Dynamics of functional networks for syllable and word-level processing

Rimmele, J.M.[1*], Sun, Y.[1], Michalareas, G.[1], Ghitza[1,2], O., Poeppel, D.[1,3]

[1]Department of Neuroscience
Max-Planck-Institute for Empirical Aesthetics
Grüneburgweg 14
D - 60322 Frankfurt am Main, Germany

[2]College of Biomedical Engineering & Hearing Research Center
Boston University
44 Cummington Mall,
Boston, MA, 02215, USA

[3]Department of Psychology and Center for Neural Science
New York University
6 Washington Place
New York , NY 10003, USA

Correspondence: johanna.rimmele@ae.mpg.de (J.M. Rimmele)

Abstract

Speech comprehension requires the ability to temporally segment the acoustic input for higher-level linguistic analysis. Oscillation-based approaches suggest that low-frequency auditory cortex oscillations track syllable-sized acoustic information and therefore emphazise the relevance of syllabic-level processing for speech segmentation. Most linguistic approaches, however, focus on mapping from acoustic-phonemic representations to the lexical level. How syllabic processing interacts with higher levels of speech processing, beyond segmentation, including the anatomical and neurophysiological characteristics of the networks involved, is debated. Here we investigate the effects of lexical processing and the interactions with syllable processing by examining MEG data recorded in a frequency-tagging paradigm. Participants listened to disyllabic words presented at a rate of 4 syllables/sec. Two hypotheses were evaluated: (i) lexical processing of words activates a 2 Hz network that interacts with syllable processing at 4 Hz; and (ii) syllable transitions contribute to word-level processing. We show that lexical content activated a left-lateralized frontal and superior and middle temporal network and increased the interaction between left middle temporal areas and auditory cortex at 2 Hz (phase-phase coupling). Mere syllable-transition information, in contrast, activated a bilateral superior-, middle temporal and inferior frontal network and increased the interaction between those areas at 2 Hz. Lexical and syllable processing interacted in superior, middle temporal, and inferior frontal areas (cross-frequency coupling), whereas syllable tracking decreased when lexical information was present. The data provide a new perspective on speech comprehension by demonstrating a contribution of an acoustic-syllabic to lexical processing route.

Keywords: speech; functional networks; lexical; syllable transition; oscillation

Significance statement

The comprehension of speech requires integrating information at multiple time scales, including phonemic, syllabic, and word scales. Typically, we think of decoding speech in the service of recognizing words as a process that maps from phonemic units to words. Recent neurophysiological evidence, however, has highlighted the relevance of syllable-sized chunks for segmenting speech. Is there more to recognizing spoken language? We provide neural evidence for brain network dynamics that support an interaction of lexical with syllable-level processing. We identify cortical networks that differ depending on whether lexical-semantic information versus low-level syllable-transition information is processed. Word- and syllable-level processing interact within MTG and STG. The data enrich our understanding of comprehension by implicating a mapping from syllabic to lexical representations.

Introduction

Speech comprehension requires segmenting acoustic information at multiple time scales and matching extracted chunks to approximately corresponding linguistic units (e.g. phonemes, syllables, words etc.). Recent oscillation-based approaches posit that the transformation into perceptually relevant acoustic chunks is achieved by temporally segmenting the continuous input signal (1–6). At the neuronal level, temporal segmentation is argued to be realized through phase-alignment of low-frequency (< 8 Hz; delta-theta) neuronal oscillations in auditory cortex to the slow (quasi-rhythmic) fluctuations of the speech signal at the syllabic scale ('speech tracking') (7–12). This speech tracking possibly indicates the alignment of oscillatory excitability cycles of neuronal ensembles, which can result in perceptual constraints (13) such that speech perception is optimal at syllabic rates that fall within the range of intrinsic auditory cortex oscillations in the delta-theta range (2, 14, 15 see also 16).

Several experiments have exploited oscillation-based approaches by testing hierarchical processing at multiple linguistic processing levels, such as word, phrase, and sentence comprehension (17–19), for review (20, 21). Emerging evidence from neurophysiological studies points to interactions of higher level speech processing and syllable processing. For example, predictions arising from multiple processing levels such as phonological or syntactic/semantic processing (22–26) might inform lower levels. The top-down modulation of the phase-alignment of low-frequency auditory cortical oscillations to speech acoustics is suggested by increased speech-tracking for intelligible versus unintelligible speech (27, 28, 12, 29), which was accompanied by increased connectivity between auditory cortex and higher-level processing areas, such as frontal and motor cortex (29, 30). However, at what linguistic level, e.g. syllable, word, phrase (31) etc., the processes interact - as well as whether this is due to lexical-semantic (32, 33) or phonological processing (34, 35) - is not fully understood. A range of puzzles remain both at the implementational (neural) and the computational (linguistic and psycholinguistic) levels. Regarding the latter, there are effects (e.g. there are interactions between lexical and phonemic representations) and debates (e.g. are there interactions with syllables, contentious units of representation?). Regarding the former, many of the spectral characteristics of the network dynamics are unknown. Low-frequency delta-theta neuronal oscillations are supposedly crucial for tracking syllable-sized acoustic chunks; however, whether the brain areas implicated in syllable processing interact with higher-level processes through phase-synchronization at these frequencies (29, 36) or through cross-frequency coupling

(31), and whether networks involved in higher-level processing synchronize in a similar spectral range, is unknown. Understanding spectral characteristics of networks provides important insights into the coupling mechanisms, i.e. which brain areas communicate how through synchronized activity (37–40), and possibly the type of connection, such as feedback or feedforward (10, 41, 42).

Based on imaging research, distinct networks have been identified for phonological and lexical-semantic processing of speech. There is evidence for widely distributed semantic representations (43). That being said, the posterior MTG (perhaps left lateralized (44)) has been suggested to provide a sound-to-meaning interface for the mapping of phonological representations in STS to lexical representations (45, 46, 44, 47, 48). Lexical-semantic compositional processing as well as sentence level processing additionally has been shown to activate the anterior temporal lobe (ATL) (49–51). *Sublexical* contingencies, such as syllable transitions, contribute to word processing, and have been shown to activate parts of the STS (52, 53) and a dorsal-path network (54). Additionally, parts of the IFG have been shown to contribute to word-level processing, in sublexical processing tasks, supposedly reflecting sensory-motor integration as part of a dorsal articulatory network (45, 54) or in tasks that elicit competition between lexical competitors (48, 55, 56). Here we show how these networks may form the basis for the interaction between syllable-level and lexical-level neural processing.

We used a frequency-tagging paradigm (57, 17) to investigate the functional networks involved in processing lexical (Exp. 1) and syllable transition (Exp. 2) cues of words as well as the network interaction underpinning syllable and word processing. In two MEG experiments, native German speakers listened to isochronously presented syllables of a native (German) and a foreign (Turkish) language, presented at a rate of 4 Hz (4 syllables/sec), resulting in a rate of 2 Hz for 'disyllabic units', i.e. words or pseudowords, while performing a simple target discrimination task (Fig. 1, A). In Exp. 1 (n = 18; Fig. 1 B, C), the lexical processing of words in the native versus the foreign language was compared. In Exp. 2 (n = 19; Fig. 1 B, D), the effects of lexical plus syllable transition information on the grouping of syllables into words was investigated by comparing (i) the processing of native words that contained syllable transition information and lexical content with (ii) foreign pseudo-words ('Non-Turkish') with neither content. Additionally, a cross-experiment comparison of foreign words with syllable transition content and foreign pseudowords allowed us to investigate syllable transition processing (Turkish versus Non-Turkish). We hypothezised that the lexical processing of words results in neuronal power increases at the word presentation rate (2

Hz), within a network that additionally to the STG includes, particularly, the left MTG. Interactions of word- and syllable-level processing were expected to result in coupling between the 2 Hz putatively word-related and the 4 Hz syllable-related activity. Additionally, syllable transition information was expected to affect the grouping of syllables into words, resulting in increased power at the word rate, and involving a bilateral fronto-temporal network including IFG.

