with whether or not an input phrase-sized chunk rate is inside the delta range**.** This suggests that at the phrase time scale, a delta oscillator could play a role analogous to that of the theta oscillator at the syllable level. The segmentation process is realized by a *flexible delta oscillator* locked to the input rhythm, with delta cycles constituting phrase-sized windows.

Introduction

Naturally spoken language is a stream of connected sounds, and although the speech acoustics contain no cues regarding the beginning or end of linguistic units a combination of interleaved cues (e.g., pauses, intonation, syllabic stress) are embedded in the acoustic stream. Information, broadly termed 'accentuation' (e.g. intonation, stress, pauses), is used by listeners to indicate boundaries associated with linguistic units (Aubanel et al., 2016; Oganian and Chang, 2019). The process by which the input signal is partitioned into temporal segments to be linked to a variety of linguistic levels of abstraction (ranging from phonetic segments to syllables to words and, ultimately, prosodic and intonational phrases) is called 'segmentation'.

The segmentation process has been shown to operate on time intervals associated with syllables (up to about 250 ms) (Brungart et al., 2007; Doelling et al., 2014; Ghitza and Greenberg, 2009; Kösem et al., 2018), and a similar process has been suggested to operate on the phrasal level (0.5–2 s) (Ding et al., 2016; Ghitza, 2017; Keitel et al., 2018; Martin and Doumas, 2017). At the syllabic level, perceptual segmentation—or chunking—is by and large a pre-lexical process. Oscillation-based models propose that this segmentation is realized by flexible theta oscillations aligning their phase to the input syllabic rhythm ('speech tracking'), where the theta cycles mark the speech chunks to be decoded (e.g., Ahissar and Ahissar, 2005; Assaneo et al., 2020; Ding and Simon, 2009; Ghitza, 2011; Giraud and Poeppel, 2012; Gross et al., 2013; Haegens and Zion Golumbic, 2017; Hovsepyan et al., 2020; Lakatos et al., 2019, 2005; Peelle and Davis, 2012; Pittman-Polletta et al., 2020; Poeppel, 2003; Rimmele et al., 2018). At the phrase level, phrase rhythm can affect segmentation (Gee and Grosjean, 1983; Martin, 2015; Deniz and Fodor, 2019; Hilton and Goldwater, 2020). There have been various studies aiming to quantify phrase length and rhythmicity (e.g., Breen, 2018; Clifton et al., 2006; Deniz and Fodor, 2019), suggesting that

typical intonational phrases are about one second in duration (Auer et al., 1999; Inbar et al., 2020; Stehwien and Meyer, 2021). More specifically, the duration of intonational phrases spans a range between approximately 0.5 to 1 sec in English (slightly faster in some other languages) (Inbar et al., 2020; Stehwien and Meyer, 2021). Prosodic segmentation (here also termed 'prosodic chunking') is based on intonation units that contain specific prosodic cues (such as pauses or pitch contour), which can pace the information flow at the phrasal time scale (Inbar et al., 2020; Shattuck-Hufnagel and Turk, 1996). The extent to which phrase level rhythmic structure supports segmentation and structural parsing was not widely studied. Here, we investigate the neural processing of rhythmic structure at the phrasal scale by analyzing how individuals' group single digits into 'phrase-sized' digit chunks. What kind of neuronal mechanism can realize this chunking process?

Cortical delta oscillators – with a frequency range (~ 0.5 – 2 Hz) that corresponds to the phrasal time scale – were shown to be elicited during phrasal processing of speech or chunking processes at the phrasal scale (Buiatti et al., 2009; Ding et al., 2016; Meyer et al., 2016; Bonhage et al., 2017; Keitel et al., 2018; Boucher et al., 2019). Delta was observed in the posterior superior temporal, the inferior frontal, precentral, and temporal-parietal cortex using ECoG (Ding et al., 2016), and using EEG at bilateral middle and superior temporal areas (also fusiform gyrus) (Bonhage et al., 2017) and at fronto-temporal sites (Boucher et al., 2019). (Recall the ambiguous definition of the delta range in the literature, which covers a range of overlapping frequency bands inside the 0.5 to 4 Hz frequency range (e.g, Bonhage et al., 2017; Bröhl and Kayser, 2020; Keitel et al., 2018). Since we are interested in the segmentation of phrasal chunks, which in English are about 0.5 to 1 sec long (e.g., Miller, 1962; Inbar et al., 2020), we opted to define the delta frequency band to be 0.5 - 2 Hz).

And behaviorally, it has been shown that performance is impaired when the chunk rate is outside of the delta range (Ghitza, 2017). These findings suggest a role of neuronal oscillatory mechanisms in the delta range in chunking at a phrasal time scale (see also: Ghitza, 2020; Martin and Doumas, 2017). Little is known, however, about the brain areas that may recruit chunking related delta oscillations.

Here we focus on the cortical mechanism that may be involved in *acoustic-driven* segmentation at a phrasal time scale, using sequences of digit chunks (with a minimal amount of content). We test the hypothesis that the decoding process is guided by a delta oscillator locked to the accentuation acoustic cues (Ghitza, 2017) by recording MEG data while participants performed a digit retrieval task. The digits in the string were grouped into chunks, with chunk rates either inside or outside to the delta frequency range (Fig. 1). The experiment addresses two questions: (1) Do elicited delta brain waves correlate with behavior, such that impaired performance in digit retrieval occurs if the chunk rate is outside of the delta range? (2) Where in the auditory pathway do those neuronal oscillations occur?

Our data show that in superior and middle temporal areas and in speech-motor planning and integration areas (IFG, PC, SMG), robust neural delta periodicities were elicited by chunk rates inside of the delta range but were diminished when the chunk rate was outside of the delta range, in line with behavioral performance. In speech-motor integration areas (SMG) and areas implicated in processing word form and meaning (MTG), periodicity was present albeit diminished even for chunk rates inside the delta range. The delta periodicities were acoustically driven, in sync with the input as long as the acoustic chunk rate was inside delta. Delta periodicities were diminished for chunk rates outside delta, even though sufficient acoustic cues for chunking were present in all conditions. Thus, the failure to track the input-chunk-rate when it

was outside of the delta range was not caused by insufficient acoustic cues but seems due to neuronal circuitry characteristics constraining the tracking of the chunks.

Materials and Methods

Participants. The data from 19 healthy right-handed (Oldfield, 1971) mean score: 75.22, SD: 18.08) participants were included in the study (mean age: 24.89 years, SD: 3.54; f = 14). Human subjects were recruited from the local community in Frakfurt. Two participants were excluded because of technical issues, and one participant because of outlier performance (i.e., performance < mean performance -2SD). Individual MRI scans were recorded for all except for two participants who did not fulfill the MRI scanning criteria. All participants gave written informed consent for participating in the study and received monetary compensation. The study was approved by the local ethics committee of the University Hospital Frankfurt (SG2-02-0128).

Digit string stimuli. We used 10-digit long stimuli where we manipulated the pauses in-between digits according to the experimental conditions. We opted for digit sequences –material that is semantically unpredictable at the digit-chunk level (i.e., while semantic information is present at the single digit level, no semantic/contextual information is present at the digit-chunk level) – in order to minimize the bottom-up/top-down interaction that is in play in setting perceptual boundaries for digit-chunks. The digit sequences were grouped into chunks, with a chunk pattern termed 2222. For example, the 2222 pattern of the sequence 4259522560 is [42 59 52 25 60]. Digits were presented as single digits, i.e. 42 was read as four-two and not as forty-two.

We used two chunk rates: 1.8 Hz (inside the delta range) and 2.6 Hz (at the outside border of the delta range, referred to as “outside”), termed conditions ‘1.8Hz’ and ‘2.6Hz’ (Fig. 1). (Note that a third condition was used, which is not reported here. The condition was a ‘no-chunk’ condition where digit chunks were presented at the rate of 2.6Hz. However, besides top-

down chunking information (provided by the instructions), there were no acoustic chunking cues. The neuronal findings resemble that of the 2.6Hz chunking condition, confirming the main claims of this paper. They are reported elsewhere, Rimmele et al., 2020).

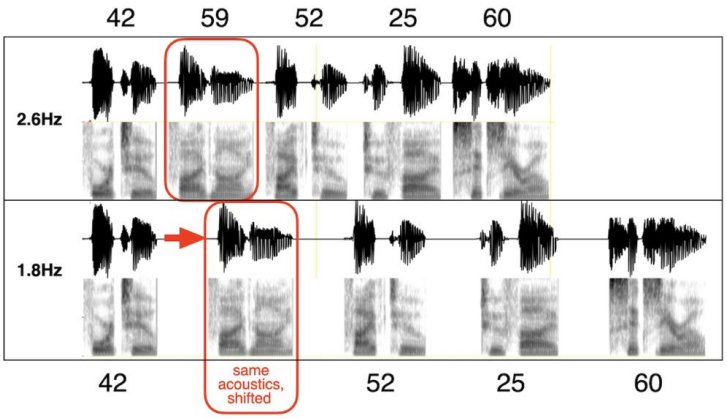


Figure 1. Chunk patterns and chunk rates for the 10-digit digit string 4259522560. The chunk pattern is 2222, with chunk rates of 1.8 Hz (inside delta) and 2.6 Hz (outside). Each chunk was synthesized as a 2-digit unit, using the AT&T Text-to-Speech System accentuation (see text). Note that a particular 2-digit chunk has the same acoustics, regardless of whether it occurs in the 1.8Hz or 2.6Hz 2222 chunk condition (red box). The 1.8Hz stimulus is generated by increasing the gap between the chunks (with identical chunk acoustics).

