The Role of Entrained Oscillations in Segmenting Rhythmic Sentences during Foreign Language Listening

Sophie Arana

Supervisors:
Anne Koesem[1,2], Tineke M. Snijders (Educational supervisor)[1,2,3]

Affiliation(s):
1: Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands
2: Donders Centre for Cognitive Neuroimaging, Nijmegen, Netherlands
2: Center for Language Studies, Nijmegen, Netherlands

Corresponding author:
Arana Sophie, sophiearana@hotmail.com

Abstract: When hearing a foreign language, listeners often have difficulties segmenting the continuous speech signal into individual words The mechanisms by which word segmentation occurs are largely unknown. Recently, neural oscillatory entrainment to the speech envelope has been proposed as a possible mechanism underlying speech processing. In the current study, we investigated whether the rhythm structure of foreign speech provides cues for word segmentation and whether neural oscillatory entrainment is involved during parsing of the signal. English natives listened to highly rhythmic and highly non-rhythmic Dutch stories. As an index of neuronal entrainment, we computed coherence between the neural activity measured with magnetoencephalography (MEG) and the speech envelope. In order to quantify word segmentation ability, participants performed a forced-choice lexical decision task after each story. They were prompted to recognize a word that had been repeated within each story and distinguish it from a novel word. Further, as a concurrent measure of successful word segmentation, we analyzed event-related fields (ERFs) in response to familiarized versus novel words. High response accuracies as well as a suppression of the ERFs in response to familiarized versus novel words indicate that participants were able to segment words in the foreign speech stream. Moreover, we observed a significant increase in speech-brain coherence for rhythmic versus non-rhythmic speech input. Contrary to our predictions, we found no effect of rhythmicity on word segmentation. Thus, our results suggest that word segmentation of foreign speech does not solely rely on rhythmic regularities in the signal. Further research is required to elucidate the underlying mechanisms of processing both rhythmic and non-rhythmic speech and to integrate the role of neuronal entrainment.

Keywords: Neural entrainment, Oscillations, Foreign language listening, Word segmentation

Introduction

When learning a foreign language, one of many challenges is the transition from classroom to real world. Language learners can have the vocabulary of a foreign language well memorized but still struggle to recognize individual words out “in the wild”. For instance, a foreign listener’s recognition of a word in a sentence context is delayed or even inhibited compared to that of native listeners (Snijders, Kooijman, Cutler, & Hagoort, 2007). This difficulty is caused by the continuous nature of the speech stream. In order to derive sentence structure or even meaning, we have to segment the incoming stream into smaller units, such as words. This can be quite challenging due to the lack of robust physical word boundaries. Therefore, it is astonishing that infants start to tackle the segmentation problem in their native language already at 7.5 month by using cues such as stress and statistical distributions (Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999; Saksida, Langus, & Nespor, 2016).

Importantly, the comprehension of a foreign language later in life often relies on cues distinct from those used in our mother tongue’s speech recognition. Information about sentence context and word frequencies that can help natives predict upcoming word candidates (Connine, Mullennix, Shernoff, & Yelen, 1990; Norris & McQueen, 2008) is not yet available to low proficient learners. Therefore, adult foreign language learners find themselves in a similar situation as infants. They have to resort to information that becomes available after short exposure. For instance, many studies have reported that statistical distributions of syllables and phonemes can inform naïve language learning (Buiatti, Peña, & Dehaene-Lambertz, 2009; Saffran, Newport, & Aslin, 1996). For these statistical distributions, the probabilities of adjacent syllables within words differ from those between words. Similarly, phonemes do not randomly concatenate into words but their order follows certain statistical rules. Consequently, these statistical cues, which are learnable after short exposure, can be used in foreign speech recognition.

In addition to statistical cues, acoustic properties of the signal, including co-articulation, subliminal pauses and prosodic cues, can provide relevant information for foreign speech segmentation (for an overview see Mattys, White, & Melhorn, 2005). Generally, acoustic speech cues are exploited quickly (Steinhauer, Alter, &

Friederici, 1999) and can even be preferred over statistical cues, when signal quality is low (Fernandes, Ventura, & Kolinsky, 2007). With subliminal pauses introduced into the word stream, foreign listeners seem to better segment the speech input; i.e. their accuracy in familiarity judgments increases (Peña, Bonatti, Nespor, & Mehler, 2002) or their ERPs in response to familiarized words become more native-like (Snijders et al., 2007). Buiatti et al. (2009) also found an effect of subliminal pauses, using an artificial language learning design. With subliminal pauses added to the artificial word stream after every three syllables, the number of reported tri-syllabic words increased and N400-like potentials were recorded.

Finally, a much investigated acoustic cue for speech segmentation is prosody (Cutler, Dahan, & van Donselaar, 1997). The rhythmic stress pattern of a language, for example carried by the syllable rhythm, has been shown to be a necessary and sufficient cue in language discrimination tasks (Ramus & Mehler, 1999). For segmentation, acoustic speech cues weigh even more than statistical cues in the case of infants (Johnson & Jusczyk, 2001). Similarly, adults seem to segment more efficiently when stress information is available (McNealy, Mazziotta, & Dapretto, 2006) and stress cues can outweigh co-articulatory cues in impoverished listening conditions (Mattys, 2004). Crucially, the underlying brain mechanisms for processing of rhythmic segmentation cues, such as stress, are still unknown.

