Behavioral/Cognitive

θ-Band and β-Band Neural Activity Reflects Independent Syllable Tracking and Comprehension of Time-Compressed Speech

Maria Pefkou,[1]* Luc H. Arnal,[1]* Lorenzo Fontolan,[2] and Anne-Lise Giraud[1]

[1]Auditory Language Group, Department of Neuroscience, University of Geneva, Biotech Campus, 9 Chemin des Mines, 1202 Geneva, Switzerland, and [2]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147

Recent psychophysics data suggest that speech perception is not limited by the capacity of the auditory system to encode fast acoustic variations through neural γ activity, but rather by the time given to the brain to decode them. Whether the decoding process is bounded by the capacity of θ rhythm to follow syllabic rhythms in speech, or constrained by a more endogenous top-down mechanism, e.g., involving β activity, is unknown. We addressed the dynamics of auditory decoding in speech comprehension by challenging syllable tracking and speech decoding using comprehensible and incomprehensible time-compressed auditory sentences. We recorded EEGs in human participants and found that neural activity in both θ and γ ranges was sensitive to syllabic rate. Phase patterns of slow neural activity consistently followed the syllabic rate (4–14 Hz), even when this rate went beyond the classical θ range (4–8 Hz). The power of θ activity increased linearly with syllabic rate but showed no sensitivity to comprehension. Conversely, the power of β (14–21 Hz) activity was insensitive to the syllabic rate, yet reflected comprehension on a single-trial basis. We found different long-range dynamics for θ and β activity, with β activity building up in time while more contextual information becomes available. This is consistent with the roles of θ and β activity in stimulus-driven versus endogenous mechanisms. These data show that speech comprehension is constrained by concurrent stimulus-driven θ and low-γ activity, and by endogenous β activity, but not primarily by the capacity of θ activity to track the syllabic rhythm.

Key words: **β oscillations; EEG; speech comprehension; θ oscillations; time-compressed speech**

Significance Statement

Speech comprehension partly depends on the ability of the auditory cortex to track syllable boundaries with θ-range neural oscillations. The reason comprehension drops when speech is accelerated could hence be because θ oscillations can no longer follow the syllabic rate. Here, we presented subjects with comprehensible and incomprehensible accelerated speech, and show that neural phase patterns in the θ band consistently reflect the syllabic rate, even when speech becomes too fast to be intelligible. The drop in comprehension, however, is signaled by a significant decrease in the power of low-β oscillations (14–21 Hz). These data suggest that speech comprehension is not limited by the capacity of θ oscillations to adapt to syllabic rate, but by an endogenous decoding process.

Introduction

As continuous speech unfolds in time, the neural language system must segment the acoustic signal into meaningful linguistic units (Poeppel et al., 2008; Giraud and Poeppel, 2012; Christiansen and Chater, 2016). This process is facilitated by iteratively generating predictions about what is going to be said next, and it is hence easy for our brain to individualize and understand words in connected speech even in adverse listening conditions (Davis et al., 2005). θ (~4–8 Hz) Oscillatory entrainment to the acoustic envelope could play a crucial role in speech encoding by both enabling the detection of syllable boundaries (Hyafil et al., 2015) and by providing for syllables a phase-informed neural code that

Received Sept. 14, 2016; revised May 24, 2017; accepted May 31, 2017.

Author contributions: M.P., L.H.A., and A.-L.G. designed research; M.P. performed research; L.F. contributed unpublished reagents/analytic tools; M.P. and L.F. analyzed data; M.P., L.H.A., and A.-L.G. wrote the paper.

This work was supported by the Swiss National Fund (Personal Grant 320030-163040 to A.L.G.), and by the Language and Communication thematic network of the University of Geneva. We thank Oded Ghitza for useful discussions.

*M.P. and L.H.A. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Anne-Lise Giraud, Department of Neuroscience, University of Geneva, Biotech Campus, 9 Chemin des Mines, 1202 Genève, Switzerland. E-mail: anne-lise.giraud@unige.ch

DOI:10.1523/JNEUROSCI.2882-16.2017

 0270-6474/17/377930-09$15.00/0

facilitates higher-level linguistic parsing (Luo and Poeppel, 2007; Luo et al., 2010; Ghitza, 2012, Peelle et al., 2013; Doelling et al., 2014). Yet, whether speech is comprehensible or not, θ phase-locking between stimulus and neural responses is present, notably when speech is time-reversed (Howard and Poeppel, 2010). Hence, the importance of speech-envelope tracking by θ oscillations in comprehension remains unclear.

Psychophysics data have recently put the syllable temporal format at the center of the speech-comprehension process. Speech becomes unintelligible when compressed by a factor of ≥3 (Nourski et al., 2009). However, if time-compressed speech is chunked into 40 ms (low-γ range) units, and silent gaps are inserted between the compressed speech segments, comprehension can be partly restored (Ghitza and Greenberg, 2009). This effect occurs even when speech is heavily compressed, e.g., by a factor of 8, provided that the reconstructed syllable-like rhythm stays <9–10 syllable-like units/s (Ghitza, 2014). Speech comprehension seems therefore limited by syllabic decoding at a maximal θ-range rate of 9–10 Hz, but not by the acoustic information encoding capacity. Importantly, these findings suggest that decoding time is essential for speech comprehension.

However, that the decoding process operates optimally when the syllabic rate is within the θ range of neural activity (Ghitza, 2014) does not mean that it is underpinned by θ activity. Using a neurocomputational model, Hyafil et al. (2015) showed that θ activity can provide a reliable on-line signaling of syllable boundaries. Yet, in this model, it is the information conveyed by low-γ (25–40 Hz) activity within a θ cycle that is informative about speech content. The model is instructive but arguably incomplete, as it does not emulate top-down, predictive, decoding processes informed by our linguistic knowledge (Sohoglu et al., 2012; Davis and Johnsrude, 2007). Using magnetoencephalography, Park et al. (2015) showed that neural activity originating in prefrontal areas influences oscillation coupling in the auditory cortex during speech perception. Such top-down processes in speech perception likely involve yet another brain rhythm, distinct from those that encode syllable boundaries (θ) and the phonemic content (low-γ), that is the β (15–25 Hz) range. These findings are highly plausible as the β rhythm is generically involved in top-down perceptual processes (Engel and Fries, 2010; Arnal and Giraud, 2012), in particular during speech perception (Fontolan et al., 2014).