Corpus. The text material comprised 100 10-digit long text strings. Stimuli were generated by using the AT&T Text-to-Speech System with the American English female speaker Crystal. [The AT&T-TTS system (http://www.wizzardsoftware.com/text-to-voice.php) uses a form of concatenative synthesis based on a unit-selection process, where the units are cut from a large, high-quality, pre-recorded natural voice fragments. The system produces natural-sounding, highly intelligible spoken material with a realistic prosodic rhythm—with accentuation defined by the system-internal prosodic rules—and is considered to have some of the finest quality synthesis of any commercial product.] To generate stimuli with a 2222 chunk pattern, we first created a 2-

digit chunk vocabulary as follows. For each doublet of digits that exists in the 100 text strings, a naturally sounding 2-digit chunk waveform was generated (naturalness was obtained by the AT&T system accentuation rules) resulting in a chunk-vocabulary. For a given text string, a 2222 10-digit stimulus was generated by concatenating the corresponding five 2-digit chunk waveforms pulled from the chunk-vocabulary. The *chunk rate* was set by adjusting the gap duration in between two successive chunks, resulting in a stimulus with a temporal structure but without any contextual cues at the digit-chunk level. To enable the generation of stimuli with chunk rates greater than the delta frequency upper bound (at 2.6 Hz), the waveforms in all conditions were time compressed by a factor of 2.5, just below the auditory channel capacity (Ghitza, 2014). The duration of the 10-digit stimuli varied across conditions; for the 1.8Hz condition: mean = 2.61 sec (VAR = 85.6 msec); and for the 2.6Hz condition: mean = 1.99 sec (VAR = 85.6 msec).

For each of the 200 10-digit stimuli (100 stimuli for each of the 1.8Hz and 2.6Hz conditions) a trial was created by concatenating the following waveform sequence: [one digit trial-count] [20-msec long gap] [10-digit string] [500-msec long gap] [2-digit target], resulting in one concatenated waveform per trial with durations that varied across trials and conditions. The 200 trials were scrambled, and the resulting pool of trials was divided into blocks, 50 trials per block. A jittered intertrial interval of 3-4.5 sec was presented between trials. Overall, two different sets of stimuli were used.

***Task*.** Behavioral and MEG data were collected while participants performed a digit retrieval task, in the form of an adapted Sternberg target identification task (Sternberg, 1966) (target ID task from here on): listeners heard a 10-digit stimulus followed by a 2-digit long target, and were asked to indicate whether or not the target was part of the utterance. A target position was always within a chunk. Note that the task is suitable for probing the role of acoustic segmentation in a

memory retrieval task: a successful yes/no decision depends on how faithful the recognized chunk objects are, generated by a decoding process that, by hypothesis, depends on the goodness of segmentation.

Procedure and Paradigm. Participants were seated in front of a board for instructions in the MEG testing booth. Binaurally insert earplugs (E-A-RTONE Gold 3A Insert Earphones, Ulrich Keller Medizin-Technik, Weinheim, Germany) were used for stimulus presentation. Two button boxes (Current Designs, Inc.) were used to record participants' responses. The Psychophysics Toolbox (Brainard, 1997) was used to run the experiment. During the experiment, on each trial participants fixated the screen center (fixation cross) while listening to the digit sequences. The sounds were presented at a comfortable loudness level (~70 dB SPL), which remained unchanged throughout the experiment. Overall, the experiment lasted about 2.5 hours, including preparation time, recording time, breaks, and post-recording questionnaires. Participants were presented with the task requirements. They were instructed that all sequences comprise concatenated chunks of two-digits. Prior to the experiment, all participants performed a short training of three trials (with feedback) in order to familiarize themselves with the stimuli and task. Participants were asked to indicate by button press (yes/no response; with the response hand balanced across participants; yes-hand right: N = 12) whether or not the target was part of the preceded utterance.

MRI and MEG Data Acquisition. A 3 Tesla scanner (Siemens Magnetom Trio, Siemens, Erlangen, Germany) was used to record individual T1-weighted MRIs. MEG recordings were performed on a 269-channel whole-head MEG system (Omega 2000, CTF Systems Inc.) in a magnetically shielded booth. Data were acquired with a sampling rate of 1200 Hz, online denoising (higher-order gradiometer balancing) and online low pass filtering (cut-off: 300 Hz) was applied. Continuous tracking of the head position relative to the MEG sensors allowed

correction of head displacement during the breaks and prior to each file saving of a participant, using the fieldtrip toolbox (http://fieldtrip.fcdonders.nl) (Stolk et al., 2013).

Behavioral Analysis. A "yes–no" model for independent observations was used to compute dprime (Green and Swets, 1966). Four classes of response are considered: (1) Hit: a "yes" response when the target chunk is present in the digit sequence, (2) Correct Rejection: a "no" response when the target chunk is absent, (3) Miss: a "no" response when the target chunk is present, and (4) False Alarm: a "yes" response when the target chunk is absent. Nonparametric Wilcoxon signed-rank tests (two-sided) were used to test differences in the mean dprime across conditions. The Bayes factor BF_{10} (Schönbrodt and Wagenmakers, 2018), which reflects the likelihood data arose from the alternative model, was computed using the software JASP (JASP Team, 2020) (10000 samples) and default priors.

MRI Analysis. The FieldTrip toolbox (http://fieldtrip.fcdonders.nl) (Oostenveld et al., 2011) was used for the MRI and MEG data analyses. The standard Montreal Neurological Institute (MNI) template brain was used for participants where an individual MRI was missing. Probabilistic tissue maps (cerebrospinal fluid gray and white matter) were constructed from the individual MRIs. Next, a single shell volume conduction model (Nolte, 2003) was applied to retrieve the physical relation between sensors and sources. Between the individual T1 MRI and the MNI template T1 a linear warp transformation was computed. A 8 mm template grid, defined on the MNI template T1, was warped on the individual head space by inversely transforming it, based on the location of the coils during the MEG recording and the individual MRI. Next, based on the warped MNI grid and the probabilistic tissue map a forward model was computed, and applied for source reconstruction. This allowed aligning the grids of all participants to each other in MNI space for the across participants statistics.

MEG Preprocessing. Line-noise was removed using bandstop filters (49.5-50.5, 99.5-100.5, two-pass; filter order 4) and the data were band-pass filtered off-line (0.1–100 Hz, Butterworth filter; filter order 4). A common semi-automatic artifact detection procedure was applied: for artifact rejection, the signal was filtered to identify muscular artifacts (band-pass: 110-140 Hz) or jump artifacts (median filter) and z-normalized per time point and sensor. The z-scores were averaged over sensors, in order to accumulate evidence for artifacts that occur across sensor. Trials that exceeded a predefined z-value (muscular artifacts, z = 15; jumps, z = 30) were rejected. Trials were the range (min-max difference) in any sensor exceeded a threshold (threshold = 0.75e-5) were identified as containing slow artifacts and rejected. Down-sampling to 500 Hz was applied. The data were epoched (-3.5 to 5 sec). Furthermore, when head movements exceeded a threshold (5 mm) a trial was rejected. Next, all blocks of recorded MEG data were concatenated. If high variance was detected at any sensor, the sensor was rejected. Finally, independent component analysis (infomax algorithm; Makeig et al., 1996) was used to remove eye-blink, eye-movement and heartbeat-related artifacts based on cumulative evidence from the component topography and time course.

MEG source-level analysis. In a first step, the data were epoched (0-5 sec). For the main analyses, only trials in which participants showed Correct responses (i.e. hits and correct rejections) were selected. Next, the sensor-space measurements were projected and localized in source-space inside the brain volume (Van Veen et al., 1997) using Linearly Constrained Minimum Variance (LCMV) beamforming. A spatial filter was computed based on the individual leadfields for each participant and condition (lambda = 10%; 0.8 cm grid). Next, all trials were epoched to the minimum stimulus duration in the corresponding condition (condition 1.8Hz: 2.38 sec; condition 2.6Hz: 1.68 sec).