A common rhythm in neural oscillations and speech

Recently, neuronal oscillations have been proposed as neural correlates underlying the parsing of speech (Giraud & Poeppel, 2012; Peelle & Davis, 2012). This proposal is based on neuroimaging studies using magnetoencephalography (MEG) or electroencephalography (EEG). In these studies, strong neural oscillatory activity is observed when participants listened to speech. These oscillations follow the dynamics of the speech envelope at frequencies matching the timescales of basic linguistic units. Most prominent is the tracking in the theta band that corresponds to the syllable rhythm (4-8 Hz). Increased activity has also been observed at higher frequencies correlating with attributes at the phonemic scale (30-40 Hz) and low frequencies correlating with phrasal units, such as noun phrases like “the dog” or “ a small house”, at around 1-2 Hz (Ahissar et al., 2001; Buiatti et al., 2009; A.-L. Giraud & Poeppel, 2012). Due to the strong correspondence between oscillatory activity during listening and the temporal fluctuations in the speech signal, neural oscillations are said to “entrain” to the speech signal.

Neural entrainment has been hypothesized to be actively involved in speech parsing mechanisms. One account of a mechanistic role proposes that synchronization of signal and oscillation serves to align phases of high neuronal sensitivity with the most informative acoustic structures of the input (Schroeder & Lakatos, 2009). Ongoing neuronal oscillations directly interact with neural firing rates leading to alternating fluctuations of neuronal excitation and inhibition (Haegens, Nácher, Luna, Romo, & Jensen, 2011; C. Kayser, Wilson, Safaai, Sakata, & Panzeri, 2015; Lakatos et al., 2005, 2012). It has been suggested that during phases of neuronal excitation the stimulus is optimally encoded, leading to an enhancement of the perceptual sensitivity to visual (Spaak, de Lange, & Jensen, 2014; for a review see VanRullen, Busch, Drewes, & Dubois, 2011) as well as auditory sensory input (Molly J Henry & Obleser, 2012; Stefanics et al., 2010). In contrast, phases of inhibition lead to a suppression of signal encoding. The alternating pattern of excitation and inhibition generate the rhythmic sinusoidal pattern inherent to oscillations and may lead to a discretized parsing of the sensory input (A.-L. Giraud & Poeppel, 2012). Speech is an ideal candidate signal for this parsing mechanism due to the inherent quasi-rhythmic temporal fluctuations in the speech envelope. When phases of excitation align to the most informative bursts in the speech signal, parsing should become most efficient. Crucially, within this account neuronal oscillations are thought to establish temporal predictions about the upcoming speech signal and discretize it accordingly into informative chunks (Peelle & Davis, 2012).

In line with the oscillatory theory of speech parsing, several studies have reported that when neuronal oscillations fail to entrain, the signal is no longer intelligible for the listener. Evidence for this link comes from studies with manipulated speech rate or envelope. Listeners are generally able to quickly adapt to different speakers’ speech rates. Nonetheless, when speech is compressed by more than half, comprehension and speech envelope tracking both suffer (Ahissar et al., 2001). A different way of manipulating intelligibility is by pairing speech with noise, creating a so-called speech-noise chimera. Depending on the number of synthesized frequency bands those stimuli can be rendered highly unintelligible. Again, comprehension decreased when theta oscillations were more dissimilar from the signal’s syllable rhythm in response to these speech-noise chimera (Luo & Poeppel, 2007). Concluding, there is evidence that neuronal entrainment to the speech signal positively correlates with successful processing.

Theta entrainment – A universal mechanism to encode non-linguistic cues?

Given the strong correlation between neuronal entrainment and speech processing, the question about the functional role of neuronal entrainment is currently under debate (Ding & Simon, 2014). Some studies have observed top-down influences on entrainment to speech (Mesgarani & Chang, 2012) and argue for a function specific to higher-order language processing. Alternatively, entrainment could be a universal response to the physical properties of acoustic stimuli. As a matter of fact, traces of theta oscillations entraining to the speech envelope have been also observed during foreign language processing. For Spanish native speakers, theta entrainment has been found in response to their native language, a closely related foreign language (Italian), as well as a distant foreign language (Japanese) (Peña & Melloni, 2012). Similarly, theta tracking is a robust finding in English natives processing both English and Mandarin (Ding, Melloni, Zhang, Tian, & Poeppel, 2015). Additionally, the average syllabic rate (4-8 Hz) at which theta waves oscillate applies across many languages (Ding et al., 2016; Greenberg, Carvey, Hitchcock, & Chang, 2003; Pellegrino, Coupé, & Marsico, 2016). Thus, theta entrainment is especially appropriate for universal entrainment to speech. The above findings point to an acoustic-driven mechanism of theta oscillations in foreign listeners that is independent from higher-level information such as sentence context or semantic prediction.