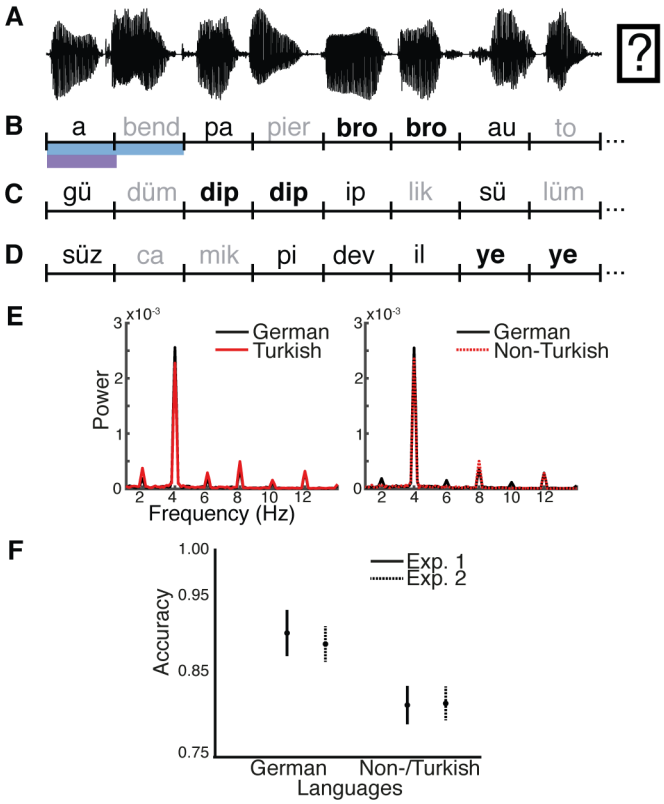

Figure 1. Schematic of the Paradigm and Material. (A) Illustrates the structure of a trial: participants indicated after listening to syllable sequences whether a target (bold) was present; In both experiments syllables were presented at a rate of 4 Hz (250 ms; purple bar); (B) In Exp. 1 in the German condition syllables can be grouped into disyllabic words (black: first syllable, gray: second syllable) at a rate of 2 Hz (500 ms; blue bar); (C) In Exp. 1, in the Turkish condition the syllables presented at 4 Hz cannot not be grouped into words based on semantic knowledge; In Exp. 2, the German condition corresponded to that of Exp. 1; (D) In Exp. 2, in the Non-Turkish condition syllables, presented at 4 Hz, cannot be grouped into words based on semantic knowledge or syllable transition information; (E) The speech acoustics show a strong peak at the frequency-tagging (syllable) rate (4 Hz) in Exp. 1 (left column) and Exp. 2 (right column); (F) Overall, the performance (accuracy) did not differ across experiments, however, overall was higher for the German compared to the Turkish/Non-Turkish conditions (error bars: +- 1 SEM).

Results

Behavioral Measures. A repeated-measures ANOVA tests the effect of lexical and syllable transition cues on target discrimination accuracy (Fig. 1 F) in Exp. 1 and 2 (between-subject factor:

experiment; within-subject factor: condition). Accuracy was higher in the German compared to the Turkish/Non-Turkish conditions ($F(1, 35) = 100.214$; $p < 0.001$; $\eta^2 = 0.374$; German: 92 % vs Turkish/Non-Turkish: 83%). There was no main effect of experiment ($F(1, 35) = 1.794$; $p > .1$; $\eta^2 = 0.039$) or interaction ($F(1, 35) = 1.303$; $p > .1$; $\eta^2 = 0.008$). The results indicate that the presence of lexical cues (in the German conditions) facilitated performance.

Lexical and Syllable Transition Processing Activate Functional Networks at 2 Hz. In the sensor-space MEG analysis, lexical access effects are reflected in Exp. 1 in power increases in the German compared to the Turkish condition at 2 Hz, at a left frontal and left temporal cluster (p = 0.0040; Fig. 2 A). In source space, the comparison revealed differences at 2 Hz at a left fronto-centro-temporal cluster (p = 0.0040; including STG, MTG, insula, supramarginal gyrus) and a right hemisphere cluster (p = 0.0420; including STG, MTG, MFG, frontal operculum, postcentral gyrus, putamen; Fig. 3 A). Non-parametric comparisons performed at 2 Hz separately for the left and right hemisphere and at the pSTG, aMTG, MTG, SMG, IFG, PMC ROIs revealed condition differences at the left PMC (p = .0038; Bonferroni corrected alpha = .0042) and a trend in the left aMTG (p = .0096; $alpha_{trend}$ = .0167) (Fig. 3 A).

The cross-experiment comparison shows sensor-space syllable transition processing effects (Turkish vs. Non-Turkish) within a broad left hemispheric cluster (p = 0.0040) and a broad right hemispheric cluster (p = .0080; Fig. 2 B). In source space, syllable transition processing resulted in increased power in the Turkish compared to the Non-Turkish condition at a bilateral frontal, central and temporal cluster (p = .0020; including the STG, MTG, Heschl's gyrus, precentral gyrus, supramarginal gyrus; Fig. 3 B). Non-parametric comparisons performed at 2 Hz revealed condition differences at the left pSTG, aMTG, MTG, SMG, IFG (ps < .0007; Bonferroni corrected alpha = .0042). In the right hemisphere condition differences were significant at the MTG, SMG, PMC (ps < .0030; alpha = .0042) and there was a trend at the pSTG, aMTG, IFG (p = .0045; p = .0121; p = .0122; $alpha_{trend}$ = .0167).

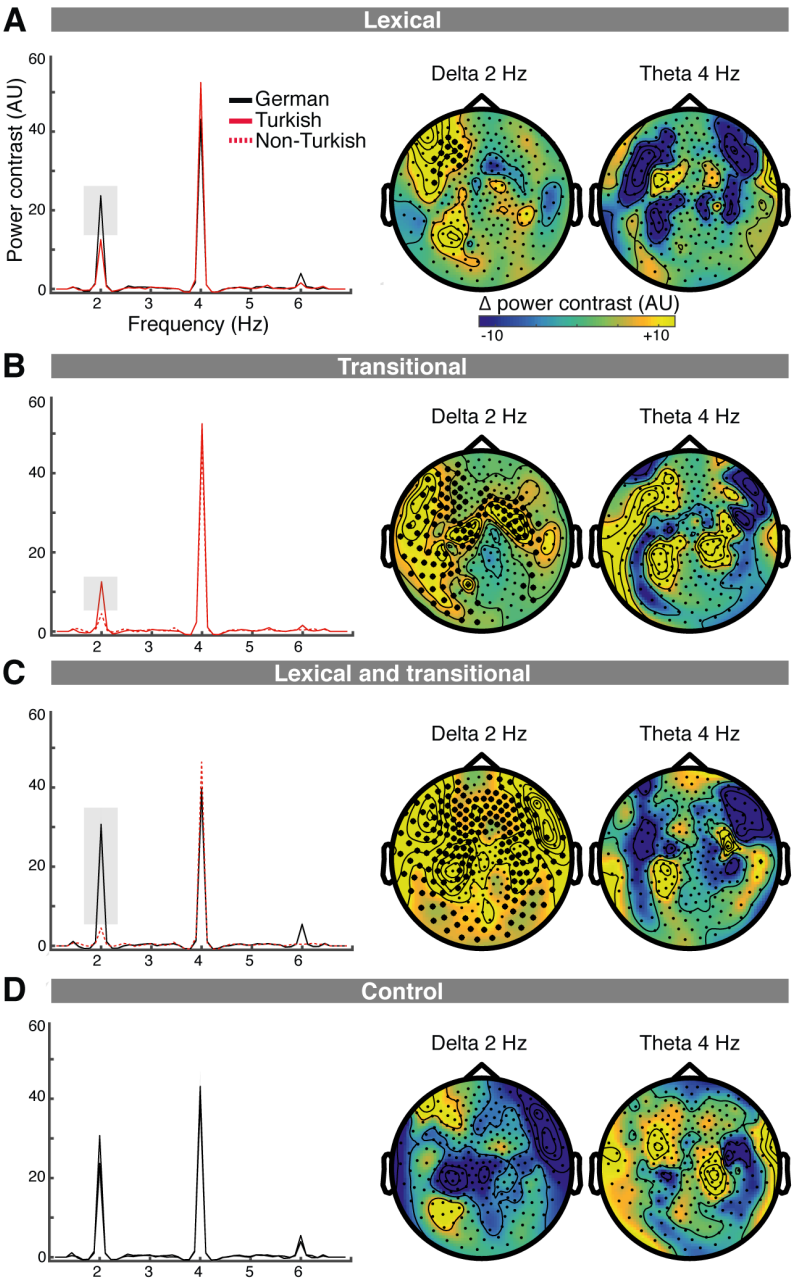

Figure 2. Lexical and Syllable Transition Processing Increases Sensor Space Power. (A-D left column) Power contrasts (with the neighbouring frequency bins) averaged across the individual M100 sensors are displayed; (A-D right column) The topography of the power contrast differences between conditions is displayed at 2 Hz and 4 Hz. Clusters that showed significant differences are marked with black dots; (A) Lexical processing, in Exp. 1, resulted in increased power at 2 Hz for the German compared to the Turkish condition at a left frontal cluster; (B) Syllable transition processing, resulted in increased power in the Turkish compared to the Non-Turkish condition at 2 Hz at a left fronto-central and temporal, and a right fronto-central cluster; (C) Lexical plus syllable transition processing, in Exp. 2, resulted in increased power at 2 Hz in the German compared to the Turkish condition at a broadly distributed cluster; (D) No differences were detected for the across experiment comparison of the German conditions.