Cortical regions of interest (ROIs). The automated anatomical labeling atlas (AAL; (Tzourio-Mazoyer et al., 2002) was used to select the regions of interest (ROIs) as follows (Fig. 2):

- STG (Temporal_sup_L/R): Auditory association areas (Binder et al., 2009; Hickok and Poeppel, 2007)
- MTG (Temporal_Mid_L): Implicated in processing word form and meaning
- IFG (Frontal_Inf_Tri_L/R): Involved in speech-motor planning
- PC (Precentral_L/R), SMG (SupraMarginal_L/R): Speech-motor integration
- Calcariane (Calcarine_L/R): Primary visual cortex (as a control region)

We opted to omit Heschl's Gyrus (primary auditory cortex area) from the list of ROIs because of the very small number of voxels (3 in the Left, 2 in the Right).

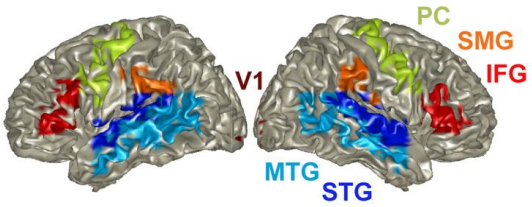


Figure 2. Cortical regions of interest (ROIs). The automated anatomical labeling atlas (AAL; (Tzourio-Mazoyer et al., 2002) was used to select ROIs in left and right superior temporal gyrus (STG), middle temporal gyrus (MTG) and speech-motor planning and integration areas (IFG, PC, SMG). V1 was used as control region. ROIs are color coded.

Periodicity density function (PDF) within ROI. We aim to determine whether the elicited brain signal measured at any given voxel within a specific ROI shows periodicity, and if so, to extract the frequency. Ultimately, we seek to characterize the density function of the periodicities across all voxels in the ROIs of interest.

The aggregated cross-correlation measure (XCOV) of periodicity. To measure the neural response periodicity in individual voxels one could use one of several widely used measures, e.g., autocorrelation—where the first nontrivial peak indicates the period, or the intertrial phase coherence (ITPC)—where the outcome would be the frequency distribution of the coherence function. Importantly, these measures build on the number of trials, M. The trial signals are noisy, both due to the SNR and due to the brain wave irregularity (which is why these methods average over trials). But what if M is too small? Here we used a newly proposed measure, termed 'Aggregated cross-correlation' (abbreviated XCOV), to measure periodicity across M trials. Broadly, we suggest taking advantage of the fact that, for M trials, we can generate about $M^2/2$ cross-correlation functions. Recall that, unlike autocorrelation, the first peak of a cross-correlation function does not indicate the period but rather the delay between the two signals. Therefore, we run each of the $M^2/2$ candidate pairs through a "match filter", which determines whether the corresponding two signals have a "zero" delay. Such a pair will have a cross-correlation function similar to that of an autocorrelation function, i.e., its peak is at zero and its earliest nontrivial peak is at the period. Only the pairs that pass the test are cross-correlated and aggregated. Obviously, the number of cross-correlation functions qualified for aggregation is between M and $M^2/2$, depending on how strict the match filter is. (For example, in the STG ROI, the mean number of trials over subjects for the 'Hit' response was M=38, with a mean number of pairs of 703. The mean number of pairs that passed the test was 433 for the 1.8Hz condition and 378 for the 2.6Hz condition – about one degree of order bigger than M. A similar trend was observed for all ROIs.) We term the outcome of this measure as the 'XCOV' function.

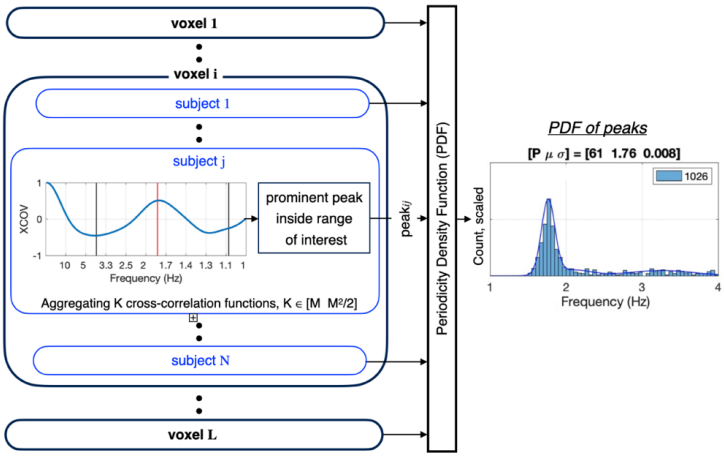


Figure 3. Analysis pipeline for deriving the periodicity density function (PDF) within a particular ROI. Shown is the resulting periodicity PDF for a given condition (say, the 1.8Hz condition), a given response class (say, Hit), and a given ROI (say, STG). L voxels, N subjects, and M trials per subject are considered. For the i-th voxel and the j-th subject, periodicity is computed using a newly proposed measure method termed 'aggregated cross-correlation', abbreviated XCOV. First (not shown), each brain signal is filtered to the frequency range of interest. Cross correlations were computed using the filtered signals. Note that as a cross-correlation function, XCOV is computed against time-lags; the abscissa here shows the time-lag *inverse*, in frequency, hence going from right to left. The *periodicity density function* (PDF) is derived by (i) forming a histogram of the XCOV peak locations inside the frequency range of interest, (ii) normalizing the histogram to L×N, the total number of data points, and (iii) building a 3rd order Gaussian Mixture Model (GMM) that fits the histogram. The GMM model is the desired PDF. The "goodness" of the PDF is quantified by in terms of P value, the percentage of datapoints inside the frequency range of interest with respect to the total number of datapoints (L×N); and the mean μ and variance σ of the prominent Gaussian component of a 3rd order GMM. (The total number of data points is shown in the inset of each entry.)

Periodicity density function (PDF) within a ROI. Fig. 3 details the analysis pipeline for deriving the periodicity density function of the periodicities within a particular ROI. L voxels, N subjects, and M trials per subject are considered. First (not shown), each brain signal is filtered to the frequency range of interest (low pass filter with cutoff frequency of 6 Hz for the (inside/outside) delta chunk rate analysis (Figs. 5 and 6 and 9a); and a bandpass filter with a [2-10] Hz frequency range, for theta (single digit rate) analysis (Fig. 7)). (The filters were chosen with a bandwidth wider than the expected mean periodicities (1.8- and 2.6 Hz for delta, about 4 Hz for theta), in

order to let the XCOV analysis determine the periodicity PDFs without any bias.) Cross correlations were computed using the filtered signals. Shown is the XCOV function at the i-th voxel, for the j-th subject, obtained by aggregating K cross-correlation functions. (Note that as a cross-correlation function, XCOV is computed against time-lags; the abscissa here shows the time-lag *inverse*, in frequency, hence going from right to left). The particular XCOV function shown in Fig. 3 has a single peak at 1.76 Hz but note that, in general, an XCOV may have more than one local peak. Next, the location of the prominent peaks is extracted, with the number of prominent peaks as a parameter. (The prominence of a peak measures how much the peak stands out due to its intrinsic height and its location relative to other peaks in the range of interest.) In our analysis one prominent peak per XCOV is considered. Hence, for L voxels and N subjects, a maximum of L×N data points are available to construct a histogram, from which only those inside the frequency range of interest are used, and the resulting histogram is normalized to L×N. A 3rd order Gaussian mixture model (GMM) that fits the histogram is the desired PDF. The "goodness" of the periodicity is quantified by in terms of P, the percentage of datapoints inside the frequency range of interest with respect to the total number of datapoints (L×N); and the mean μ and variance σ of the prominent Gaussian component of a 3rd order GMM. (The total number of data points is shown in the inset of each entry.)

Results

***Behavioral results*.** Dprime scores were the highest in the 1.8Hz condition (mean = 2.19, SD = 0.49) (Fig. 4), i.e., when the chunk rate is inside the delta frequency range. Lower dprime scores were registered in the 2.6Hz condition (mean = 1.74, SD = 0.42), when the chunk rate was just at the outside edge of the delta range. The difference in scores was significant (1.8Hz condition vs. 2.6 Hz: $W = 177$, $p < .001$, $r = .863$; $BF_{10} = 199.6$).

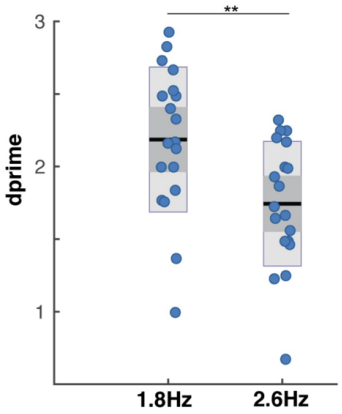


Figure 4. Behavioral performance in the digit retrieval task. Dprime values are displayed, as measure of performance accuracy, separately for each condition. Blue dots indicate individual dprime scores, black lines indicate the mean dprime scores, dark gray bars indicate the +/-1 standard error of the mean and light gray bars the confidence interval. Significance is indicated by **($p < .01$). The performance was higher in the 1.8Hz acoustical chunk (inside delta chunking), compared to the 2.6Hz acoustical chunk condition (outside of delta) (replicating findings in [7]).