Importantly, if neural entrainment is truly following non-speech specific temporal regularities available to non-speakers, its function during speech processing should be driven by elemental temporal speech cues, such as prosody. As discussed above, the oscillatory parsing mechanisms could generate temporal predictions about the incoming signal and ultimately facilitate processing when sufficiently entrained to the signal. We therefore expect that a highly rhythmic, temporally predictable signal should lead to stronger entrainment and therefore facilitate speech processing.

In this study, we investigate the effects of rhythmicity on neural entrainment to the speech input as well as on successful word segmentation during foreign speech listening. We aimed at enhancing and inhibiting oscillatory entrainment in monolingual English native speakers by presenting them with highly rhythmic and non-rhythmic Dutch stimuli. Both Dutch and English have a similar metrical pattern with preferred word-initial stress (Cutler & Butterfield, 1992). Because both languages are considered stress-based languages, the foreign listeners could partly rely on their native segmentation strategies. At the same time, they should be particularly sensitive to the rhythmic stimuli as rhythmicity was created through the stress pattern. During the experiment, the English natives listened to Dutch short stories in two conditions (rhythmic and non-rhythmic). Crucially, one target word was repeated throughout the stories (familiarized word). To assess whether participants successfully segmented the familiarized word, we compared event-related fields (ERFs) in response to this word versus a non-familiarized matched control word (novel word). Thereby, we made use of the previously established ERP word repetition effect: a positive shift in the response to the repeated presentation of a word (here the familiarized word) compared to its first presentation (here the novel word) (Doyle, Rugg, & Wells, 1996; M. D. Rugg, Doyle, & Wells, 1995; M. Rugg, Furda, & Lorist, 1988; Snijders et al., 2007). In an EEG study, Snijders et al. (2007) used the word repetition effect to assess English natives' ability to segment foreign speech. In their study, English natives were exposed to repeated presentation of a Dutch target word. For repeated words, they found a similar ERP response in the English native listeners compared to Dutch natives, i.e. a positive shift in a 400 – 900 millisecond time window. In the current study, we additionally included a word decision task following the presentation of familiarized and novel words as an offline measure. This behavioral task mainly served to maintain participants' attention during the experiment. At the same time, the data could inform us about the relevance of successful entrainment for behavior. Finally, in order to establish the link between neuronal entrainment and word segmentation, we analyzed oscillatory activity during the familiarization phase and correlated this measure with the word repetition effect.

Based on the proposed models of neuronal entrainment and its role during speech processing, we made several predictions: (1) We expected neuronal entrainment to the speech envelope to be stronger in rhythmic versus non-rhythmic conditions due to the temporal predictability of a rhythmic signal. (2) We also expected to find a greater word repetition effect, as indexed by a larger amplitude difference of familiarized and novel word ERFs when familiarized with rhythmic compared to non-rhythmic sentences. (3) Further, we predicted stronger tracking of the speech signal to result in better segmentation of the familiarized word. Thus, the strength of entrainment at the relevant frequencies was predicted to positively correlate with the word repetition effect.

Summarizing, while our participants had extremely limited access to lexical, syntactic, or semantic information due to their low proficiency in Dutch, stimuli in the rhythmic condition provided strong acoustic cues that could provoke oscillatory tracking of the speech envelope. If neuronal entrainment indeed underlies low-level speech processing, we expected a rhythmic speech signal to enhance and a non-rhythmic speech signal to inhibit this mechanism.

Methods

Participants

Eighteen English native speakers participated in this study. Two participants had to be excluded; one due to chance performance at the word decision task and one due to excessive movement. This resulted in 16 participants (mean age = 22.63, range = 19-27 years, seven female), whose data were submitted to analysis. At the time of testing, participants had been exposed to Dutch for maximally one year. They had gained zero or very low proficiency in Dutch as determined through self-report of language experience and training (none of the participants had received formal language training). Language proficiency was further assessed with the lexTALE test (Lemhöfer & Broersma, 2012) in which participants scored at chance level (mean score = 53 +- 5.42 s.d.). All participants had normal hearing and were right-handed. They provided written informed consent before the experiment and were offered financial compensation. The study was approved by the Radboud University Institutional Review Board.