In summary, speech processing by stimulus-driven θ and γ neural activity accounts, both experimentally and theoretically, for how speech is encoded in the auditory cortex, but not for how it is decoded by higher-order brain structures. In this study, we examined whether speech comprehension is constrained by the capacity of θ rhythm to track the syllabic rhythm and segment speech into decodable elements, or by the capacity of endogenous oscillatory mechanisms to decode the speech elements. We challenged both stimulus-driven speech-tracking and endogenous speech-decoding processes by using time-compressed speech, which increases the amount of input information per time unit, while leaving all other stimulus parameters unchanged. Using these stimuli, we contrasted the neural dynamics associated with a linear tracking of syllable duration, with speech comprehension dynamics that show a sudden drop at a compression rate of three.

Materials and Methods

Participants. Nineteen native French speakers (eleven females) participated in a study involving behavioral and EEG measurements (data from two participants could not be analyzed due to excessive movement and low impedances during EEG recordings). All participants were right-handed (mean age: 24.9; ±5 SD) without history of hearing impairment or dyslexia, gave written informed consent to participate in the study, and received financial compensation for their participation. The study was approved of by the local ethics committee (Commission Cantonale d'Ethique dans la Recherche).

Stimuli and procedure. The stimuli belonged to a corpus of 356 French sentences that were constructed such that every set of four sentences amounted to a small story. Each sentence was semantically meaningful, grammatically simple, with no more than one embedded clause (Hervais-Adelman et al., 2015). The sentences were recorded in a soundproof room by a male native French speaker and digitized at a 44.1 kHz sampling rate. The root mean square (RMS) of all stimuli was computed and stimuli whose RMS was higher or lower than 3 SDs from the mean were normalized to the mean RMS of the initial set. They were then time-compressed by factors of 2 or 3 using the Waveform Similarity Based Overlap-Add algorithm (Verhelst and Roelands, 1993), which maintains maximal similarity of the time scale of the modified waveform to the original one without modifying the pitch (Fig. 1*A*). The duration of the time-compressed versions of each sentence was, therefore, equal to half and one-third of the natural sentence duration, with a slight jitter (0.13 ms for compression rate 2 and 0.02 ms for compression rate 3, on average) across sentences. Finally, all stimuli were low-pass filtered at 10 kHz using a fifth-order Butterworth filter.

Each participant performed both the behavioral experiment and the EEG one (Fig. 1*C*,*D*) performed on the same day, in a randomized order. To control for a potential task presentation order effect, eight participants started with the behavioral experiment and nine with the EEG experiment.

During the behavioral experiment (Fig. 1*C*), participants were asked to repeat each of the 20 heard sentences per condition (compression factor 1–3). Following the method described by Davis and colleagues (2005), speech comprehension was measured by computing the percentage of correctly repeated words per trial, which reflects how well participants understood the sentence overall. A word was scored as correct if it was pronounced correctly, but scored as incorrect if reported in the wrong order. Words reported in the correct order were scored as correct even if intervening words were absent or incorrectly reported.

In the EEG experiment (Fig. 1*D*), participants manually rated subjective comprehension on a 1–5 scale (totally incomprehensible to fully comprehensible) This enabled us to minimize muscle artifacts (no oral response) and to maximize the number of trials. Each participant listened to 64 sentences per condition in the EEG experiment. Stimuli appeared as single sentences in the behavioral experiment but were concatenated into sets of four story-like combinations for the EEG experiment. Each stimulus appeared only once without repetitions throughout the two experiments (behavior and EEG) and was presented in a pseudorandomized order for each participant. Speech rate was blocked into groups of four sentences in both the behavioral and the EEG task. Individual comprehension scores (accuracy and reports in behavioral and EEG experiments, respectively) were averaged per subjects and condition and entered in a repeated-measures ANOVA with compression rate (three levels: uncompressed, compressed by 2, and compressed by 3) as a within-subjects factor. To assess whether subjects' ratings reliably reflected comprehension, we also computed the correlation between the subjective comprehension scores collected during the EEG experiment and the percentage of correctly reported words during the behavioral experiment.

EEG recording and data analysis. Brain electrical responses were recorded with a 256-electrode Electrical Geodesics HydroCel system (Electrical Geodesics). The signal was recorded continuously and digitized at a sampling rate of 1000 Hz. By default, the recording system sampled at 20 kHz before applying an analog hardware antialiasing filter with a cutoff frequency of 4 kHz and down-sampling the signal to 1000 Hz, and applying a low-pass Butterworth filter with a cutoff of 400 Hz. The reference electrode was the Cz, situated at the vertex. Electrode impedances were checked at the beginning of the session and after the end of each recording blocks, and were <30 kΩ at the beginning of each block.

EEG data were analyzed using custom Matlab (Mathworks) scripts as well as the Cartool software (Brunet et al., 2011). Data were first down-