Periodicity density function (PDF) of elicited brain waves. We used the aggregated cross-correlation measure (XCOV) of periodicity across M trials to determine whether the elicited brain signal measured at any given voxel within a specific ROI shows periodicity, and if so, to extract the frequency. Then, we derived the periodicity density function of the periodicities across all voxels in the ROIs of interest (Fig. 2). The "goodness" of the periodicity is quantified by in terms of P, the percentage of datapoints inside the frequency range of interest with respect to the total number of datapoints (L voxels ×N subjects); and the mean μ and variance σ of the prominent Gaussian component of a 3rd order GMM. Fig. 5 shows the periodicity PDFs in the [1 4] Hz frequency range for the STG region of interest (ROI) in the left hemisphere. For the 1.8Hz condition, a strong periodic response at about 1.8 Hz was recorded for the Hits and Correct Rejections, with the P over 50%. Much weaker presence of periodicity was recorded for the

Misses and False Alarms. A similar trend is shown for the 2.6Hz condition, albeit with much weaker periodicity compared to the 1.8Hz condition, and with a smaller P (of below 30%). Notice that, across chunk conditions, the PDF patterns for hits and correct rejections are similar, as are the patterns for misses and false alarms. Such similarities were observed for all ROIs. Therefore, in presenting the rest of the data, the hits and correct rejections are combined to indicate *Correct responses*, and the misses and false alarms are as *Erroneous responses*.

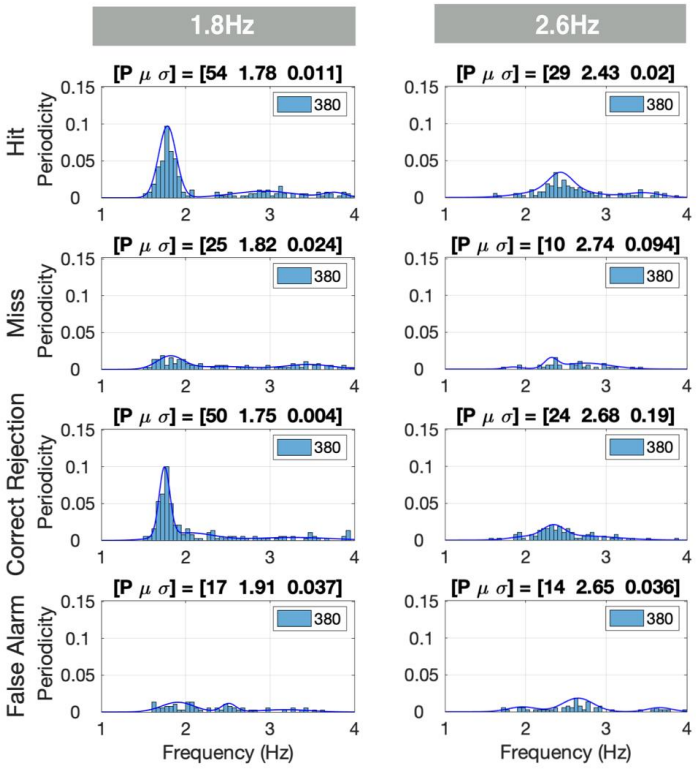


Figure 5. Periodicity density functions (PDFs) of delta periodicities per response class. The PDFs are displayed for the left hemisphere STG region of interest (ROI). The number of voxels in this ROI is 20 and the number of participants 19. Per voxel and subject, one XCOV peak inside the delta [1 4] Hz range was selected. The rows indicate the response classes (Hit, Miss, etc.), and the columns – the chunking conditions. Each entry shows the histogram (with the periodicity count scaled to L x N), and the corresponding PDF. The inset of each entry shows the total number of data points L x N analyzed (20 voxels × 19 subjects = 380 incidences). The

"goodness" of the PDF is quantified in terms of the percentage (P value) of datapoints inside the frequency range of interest with respect to the total number of datapoints, and the mean μ and variance σ of the prominent Gaussian component of a 3rd order GMM. For the 1.8Hz condition, a strong periodicity presence at 1.8 Hz was recorded for the Hit and Correct Rejection responses, with P over 50%. A much weaker presence was recorded for the Miss and False Alarm responses. A similar trend is shown for the 2.6Hz condition, albeit with much weaker periodicity presence compared to the 1.8Hz condition, and a smaller number of datapoints (P of below 30%).

In the following figures, the data are presented as follows. Each figure contains 6×2 entries organized in six rows (ROIs) and two columns (chunking conditions). Each entry shows the periodicity PDF, and the "goodness" of the periodicity is quantified in terms of P, μ and σ. In some selected entries, the upper left corner shows the Kullback-Leibler Divergence (KLD) of the entry's PDF with respect to a reference PDF defined in the respective figure caption. Finally, in some entries, no μ and σ values are present. This is so because of a failure of the 3rd order GMM to converge due to the small P value.

Figures 6A and 6B show the elicited responses in the [1 4] Hz frequency band for Correct responses (i.e., Hits and Correct Rejections combined), and Erroneous responses (i.e., Misses and False Alarms combined), respectively. We term these elicited responses *delta responses*. For Correct responses in the 1.8Hz condition a strong periodicity presence at about 1.8 Hz is recorded. A similar pattern is shown for the 2.6Hz condition, albeit with much weaker periodicity presence compared to the 1.8Hz condition (lower P value and wider σ). For Erroneous responses, for all ROIs, no presence of periodicities is recorded, for any condition. More specifically: for Correct responses, in the chunked conditions, the auditory association ROI (STG) shows a compelling periodicity presence at 1.8 Hz in the 1.8Hz condition and a weaker presence at 2.6 Hz in the 2.6Hz condition. At the middle temporal ROI (MTG), periodicity exists for the chunked conditions, albeit with 1.8 Hz periodicity stronger than that of 2.6 Hz. Similar patterns are observed in the speech-motor planning and integration ROIs (IFG, SMG, PC), whereas periodicity is present at 1.8 Hz, and is absent in the 2.6Hz condition. Note that in the visual ROI

(Calcarine), delta periodicities are absent for all conditions. Finally, the 1.8Hz condition column of Fig. 6A also shows the Kullback-Leibler Divergence (KLD) for all ROIs, with respect to the STG ROI (highlighted in red). The KLD values suggest similar patterns of elicited delta periodicities observed in the temporal brain regions (STG and MTG ROIs, with KLD value of 0.15 for MTG), and the frontal motor and temporal-parietal regions (IFG, SMG and PC ROIs, with KLD values of 0.15, 0.25 and 0.11, respectively).

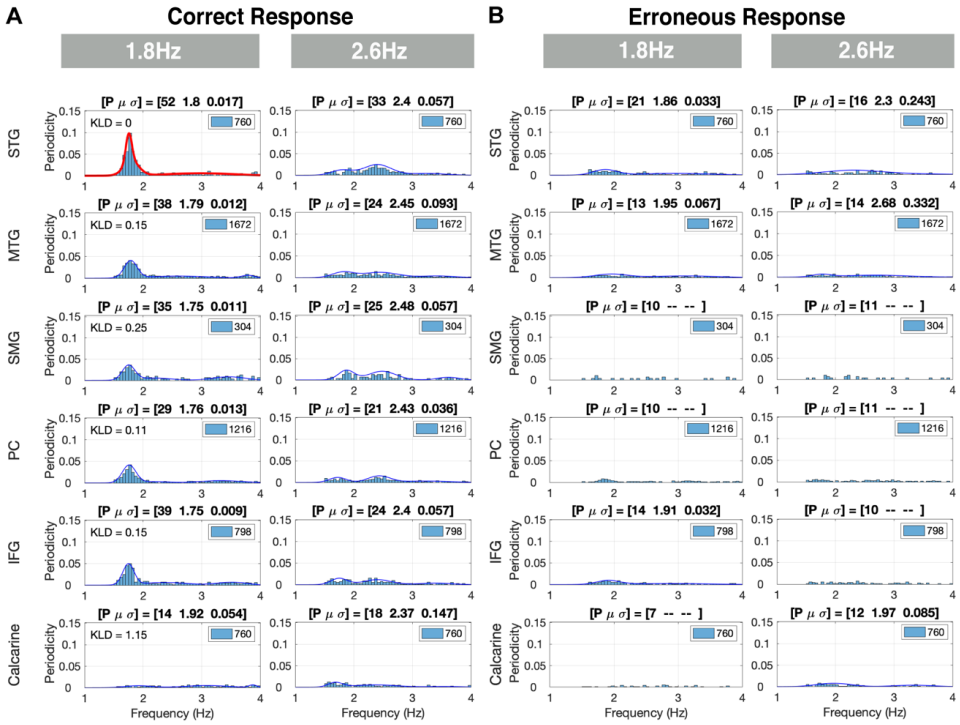


Figure 6. Delta periodicities for Correct and Erroneous responses in the left hemisphere. (A) Periodicities for Correct responses: Rows indicate the regions of interest (ROIs) and columns the chunking conditions. Each entry shows the histogram (with the periodicity count scaled to the L x N), and the PDF, quantified in terms of the percentage (P value) of datapoints inside the frequency range of interest with respect to the total number of datapoints (L x N, see inset). For the 1.8Hz condition a strong periodicity presence at about 1.8 Hz is recorded. A similar trend is shown for the 2.6Hz condition, albeit with much weaker periodicity presence compared to the 1.8Hz condition. The 1.8Hz condition column shows the Kullback-Leibler Divergence (KLD) computed for this condition at all ROIs, with respect to the STG ROI highlighted in red (upper left corner of the ROIs). The KLD values suggest similar patterns of elicited delta periodicities in the

temporal brain areas (STG and MTG ROIs), and in the frontal motor and temporal parietal areas (IFG, SMG and PC ROIs). (B) Periodicities for Erroneous responses: No presence of periodicities is recorded for any condition.