Stimuli

Stimuli consisted of 40 spoken stories, each composed of eight sentences. Each sentence consisted of 13 syllables arranged in a trochaic stress pattern. In all stories, one word (familiarized word) was embedded within every sentence, thereby earning high frequent word status. The position of the familiarized word in a sentence was varied in order to prevent predictions about timing, but was never placed at the beginning of a sentence. A female speaker, who was not involved in the compilation of the stimulus material, recorded the stories in a sound-attenuating booth. The speaker recorded the sentences while listening to a 120 beats per minute (bpm) metronome and was instructed to pronounce every syllable on the beat to ensure a rhythmic speech pattern. At the same time, the speaker was encouraged to vary her intonation across sentences, so as to preserve a natural variation throughout each story (Fig. 1A). Sentences were repeatedly recorded until the recordings were noise-free. The order of sentences during the recording session was randomized across all stories in order to avoid the speaker becoming strongly familiarized with the repeating words. Using PRAAT software (Boersma & Weenink 2016), the recordings were cut and normalized to 70 dB. Because the present experiment is part of a comparative study of speech processing in adults and infants, the stimuli were slowed down to 90 bpm for a more infant-directed speech rate. The resulting recordings consisted of highly rhythmic speech signals with a syllabic rate at 3.2 Hz and a word rate at 1.6 Hz. To create the stimuli for the rhythmic condition, the eight individual sentences of each story were then concatenated in a pseudo-randomized order. For the non-rhythmic condition, an additional manipulation consisted of slowing down and speeding up alternating parts within a sentence by a factor of 0.5, resulting in a non-rhythmic speech pattern (Fig. 1A). The manipulated segments were 666 milliseconds long, which correspond to approximately two syllables. Whether the manipulation started with a lengthened or shortened fragment, was randomized within each story. The manipulation of local speech rate did not further affect quality of the speech signal, such as its pitch. Crucially, the repeated word within each sentence, together with its preceding and following syllable, remained at the original speech rate to guarantee comparability across conditions. Power spectra were created to confirm the success of the manipulation (Fig. 1B). As expected, large peaks of power at the relevant frequencies were present in the rhythmic condition and attenuated in the non-rhythmic condition. Further, the 3.2 Hz peak was more prominent than the 1.6 Hz peak in the original stimuli.

<Figure 1 >

For each story, the repeated word was randomly chosen across two words. In total, 40 word pairs were selected using the CELEX Dutch lexical database

(WebCELEX of the Max Planck Institute for Psycholinguistics, 2001). All words were disyllabic, low frequent (log < 1.3) nouns with stress on the initial syllable. Four stimuli lists were constructed, such that across lists each story appeared in both the rhythmic and the non-rhythmic condition and both possible words were familiarized equally often over participants. For example, for half the listeners the *hommels* story (Table 1) was presented in the rhythmic condition and for the other half in the non-rhythmic condition. In both cases, for half of the listeners the word *hommels* was familiarized and the word *kevers* was not, while for the other half *kevers* was familiarized and *hommel*s was not. Overall, every story was only presented once per subject. Although all familiarized words were carefully chosen to avoid Dutch-English cognate words, several provisions were made to ensure that prior knowledge would not facilitate word recognition. First, all familiarized and novel words were rated by five native English speakers, who did not take part in the experiment (Mean age = 27.20). The English native speakers would listen to each word and note down any English word it resembled. When there was too much phonetic overlap between the stimulus word and the associated word, we replaced it with a phonetically more different noun. Second, in a debriefing questionnaire after the experiment all participants were asked to note down any words they had heard that had been familiar (Dutch or English). Only three participants noted down words that were part of the familiarized-novel word pair list (toekans, kantoor, sultan). Because of the low number of participants having recognized the words and in order to avoid further reduction in amount of trials, the three stimuli were nonetheless included in all further analysis.

< Table 1>

Procedure

Participants were seated comfortably in the MEG and were given time to read through the task instructions and ask questions if necessary. Speech stimuli were presented via earphones. The experimental trials were presented in 40 blocks, each consisting of three phases. An example for one trial is shown in Table 1. Specifically, participants first listened to a Dutch story that could be either highly rhythmic or highly non-rhythmic. Subsequently, four words were presented in isolation with an inter-stimulus interval (ISI) varying between two and three seconds. The data recorded during this phase were later used to analyze ERFs for the word recognition effect. Two of the isolated words were presentations of the familiarized word and two were presentations of the novel word. Participants were instructed to passively and attentively listen to all presented stimuli. To restrict visual stimulation to a minimum and to reduce head movement, participants were asked to fixate a central fixation-cross that appeared two seconds before story onset and remained on the screen during the presentation of stories and isolated words. The isolated word presentation was followed by a word decision task. Participants heard the familiarized and novel word once again with an ISI between two and three seconds. Once they had heard both words, response options (“First word” or “Second word”) appeared on the screen. Participants were then asked to respond as fast as possible which of the two words they had heard in the preceding story. Responses were given via a button press of the left middle or index finger.

MEG acquisition

We recorded ongoing brain activity with a whole-head MEG system with 275 axial gradiometers (VSM/CTF Systems, Coquitlam, BC, Canada). The MEG system was located in a magnetically shielded room and instructions were projected from outside via mirrors onto a screen. Both horizontal and vertical electro-oculograms (EOGs) as well as an electrocardiogram (ECG) were recorded to facilitate removal of artifacts stemming from eye-movement or heart beat. The reference electrode was placed on the left mastoid. Data were recorded at a sampling frequency of 1200 Hz. Head location of the participant was monitored with marker coils placed at the nasion and in both ear canals via earplugs and monitored online (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). Rubber tubes for transmission of the audio signal were also connected to the earplugs.

Anatomical T1-weighted magnetic resonance (MR) images of each participant’s brain were acquired using 3 Tesla Siemens PrismaFit and Skyra scanners (Erlangen, Germany). All scans covered the entire brain and had a voxel size of 1x1x1mm^3. During MRI acquisition, earplugs - similar to the ones used during MEG but with a drop of vitamin E in place of the coils - were used to allow for co-registration of the MRI and MEG data. Additionally, each participant’s head shape was recorded using a Polhemus Isotrack system for more accurate co-registration.