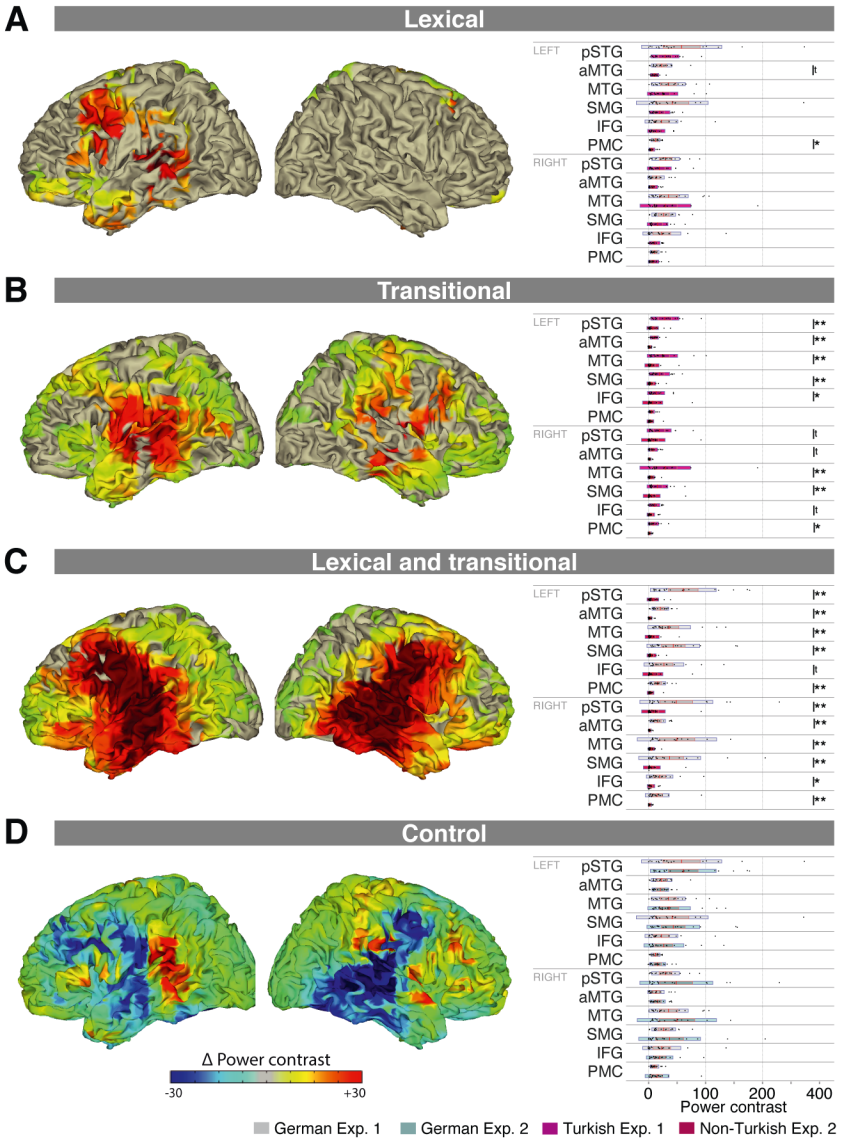

Figure 3. Lexical and Syllable Transition Processing Activates Frontal and Temporal Cortex. (A) (left column) In Exp. 1, lexical processing resulted in increased power at 2 Hz in the German compared to the Turkish condition in two fronto-central-temporal clusters. (right column) Condition differences were significant in the left SMG and there was a trend in the left PMC; (B) (left column) Syllable transition processing resulted in a broad left and a broad right hemispheric cluster showing power increases at 2 Hz; (right column) Condition differences were significant in several left and right hemispheric ROIs; (C) (left column) Lexical plus syllable transition processing resulted in a broad bilateral cluster showing power increases; (right column) Condition differences were significant in several left and right hemispheric ROIs; (D) No significant differences were revealed in the German conditions across experiments;

Lexical plus sublexical processing, in Exp. 2, resulted in sensor power increases in the German compared to the Non-Turkish condition at a bilateral widespread cluster (p = .0040; Fig. 2 C). In source space, connectivity was increased in the German compared to the Non-Turkish condition at a bilateral frontal, central and temporal cluster at 2 Hz (p = .0020; including the STG, MTG,

precentral/postcentral gyrus, frontal inferior operculum; Fig. 3 C). Non-parametric comparisons performed at 2 Hz revealed significant condition differences in the left hemisphere at the pSTG, aMTG, MTG, SMG and PMC (ps < .0001; Bonferroni corrected alpha = .0042) and a trend at the IFG (p = .0062; $alpha_{Trend}$ = .0083). In the right hemisphere condition differences were significant at all ROIs (ps < .0022).

There were no significant differences detected at any cluster for the cross-experiment control comparison of the German conditions (sensor space: Fig. 2 D; source space: Fig. 3 D) and no condition differences at any ROI of the two hemispheres (ps > .2674). Likewise, there were no effects at 4 Hz at any comparison in sensor or source space.

Lexical and Syllable Transition Processing Increase Connectivity Within a 2 Hz Network. In sensor space, lexical access (Exp. 1) resulted in connectivity increases between the M100 ROI and a left fronto-temporal cluster at 2 Hz (p = .0020; Fig. 4 A). Source space analysis revealed no cluster with significant differences in the STG connectivity that would indicate lexical access effects (Exp. 1). More specific hypotheses were tested by analyzing connectivity increases between the ROIs of interest (STG, MTG, IFG). In Exp. 1, the connectivity increases between the German and Turkish condition did not reach significance at any ROI pair and frequency (Fig. 4 B; 2 Hz: p-values > .0936 and 4 Hz: p-values > .0156; Bonferroni corrected alpha = .0083).

Lexical plus syllable transition processing (Exp. 2), resulted in connectivity increases in the German compared to the Non-Turkish condition between the M100 sensors and at a bilateral widespread cluster (p = .0020; Fig. 4 C). In source space, connectivity increased between STG and a right fronto-centro-temporal network (p = .0200; including right MTG, SMG, pars opercularis and pre- and postcentral areas) and a left lateralized one (p = .0360; including MTG, SMG, pars opercularis and MTG). A more specific analysis revealed, in Exp. 2, at 2 Hz connectivity was increased between all ROIs (IFG-MTG: p = .0022; STG-MTG: p = .0033; IFG-STG: p = .0025; Bonferroni corrected alpha = .0083; Fig. 4 D), no differences were detected at 4 Hz (p-values > .03547).

No clusters were detected at 4 Hz for any comparison.