Furthermore, we compared the elicited delta responses in all ROIs in the Left versus the Right hemispheres for Correct responses. Similar periodicity PDFs are observed for all ROIs in all chunking conditions. The KLD was calculated for each ROI in the Right hemisphere against the corresponding Left ROI. The KLD values show a closer similarity between the periodicity PDFs of the left and right hemisphere of the temporal brain regions (STG and MTG, with KLD values of 0.1 and 0.11, respectively). In contrast, in the frontal motor and temporal-parietal regions periodicities were more prominent in the left compared to the right hemisphere (IFG, SMG, and PC, with KLD values of 0.28, 0.15 and 0.47, respectively).

Figures 7A and 7B show the elicited responses in the [2 6] Hz frequency band for the Correct and Erroneous responses, respectively, for ROIs in the left hemisphere. We term responses in this frequency band *theta responses*. For the Correct behavioral responses, strong theta was elicited in all ROIs and for all chunking conditions. Such elicited neural response patterns reflect the single digit presentation rate. Two observations are noteworthy, the bimodal characteristic of the PDFs for all chunking conditions, in particular for the 1.8Hz chunking condition, and the strong, unexpected, theta periodicity presence in the Calcarine ROI. For the Erroneous responses, a weaker more dispersed periodicity presence was observed. Finally, for the Correct responses, the periodicity PDFs were similar in shape across conditions, as was quantified by the KLD values comparing the periodicity PDFs in the 1.8 Hz condition with respect to the 2.6 Hz condition (KLD values between 0.13 to 0.2 across ROIs). The similarity of the PDFs across chunking conditions confirms that the decoding time at the digit level was sufficient across conditions.

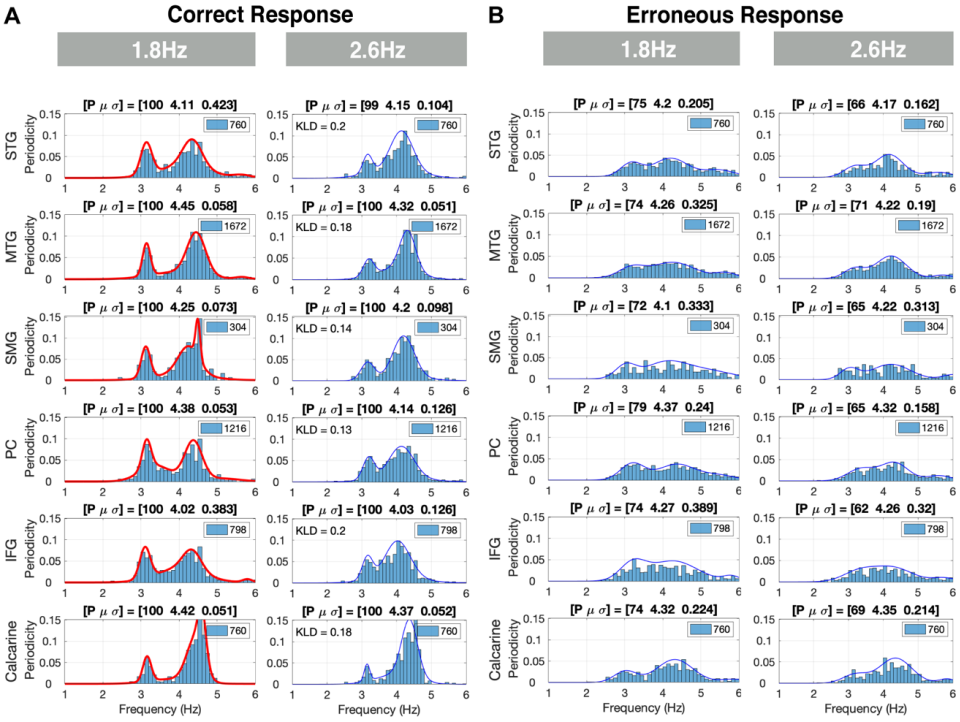


Figure 7. Theta periodicities for Correct and Erroneous responses in the Left hemisphere. (A) Periodicities for Correct responses: Strong theta periodicities were present in all ROIs and for all chunking conditions. Such elicited neural response patterns reflect the single digit presentation rate. The histograms are scaled to L x N (see inset). The PDFs are similar in shape across conditions, as is quantified by the KLD values comparing the PDFs in the 1.8 Hz condition with respect to the 2.6 Hz condition. (B) Periodicities for Erroneous responses: A weaker more dispersed presence of theta periodicities is recorded for all conditions (lower P value and wider σ).

***Correspondence between behavioral data and electrophysiological data*.** Fig. 8 quantifies the correspondence between the elicited delta periodicity patterns and the behavioral data. Shown are the 3rd order GMMs computed for the Correct responses in the left hemisphere and the two stimulus conditions. Unlike Fig. 6A, which shows PDF in terms of scaled periodicity count, shown here are the actual *probability densities* (with the $\int p(x)dx = 1$). The title of each panel

shows three measures: (i) [*dprime* σ] – the behavioral performance indicated by mean *dprime* values and the variance across subjects; (ii) [*Bias* σ] – the average of the absolute difference (termed Bias) between the mean of the prominent Gaussian component of the GMM and the acoustic chunk rate, and the variance across the ROIs; and (iii) [*P* σ] – the average P value and the variance across the ROIs. Two observations are noteworthy. First, the tightness of the PDFs in the 1.8Hz condition as reflected in the high probability value at the periodicity frequency, compared to the pseudo-uniform shape of the PDFs in the 2.6Hz condition. And second, the decrease in *dprime* accompanies the increase in *Bias* and the decrease in *P*. These data support the hypothesis that perceptual chunking at the time scale of phrase is derived by acoustic-driven delta oscillators.

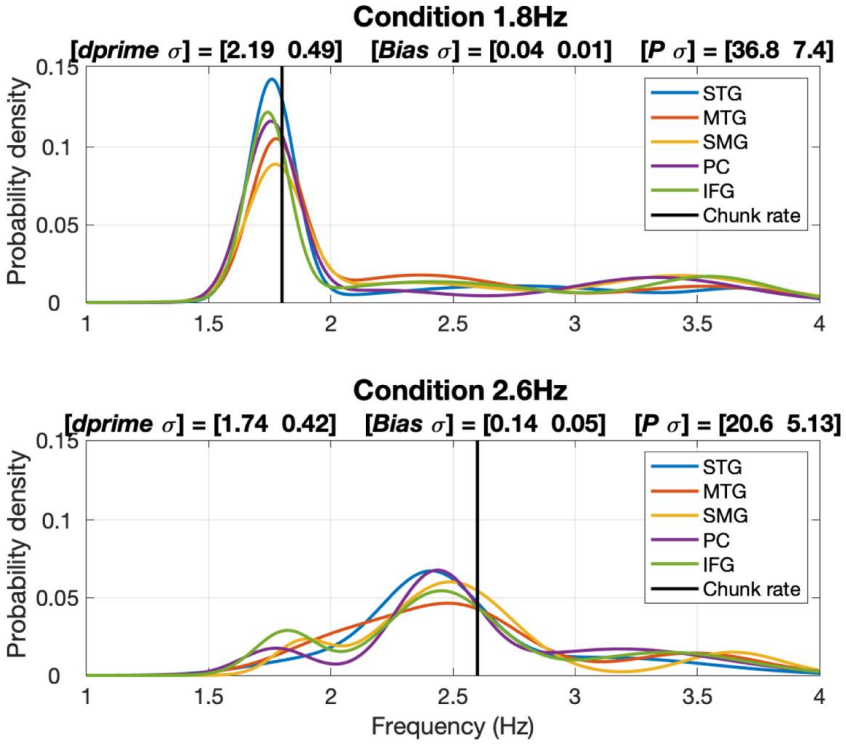


Figure 8. ***Correspondence between behavioral data and electrophysiological data*****.** Shown are the 3rd order GMMs Correct responses in the left hemisphere. Unlike Fig. 6A, which shows PDF scaled to P (the percentage of datapoints inside the frequency range of interest with respect to the total number of datapoints), shown here are the actual *probability densities* (with the $\int p(x)dx = 1$). The title of each panel shows three measures: (i) [*dprime σ*] – the behavioral performance; (ii) [*Bias σ*] – the average of the absolute difference between the mean of the prominent Gaussian component of the GMM and the driving acoustic chunk rate, and variance across the ROIs; and (iii) [*P σ*] -- the average P-value (defined in Fig. 6) and variance across the ROIs. Note the tightness of the PDFs in the 1.8Hz condition compared to the pseudo-uniform shape of the PDFs in the 2.6Hz condition, and the correlation between the decrease in *dprime* and the increase in *Bias* and the decrease in *P*.