Data Analysis

Behavioral data.

During the word decision task, reaction time (RT) and accuracy were recorded. RTs were measured starting at the beginning of the second stimulus presentation. Stimuli had an average length of 664 milliseconds (SD = 89.6 ms). To evaluate differences in effects of behavior for our two conditions we performed paired t-tests using SPSS (IBM Corp. Released 2010. IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY: IBM Corp.).

MEG preprocessing.

The data preprocessing and analysis was performed using the Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld, Fries, Maris, & Schoffelen, 2011; Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands. See http://www.ru.nl/neuroimaging/fieldtrip). At the first step of preprocessing, line noise and its harmonics were removed from the signal using a dft filter. For the ERF analysis, the data recorded during the “isolated word phase” were segmented into epochs from 300 milliseconds pre-stimulus to 1.5 seconds post-stimulus. Only the first presentation of each familiarized or novel word was analyzed. For the speech-brain coherence analysis, the data recorded during the “familiarization phase” were segmented into epochs comprising individual sentences (4.666 s). Based on each epoch’s variance, outliers were selected and all epochs with strong, irregular artifacts or SQUID jumps were rejected. Further, all epochs were inspected individually for potential artifacts. For the isolated word phase, this resulted in the exclusion of 12.97% of the epochs (SD = 7.02%), leaving on average 35 epochs per participant (SD = 2.88) for the ERF analysis. For the familiarization phase, 5.20% (SD = 1.84%) of the epochs were excluded and on average 303 epochs per participant (SD = 5.90) were left for the coherence analysis. About the same amount of epochs were rejected per rhythmicity condition in both familiarization phase (mean rhythmic = 5.62%, mean non-rhythmic = 4.77%) as well as in the isolated word phase (mean rhythmic = 14.06%, mean non-rhythmic = 11.88%, mean familiarized word = 13.44%, mean novel word = 12.50%)

Subsequently, an independent component analysis (ICA) was used to remove artifacts stemming from the cardiac signal and eye blinks. For each participant, the time course of the independent components was correlated to the horizontal and vertical EOG signals and to the ECG signal to identify and remove potentially contaminating components. After removal of the ICA components, the data were again cleaned of smaller artifacts by rejecting epochs with highest variance

Event-related field analysis.

For each participant, the neural response was obtained separately for trials of the rhythmic and non-rhythmic condition as well as for responses to familiarized words and novel words. First, the data were filtered with a low-pass filter of 35 Hz and a high-pass filter of 0.5 Hz using a butterworth filter. Subsequently, ERFs were averaged over trials and baseline corrected with a 300 milliseconds time window prior to stimulus onset.

Further, we calculated the planar gradients, on which all subsequent sensor-level analyses were performed. For the combined planar gradient representation of the magnetic fields, we know that neuronal sources of the signal are located directly below the strongest signal on the scalp (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). This provided us with a better topographical description of the ERFs and facilitated group analysis. We compared grand-average ERFs to novel versus familiarized words pooled over rhythmicity conditions for the repetition effect. Further, we compared the magnitudes of the repetition effect in the rhythmic versus non-rhythmic condition for the effect of rhythmicity.

Speech-brain coherence.

In order to quantify phase locking between the acoustic signal and neural oscillations during the familiarization phase, we used the measure of coherence. This frequency-domain measure reflects the degree to which the phase relationships of two signals are consistent across measurements. In the context of entrainment to speech, our coherence measure indicates the consistency of phase locking of the speech signal and the neuronal oscillatory activity across trials.

Before computing the speech envelope of the stimuli, we removed all sentences corresponding to trials that were rejected during preprocessing. Subsequently, the broadband envelope of the speech stimuli was extracted by filtering the speech signal between 250 and 1000 Hz and extracting the temporal envelope using the Hilbert transform.

Finally, both the speech envelope and the MEG data were transformed from the time to frequency domain using a fast Fourier transform (FFT) with a Hanning window. The Fourier transform was restricted to a frequency range of 0 to 10 Hz and applied over non-overlapping, sentence-long segments (4.666 s each), resulting in a frequency resolution of approximately 0.2 Hz. Based on the hypothesis that neuronal entrainment is driven by auditory cues in the signal, we focused on a region of interest comprising sensors that pick up on the auditory response as reflected in the M100 (Eulitz, Diesch, Pantev, Hampson, & Elbert, 1995). Thus, the cross-spectral density was computed for each combination of those sensors and the speech signal. Given the cross-spectral density averaged over segments, we computed the speech-brain coherence for every participant. Coherence was calculated with the following formula (Rosenberg, Amjad, Breeze, Brillinger, & Halliday, 1989):

$$\mathrm{Coh}_{x,y} = \mathrm{I\ mean}(S_{xy})\mathrm{I}\ /\ \mathrm{sqrt}(\mathrm{mean}(S_{xx}) * \mathrm{mean}(S_{yy}))$$

Where S_{xy} denotes the cross-spectrum between sensor signal x and speech signal y and S_{xx},S_{yy} the power spectra of those signals, respectively.