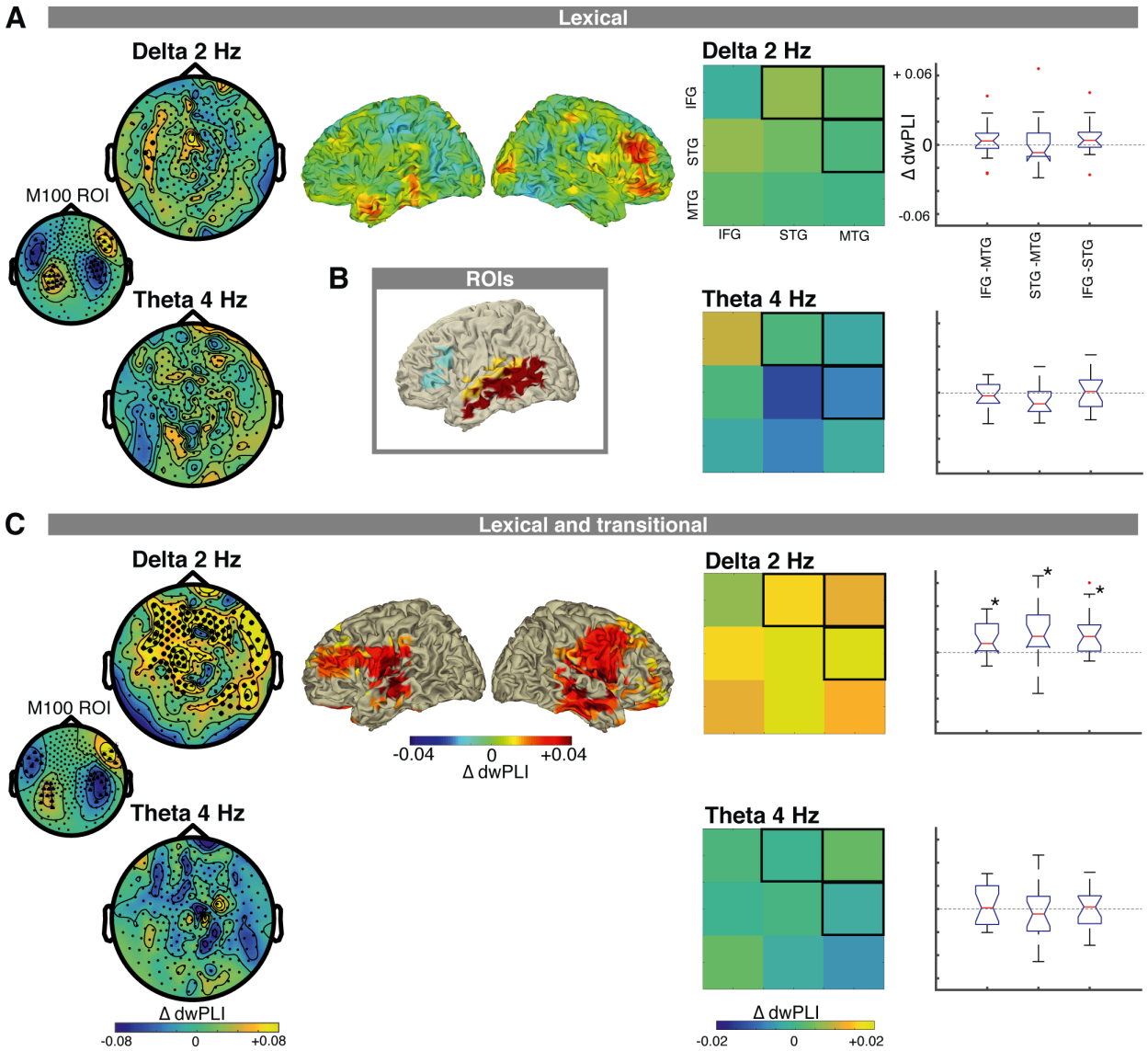

Figure 4. Frontal and Temporal Brain Areas Interact During Lexical and Syllable Transition Processing. (A) Lexical processing: (Left column) The individual M100 sensors were used as region of interest (ROI) for the sensor space connectivity analyses (Exp. 1: the grand average M100 sensors are displayed). In Exp. 1, in the German compared to the Turkish condition connectivity increased at 2 Hz between the M100 ROI and a left fronto-temporal cluster; (Middle column) No significant clusters were detected that showed differences in source connectivity for this comparison (right column) The source connectivity matrix displays condition differences between the STG, MTG, IFG (black boxes mark relevant connectivity pairs). The contrasts are also plotted as bar graphs. (B) Three ROIs including the STG (yellow), MTG (red) and IFG (blue) were selected; (C) Lexical and syllable transitional processing: (Left column) The M100 ROI for Exp. 2 is displayed. In Exp. 2, in the German compared to the Non-Turkish condition connectivity increased at 2 Hz between the M100 ROI and a fronto-temporo-central cluster; (Middle column) The source connectivity between STG and a bilateral middle temporal, and fronto-central cluster increased at 2 Hz for the German vs. the Non-Turkish condition; (Right columns) source connectivity was significantly increased between the IFG and MTG, STG and MTG and IFG and MTG ROI at 2 Hz.

Word-Level Processing at 2 Hz Affects Syllable Processing at 4 Hz. The findings suggest increases in cross-frequency coupling (i.e. the 4 Hz power was computed for 4 different phase-bins of the 2 Hz activity) in different brain areas when lexical and when lexical+syllable transition information is present. In Exp. 1 (Fig. S1 A), power increased in the German compared to the Turkish condition in the MTG ROI for phase bin 1 compared to bin 3 ($p = .0005$; Bonferroni corrected alpha = .0028) and bin 3 compared to 4 ($p = .0001$). In Exp. 2, power increased in the German compared to the Non-Turkish condition at the STG ROI in bin 1 compared to bin 3, however, survived the multiple comparison control only as a trend ($p = .0055$; Bonferroni corrected alpha-trend-0.1 = .0056).

Additional analyses performed separately for all conditions showed cross-frequency coupling in conditions with lexical and transitional and syllable information (German), with transitional and syllable information (Turkish), but as expected, not with only-syllable information (Non-Turkish) (Fig. S1 B). In the German conditions (merged across experiments), power varied across phase-bins 1 and 2 in STG ($p = .0004$; Bonferroni corrected alpha = .0028; a trend for phase-bin 1 and 3, $p = .0053$), and in IFG between bin 1 and 2, bin 1 and 3 ($ps < .0001$) and bin 3 and 4 ($p = .0009$); in the Turkish condition across the phase-bins 1 and 2 and bin 1 and 3 in MTG ($ps < .0005$; a trend for phase-bin 3 and 4, $p = .0042$). There was no significant coupling in the Non-Turkish condition. The findings suggest cross-frequency coupling was present in all but the Non-Turkish condition, which did not require word-level processing.

An analysis of cerebro-acoustic coherence in STG suggests, interestingly, that word-level processing affects syllable tracking at 4 Hz in STG by decreasing it. As the direction of changes in cerebro-acoustic coherence in the German conditions and Turkish/Non-Turkish conditions were the same, the merged German conditions were tested against the merged Turkish/Non-Turkish conditions, showing significant decreases in the German conditions ($p = .0062$).

Discussion

In two MEG experiments, we show that the frequency-tagging paradigm can be used to tease apart aspects of lexical-level and syllable-transition information processing by differentiating neuronal networks activated at 2 Hz (Fig. 5). Speech comprehension models have focused on mapping of acoustic-phonemic to lexical processing, e.g. (58), supported by important cognitive neuroscience evidence for the processing and encoding of phonemic level information (53, 35). Other models

have suggested, however, that suprasegmental information such as stress or syllable rhythm can facilitate lexical access (59, 60). The data we provide enrich our understanding by implicating contributions of an acoustic-syllabic to lexical processing route. Specifically, we show interactions between lexical and syllable-transition processing of words (here at 2 Hz) and syllable processing (here at 4 Hz), first, in terms of phase-amplitude coupling, and second, in terms of a decreased tracking of syllables (cerebro-acoustic coherence) at 4 Hz in STG when lexical or syllable transition word-level content was present.

Lexical and Syllable Transition Processing of Words Activate Networks at 2 Hz. Consistent with previous research, lexical processing activated left supramarginal gyrus, fronto-central and middle temporal brain areas, which have been argued to support lexical-semantic processing of words (Fig 2A, 3A) (46, 54, 32, 47, 51). 'Mere' syllable transition processing, in contrast, activated fronto-centro-temporal brain areas in both hemispheres (Fig 2B, 3B). Previously, a functional subdivision of the temporal cortex has been proposed, with bilateral STS activations during lower level acoustic speech processing and a left-lateralized activation of the more ventral temporal-parietal cortex during lexical-semantic processing (61, 44). In line with this subdivision, our findings further suggest that, beyond acoustic processing, sublexical syllable-transition processing occurs bilaterally. The functional connectivity increases between left middle temporal brain areas and auditory cortex during lexical processing of words, might indicate interactions of higher-level word processing in MTG with lower-level phonological or acoustic processing of words, in line with the view of the pMTG as hub for the mapping between words and meaning that links acoustic phonetic representations in the bilateral pSTG to semantic representations (62, 63, 54, 47). Brain areas involved in articulatory processing and motor predictions, such as the IFG, have been shown to be particularly activated during sublexical processing, but might also indicate articulatory influences during lexical processing (47). Thus, the bilateral interactions of STG, MTG and IFG when additionally syllable transition information was present, indicate an increased recruitment of articulation-related brain areas, possibly indicating motor predictions (45, 54).