Discussion

In this study we adopted a reductionist approach to test, in electrophysiological terms, the hypothesis that the speech decoding process at the phrasal time scale is guided by a flexible,

acoustic-driven neuronal delta oscillator locked to phrase-size acoustic cues (Ghitza, 2017). The proposal suggests an analogue role of a delta oscillator, at the phrasal time scale, to the role played by neuronal theta-band oscillations at the syllabic time scale. The study is reductionist in the sense that it is confined to the perceptual chunking of digits sequences, where the digits in the sequence are grouped into phrase-size chunks. We collected, concurrently, behavioral and MEG data during a digit retrieval task, in which the digit sequences were either presented with an acoustic chunk pattern inside or outside of the delta range. Stimuli with a chunk rate inside the delta range elicited considerable neuronal periodicity at the chunk rate in STG, MTG ROIs and IFG, SMG and PC ROIs. Critically, this pattern of detected periodicities was directly related to Correct behavioral responses. In contrast, stimuli with a chunk rate outside of the delta range elicited weak periodicity, aligned with observed declines in behavioral performance. In the calcarine ROI (early visual cortex), considered a 'control area' for our analyses, no periodicities at the chunk rate were elicited.

Presence of delta periodicities in the auditory pathway

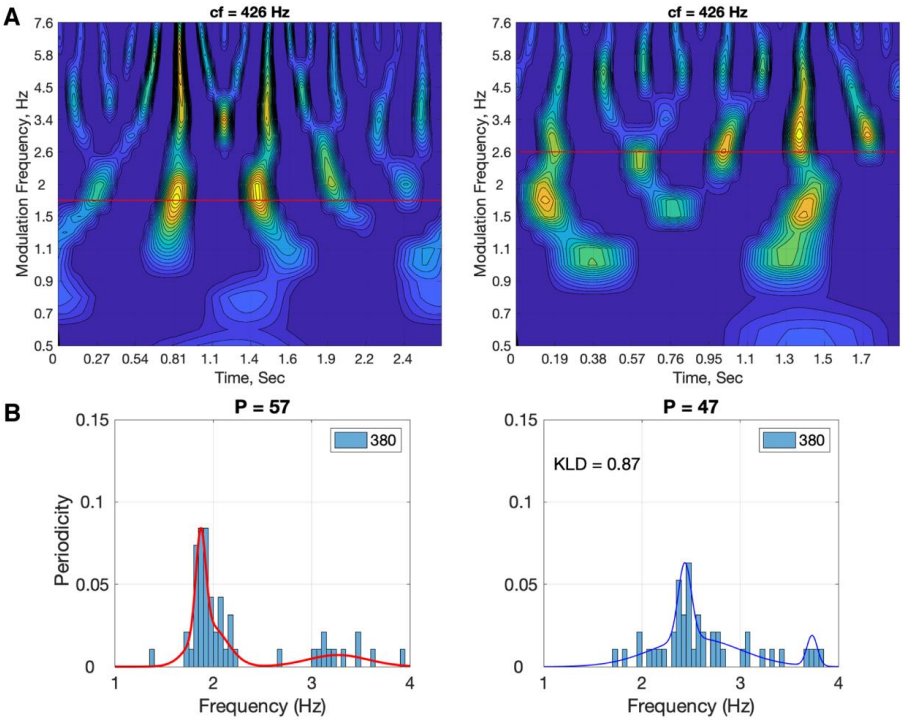
How should these activity patterns of neuronal delta and theta periodicities, be interpreted? In the temporal cortex (STG and MTG), robust periodicities were recorded mainly by stimuli with a chunk rate inside the delta range, and only for Correct behavioral responses. Periodicities in these brain areas were present even for acoustic chunk rates at the edge of the delta range, albeit considerably weaker. A similar pattern of periodicities was observed in the speech-motor planning and integration areas (IFG, SMG and PC), where periodicities were absent for acoustic chunk rates outside the delta range. Note that the observed lack of hemispheric lateralization in auditory cortex in our study is in line with previous reports on bilateral theta/delta activity elicited to more complex speech stimuli (Assaneo et al., 2019; Flinker et al., 2019). Interestingly, in

contrast to the temporal brain areas, in the speech-motor planning and integration areas more divergence between the left and right hemisphere was observed, with more prominent delta periodicities in the left hemisphere. The left hemisphere more tightly followed the chunking rate compared to the right. These findings suggest an important role for superior and middle temporal and speech-motor planning and integration areas in chunking at the phrasal scale. Importantly, and quite remarkably, the delta-band activity in these areas was fully aligned with behavioral performance (i.e. delta activity was only elicited in Correct, but not in Erroneous responses). Previously, EEG studies showed delta in bilateral middle and superior temporal areas (also fusiform gyrus) (Bonhage et al., 2017) and at fronto-temporal sites (Boucher et al., 2019) was related to chunking during phrase and sentence processing. Delta might reflect the chunking of ordered sensorimotor events as articulated sound, rather than syntactic phrasal/sentential processing directly (Boucher et al., 2019). Furthermore, Keitel et al. (2018) and Morillon et al. (2019) recently proposed that delta oscillations in the motor cortex are involved in temporal predictions, affecting speech processing in the auditory cortex at a phrasal scale (for a predictive account of delta see also: Breska and Deouell, 2017; Daume et al., 2021; or a statistical learning account: Henin et al., 2021). A possible interpretation of their findings through the lens of our results is that acoustic-driven segmentation of phrase-size chunks takes place in STG, and the recorded behavioral performance with respect to chunk rate is a consequence of the goodness of segmentation. When the chunk rate is inside the delta band, successful segmentation results in delta activities in speech-motor integration areas (SMG, PC, IFG) that may reflect decoding processes and possibly auditory-motor mapping related processes (Park et al., 2015). In contrast, chunk rates outside of the delta band might result in bad segmentation in STG, and in turn suppressed periodicities in speech-motor integration areas (SMG, PC, IFG) due to unreliable decoding and audio-motor mapping. This interpretation is in line with another study (Donhauser

and Baillet, 2020) that reports strong delta activity in STG when the speech input was 'informative', which may be the consequence of appropriate segmentation.

It could be argued that one cannot draw a conclusive relationship between 'chunking' and the neural periodicity in the delta range. In particular, the drop in intelligibility for the 2.6Hz condition may be due to the fact that the silent gaps in-between the two-digit chunks are shorter. This argument raises three points that merit discussion. First, a question arises whether or not a 2.6 Hz rhythm in the acoustics is present at the cochlear output level. Figure 9A shows a simulation of the cochlear modulation spectrum (e.g., Jepsen et al., 2008) for a 1.8Hz (left) and a 2.6Hz (right) stimuli, taken at a characteristic frequency of 426 Hz (this cochlear place was selected at random, for demonstration). A robust modulation presence is observed for both stimuli, at their respective acoustic input rhythm. Second, it could be argued that the shorter silent gaps result in weaker acoustic cues for chunking. Recalling that neural activity in primary auditory cortex represents sensory representations of the acoustics with a minimal information loss (see, e.g., Nourski et al., 2009), a weakening in acoustic cues should be reflected in terms of a weaker periodicity presence at primary cortex (e.g., the Heschl's Gyrus). As mentioned earlier (see Fig. 2), we opted to omit the Heschl's Gyrus from our rigorous periodicity analysis because of the small number of voxels present (3 in the Left, 2 in the Right). Figure 9B shows the XCOV periodicity PDF for all five available voxels, for Correct and Erroneous responses combined. Keeping in mind the concern over the validity of the results due to the limited number of voxels, we observe a strong periodicity presence for both chunking conditions at their respective chunk rates, suggesting no weakening of the acoustic cues for chunking. In contrast, and as early as the STG level, we observe strong periodicities only for chunk rates inside the delta frequency range (Fig. 6). The findings suggest that the neuronal circuitry of the delta oscillator resides at the STG level and constrains prosodic chunking. Third, it could also be argued that the shorter silent gaps

result in an insufficient decoding time at the single digit level. However, our data show that this is not the case, as at the digit level, for all chunking conditions and at all ROIs, strong theta periodicities (at the single digit rate) were elicited regardless of the level of behavioral chunking performance. Thus, the drop in performance for the 2.6Hz condition – with a chunk rate just outside the delta frequency range – is due to the lack of decoding time at the chunk level but not due to digit decoding time. Recall that for both the 2.6Hz and the 1.8Hz stimuli, the two-digit chunks themselves have an identical time-compressed acoustics; the only difference is the duration of the silent gaps between the chunks (see Fig. 1). Performance is recovered by bringing the chunk rate back inside the delta range, hence providing the extra decoding time needed. As a whole, therefore, our data suggest that segmentation of phrase-sized chunks is realized by neuronal delta oscillators, and that the chunk's decoding time is determined by delta, in analogy to the role of theta in determining the decoding time at the syllable level (Ghitza, 2014).