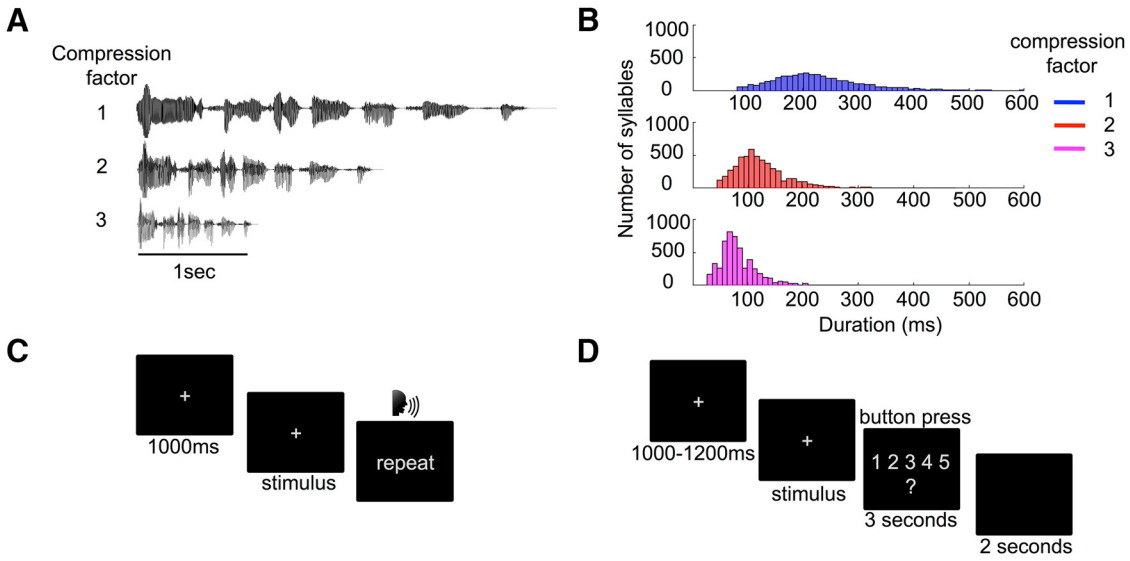

Figure 1. Stimuli and experimental design. ***A***, Stimulus waveforms. The waveform of the French phrase "*Le ministre a visité le pays pour la premiere fois*" ("The minister visited the country for the first time") in the three experimental conditions, namely in its original form (i.e., compression factor equal to 1) and time-compressed by a factor of 2 and 3. ***B***, Syllable duration. Distribution of the duration of syllables across the three compression factors. As expected, this distribution shifts toward the left as the compression factor increases. ***C***, Experimental design (behavioral task). Experimental design for the behavioral task. Each trial began with a fixation cross, followed by the presentation of the auditory stimulus. Participants were asked to repeat the stimulus. ***D***, Experimental design (EEG task). Experimental design of the behavioral task during the EEG recording. Each trial began with a fixation cross, followed by the presentation of the auditory stimulus. Participants were asked to rate the comprehension of the stimulus on a 1-to-5 scale.

sampled to 200 Hz and a set of 204 electrodes was selected for further analysis (channels covering the cheeks were excluded). All blocks of the EEG experiment were concatenated and an independent components analysis (ICA) was computed on the whole dataset using the Infomax routine from the Matlab-based EEGLAB toolbox (Delorme and Makeig, 2004). Components corresponding to eye blinks and electrical line-noise were removed. The resulting data were filtered between 0.75 and 80 Hz using fifth-order Butterworth bandpass filters. An additional bandstop filter between 49.9 and 50.1 Hz was applied to remove residual 50 Hz line noise. Noisy electrodes showing poor contact with the scalp were identified through visual inspection and interpolated (4 over 204 electrodes on average across participants). To obtain epochs of equal length for all conditions, the EEG data were segmented into epochs starting 1 s before the sound onset and ending 5 s after, which corresponded to the shortest time-compressed stimulus. Importantly, in the time-compressed conditions, the amount of linguistic information contained in the fixed-length 5 s poststimulus epoch was therefore two or three times higher than in the uncompressed speech, depending on the compression factor. The segmented data were then rereferenced to the average reference.

Stimuli and data analysis. To assess the effect of speech compression on oscillatory neural dynamics across frequencies, we first extracted the time-frequency (TF) decomposition of both the acoustic stimuli and the EEG data using Morlet wavelets ($m = 7$) in a frequency range between 1 and 40 Hz, with a 1 Hz resolution. The envelopes of the acoustic stimuli were extracted by applying full-wave rectification and a second-order Butterworth low-pass filter at 30 Hz. Power values were then computed by taking the modulus of the complex TF values. EEG data were baseline-corrected for each frequency using the average power across a time window from −750 to −100 ms before the stimulus onset.

To assess the effect of compression rate on the power and phase of neural responses, we used two distinct measures: cross-spectral density (CSD) and phase coherence. CSD provided a measure of the power shared by the acoustic and the EEG signals at each frequency, while phase coherence enabled us us to look at whether the two signals were phase-locked in time at a given frequency. Both CSD and phase coherence analyses were performed on a subset of 5 "auditory" electrodes, selected on the basis of a functional localizer experiment, performed after the main experiment. During the functional localizer experiment, participants listened to 200 pure tones and were instructed to rest but keep their eyes open and fixate on a cross on the screen. These data were preprocessed in the same way as the data collected for the main task and the evoked response was computed by averaging all clean trials. For each participant, we then identified the five electrodes showing the largest N100 amplitude. These "auditory" electrodes were used for the CSD and phase coherence analyses under the assumption that the on-line tracking of syllabic onsets takes place in primary auditory regions (Nourski et al., 2009). This selection of electrodes maximized the sensitivity of subsequent analyses to exogenously, stimulus-driven signals.

CSD was estimated with the power spectral density function (cpsd.m) in Matlab, using Welch's averaged modified periodogram method in steps of 0.33 Hz. For each trial and electrode, we computed the CSD between the acoustic stimulus and the corresponding brain responses for each frequency step. We then defined three frequency ranges of interest, corresponding to the mean syllabic rate of each condition. Within each frequency range, we used one-tail t tests to check whether averaged CSD values were statistically higher for the corresponding condition compared with each one of the other two (see Fig. 3*A*).

To compute phase coherence, we first determined the stimulus syllable rate by counting the average number of peaks in the envelope of the stimulus waveform per second, rounded to the closest integer. We then filtered both the stimulus and the EEG data using a third-order Butterworth bandpass filter at this frequency (±1 Hz) and computed the Hilbert transform of both signals and the mean resultant vector length of their phase difference. This yielded one value per subject for discrete syllable rates ranging from 4 to 14 Hz with 1 Hz steps (see Fig. 3*B*; values from all subjects are plotted as the distribution of syllabic frequencies was not the same across subjects). We then investigated whether neural responses entrained more at specific frequencies (e.g., in the θ band) or whether they similarly entrained to stimuli regardless of the syllable rate. We tested a linear and a quadratic relationship between the stimulus-averaged syllable frequency rate and the phase coherence at the same frequency. We reasoned that a significant linear relationship between syllabic rate and phase coherence should indicate privileged phase alignment for stimuli whose syllable rate is low or high. Furthermore, a negative quadratic relationship (inverted u-shape) between syllabic rate and phase coherence should indicate a privileged phase alignment for stimuli whose syllable rate falls within the θ band. We thus fitted two polyno-