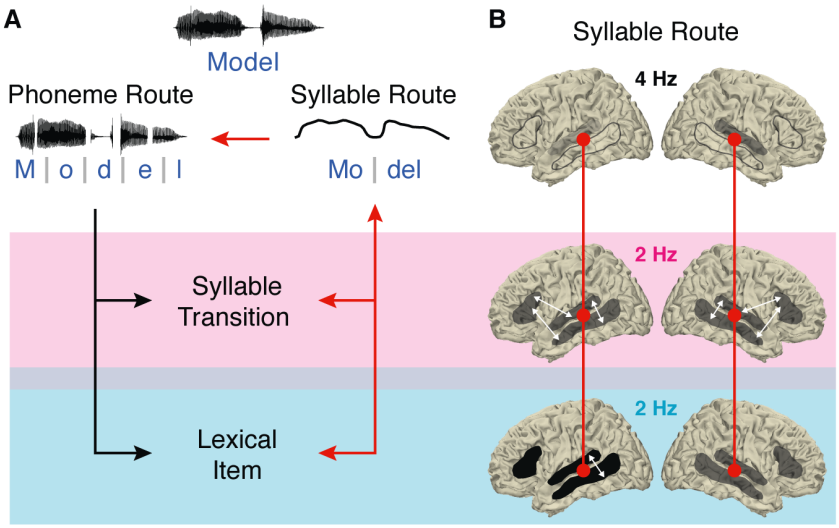

Figure 5. Schematic of the Interaction of Syllable and Word-Level Processing. (A) Extending classic 'Phoneme Route' models, we propose contributions of a 'Syllabic Route' to word-level processing. (B) (first row) Low-frequency tracking of the syllabic rate in auditory cortex including STG (4 Hz activity) affects (second row) higher level syllable transition processing in a network including bilateral STG, MTG and IFG (2 Hz activity) and (third row) lexical processing in a network including STG and left MTG and IFG and interactions between STG and left middle temporal areas (2 Hz activity). The interactions between syllable and word-level processing occur in STG and MTG (red line). (B, C, D) The intensity of the gray shades indicates the activation level during the respective process.

Word-Level Networks at 2 Hz and Syllable Networks at 4 Hz Interact. Previous speech comprehension models have focused on mapping of acoustic-phonemic to lexical processing [e.g. (58)]. Neurophysiological data, however, provide compelling evidence for the extraction of information at the syllable level (64, 8, 3). What does that mean for our understanding of speech comprehension? We provide evidence for interactions between word-level and syllable level processing (2 Hz - 4 Hz phase-amplitude coupling, Fig. S1, and decreased cerebro-acoustic coherence when word information is present), suggesting that syllable information is not only extracted to inform syllable transitions but exploited for lexical-level processing. Thereby, word- and syllable-level processing resulted in increased interactions in the MTG and the STG during lexical and lexical-plus-syllable transition processing, while no such interactions occurred when only syllable content was present (Fig. S1). The findings suggest, in line with proposals of a crucial role of the MTG as an interface between phonetic and semantic representations (47), that additionally to the STG, the MTG is involved in communicating information between syllable and word-level processing. It is likely that our findings indicate both feedforward communication from STG to higher-level processing areas and feedback from the word-level (2 Hz) to the syllable-level (4 Hz) (cf. SI Appendix Granger Causality), for example the first syllable might provide (temporal

and/or semantic) predictions of the second syllable. Note that the interaction effect we observed cannot be explained by mere differences in the ERPs elicited by the first and second syllables (Fig. S 4). Interactions between lexical and phonological processing have been shown to involve feedback from pMTG to pSTG (63), for review (66). Furthermore, several electrophysiological studies suggest interactions/feedback from sentential (67) or phrasal processing (31), or possibly both (29) to syllable processing. However, research that is particularly designed to investigate the interactions at the word level is rare (34, 67, 31). The frequency-tagging paradigm that we used might be particularly sensitive to revealing interaction effects by generating a higher consistency in the frequency of word processing. Interestingly, in contrast to lexical content, different brain areas were involved in the word-syllable level interactions when lexical-plus-syllable transition content was present. A possibility is that during the lexical processing of words, the association between syllables and words are processed at different hierarchical levels (in the STG vs. MTG), compared to the processing of syllable transitions. Future research, however, would need to test this claim. Additionally, our findings of decreased speech tracking at 4 Hz in STG (i.e. cerebro-acoustic coherence) when word-level content is present indicate an interaction between the word-level and syllable-level processing and support the notion of feedback communication. While several studies found increased speech tracking during the processing of intelligible compared to unintelligible speech (27, 12, 29), a previous artificial word learning study revealed reduced power elicited by syllable processing when words were intelligible (57, 68), suggesting that the effects of the interaction can vary depending on the paradigm, performed process etc. This, too, will need to be assessed in further experiments.

Processing of Syllable Transition Cues. Behavioral research suggests that sequencing of phonemes - because the distribution of phonemes varies across syllables - can be used to detect syllable transitions and word boundaries (69), as well as the position of syllables within words (60, 70). We showed the neuronal networks involved in using syllable transition information to process disyllabic words (Fig. 2 B, 3 B) and their interaction with syllable-level processing (Fig. S1). Interestingly, our findings suggest that even the syllable transition information present in a foreign language (Table 1), that is, cues such as the onset of a syllable or the consonant-vowel pattern, can be extracted and used to group syllables into words. In the present study, the stimuli were recorded and preprocessed so that acoustical cues at the word level were minimized, resulting in a prominent power peak only at the syllable rate (Fig. 1 E). Thus, the increased power peak at 2 Hz (in the

Turkish compared to the Non-Turkish) condition reflects the processing of syllable-transition features rather than the processing of acoustic cues. Statistical learning of the contingencies throughout the experiment might have additionally contributed to our findings, as the power peak at 2 Hz showed a tendency to vary across consecutive blocks of the experiment (SI Appendix; Fig. S2). The rapid statistical learning of such cues would be in line with previous studies (71–74). Our findings raise new questions. In the current study, we carefully matched the German and Turkish stimulus material with regard to sublexical cues that can be used to group syllables into words. Possibly this enhanced the ability of participants to quickly extract (and possibly learn) the sublexical contingencies of a foreign language. If the ability to extract sublexical contingencies at the word level depends on the similarity of these features between languages, the frequency-tagging paradigm could be used as a neurophysiological tool to investigate the phonological similarity between languages, without requiring explicit feedback from participants.

Conclusions. Our study provides novel evidence for the interaction of syllable and word-level functional networks. The increased feedback between word- and syllable-level processing, when lexical or syllable transition cues are present, suggests that the tracking of speech at the syllabic and word-level processing might be more interactive than previously thought. Furthermore, our findings open up the exciting possibility of using the frequency-tagging paradigm as neurophysiological indicator of phonological similarity.

Methods

Participants. German native speakers with no previous knowledge of Turkish participated in the two MEG experiments. In the first experiment the data of 18 (mean age: 24.32 years; SD: 3.30 years; female: 10) healthy right handed (Oldfield mean score: 91.38, SD: 15.44) participants are included in the analysis (n=9, right hand "yes" button). In the second experiment the data of 19 (mean age: 24.46 years; SD: 3.73 years; female: 10) healthy right handed (Oldfield mean score: 92.74, SD: 14.66) participants are included in the analysis (n = 10, right hand "yes" button). Several participants were excluded: because of outlier behavioral performance (accuracy < mean – 2 *SD; Exp. 1: n = 2; Exp. 2: n = 1) and because of technical issues (no triggers, audio problems; Exp. 1: n=2; Exp. 2: n = 1). Individual T_1-weighted MRI scans were conducted for all participants, except for some participants who either did not match the MRI criteria or did not show up to the MRI scan session (Exp. 1: n = 5; Exp. 2: n = 3). The study was approved by the local ethics

committee of the University Hospital Frankfurt. All participants gave written informed consent for participating in the study and received monetary compensation.

Paradigm and Stimuli

Paradigm. Participants were asked to listen to sequences of disyllabic words, that contained either German, Turkish, or pseudo-words. In Exp. 1, sequences of German words (Fig. 1 B) and Turkish words (Fig. 1 C) were used. In Exp. 2, sequence of German words (Fig. 1 B) and Turkish pseudo-words (Fig. 1 D) were used. Each sequence contained 38 syllables that formed 19 'disyllabic units' (word or pseudo-words). By presenting isochronous syllables, the presentation rate for syllables was fixed at 4 Hz, resulting in an occurrence of disyllabic units at 2 Hz. In order to maintain participants' attention on the auditory stimuli, a target stimulus, which consisted of a disyllabic unit that was made of a repetition of the same syllable, was inserted in 28.57% of the trials. Each block contained 29% trials with a target stimulus (equally distributed across conditions). Participants were asked to indicate with a button press whether a target stimulus was present after each trial (Fig. 1 A; index finger left and right hand; the response hand was counterbalanced across participants). Each trial was followed by a jittered intertrial interval (2 - 3.5 sec).

Stimulus Selection. German disyllabic words were selected from the CELEX lexical database (75). In order to maximize the speed of lexical access of the word as well as the predictability of the second syllable within a word, we selected words with a high frequency (CELEX spoken word frequency: MannSMln ≥ 10; Celex zeros replaced by Leiziger Wortschatz Corpus; LWC ≥ 4000) and with a high transition probability between the two syllables (STP ≥ 0.3 %). The STP was calculated for all disyllabic lemmas in the corpus, by dividing the wordform frequency of each lemma by the sum of the wordform frequencies of each wordform that contained the first syllable of the token (76). Laplace transformation of zero frequencies was used (77). Turkish disyllabic words were selected from the TELL database (http://linguistics.berkeley .edu/TELL/; e.g. (78)) and manually checked by a Turkish native speaker (for wordness). In total, 134 German words and 134 Turkish words were selected (noun, verb, adjectives). German and Turkish syllables were maximally matched (Table 1) with respect to the overall distribution of the manner of articulation for the onset phoneme of each syllable and the percentage of syllabic consonant-vowel CV structure.