Source analysis.

For the analysis of speech-brain coherence, we calculated the sources that were most coherent to the speech signal with a spatial beamformer using dynamic imaging of coherent sources (DICS) (Gross et al. 2001). The DICS algorithm computes a spatial filter from the forward solution at each voxel and the cross-spectral density described above.

Source analysis was performed on the axial gradiometer data. We created the source model by dividing each participant’s brain volume into a regular grid of points spaced 8mm apart and warping it to a template brain (Montreal Neurological Institute). A single-shell head model was derived from each participant’s structural magnetic resonance image (Nolte, 2003) to compute the lead fields. Beforehand, MRIs were spatially co-registered to the MEG sensors according to the position of fiducials on the nasion and the two ears and taking into account Polhemus information.

Statistical analysis.

We statistically tested the word repetition effect observed on the sensor level using a nonparametric cluster-based permutation approach (Maris & Oostenveld, 2007) to control for multiple comparisons. For this statistical analysis, we included all time points and sensors. To test the difference between the rhythmic and the non-rhythmic condition we further restricted the analysis to those sensors that we had previously found to contribute to a significant cluster when analyzing the word repetition effect.

Similarly, to evaluate whether speech-brain coherence was significantly larger in the rhythmic compared to the non-rhythmic condition, we used a cluster-based permutation test. Specifically, we conducted a whole-brain analysis including only frequencies below 10 Hz, since we were most interested in entrainment to amplitude modulations (i.e. stress) of the speech signal.

Results

Word decision task

Participants listened to Dutch spoken texts with a word being repeated eight times within one story, thereby becoming a high frequent target. To test if participants detected the familiarized word, they were asked to perform a word decision task. At the end of each trial, participants were presented with the familiarized word and a novel word, and responded in a two-alternative forced choice task which word they thought they had heard in the story before. Overall, participants performed far above chance level in the word decision task (mean accuracy = 92 %, SD = 0.09 %). Their high performance implicates that, in general, participants were able to segment individual words from continuous foreign speech without difficulty. RTs had a large variance within and across participants (Mean = 1209.87 ms SD = 314.67 ms). Incorrect trials were removed previous to analysis of the RTs. For all subsequent analyses of accuracies or RTs, trials with RTs larger or smaller than three standard deviations from the mean were removed. With regard to our manipulation, participants' ability to detect the repeated word did not differ whether the spoken passages were rhythmic or non-rhythmic: paired samples t-tests did not reveal any significant effects of rhythmicity on accuracy (rhythmic: 92 %; non-rhythmic: 93 %; $t(15) = - 0.40$, $p > .69$) or RTs ($t(15) = 1.22$, $p > .24$). This result suggests that rhythmic cues in the speech signal did not facilitate word segmentation in comparison to a non-rhythmic signal.

Event-Related Field analysis

Repetition effect.

To confirm previous findings of the word repetition effect, we compared the ERFs in response to the familiarized word versus a novel word. ERFs elicited by the familiarized word were attenuated in amplitude compared to the ERFs elicited by the novel word. Spatio-temporal cluster-based permutation tests revealed a significant difference between the response to familiarized words and the response to novel words (p = .004). Figure 2A shows the topographical map of the amplitude difference between familiar and novel response. The difference was most pronounced over right fronto-temporal sensors (marked with white circles) and between 187 and 592 milliseconds after stimulus onset. A grand average of both ERFs, averaged over the sensors significantly contributing to the cluster, is shown in figure 2B. We further asked whether the repetition effect could predict behavior in the word decision task. The analysis was inconclusive, as there was no significant correlation between the size of the repetition effect in ERFs and measures of accuracies or RTs (Pearson correlation: RTs, r = + 0.17, accuracy, r = - 0.15, both p > 0.5). However, one must note that participants had a very good performance in the task and the variation in RTs across participants was large.

< Figure 2 >

Rhythmicity effect.

We further expected rhythmicity to influence successful detection of the familiarized word. In order to quantify this effect, we compared the strength of the repetition effect in the rhythmic and non-rhythmic trials (Fig. 3A). A spatio-temporal cluster based permutation test did not reveal any effect of rhythmicity. Moreover, the effect did not reach significance when we restricted the cluster based permutation test to right fronto-temporal sensors. In agreement with the absence of the rhythmicity effect, topographies were similar in both conditions with peak activity over right lateralized fronto-temporal sensors. Figure 3C shows topographies of the difference between novel and familiarized words in both the rhythmic and the non-rhythmic conditions. Across participants, ERFs varied with respect to the direction of the rhythmicity effect. In some participants, a larger repetition effect (positive difference for control - target) was found in the rhythmic condition (Fig. 3B lower panel, blue line), as predicted. Other participants displayed the opposite pattern: a large repetition effect in the non-rhythmic condition only (Fig. 3B upper panel, red line).