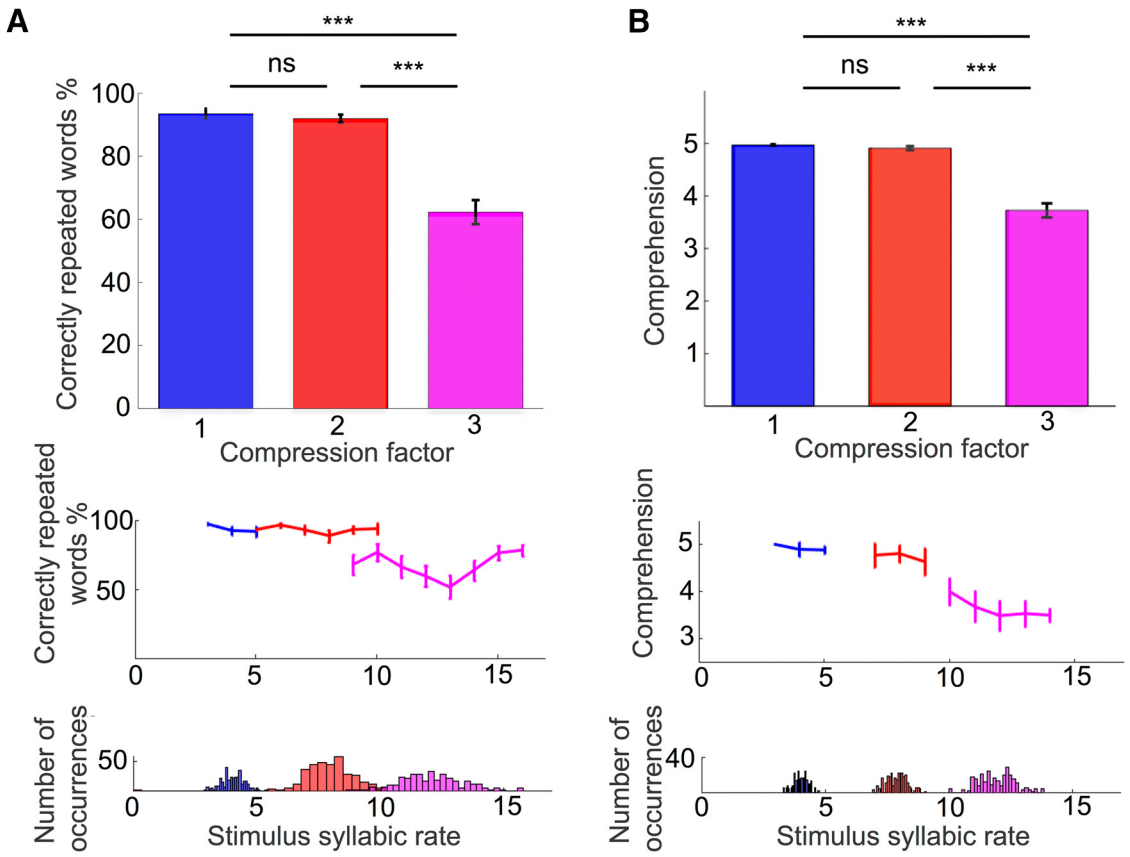

Figure 2. Behavioral results. ***A***, Behavioral task. Mean percentage of correctly reported words as a function of compression factor in the behavioral task. ***B***, Behavioral task during EEG recording. Mean comprehension ratings given by subjects to the sentences presented during the EEG recording. Left and right middle panels respectively represent the mean performance and comprehension ratings plotted as a function of the average syllabic rate. Bottom panels represent the number of occurrences of each syllabic rate for each task. Error bars indicate SEM. ***$p < 0.001$. n.s., Nonsignificant.

mials, of degree 1 and 2, to the individual data and, using *t* tests, tested whether linear and quadratic coefficients were reliably different from zero across subjects.

To disentangle the effect of linear syllabic rate increase and nonlinear comprehension decrease (see Fig. 4*A*) in the power of neural responses in each frequency band, we used a general linear model (GLM). This analysis was performed on single-trial power values, averaged across time, per frequency for each subject and electrode, using the two regressors: mean syllable rate and comprehension. The mean syllable rate was measured for each specific trial, estimated by the average number of peaks in the envelope of the stimulus waveform. Comprehension corresponded to the score provided by the subject for that specific trial. Both regressor values were *z*-scored per subject before computing the GLM. The GLM analysis assessed the part of the variance in neural responses explained by each regressor across electrodes and frequencies, independent of the effect of the other one (see Fig. 4*C*,*D*). As comprehension ratings are partly explained by syllable rate, we sought to explore the part of the variance solely associated with comprehension. To do so, we regressed out the effect of syllable rate from the comprehension ratings and computed the GLM using the residual scores. To reduce the dimensionality of the data, we averaged β values across electrodes and performed *t* tests across subjects for each regressor and each frequency (see Fig. 4*C*,*D*). We identified frequency-selective clusters reaching a $p < 0.05$ significance threshold (corrected for multiple comparison using nonparametric statistics; see below). To control whether the resulting effects of syllable rate and comprehension significantly differed from each other, we compared the respective β values against each other (see Fig. 4*D*, bottom). We then investigated how these frequency-selective effects evolved across time by computing the GLM per frequency band and time bin using an averaged sliding window of 50 ms, with 20 ms steps (see Fig. 5*A*). The time course of the corresponding parameter estimates—i.e., the normalized best-fitting regression coefficients, expressed in β values—measured the sensitivity of single-trial EEG signals to each of the regressors across time. β Values were tested against zeros per time window for each regressor and frequency band of interest using *t* tests across subjects. β Values for syllable rate and comprehension were also tested against each other using *t* tests across subjects (see Fig. 5*A*).

Statistical tests performed on the CSD and phase coherence analyses, as well as the parameters estimated with the GLM, were corrected for multiple comparisons using a cluster-based nonparametric approach (Maris and Oostenveld, 2007). Each statistical test performed was compared with 1000 permutations of the same test where we randomly shuffled the condition labels within subject.

Results

We primarily assessed the effect of time compression on sentence comprehension. The analysis revealed a significant main effect of compression rate ($F_{(2,50)} = 65.45$, $p < 0.001$, partial $\eta^2 = 0.4$) on the accuracy of reported word in the behavioral task. *Post hoc* tests revealed that while there was no significant difference between compression factors 1 and 2 ($t_{(16)} = 1.0929$, $p = 0.29$) performance significantly differed between compression factors 1 and 3, as well as between 2 and 3 ($t_{(16)} = 8.35$, $p < 0.001$ and $t_{(16)} = 8.41$, $p < 0.001$, respectively; all *post hoc* tests are Bonferroni-corrected). These results confirmed a drop in performance for speech compressed by a factor of 3, compared with a factor of 1 or 2 (Fig. 2*A*, top). Plotting these data as a function of the syllabic rate (Fig. 2*A*, bottom) confirmed that comprehension drops when speech rate exceeds ~10 syllables per second (Ghitza, 2014).