Stimulus Processing. Syllable stimuli produced by a female German/Turkish bilingual speaker were recorded using an external audio card (Brand) (44100 Hz sampling rate). We recorded individual syllables in isolation (randomly presented to the speaker). The mean duration for German syllables was 358 ms (sd: 72 ms), for Turkish 334 ms (sd: 58 ms). Using this method, we eliminated any co-articulation and prosodic modulation between the two syllables within each word, such that a first syllable contains no acoustic cues that would allow the prediction of the second syllable (Fig. 1 E). Consequently, the prediction of the second syllable relies on higher level linguistic processing (e.g. lexical access of the word and syllable transition information).

Using Praat Vocal Toolkit (http://www.praatvocaltoolkit.com/; Boersma, 2002), the selected syllables tokens were high-pass filtered at 60 Hz, compressed in duration (250 ms; Note that syllables starting with a plosive consonant were compressed to 230 ms and a 20 ms silence period was added in the beginning to simulate the oral occlusion before the burst of plosive consonants), and normalized in peak-amplitude and pitch contour (at 250 Hz). Word stimuli were created by concatenating the two syllables of each word. For the Non-Turkish condition, disyllabic pseudo-word stimuli were created by concatenating two syllables that were quasi-randomly selected from all Turkish syllable stimuli (equal probability of first/second syllable position). Each sequence was created by concatenating randomly selected disyllabic stimuli. For the German conditions, since the word list contained several grammatical classes, we specifically checked the word order of each sequence to eliminate all possible formation of linguistic phrases as well as compound words by consecutive words. For Turkish and Non-Turkish conditions, sequences were checked to avoid 'German-like' homophones. Overall, three different sets of sequences were created for each condition.

Procedure. Participants were seated in the MEG testing booth in front of a board for instructions. Stimuli were presented binaurally via insert ear-plugs (E-A-RTONE Gold 3A Insert Earphones, Ulrich Keller Medizin-Technik, Weinheim, Germany). Participants' responses were collected with two button boxes (Current Designs, Inc.). Prior to the experiment, the MEG was recorded while participants listened to a sequence of 'auditory cortex localizer sounds' and held their gaze at a fixation cross (pure tones : 0.4 sec tone duration ; 250 Hz and 1000 Hz, 100 repetitions, jittered intertrial interval 0-5-1.5 sec). In the experiment, during each trial participants held their gaze at a fixation cross while listening to the auditory sequences. The experiment was run using Psychophysics Toolbox (79).

Both experiments contained 210 trials (105 per condition). Trials were grouped into 15 blocks. In total, each German, and Turkish disyllabic unit was repeated 15 times. The overall duration for the experiment was 150 minutes, including 90 minutes of recording time and 60 minutes of preparation time, breaks, and post-recording questionnaires.

MRI and MEG Data Acquisition

The MRI scanning was acquired on a 3 Tesla scanner (Siemens Magnetom Trio, Siemens, Erlangen, Germany). Vitamin-E capsules were used to mark anatomical landmarks (nasion, left and right pre-auricular points). For the MEG recordings, a 269-channel whole-head system (Omega 2000, CTF Systems Inc.) situated in a magnetically shielded room was used. Data recording was performed with a sampling rate of 1200 Hz, an online low pass filtered (cut-off: 300 Hz) and online denoising (higher-order gradiometer balancing). The head position relative to the MEG sensors was continuously tracked and head displacement was corrected in the breaks using the fieldtrip toolbox (http://fieldtrip.fcdonders.nl) (80).

Analysis

Acoustic Analysis. The speech envelope was computed separately for each sentence. The acoustic waveforms were filtered in 8 frequency bands that are equidistant on the cochlear map (between 100 and 8,000 Hz; third-order Butterworth filter; forward and reverse) (81). The speech envelop was computed by averaging the magnitude of the Hilbert transformed signal of the 8 frequency bands separately for each sentence. The envelope was resampled to 500 Hz to match the MEG data sampling rate. Next the data was averaged across all trials of a sequence, separately per condition. The mean spectral power across all sequences of an experiment, as well as the standard error across sequences, is plotted separately for the conditions.

MRI and MEG Data Analysis

For MRI and MEG data analyses, we used the FieldTrip toolbox (http://fieldtrip.fcdonders.nl) (82).

MRI Data Analysis. From the individual MRIs of all participants probabilistic tissue maps (including cerebrospinal fluid white and gray matter) were retrieved. In case an individual MRI was missing, the standard Montreal Neurological Institute (MNI) template brain was used. In a next step, the physical relation between sensors and sources was obtained using a single shell

volume conduction model (83). The linear warp transformation was computed between the individual T1 MRI and the MNI template T1. The inverse of that transformation was computed, that is, a template 8 mm grid defined on the MNI template T1 was inversely transformed so that it was warped on the individual head space, based on the individual MRI and the location of the coils during the MEG recording. A leadfield (forward model) was calculated based on the warped MNI grid and the probabilistic tissue map, and used for source reconstruction. This allowed computing statistics across subjects in the MNI space with the grids of all subjects being aligned to each other.

MEG Data Analysis – Preprocessing. For preprocessing, the data were band-pass filtered off-line (1–160 Hz, Butterworth filter; filter order 4) and line-noise was removed using bandstop filters (49.5-50.5, 99.5-100.5, 149.5-150.5 Hz, two-pass; filter order 4). In a common semi-automatic artifact detection procedure (i.e. the output of the automatic detection was monitored), the signal was filtered in a frequency range that typically contains muscular artifacts (band-pass: 110-140 Hz) or jump artifacts (median filter) and z-normalized per time point and sensor. To accumulate evidence for artifacts that typically occur in more than one sensor, the z-scores were averaged over sensors. We excluded trials exceeding a predefined z-value (muscular artifacts, z = 15; jumps, z = 30). Slow artifacts were removed by rejecting trials were the range (min-max difference) in any channels exceeded a threshold (threshold = 0.75e-5). The data was down-sampled to 500 Hz. And epoched (-2.1 to 9.6 sec). Trials with head movements that exceeded a threshold (5 mm) were rejected. Afterwards, the different blocks of recorded MEG data were concatenated (Note that for each block, during the recording, the head position was adjusted to the initial position of the first block). Sensors with high variance were rejected.

Eye-blink, eye-movement and heartbeat-related artifacts were removed, using independent component analysis (infomax algorithm; (84)). Components were first reduced to 64 components using principal component analysis. Only in the case of a conclusive conjunction of component topography, time course and variance across trials components were rejected. For the sensor space analysis, spherical spline interpolation was used to interpolate the missing sensors (85).

Trials with correct responses were selected and the trial number was matched between the conditions by randomly selecting trials of the condition with less trials (trial number, Exp. 1: mean = 73.22, SD= 11.02; Exp. 2: mean = 68.68, SD=10.27).

MEG Data Analysis – Power. For the sensor space analysis, the data was interpolated towards a standard gradiometer location based on the headmodel. It was epoched using a time window of 0.5-9.5 sec (0-0.5 sec after stimulus onset were excluded to avoid onset-related contamination) and averaged across all trials of a condition. Evoked power was computed using singletaper frequency transformation (1-7 Hz) separately for each participant of the two experiments at each condition (frequency resolution: 0.1111 Hz). At each frequency the power was contrasted by the neighbouring frequency bins (+- 2-3 bins). Cluster-based permutation tests using Monte-Carlo estimation (86) were performed to analyze differences between the conditions within each experiment (German vs. Turkish/Non-Turkish; dependent-sample T-statistics) and across experiments (German vs. German and Turkish vs. Non-Turkish; independent-sample T-statistics) with an iteration of the condition affiliation (500 iterations). Adjacent sensor points (minimum number of neighbourhood sensors = 2) with t-values larger .05 were spatially clustered. Clusters were considered significant when the summed t-values with the clusters exceeded 95% of the permutation distribution ($p < .05$; two-sided).