605

606 **Figure 9.** ***Cochlear modulation spectrum (A) and delta periodicities at Heschl's Gyrus (B)***
607 ***for the digit-sequence stimuli shown in Fig. 1.*** (A) Cochlear output in terms of cochlear
608 modulation spectrum (Jepsen et al., 2008). Shown are modulation spectra of the digit-sequence
609 stimuli shown in Fig. 1, for the 1.8Hz stimulus (left) and for the 2.6Hz stimulus (right). The
610 modulation spectra shown are snapshots at the cochlear characteristic frequency (CF) of 426 Hz.
611 Abscissae represent time (duration of 2.7- and 1.9 seconds, for the 1.8Hz and the 2.6Hz stimuli,
612 respectively) and the ordinate represents the modulation frequencies (0.5 - to 7.6 Hz). Note the
613 strong presence of modulations at 1.8 Hz for the 1.8Hz stimulus and at 2.6 Hz for the 2.6Hz
614 stimulus. (B) Delta periodicities at Heschel Gyrus ROI for the Correct and Erroneous responses,
615 combined. Note that the total number of datapoints is 380: the number of voxels (left and right
616 combined) is 5, the number of participants 19, and the number of response conditions (correct
617 and erroneous) is 4. The KLD value of the 2.6Hz probability density function (shifted to 1.8Hz)
618 with respect to the 1.8Hz probability density function is 0.87. Keeping in mind the concern over
619 the validity of the results due to the limited number of voxels, the strong periodicity presence for
620 both chunking conditions suggest that the diminished periodicity for the 2.6Hz condition is due to
621 neuronal circuitry characteristics at the STG level and not due to weakening of acoustic cues for
622 chunking.
623

624 **Presence of theta periodicities in all chunking conditions**

Our data show strong theta periodicities in all ROIs and for all chunking conditions. Such elicited neural response patterns reflect the single digit presentation rate. A bimodal characteristic of the PDFs is observed for all chunking conditions, but in particular for the 1.8Hz condition. The bimodality arises from the acoustic properties of the stimuli. Consider, for example, the stimulus shown in Fig. 1. Three intra-digit durations can be identified: (i) the duration between the onset of the first digit of a chunk and the first digit in the following chunk, which gives rise to the chunking rate, (ii) the duration between the onset of the first digit and onset of the second digit in a chunk, and (iii) the duration between the onset of the second digit in a chunk and the onset of the first digit in the following chunk. This plurality in intra-digit durations give rise to a bimodal duration distribution with a skewness determined by the prescribed chunking rate. The skewness is accentuated, in particular, in our 1.8Hz stimuli. The bimodal nature in the acoustics drives the elicited neural response seen in our data (Fig. 7A).

Oscillations versus evoked responses

Our data show strong delta cortical periodicities while listening to the 1.8 Hz chunked stimuli. Are these brain waves generated by a neuronal oscillator locked to the acoustic chunk rhythm or do they reflect the evoked response to the corresponding acoustic cues? The answer to this question at the syllabic level has been difficult to determine, because the impulse response of the neuronal circuitry to discrete acoustic cues associated with syllables (e.g., acoustic edges, vocalic nuclei) corresponds, in duration, to the theta-cycle range (about [125 330] msec). Doelling et al. (Doelling et al., 2019) addressed this conundrum by generating simulated outputs of an oscillator model and of an evoked response model, and comparing the quantitative predictions of phase lag patterns generated by the two models against recorded MEG data. They showed that, compared to the evoked response model, a model that includes oscillatory dynamics better predicted the MEG

data. Our data provides additional support for the oscillator interpretation. Can the observed, robust periodic responses to a 1.8 Hz chunked stimulus reflect evoked responses elicited by discrete acoustic cues at the phrase time scale? Indeed, steady-state evoked responses to slow dynamics have been observed in both visual and auditory sensory regions (e.g. Capilla et al., 2011; Wang et al., 2011). However, only a model of oscillatory dynamics can explain the fact that neural response at the delta range is only present when the acoustic chunk rate is inside, but is absent for rates outside the delta range.

Generalizability of the neuronal chunking mechanism

Scaling up to real speech. The studies discussed above (Meyer et al., 2016; Bonhage et al., 2017; Keitel et al., 2018; Boucher et al., 2019; Morillon et al., 2019) suggest a presence of delta brain waves in phrasal chunking for continuous speech, beyond the digit retrieval paradigm used here. Extending our results to naturalistic speech has important implications for what would constitute optimally sized acoustic chunks for the sentential decoding – or parsing – process. If the information 'bound' within windows of roughly a delta cycle are integrated as phrases (intonation phrases and perhaps structural phrases, depending on the specific relation), it suggests that there are natural patterns of spoken phrase rhythms or phrase durations that are best suited for decoding spoken language, driven by the necessity to match a cortical function. Deploying the experimental analysis approach, we describe here to real speech can elucidate the temporal aspects of spoken language comprehension.

Infra-delta chunking rate. As discussed earlier we define the relevant delta range to be between 0.5 to 2 Hz, and chose the 1.8Hz condition to represent the case where the input chunking rate is inside delta, and the 2.6Hz condition to represent the outside delta case. The main research question of our study was whether elicited delta cortical oscillations correlate with behavior. In particular, does performance deteriorate if the chunk rate is outside the delta range? We addressed

this question by looking at an above-delta chunking rate (2.6 Hz), but we didn't look at infra-delta rates (e.g., 0.3 Hz). The reason to skip the effects of infra-delta rates stemmed from the fact that the decay time of sensory memory – about 2 sec long (e.g. Cowan, 1984) – roughly coincides with the lower bound of the delta-cycle duration. Consequently, the dominant factor at the origin of a possible deterioration in performance may very well be an internal time constraint on processing spoken material (due to echoic memory span) rather than prosodic segmentation.

Conclusion

Oscillation-based models of speech perception (Ghitza, 2011; Giraud and Poeppel, 2012; Gross et al., 2013; Haegens and Zion Golumbic, 2017; Lakatos et al., 2019; Martin and Doumas, 2017; Rimmele et al., 2018) postulate a cortical computational principle by which decoding is performed within a time-varying window structure, synchronized with the input on multiple time scales. The windows are generated by a segmentation process, implemented by a cascade of oscillators. At the pre-lexical level, the segmentation process is realized by a *flexible theta oscillator* locked to the input syllabic rhythm, where the theta cycles constitute the syllabic windows. Doelling et al. (2014) provided MEG evidence for the role of theta, showing that intelligibility is correlated with the existence of acoustic-driven theta neuronal oscillations.

Our major finding – that phrase-size chunking of digit strings is correlated with acoustic-driven delta oscillations – suggests that the role played by neuronal theta-band oscillations in syllabic segmentation can be generalized to the phrasal time scale. The segmentation process is realized by a *flexible delta oscillator* locked to the input phrase-size chunk rhythm, where the delta cycles constitute the phrase-size chunk windows.

Future research is required to investigate whether our findings can be generalized to continuous speech (i.e., beyond digit strings). That is, whether the intonational phrase patterns of language could be constrained by cortical delta oscillations. Adopting the view that the strategy of

composing syllables and words into phrasal units is the result of an evolutionary trajectory to match a cortical function (Bosman and Aboitiz, 2015; Patel and Iversen, 2014), we hypothesize that the phrases of language are constrained by delta oscillations, and the rules of chunking in speech production may be the product of common cortical mechanisms on both motor and sensory sides, with delta at the core.

Software Accessibility statement: Analysis code will be made available upon request [to corresponding author].

References

Ahissar E, Ahissar M (2005) 18. Processing of the temporal envelope of speech. Audit Cortex Synth Hum Anim Res 295.

Assaneo MF, Rimmele JM, Orpella J, Ripollés P, de Diego-Balaguer R, Poeppel D (2019) The Lateralization of Speech-Brain Coupling Is Differentially Modulated by Intrinsic Auditory and Top-Down Mechanisms. Front Integr Neurosci 13:28.