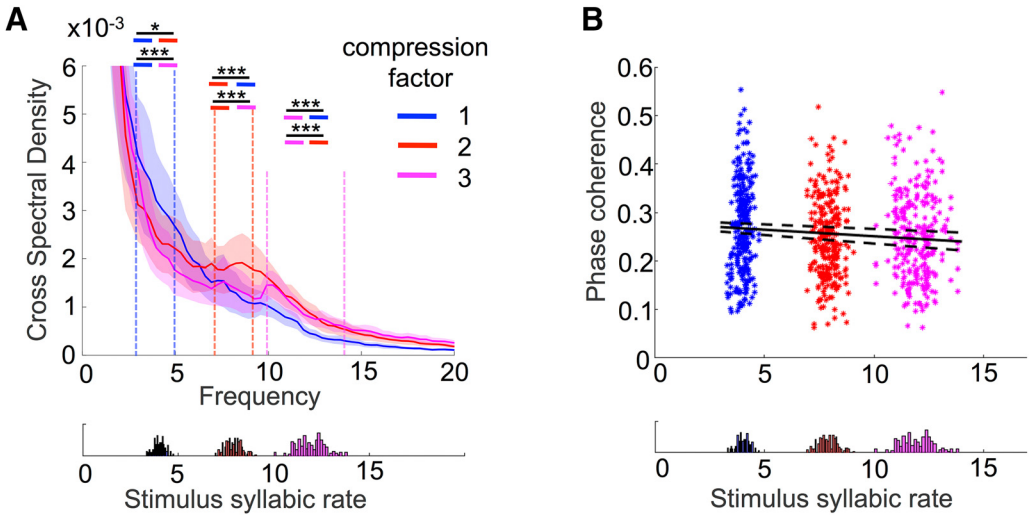

Figure 3. Linking the acoustic and the EEG data. ***A***, Cerebroacoustic CSD. Cross-spectral density values computed for each pair of acoustic stimuli and simultaneously recorded EEG data, for a frequency range from 1 to 20 Hz. The shaded areas correspond to the SEM, computed across participants. Dashed lines delimit the frequency ranges of interest, which for each compression factor correspond to its syllabic rates. Colored lines within each frequency range indicate the comparison of CSD values between the corresponding condition and the other two. ***B***, Phase coherence. Mean phase coherence value between the acoustic and EEG data at stimuli-averaged syllabic rates. Data from all subjects are plotted. The fitted line corresponds to the average of the fitted polynomials of degree 1 and dashed lines correspond to 1 SD of the same polynomials, divided by the square root of the sample size. Bottom panels represent the number of occurrences of each syllabic rate for each task. $^*p < 0.05$, $^{***}p < 0.001$. n.s., Nonsignificant.

We also analyzed the comprehension ratings made by participants at the end of each trial, during the EEG experiment. As expected, this analysis revealed a main effect of compression rate ($F_{(2,50)} = 80.54$, $p < 0.001$, partial $\eta^2 = 0.49$). Similarly to the analysis described above, we performed *post hoc* tests (Bonferroni-corrected), and found that there was no significant difference in perceived comprehension between compression factors 1 and 2 ($t_{(16)} = 1.85$, $p = 0.08$), but a significant difference between compression factors 1 and 3, as well as between factors 2 and 3 ($t_{(16)} = 9.06$, $p < 0.001$ and $t_{(16)} = 9.16$, $p < 0.001$ respectively; Fig. 2*B*, top). Comprehension (subjective) reports followed the same pattern as comprehension measures (Fig. 2*B*, bottom), confirming the reliability of this metric to assess the correlates of comprehension in our EEG data. This qualitative observation was supported by the strong correlation between the two metrics across stimuli ($r = 0.78$, $p < 0.001$).

To assess whether θ neural activity could track the syllabic structure in speech, we first computed the CSD between the acoustic stimuli and the neural data (Fig. 3*A*). We then compared CSD values between conditions in their respective syllabic rates. CSD was higher for noncompressed sentences compared with sentences compressed by a factor of 2 ($t_{(16)} = 3.59$, $p < 0.001$) and 3 ($t_{(16)} = 4.45$, $p < 0.001$) at 3–5 Hz. Similarly, sentences compressed by 2 yielded higher CSD values between 7 and 9 Hz, compared with noncompressed stimuli ($t_{(16)} = 4.01$, $p < 0.001$) and stimuli compressed by a factor of 3 ($t_{(16)} = 3.85$, $p < 0.001$). Finally, for frequencies between 10 and 14 Hz, CSD values were higher for sentences compressed by 3 compared with noncompressed sentences ($t_{(16)} = 6.34$, $p < 0.001$) as well as sentences compressed by 2 ($t_{(16)} = 3.88$, $p < 0.001$). To rule out the possibility of preferential phase-locking for one specific subfrequency band (e.g., the 4–9 Hz θ band), we tested the significance of the inverted U-shaped distribution on the phase-coherence values (see Materials and Methods), which indicated no significant trend ($t_{(16)} = 1.03$, $p = 0.12$). We also tested for linear behavior and found a significant negative linear trend ($t_{(16)} = -3.24$, $p = 0.002$), suggesting better phase alignment for stimuli with a low versus high syllable rate. Finally, for each participant, we regressed comprehension scores against the phase-coherence values, to test whether comprehension accounted for the negative linear relationship between phase-coherence and syllabic rate. We found that this was not the case, as the resulting regression parameter estimates failed to reach significance when tested against zero across subjects ($t_{(16)} = 1.56$, $p > 0.1$).

Using a linear model approach, we then investigated the relative effect of time compression and the correlate of comprehension on the neural responses across frequencies. We first observed a significant effect of syllabic rate in the θ (4 Hz) and low-γ (27–28 Hz) range (Fig. 4*C*), showing that increasing the speech compression rate resulted in an increase in θ power and a reduction in low-γ power. On the other hand, the comprehension regressor revealed that neural responses in the β (14–21 Hz) range were larger for sentences compressed by a factor of 1 and 2 (comprehensible), than for sentences compressed by a factor of 3 (incomprehensible; Fig. 4*D*). Despite the presence of power in α frequencies (Fig. 4*B*; centered around 10 Hz), no significant effect of syllable duration or comprehension was observed in this frequency band.

To further investigate the temporal dynamics of these effects in the θ (4 Hz), β (14–21 Hz), and low-γ (27–28 Hz) frequency bands, we ran a GLM with syllable rate and comprehension as regressors across time (Fig. 5*A*). We first observed that θ power increased (relative to prestimulus baseline) linearly with syllable rate during the first 2 s of auditory stimulation (Fig. 5*A*, left). We also found that low-γ power decreased as a function of compression rate. This effect started shortly (~1 s) after the beginning of the stimulus and did not last (Fig. 5*A*, middle). The time course of the regression parameter estimates in the β band (Fig. 5*A*, right) on the other hand, demonstrates that neural responses in this band increased for the least-compressed conditions compared with sentences compressed by a factor of 3. This shows that when speech is comprehensible, there is an increase in β power (starting ~1 s after the onset and lasting until the end of the sentences), whereas this effect is abolished when sentences are too heavily compressed.