Dynamic Imaging of Coherent Sources (DICS) was used to localize neuronal power (63). First, based on the individual leadfields a common source filter (1.3332-4.6662 Hz) was computed across conditions for each participant (lamda = 100%; 0.8 cm grid). Second, based on the filter and fourier transformed data (multi-taper frequency transformation; 0.1111 Hz resolution) the power at 2 Hz and 4 Hz was localized and contrasted with the neighbouring frequency bins (+- 2-3 bins). Differences in source power at 2 Hz and 4 Hz were tested using cluster-based permutation tests (1000 iterations; two-sided) to analyze differences between the conditions within each experiment (German vs. Turkish and German vs. Non-Turkish; dependent-sample T-statistics) and across experiments (German vs. German and Turkish vs. Non-Turkish; independent-sample T-statistics) with an iteration of the condition affiliation. Adjacent voxels with t-values larger .05 were spatially clustered. Clusters were considered significant when the summed t-values with the clusters exceeded 95% of the permutation distribution ($p < .05$).

The Brainetome atlas (87) was used to define ROIs (the left and right STG A41/42, MTG A21r, MTG A21c, IPL A40rv, IFG A40v, PCG A6cdl) to further test the condition differences at 2 Hz revealed in the cluster-test analysis . Differences between conditions at each ROI were tested separately for the hemispheres and the comparisons within each experiment (German vs. Turkish and German vs. Non-Turkish; Wilcoxon signed-rank tests) and across experiments (German vs.

German and Turkish vs. Non-Turkish; Mann-Whitney-Wilcoxon test's). Bonferroni correction across ROIs and hemispheres was applied to correct for inflated p-values. (Fig. 3)

MEG Data Analysis – Auditory Localizer. The M100 component elicited by the localizer sounds was analyzed in order to determine sensors that reflect individual auditory cortex activity, which were used as ROI for the sensor space connectivity analysis. After preprocessing, the epoched (0-0.5 sec) trials were baseline corrected (-0.4-0 sec). The M100 peak was determined for each participant, as the max. amplitude of the root mean squared data within a window of 0.07-0.160 sec. The 15 sensors with max. amplitude at this time point were selected separately for each hemisphere and participant. For 5 participants no localizer MEG recordings were present (Exp. 1, n = 4) and thus the M100 sensors were selected based on the grand average of the respective experiment.

MEG Data Analysis – Connectivity. For the group analysis, the data was interpolated towards a standard gradiometer location based on the headmodel. The Debiased Weighted Phase Lag Index (dwPLI) (88, 89) was computed at 2 Hz and 4 Hz based on the spectral complex coefficients (singletaper frequency transformation; 0.1111 Hz resolution) of each trial (values between 0-1). The connectivity between the individual M100 sensor ROI (Exp. 1 Fig. 5 A and Exp. 2 Fig. 5 B; Note that for this analysis the localizer data was interpolated prior to the M100 sensor selection) and all other sensors was estimated. Fischer z-transformation was applied to normalize the data prior to further analyzis (for displaying purposes the hyperbolic tangent was applied). The dwPLI values were averaged across all individual M100 sensors. The connectivity of the ROI with itself was set to zero. Differences in sensor space connectivity at 2 Hz and 4 Hz at the M100 ROI were tested using cluster-based permutation tests to analyze differences between the conditions within each experiment (German vs. Turkish and German vs. Non-Turkish; dependent-sample T-statistics) with an iteration of the condition affiliation (500 iterations). Adjacent sensors with t-values larger .05 were spatially clustered. Clusters were considered significant when the summed t-values with the clusters exceeded 95% of the permutation distribution ($p < .05$; one-sided).

For the source space connectivity analysis, the spectral complex coefficients were multiplied with a common filter (DICS; across 2 Hz and 4 Hz), computed across conditions separately for each trial. The dwPLI was computed between all voxel. Fischer z-transformation was applied to normalize the data prior to further analyzis. The connectivity between a STG ROI and all voxel

was computed by averaging across the voxel of the STG ROI (selected based on the *automated anatomical labeling* atlas (AAL; (90)'Temporal_Sup_L/R'; Fig. 4 E). The connectivity of the ROI with itself was set to zero. Source space dwPLI differences between the conditions within each experiment (German vs. Turkish and German vs. Non-Turkish; dependent-sample T-statistics) were tested using cluster-based permutation tests (1000 iterations ; see above).

For a more specific analysis of our hypothesis, based on the AAL ROIs in the STG ('Temporal_Sup_L/H'), MTG ('Temporal_Mid_L/H') and IFG ('Frontal_Inf_Tri_L/H') were selected (Fig. 4 E). The connectivity between pairs of ROIs was computed by averaging across the voxels of a ROI and the corresponding ROI pair. Wilcoxon signed-rank and Mann-Whitney-Wilcoxon tests were used to test connectivity differences between conditions separately at each ROI pair. Bonferroni correction was applied to correct for inflated p-values due to multiple-comparisons.

Cross-Frequency Coupling. In order to analyze interactions between the 2 Hz word-level and 4 Hz syllable level network, phase-amplitude coupling between the 2 Hz phase and 4 Hz power was computed. After trials were split into 2 seconds long time windows (to increase the amount of data points), the Fourier Transform was used to compute the complex spectral coefficients at 2 Hz and at 4 Hz (0.5 Hz resolution) separately for each trial, condition and participant. A common filter (DICS; lamda = 100%; 0.8 cm grid; 0.5 Hz resolution) was computed at 2 Hz and at 4 Hz and used to project each trial in source space. According to its phase at 2 Hz, the power value at 4 Hz of each trial was sorted into four phase-bins (- pi to pi) separately for each ROI (STG, MTG, IFG), condition and participant. Power was averaged, first across trials, second across the voxels of each ROI, separately for the phase-bins, conditions and subjects. Prior to the statistics each phase-bin was normalized by contrasting it with the average of the remaining phase bins. Nonparametric Wilcoxon Signed-Rank and Mann-Whitney-Wilcoxon tests were used to test differences in power distribution across phase-bins separately for each condition (the data from the German conditions of the two experiments were merged) and for the condition difference in each experiment. Bonferroni correction was applied to test for multiple comparisons of phase-bins and ROIs.

Cerebro-Acoustic Coherence: After the spectral complex coefficients at 4 Hz was computed for the speech envelope of each trial (cf. Acoustic Analysis) and the neuronal data (0.1111 Hz resolution), coherence between all sensors and the speech envelope was computed. A common

filter (DICS; lamda = 100%; 0.8 cm grid) was multiplied with the coherence, and Fischer z-transformation was applied. The cerebro-acoustic coherence was averaged across voxels of the STG ROI (using the AAL atlas). Wilcoxon signed-rank tests were used to test condition differences.

Acknowledgements

This work was funded by the Max-Planck-Institute for Empirical Aesthetics. We thank Marius Schneider for help with the data recording, Ilkay Isik for checking the Turkish stimulus material, Dr. Florencia Assaneo for discussions and Felix Bernoully for graphics support.

Table 1

Syllable Transitions

German				
Measurment	Within word	Between word	Difference	p-value
Syllable identity	0.91	0.02	0.89	< .0001
Syllable CV	0.51	0.29	0.23	< .01
Onset	0.36	0.05	0.31	< .0001
Onset manner	0.44	0.22	0.22	> .1
Rhyme	0.51	0.05	0.46	< .0001

Turkish				
Measurment	Within word	Between word	Difference	p-value
Syllable identity	0.75	0.02	0.73	< .0001
Syllable CV	0.57	0.34	0.23	< .05
Onset	0.24	0.09	0.15	< .0001
Onset manner	0.26	0.34	-0.07	> .1
Rhyme	0.47	0.05	0.42	< .0001

For each phonological measurement average transitional probabilities between consecutive syllables within word boundary and between word boundary in German and Turkish sequences are displayed. Mann-Whitney-Wilcoxon tests revealed significantly higher within word transitional probabilities for all the measurements except for onset manner (where a trend was observed). The presence of phonological patterning at the word (2 Hz) rate in the Turkish sequences, allows for temporal grouping of Turkish syllable pairs by German listeners in the absence of lexical processes.

References

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

2. Ghitza O, Greenberg S (2009) On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66(suppl 1–2)(1–2):113–126.

3. Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. *Trends Neurosci* 33(3):111–120.

4. Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front Psychol* 2:1–13.

5. Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15(4):511–517.

6. Zion Golumbic EM, Poeppel D, Schroeder CE (2012) Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain Lang* 122(3):151–161.

7. Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as [`]asymmetric sampling in time'. *Nat Speech Percept* 41(1):245–255.

8. Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54(6):1001–1010.

9. Gross J, et al. (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol* 11(12):e1001752.

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

11. 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, Part 2(0):761–768.

12. Rimmele J, Zion Golumbic E, Schröger E, Poeppel D (2015) The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* 68:144–154.

13. Rimmele JM, Morillon B, Poeppel D, Arnal LH (2018) Proactive sensing of periodic and aperiodic auditory patterns. *Trends Cogn Sci* 22(10):870–882.

14. Ghitza O (2014) Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. *Front Psychol* 5. doi:10.3389/fpsyg.2014.00652.