Assaneo MF, Rimmele JM, Sanz Perl Y, Poeppel D (n.d.) Speaking rhythmically can shape hearing. Accept Nat Hum Behav.

Aubanel V, Davis C, Kim J (2016) Exploring the Role of Brain Oscillations in Speech Perception in Noise: Intelligibility of Isochronously Retimed Speech. Front Hum Neurosci 10:430.

Auer P, Couper-Kuhlen Elizabeth, Müller Frank (1999) Language in time : the rhythm and tempo of spoken interaction. New York: Oxford University Press.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cereb Cortex 19:2767–2796.

Bonhage CE, Meyer L, Gruber T, Friederici AD, Mueller JL (2017) Oscillatory EEG dynamics underlying automatic chunking during sentence processing. NeuroImage 152:647–657.

Bosman CA, Aboitiz F (2015) Functional constraints in the evolution of brain circuits. Front Neurosci 9:303.

Boucher VJ, Gilbert AC, Jemel B (2019) The Role of Low-frequency Neural Oscillations in Speech Processing: Revisiting Delta Entrainment. J Cogn Neurosci 1–11.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Breen M (2018) Effects of metric hierarchy and rhyme predictability on word duration in The Cat in the Hat. Cognition 174:71–81.

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.

Bröhl F, Kayser C (2020) Delta/theta band EEG differentially tracks low and high frequency speech envelopes. bioRxiv 2020.07.26.221838.

Brungart D, Wassenhove V van, Brandewie E, Romigh GD (2007) The effects of temporal acceleration and deceleration on AV speech perception In: AVSP .

Buiatti M, Peña M, Dehaene-Lambertz G (2009) Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. NeuroImage 44:509–519.

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–e14543.

Clifton CJ, Carlson K, Frazier L (2006) Tracking the what and why of speakers' choices: prosodic boundaries and the length of constituents. Psychon Bull Rev 13:854–861.

Cowan N (1984) On short and long auditory stores. Psychol Bull 96:341–370.

Daume J, Wang P, Maye A, Zhang D, Engel AK (2021) Non-rhythmic temporal prediction involves phase resets of low-frequency delta oscillations. NeuroImage 224:117376.

Deniz ND, Fodor JD (2019) Timing of Syntactic and Rhythmic Effects on Ambiguity Resolution in Turkish: A Phoneme Restoration Study. Lang Speech 0023830919894614.

749 Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical
750 linguistic structures in connected speech. Nat Neurosci 19:158–164.
751 Ding N, Simon JZ (2009) Neural Representations of Complex Temporal Modulations in the
752 Human Auditory Cortex. J Neurophysiol 102:2731–2743.
753 Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta–theta
754 oscillations to enable speech comprehension by facilitating perceptual parsing.
755 NeuroImage 85, Part 2:761–768.
756 Doelling KB, Assaneo MF, Bevilacqua D, Pesaran B, Poeppel D (2019) An oscillator model
757 better predicts cortical entrainment to music. Proc Natl Acad Sci 116:10113.
758 Donhauser PW, Baillet S (2020) Two Distinct Neural Timescales for Predictive Speech
759 Processing. Neuron 105:385-393.e9.
760 Flinker A, Doyle WK, Mehta AD, Devinsky O, Poeppel D (2019) Spectrotemporal modulation
761 provides a unifying framework for auditory cortical asymmetries. Nat Hum Behav 3:393–
762 405.
763 G. Miller (1962) Decision units in the perception of speech. IRE Trans Inf Theory 8:81–83.
764 Gee JP, Grosjean F (1983) Performance structures: A psycholinguistic and linguistic appraisal.
765 Cognit Psychol 15:411–458.
766 Ghitza O (2020) “Acoustic-driven oscillators as cortical pacemaker”: a commentary on Meyer,
767 Sun & Martin (2019). Lang Cogn Neurosci 35:1100–1105.
768 Ghitza O (2017) Acoustic-driven delta rhythms as prosodic markers. Lang Cogn Neurosci
769 32:545–561.
770 Ghitza O (2014) Behavioral evidence for the role of cortical theta oscillations in determining
771 auditory channel capacity for speech. Front Psychol 5.
772 Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by
773 cascaded oscillators locked to the input rhythm. Front Psychol 2:1–13.
774 Ghitza O, Greenberg S (2009) On the Possible Role of Brain Rhythms in Speech Perception:
775 Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of
776 Silence. Phonetica 66(suppl 1–2):113–126.
777 Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging
778 computational principles and operations. Nat Neurosci 15:511–517.
779 Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York, Ny: Wiley.
780 Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech
781 Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. PLoS Biol
782 11:e1001752.
783 Haegens S, Zion Golumbic E (2017) Rhythmic facilitation of sensory processing: a critical
784 review. Neurosci Biobehav Rev.
785 Henin S, Turk-Browne NB, Friedman D, Liu A, Dugan P, Flinker A, Doyle W, Devinsky O,
786 Melloni L (2021) Learning hierarchical sequence representations across human cortex
787 and hippocampus. Sci Adv 7:eabc4530.
788 Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci
789 8:393–402.
790 Hilton C, Goldwater M (2020) Linguistic syncopation: Meter-syntax alignment and its effect on
791 sentence comprehension and sensorimotor synchronization.
792 Hovsepyan S, Olasagasti I, Giraud A-L (2020) Combining predictive coding and neural
793 oscillations enables online syllable recognition in natural speech. Nat Commun 11.
794 Inbar M, Grossman E, Landau AN (2020) Sequences of Intonation Units form a ~ 1 Hz rhythm.
795 Sci Rep 10:15846–15846.

JASP Team (2020) JASP (Version 0.12)[Computer software].
Jepsen ML, Ewert SD, Dau T (2008) A computational model of human auditory signal processing and perception. J Acoust Soc Am 124:422–438.
Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol 16:e2004473.
Kösem A, Bosker HR, Takashima A, Meyer A, Jensen O, Hagoort P (2018) Neural Entrainment Determines the Words We Hear. Curr Biol 28:2867-2875.e3.
Lakatos P, Gross J, Thut G (2019) A New Unifying Account of the Roles of Neuronal Entrainment. Curr Biol CB 29:R890–R905.
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.
Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1996) Independent component analysis of electroencephalographic data. Adv Neural Inf Process Syst 8:145–151.
Martin AE, Doumas LAA (2017) A mechanism for the cortical computation of hierarchical linguistic structure. PLOS Biol 15:e2000663.
Martin P (2015) The Structure of Spoken Language: Intonation in Romance. Cambridge: Cambridge University Press.
Meyer L, Henry M, Gaston P, Schmuck N, Friederici A (2016) Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations. Cereb Cortex 27.
Morillon B, Arnal LH, Schroeder CE, Keitel A (2019) Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neurosci Biobehav Rev 107:136–142.
Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol 48:3637.
Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, Brugge JF (2009) Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. J Neurosci 29:15564.
Oganian Y, Chang EF (2019) A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci Adv 5:eaay6279.
Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.
Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Intell Neurosci 2011:1–9.
Park H, Ince RAA, 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.
Patel AD, Iversen JR (2014) The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. Front Syst Neurosci 8:57.
Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. Front Psychol 3:320.
Pittman-Polletta BR, Wang Y, Stanley DA, Schroeder CE, Whittington MA, Kopell NJ (2020) Differential contributions of synaptic and intrinsic inhibitory currents to speech segmentation via flexible phase-locking in neural oscillators. bioRxiv 2020.01.11.902858.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as [`]asymmetric sampling in time'. Nat Speech Percept 41:245–255.
Rimmele JM, Gross J, Molholm S, Keitel A (2018a) Editorial: Brain Oscillations in Human Communication. Front Hum Neurosci 12:39.
Rimmele JM, Morillon B, Poeppel D, Arnal LH (2018b) Proactive Sensing of Periodic and Aperiodic Auditory Patterns. Trends Cogn Sci 22:870–882.
Rimmele JM, Poeppel D, Ghitza O (2020) Acoustically driven cortical delta oscillations underpin perceptual chunking. bioRxiv 2020.05.16.099432.
Schönbrodt FD, Wagenmakers E-J (2018) Bayes factor design analysis: Planning for compelling evidence. Psychon Bull Rev 25:128–142.
Shattuck-Hufnagel S, Turk AE (1996) A prosody tutorial for investigators of auditory sentence processing. J Psycholinguist Res 25:193–247.
Stehwien S, Meyer L (2021) Rhythm Comes, Rhythm Goes: Short-Term Periodicity of Prosodic Phrasing.
Sternberg S (1966) High-speed scanning in human memory. Science 153:652–654.
Stolk A, Todorovic A, Schoffelen J-M, Oostenveld R (2013) Online and offline tools for head movement compensation in MEG. NeuroImage 68:39–48.
Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. NeuroImage 15:273–289.
Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44:867–880.
Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ (2011) Sensitivity to Temporal Modulation Rate and Spectral Bandwidth in the Human Auditory System: MEG Evidence. J Neurophysiol.