< Figure 3 >

Neuronal entrainment (speech-brain coherence)

We stated the hypothesis that neural entrainment to speech rhythms would influence participants' ability to detect repeated words in continuous speech. Specifically, the ability to track rhythmic patterns in speech may be variable across participants and could explain the inter-subject variability we observed for the repetition effect in the rhythmic and non-rhythmic condition. To test this, we measured the coherence between the MEG signal and the speech envelope during the familiarization phase. First, we performed this analysis in sensor space, restricting the analysis to sensors that maximally respond to speech sounds. These sensors were found by selecting those with the largest M100 response to sound onsets during the isolated word phase, i.e. when participants were listening to the familiarized and novel word in isolation after the presentation of the spoken passage (see planar gradient data in Fig. 4A, selected sensors are marked with white circles).

We compared speech-brain coherence in both the rhythmic and the non-rhythmic condition averaged over the selected sensors and restricted to frequencies most prominent in speech (1 - 10 Hz). In almost all participants, we observed a peak in coherence for the rhythmic condition around 3.2 Hz corresponding to the syllabic rate of the stimulus. These results mirrored the power spectra derived from the stimuli and their manipulations. In contrast, no clear peak at 1.6 Hz corresponding to the word rate was observed, and it did not systematically differ between the rhythmic and the non-rhythmic condition. Figure 4C shows coherence values for both conditions across sensors of interest averaged over participants. At the source level, the increase in 3.2 Hz coherence values for the rhythmic condition was most prominent in right superior and right middle temporal areas (whole brain cluster-based permutation test, p = 0.01). In figure 4D, the localized source activity is depicted at those voxels where the cluster-based permutation test revealed a difference between conditions.

< Figure 4 >

To answer the question about the influence of neural entrainment on word segmentation, we correlated entrainment during the familiarization and neural response during the isolated word phase. Specifically, we tested if the increase in coherence strength in the rhythmic condition (as measured at the sensor level' s ROI) correlated with the strength and direction of the word repetition effect measured in ERFs. However, there was no significant correlation (Pearson's correlation: r = -0.33, p = 0.22). Given the inter-subject variability of the repetition effect between rhythmic and non-rhythmic condition, we extended this question to whether this variation could be explained by the difference in coherence strength between rhythmic and non-rhythmic condition, but this was also not significant (r = -0.29, p = 0.27).

Lastly, as for the neural repetition effect, we tested whether the difference between in coherence strength predicted Behavioral differences between rhythmic and non-rhythmic conditions. Again, there was no significant correlation with neither of the behavioral measures (RTs, r = - 0.14; accuracy, r = - 0.32; both p > 0.2).

Discussion

In this study, we asked whether the rhythm structure of foreign speech provides cues for word segmentation, and whether neural oscillatory entrainment is involved during speech parsing. To do so, we tested whether word segmentation is facilitated in the presence of a highly rhythmic speech signal compared to a non-rhythmic speech signal. As a measure of successful word segmentation, we analyzed the word repetition effect for familiarized versus non-familiarized target words measured in ERFs as well as behavior.

All participants performed with very high accuracy in the behavioral word decision task, despite their lack of proficiency in the Dutch language. This confirms that all participants listened to the stimuli attentively and that after eight sentences, they were sufficiently familiarized with the target. We also found a significant difference in the ERF response to familiarized compared to a novel words. Further, an analysis of speech-brain coherence revealed stronger coherence at 3.2 Hz in the rhythmic compared to the non-rhythmic condition. Thus, we conclude that our rhythmic stimuli successfully evoked neuronal entrainment. Unexpectedly, effects of rhythmicity on word segmentation did not reach significance, neither at the behavioral nor at the neural level. Further, the amount of neural entrainment was seemingly unrelated to successful segmentation of familiarized words.

Repetition effect

We replicated the repetition effect in foreign listeners, which is characterized by reduced activity for repeated stimuli, thereby providing additional evidence for the possibility to recognize a word repetition without prior lexical knowledge (M. D. Rugg et al., 1995; Snijders et al., 2007). While Snijders et al. (2007) found this effect in foreign listeners after the repeated presentation of isolated words, our paradigm extends their findings by familiarizing the target in sentence context. Snijders et al. (2007) found that foreign listeners, in contrast to natives, did not recognize a repeated stimulus when presented in sentence context. Here, listeners seemed to pick up on the repetition within sentences. After eight phrase-embedded word repetitions, the familiarized word evoked reduced neural activity compared to a novel word, suggesting that listeners had build up a lexical representation of the repeated word. This entails that word segmentation effectively takes place when listening to sentences in a foreign language, as long as ample repetitions are provided.

Apart from the increased number of repetitions, it is possible that the additional word decision task amplified the repetition effect. Although Snijders et al.'s participants might have been aware of the word repetition, there was no instruction that would have evoked goal-oriented listening. In our study, participants were expecting the word decision task at the end of every familiarization phase and therefore adopted listening strategies to enhance their performance. In a debriefing questionnaire, all participants reported to have noticed the repetition of the target word throughout the sentences. The majority further said they had adopted a strategy to prepare themselves for the upcoming decision.