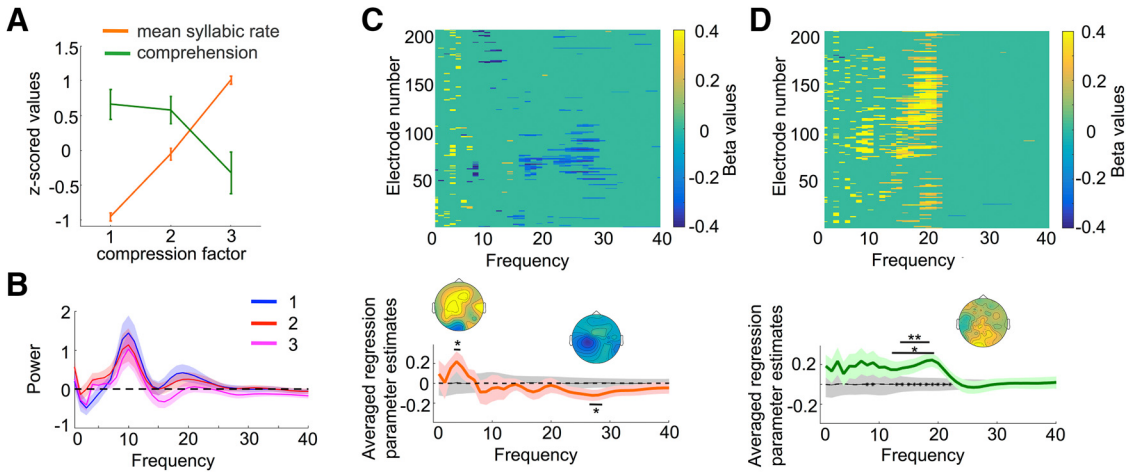

Figure 4. GLM on EEG power. ***A***, Regressors. Regressors used in the GLM, namely mean syllable rate and comprehension scores, expressed in *z*-scores averaged across subjects. Error bars represent SEM. ***B***, EEG power. Average power spectra of the EEG data for each compression factor, computed for frequencies between 1 and 40 Hz. Shaded areas correspond to the SEM, computed across participants. ***C***, Mean syllabic rate. Electrodes showing a significant effect for mean syllable rate as a function of frequency. ***D***, Comprehension. Electrodes showing a significant effect for comprehension, as a function of frequency. Effects are shown at $p < 0.05$, without correction for multiple comparisons for displaying purposes. The bottom plots in ***C*** and ***D*** represent the estimated regression parameters averaged across all electrodes for mean syllable rate and comprehension, respectively. Shaded areas indicate SEM. $*p < 0.05$, $***p < 0.001$. The dashed black line corresponds to the averaged betas estimated through 1000 permutations and the shaded gray areas represent the SEM of the parameters estimates generated by permuting the data. Black stars on the dashed line in ***D***, bottom plot, correspond to the frequencies where the estimated regression parameters for syllable rate and comprehension significantly differ from each other. Scalp topographies of the estimated parameters are shown for each significant effect.

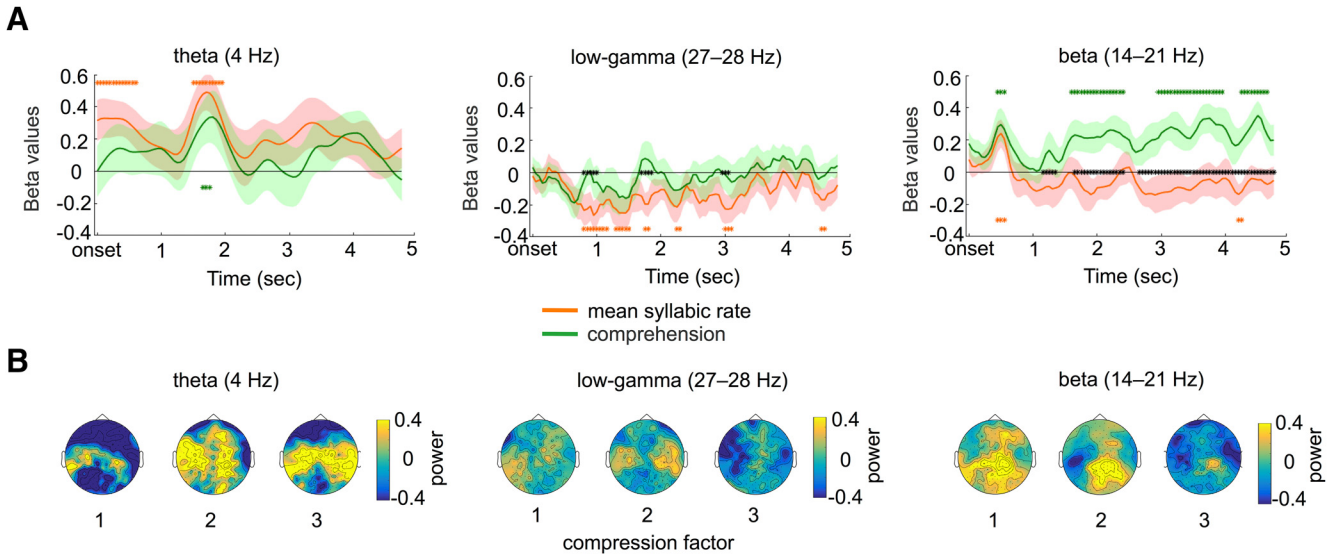

Figure 5. Time course of the effects of syllable rate and comprehension. ***A***, Regressions time courses. Time course of the estimated regression parameters averaged in the θ (4 Hz; left), low-γ (27–28 Hz; middle), and β (14–21 Hz; right) frequency bands. The shaded areas correspond to SEM. Orange and green stars denote the time windows of significance ($p < 0.05$) for mean syllable rate and comprehension regression parameters, respectively, and black stars denote the time windows where the estimated regression parameters significantly differ from each other (cluster-corrected for multiple comparisons, using 1000 permutations). ***B***, Power topographies. Scalp topographies of the power in θ (left), γ (middle), and β (right), averaged across the whole peristimulus epoch for each compression factor.

Discussion

Our data overall suggest that θ/low-γ and β oscillatory signals reflect distinct functional processes concurrently at play during speech perception. Specifically, they show that the syllable rate was consistently tracked by low-frequency neural phase patterns. Although this neural activity flexibly adapted to syllable rate beyond the upper limit of the classical θ rhythm (≤14 Hz), we assume from a functional viewpoint that it corresponds to a "θ" oscillation, because it showed distinct dynamics from the frequency-stable α peak centered on 10 Hz (Fig. 4*B*). Importantly, the syllable-tracking process was detected beyond 9 Hz, the point where comprehension started to drop. Our data confirm that θ neural activity can track the syllabic rhythm when the speed of speech varies (Ahissar et al., 2001), but does not support that 9 Hz is the upper limit of the tracking process (Ghitza, 2014). Our results, hence, do not confirm that speech decoding is limited by the ability of θ rhythm to track the syllabic structure of speech.