15. Ghitza O (2016) Acoustic-driven delta rhythms as prosodic markers. *Lang Cogn Neurosci*:1–17.

16. Teng X, Tian X, Rowland J, Poeppel D (2017) Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLOS Biol* 15(11):e2000812.

17. Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci* 19(1):158–164.

18. Martin AE, Doumas LAA (2017) A mechanism for the cortical computation of hierarchical linguistic structure. *PLOS Biol* 15(3):e2000663.

19. Molinaro N, Lizarazu M (2018) Delta(but not theta)-band cortical entrainment involves speech-specific processing. *Eur J Neurosci* 48(7):2642–2650.

20. Meyer L (2017) The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *Eur J Neurosci*. doi:10.1111/ejn.13748.

21. Rimmele J, Gross J, Molholm S, Keitel A (2018) Editorial: Brain oscillations in human communication. *Front Hum Neurosci* 12:39.

22. Altmann GT., Kamide Y (1999) Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73(3):247–264.

23. Jadoul Y, Ravignani A, Thompson B, Filippi P, de Boer B (2016) Seeking temporal predictability in speech: Comparing statistical approaches on 18 world languages. *Front Hum Neurosci* 10:586.

24. Kotz SA, Schmidt-Kassow M (2015) Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Spec Issue Predict Speech Lang Process* 68:48–60.

25. Marslen-Wilson W, Tyler LK (1980) The temporal structure of spoken language understanding. *Cognition* 8(1):1–71.

26. Scontras G, Badecker W, Shank L, Lim E, Fedorenko E (2015) Syntactic complexity effects in sentence production. *Cogn Sci* 39(3):559–583.

27. Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex* 23(6):1378–1387.

28. Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D (2013) Visual input enhances selective speech envelope tracking in auditory cortex at a "Cocktail Party." *J Neurosci* 33(4):1417–1426.

29. 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(12):1649–1653.

30. Park H, Thut G, Gross J (2018) Predictive entrainment of natural speech through two fronto-motor top-down channels. *Lang Cogn Neurosci*:1–13.

31. Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biol* 16(3):e2004473.

32. Peelle JE (2012) The hemispheric lateralization of speech processing depends on what "speech" is: A hierarchical perspective. *Front Hum Neurosci* 6. doi:10.3389/fnhum.2012.00309.

33. Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front Psychol* 3:320.

34. Mai G, Minett JW, Wang WS-Y (2016) Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage* 133:516–528.

35. Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr Biol* 25(19):2457–2465.

36. Assaneo MF, Poeppel D (2018) The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Sci Adv* 4(2):eaao3842.

37. Engel AK, König P, Kreiter AK, Schillen TB, Singer W (1992) Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends Neurosci* 15(6):218–226.

38. Varela F, Lachaux J-P, Rodriguez E, Martinerie J (2001) The brainweb: Phase synchronization and large-scale integration. *Nat Rev Neurosci* 2(4):229–239.

39. Passingham RE, Stephan KE, Kotter R (2002) The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci* 3(8):606–616.

40. Friederici AD, Singer W (2015) Grounding language processing on basic neurophysiological principles. *Trends Cogn Sci* 19(6):329–338.

41. Fries P (2015) Rhythms for cognition: Communication through coherence. *Neuron* 88(1):220–235.

42. Michalareas G, et al. (2016) Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89(2):384–397.

43. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532(7600):453–458.

44. 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(12):2767–2796.

45. Hickok, Poeppel (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4(4):131–138.

46. Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *New Funct Anat Lang* 92(1–2):67–99.

47. Gow DWJ (2012) The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain Lang* 121(3):273–288.

48. Rodd JM, Vitello S, Woollams AM, Adank P (2015) Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain Lang* 141:89–102.

49. Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain J Neurol* 123 Pt 12:2400–2406.

50. Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJS (2006) Converging language streams in the human temporal lobe. *J Neurosci Off J Soc Neurosci* 26(28):7328–7336.

51. Rice GE, Caswell H, Moore P, Lambon Ralph MA, Hoffman P (2018) Revealing the dynamic modulations that underpin a resilient neural network for semantic cognition: An fMRI Investigation in patients with anterior temporal lobe resection. *Cereb Cortex* 28(8):3004–3016.

52. Okada K, Hickok G (2006) Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *Neuroreport* 17(12):1293–1296.

53. Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science*. doi:10.1126/science.1245994.

54. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8(5):393–402.

55. Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc Natl Acad Sci* 94(26):14792–14797.

56. Kan IP, Kable JW, Van Scoyoc A, Chatterjee A, Thompson-Schill SL (2006) Fractionating the left frontal response to tools: dissociable effects of motor experience and lexical competition. *J Cogn Neurosci* 18(2):267–277.

57. Buiatti M, Peña M, Dehaene-Lambertz G (2009) Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage* 44(2):509–519.

58. Marslen-Wilson WD, Welsh A (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognit Psychol* 10(1):29–63.

59. Mehler J, Dommergues JY, Frauenfelder U, Segui J (1981) The syllable's role in speech segmentation. *J Verbal Learn Verbal Behav* 20(3):298–305.

60. Cutler A (2012) *Native listening: Language experience and the recognition of spoken words* (MIT Press, Cambridge, MA) Available at: https://www.jstor.org/stable/j.ctt5hhjc1.

61. Binder JR, et al. (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10(5):512–528.

62. Kertesz A, Sheppard A, MacKenzie R (1982) Localization in transcortical sensory aphasia. *Arch Neurol* 39(8):475–478.

63. Gow DW, Segawa JA (2009) Articulatory mediation of speech perception: A causal analysis of multi-modal imaging data. *Cognition* 110(2):222–236.

64. Gross J, et al. (2001) Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proc Natl Acad Sci U S A* 98(2):694–699.

65. Helfrich RF, Herrmann CS, Engel AK, Schneider TR (2016) Different coupling modes mediate cortical cross-frequency interactions. *Transcranial Electr Stimul TES Neuroimaging* 140:76–82.

66. Gow DW, Segawa JA, Ahlfors SP, Lin F-H (2008) Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage* 43(3):614–623.

67. Gow DW, Olson BB (2016) Sentential influences on acoustic-phonetic processing: a Granger causality analysis of multimodal imaging data. *Lang Cogn Neurosci* 31(7):841–855.

68. Sohoglu E, Davis MH (2016) Perceptual learning of degraded speech by minimizing prediction error. *Proc Natl Acad Sci* 113(12):E1747–E1756.

69. McQueen JM (1998) Segmentation of continuous speech using phonotactics. *J Mem Lang* 39(1):21–46.

70. van der Lugt AH (2001) The use of sequential probabilities in the segmentation of speech. *Percept Psychophys* 63(5):811–823.

71. Saffran JR, Newport EL, Aslin RN (1996) Word segmentation: The role of distributional cues. *J Mem Lang* 35(4):606–621.

72. Pena M, Melloni L (2012) Brain oscillations during spoken sentence processing. *J Cogn Neurosci* 24(5):1149–1164.

73. Ota M, Skarabela B (2016) Reduplicated words are easier to learn. *Lang Learn Dev* 12(4):380–397.

74. Ota M, Skarabela B (2018) Reduplication facilitates early word segmentation. *J Child Lang* 45(1):204–218.

75. Baayen R, Piepenbrock R, Gulikers L (1995) *CELEX2 LDC96L14. Web Download.* (Philadelphia: Linguistic Data Consortium).

76. Lewis G, Solomyak O, Marantz A (2011) The neural basis of obligatory decomposition of suffixed words. *Brain Lang* 118(3):118–127.

77. Brysbaert M, Diependaele K (2012) Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice. *Behav Res Methods*:422–430.

78. Scharinger M, Idsardi WJ, Poe S (2011) A comprehensive three-dimensional cortical map of vowel space. *J Cogn Neurosci* 23(12):3972–3982.

79. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.

80. Stolk A, Todorovic A, Schoffelen J-M, Oostenveld R (2013) Online and offline tools for head movement compensation in MEG. *NeuroImage* 68:39–48.

81. Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416(6876):87–90.

82. 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.

83. 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(22):3637.

84. Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1996) Independent component analysis of electroencephalographic data. *Adv Neural Inf Process Syst* 8:145–151.

85. Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol* 72(2):184–187.

86. Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164(1):177–190.

87. Fan L, et al. (2016) The Human Brainnetome Atlas: A new brain atlas based on connectional architecture. *Cereb Cortex* 26(8):3508–3526.

88. Stam CJ, Nolte G, Daffertshofer A (2007) Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp* 28(11):1178–1193.

89. Vinck M, Oostenveld R, van Wingerden M, Battaglia F, Pennartz CMA (2011) An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage* 55(4):1548–1565.

90. Tzourio-Mazoyer N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15(1):273–289.