The observation that θ-band (4 Hz) power increases with syllabic rates suggests, however, that compression selectively affects oscillatory responses in this band. One interpretation for this could be that speech compression steepens the slope of syllable onsets, enhancing evoked responses in the θ band. This is sup-

ported by the early latency of this effect (Fig. 5*A*), suggesting that high compression rates boost early evoked responses, and that the θ band essentially reflects exogenous bottom-up processes. Previous works have related θ-band activity to comprehension and top-down mechanisms (Peelle et al., 2013; Park et al., 2015), but unlike here the acoustic manipulations in those studies were such that they did not permit the distinction between comprehension impairments resulting from a failure of the auditory system to encode the stimulus and those resulting from a higher-level decoding process. In light of these previous studies, the current data could suggest that entrainment in the θ band is a necessary but not sufficient condition for speech comprehension.

Our data also indicate that speech compression linearly reduced early low-γ responses (Fig. 4*C*). Low-γ (~30 Hz) rhythm approximately corresponds to the range of phonemic rate in speech and has been proposed to contribute to the encoding of fast temporal cues in auditory cortical neurons (Giraud and Poeppel, 2012). The early latency of the effect (Fig. 5*A*) and the bilateral temporal topography (compatible with an auditory evoked pattern; Fig. 5*B*) of brain responses in that frequency range suggests that time-compressing speech degrades phonemic encoding in the γ band. That low-γ (27–28 Hz) activity appears stronger at normal speech rates could indicate that phonemic information is most efficiently represented at its natural rate in the auditory system. The low-γ (20–40 Hz) range enables the encoding of discrete acoustic events individually (Joliot et al., 1994; Miyazaki et al., 2013) in the time range of the key phonemic cues (25–50 ms), while at higher speech rates phonemic information may be fused. However, given that γ-band responses can be observed for stimulus rates ≤100–150 Hz in the auditory cortex (Brugge et al., 2009; Nourski et al., 2013), it is possible that the denser information present in compressed speech is encoded by higher γ frequencies, which are simply less easily detectable by surface EEG measurements. In this view, the γ-power decrease we observe could be explained by the limited spectral sensitivity of EEG signals >40 Hz (the EEG spectrum typically follows a power law $1/f^{\alpha}$, where $1 \leq \alpha \leq 2$). At any rate, linear changes in γ-band responses cannot explain the abrupt, nonlinear comprehension drop observed for speech compressed by a factor of 3. This suggests that speech comprehension is not limited by bottom-up sampling/encoding, but more likely by additional endogenous processes involved in the downstream processing of the encoded information.

Contrasting with the linear relationship between compression rate and low-γ power, β (14–21 Hz) power decreased in a nonlinear way with compression, and accounted for comprehension ratings at the single-trial level. Interestingly, this effect built up ~1.5 s after sentence onset and was sustained until the end of analyzed epochs. Given the established role of β oscillations in top-down predictive mechanisms (Schubert et al., 2009; Engel and Fries, 2010, Arnal and Giraud, 2012; Fontolan et al., 2014; Volberg and Greenlee, 2014; Bastos et al., 2012; Sedley et al., 2016), we interpret this effect in the context of generative models of perception (Friston, 2005), where the brain recurrently uses available sensory information to generate predictions propagating top-down in the β-frequency channel. This β-band signal might correspond to the endogenous processes required for comprehension that we conjectured earlier. This "top-down" interpretation is supported by the topography of β power, which, in contrast with the auditory evoked topography of low-γ activity (Fig. 5*B*), is compatible with a parietal or premotor cortical source. As previously hypothesized by Arnal and Giraud (2012), when speech information is conveyed at a comprehensible pace, the gradual buildup of β activity across time would reflect the use of contextual information to generate on-line predictions while the sentence unfolds. On the other hand, the absence of β activity associated with unintelligible sentences suggests that when syllabic information is presented too fast, the deployment of top-down mechanisms is disrupted. This may occur either because the time between successive syllabic information packets is too short for the speech-perception system to dynamically deploy predictions, or simply because the task is too demanding on the participant's attentional resources.

The present study extends previous knowledge by showing that comprehension not only depends on the tracking of the speech envelope by θ-range oscillations, but is also reflected in the buildup of β oscillations, which likely reflects the engagement of top-down mechanisms. The exact mechanism by which β oscillations contribute to speech comprehension remains unclear. One possibility is that they carry temporal predictions about the timing of the onset of upcoming syllables, which could facilitate the extraction of relevant information (Arnal, 2012; Arnal et al., 2015; Kulashekhar et al., 2016; Merchant and Yarrow, 2016). An alternative hypothesis could be that β oscillations carry more than timing information (Bastiaansen and Hagoort, 2006; Bastiaansen et al., 2010; Magyari et al., 2014) and provide an informational substrate for analysis-by-synthesis processes (Halle and Stevens, 1962; Poeppel et al., 2008). In this view, β signals might facilitate the predictive processing at other linguistic levels of information (phonemic, semantic, or syntactic organization) in a generative manner (Lewis and Bastiaansen, 2015; Lewis et al., 2016) by adaptively tuning lower-tier sensory areas to improve the decoding of the speech signal's content. At any rate, our data suggest that this process requires a certain amount of time (~100 ms) between syllabic information packets to read out and synthesize meaningful representations of what is being (and will be) said.

Perceptual synthesis and multiplexing

The current results suggest that β oscillations might constitute the neural substrate of top-down signals, complementing the proposed role of θ and γ oscillations in speech exogenous encoding (Giraud and Poeppel, 2012; Pasley et al., 2012; Gross et al., 2013; Hyafil et al., 2015; Ding et al., 2016). This interpretation provides additional support to the proposal that feedforward and feedback signals are transmitted through different frequency bands. Recent evidence from intracranial recordings (Fontolan et al., 2014) further suggests that these two processes might alternate in time, meaning that the generation of feed-back signals would follow the propagation of bottom-up ones every 250 ms on average (2 up/down cycles per second) in the auditory cortex. The dynamic alternation of the two processes implies that speech encoding and readout take place in a discretized manner. If this interpretation holds, it would suggest that compressing speech by a factor of 3 saturates the capacity of the speech-perception process by disrupting its top-down/descending phase. This speculation is supported by the observation that the effects of time compression (up to a factor of 8; Ghitza, 2014) on comprehension can be reduced if compressed speech packets are presented at a slower pace, separated by 80 ms periods of silence, which do not add any speech-related information but only offer more decoding time.

Conclusion

Our results confirm that the neural mechanisms of speech comprehension are disrupted when listening to time-compressed

speech. Although previous studies suggest that successful envelope tracking is necessary to comprehend speech (Nourski et al., 2009; Doelling et al., 2014), our results suggest that faithful speech encoding by θ–γ oscillators may not be sufficient for comprehension. The time-compression paradigm reveals the existence of a temporal bottleneck for comprehension that critically depends on the ability to reconstruct the incoming input in a generative manner. Speech comprehension does not seem limited by the encoding capacity (at least within the levels of compression used here), but by the time required for reading out the information after it has been encoded and deploying predictions, a process possibly instantiated by θ-based syllabification enabling β-based predictive processes (Arnal and Giraud, 2012). These mechanisms provide a plausible neurophysiological substrate for rapidly "recoding" the speech input and building linguistic representations in a predictive manner, as conjectured by recent theoretical views in psycholinguistics (Christiansen and Chater, 2016).

References

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98:13367–13372. CrossRef Medline

Arnal LH (2012) Predicting "when" using the motor system's beta-band oscillations. Front Hum Neurosci 6:225. CrossRef Medline

Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci 16:390–398. CrossRef Medline

Arnal LH, Doelling KB, Poeppel D (2015) Delta-beta coupled oscillations underlie temporal prediction accuracy. Cereb Cortex 25:3077–3085. CrossRef Medline

Bastiaansen M, Hagoort P (2006) Oscillatory neuronal dynamics during language comprehension. Prog Brain Res 159:179–196. CrossRef Medline

Bastiaansen M, Magyari L, Hagoort P (2010) Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. J Cogn Neurosci 22:1333–1347. CrossRef Medline

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711. CrossRef Medline

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. CrossRef Medline

Brunet D, Murray MM, Michel CM (2011) Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput Intell Neurosci 2011:813870. CrossRef Medline

Christiansen MH, Chater N (2016) The now-or-never bottleneck: a fundamental constraint on language. Behav Brain Sci 39:e62. CrossRef Medline

Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res 229:132–147. CrossRef Medline

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. J Exp Psychol Gen 134:222–241. CrossRef Medline

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21. CrossRef Medline

Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 19:158–164. CrossRef Medline

Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85:761–768. CrossRef Medline

Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. J Neurosci 32:1791–1802. CrossRef Medline

Engel AK, Fries P (2010) Beta-band oscillations—signalling the status quo? Curr Opin Neurobiol 20:156–165. CrossRef Medline

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud AL (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun 5:4694. CrossRef Medline

Friston K (2005) A theory of cortical responses. Philos Trans R Soc B Lond B Biol Sci 360:815–836. CrossRef Medline

Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: comprehension of speech with a manipulated modulation spectrum. Front Psychol 3:238. CrossRef Medline

Ghitza O (2014) Behavioral evidence for the role of cortical oscillations in determining auditory channel capacity for speech. Front Psychol 5:652. CrossRef Medline

Ghitza O, Greenberg S (2009) On the possible role of brain rhythms in speech perception: comprehension of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica 66:113–126. CrossRef Medline

Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517. CrossRef Medline

Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol 11:e1001752. CrossRef Medline

Halle M, Stevens K (1962) Speech recognition: a model and a program for research. IRE Trans Information Theory 8:155–159. CrossRef

Hervais-Adelman A, Moser-Mercer B, Michel CM, Golestani N (2015) FMRI of simultaneous interpretation reveals the neural basis of extreme language control. Cereb Cortex 25:4727–4739. CrossRef Medline

Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. J Neurophysiol 104:2500–2511. CrossRef Medline

Hyafil A, Fontolan L, Kabdebon C, Gutkin B, Giraud AL (2015) Speech encoding by coupled cortical theta and gamma oscillations. eLife 4:e06213. CrossRef Medline

Joliot M, Ribary U, Llinás R (1994) Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. Proc Natl Acad Sci U S A 91:11748–11751. CrossRef Medline

Kulashekhar S, Pekkola J, Palva JM, Palva S (2016) The role of cortical beta oscillations in time estimation. Hum Brain Mapp 37:3262–3281. CrossRef Medline

Lewis AG, Bastiaansen M (2015) A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex 68:155–168. CrossRef Medline

Lewis AG, Schoffelen JM, Schriefers H, Bastiaansen M (2016) A predictive coding perspective on beta oscillations during sentence-level language comprehension. Front Hum Neurosci 10:85. CrossRef Medline

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010. CrossRef Medline

Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol 8:e1000445. CrossRef Medline

Magyari L, Bastiaansen MC, de Ruiter JP, Levinson SC (2014) Early anticipation lies behind the speed of response in conversation. J Cogn Neurosci 26:2530–2539. CrossRef Medline

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190. CrossRef Medline

Merchant H, Yarrow K (2016) How the motor system both encodes and influences our sense of time. Curr Opin Behav Sci 8:22–27. CrossRef

Miyazaki T, Thompson J, Fujioka T, Ross B (2013) Sound envelope encoding in the auditory cortex revealed by neuromagnetic responses in the theta to gamma frequency bands. Brain Res 1506:64–75. CrossRef Medline

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3rd, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564–15574. CrossRef Medline

Nourski KV, Brugge JF, Reale RA, Kovach CK, Oya H, Kawasaki H, Jenison RL, Howard MA 3rd (2013) Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: an intracranial electrophysiology study. J Neurophysiol 109:1283–1295. CrossRef Medline

Park H, Ince RA, Schyns PG, Thut G, Gross J (2015) Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr Biol 25:1649–1653. CrossRef Medline

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF (2012) Reconstructing speech from human auditory cortex. PLoS Biol 10:e1001251. CrossRef Medline

Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23:1378–1387. CrossRef Medline

Poeppel D, Idsardi WJ, van Wassenhove V (2008) Speech perception at the interface of neurobiology and linguistics. Philos Trans R Soc Lond B Biol Sci 363:1071–1086. CrossRef Medline

Schubert R, Haufe S, Blankenburg F, Villringer A, Curio G (2009) Now you'll feel it, now you won't: EEG rhythms predict the effectiveness of perceptual masking. J Cogn Neurosci 21:2407–2419. CrossRef Medline

Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD (2016) Neural signatures of perceptual inference. Elife 5:e11476. CrossRef Medline

Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. J Neurosci 32:8443–8453. CrossRef Medline

Verhelst W, Roelands M (1993) An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. Proc ICASSP 2:554–557. CrossRef

Volberg G, Greenlee MW (2014) Brain networks supporting perceptual grouping and contour selection. Front Psychol 5:264. CrossRef Medline