Strong rhythmic patterns evoke lateralized speech-brain coherence

In our study, we observed oscillatory entrainment at the syllabic rate (here ≈ 3.2 Hz) in response to a highly rhythmic speech signal. This finding is in agreement with previous studies showing that cortical oscillations follow the speech envelope during listening (Ahissar et al., 2001; Buiatti et al., 2009; Ding et al., 2015; A. L. Giraud et al., 2007; Luo & Poeppel, 2007). Unlike studies exhibiting a close correlation between entrainment and speech comprehension, we did not find a correlation between the strength in coherence with the rhythmic signal and behavioral measures. For example, Ahissar et al. (2001) reported accuracy measures that strongly decreased with increasing compression of speech rate. In contrast, our rhythmic stimuli were undistorted and played at a moderate speech rate, resulting in high performance and therefore limited behavioral variation.

Statistical tests at the source level revealed that the increase in coherence was most prominent in right superior and middle temporal cortices. This lateralization of coherence has previously been observed when investigating oscillatory brain activity in response to speech (Luo & Poeppel, 2007; Obleser, Eisner, & Kotz, 2008), and especially during processing of syllabic patterns (Abrams, Nicol, Zecker, & Kraus, 2008). Right hemisphere lateralization for low-frequency speech-brain coherence can be accommodated by the asymmetric sampling in time (AST) hypothesis (Poeppel, 2003). According to AST, speech, although evoking bilateral representation in primary auditory cortices, may be processed asymmetrically in the time domain. Based on electrophysiological data, Poeppel (2003) argues that while the left non-primary auditory cortex prefers short temporal integration windows (20 – 50 ms), the corresponding right hemisphere extracts slower acoustic changes (150 –

250 ms). In order to determine, whether our results support the AST, a direct comparison between neural activity in the two hemispheres would be necessary.

Rhythmicity effect

In the non-rhythmic condition, participants displayed significantly less speech-brain coherence due to less reliable acoustic patterns to entrain to. Despite the inhibited entrainment, word segmentations was neither hindered nor delayed. This absence of an effect of rhythmicity suggests that speech rhythm as an acoustic cue and with it neuronal entrainment as a proposed underlying mechanism are not necessary for successful word segmentation in a foreign language listening paradigm. There are two different interpretations of our results that, nonetheless, may still afford a mechanistic role of neuronal entrainment in speech processing. One possibility is that the facilitative effects of rhythmic cues have been outweighed by other acoustic cues that were present in both the rhythmic and the non-rhythmic condition. Alternatively, listeners may have applied two distinct processing strategies depending on the availability of rhythmic cues, leading to successful segmentation in both conditions.

Are effects of rhythmicity outweighed by other acoustic cues?

Concerning the first interpretation, it is likely that listeners have exploited other cues apart from speech rhythm, since intonational or statistical information was intact in both rhythmic and non-rhythmic stimuli. Previous research pitting cues against each other has proposed that listeners use a hierarchically integrated speech cue approach instead of solely focusing on one type of cue (Mattys et al. 2005). Moreover, participants seem to rely on low-level information, such as word stress, more strongly when the signal is impoverished and can otherwise resort to different sub-lexical cues (Mattys et al. 2005). We expected here that the lack of Dutch proficiency constitutes an impoverished speech environment, as no linguistic cues could be used for speech segmentation. Yet, participants may have relied on several acoustic cues to perform the segmentation task. In particular, spectral cues in the speech signal generally have been shown to have a greater effect on recognition than temporal cues (Kim, Chang, Yang, Oh, & Xu, 2015). We used a very slow, infant-directed speech rate that might have enhanced the importance of spectral cues such as intonation or co-articulation. Also, the spectral content was left intact by our manipulation. This could explain the observed high performance throughout conditions in the current study. Moreover, as has been shown with non-speech stimuli, spectral fluctuations can evoke neural entrainment and thereby influence auditory perception (M J Henry, Herrmann, & Obleser, 2014; Molly J Henry & Obleser, 2012). Therefore, our experimental manipulation limits the interpretation of our results. An improved version of the paradigm should aim at reducing spectral cues and increasing the task difficulty in order to strengthen the importance of rhythmic cues as well as boosting inter-trial variability in performance. This could, for example, be achieved by using a faster speech rate and fewer repetitions or using degraded speech (e.g. by adding stationary noise).

Non-rhythmic speech profits from “vigilance” mode.

In the debriefing questionnaires, the majority of participants rated the rhythmic condition as “easier to listen to” compared with the non-rhythmic condition. Hence, a second possible interpretation assumes that participants adopted a different processing strategy altogether for the non-rhythmic stimuli and the rhythmic stimuli. Specifically, in the absence of rhythmic cues, participants might have switched to a “vigilance” mode as described by Schroeder and colleagues (2009). In vigilance mode, the processing system is thought to maintain a state of continuous high excitability by enhancing gamma amplitude and suppressing low frequencies (Fries, Reynolds, Rorie, & Desimone, 2001; Schroeder & Lakatos, 2009). Specifically, because such extended high excitable states require more cognitive resources, they should be associated with a reduction in alpha power. In line with the vigilance mode hypothesis, previous studies report alpha power to decrease with loss of acoustic regularities (S. J. Kayser, Ince, Gross, & Kayser, 2015; Obleser & Weisz, 2012; Peña & Melloni, 2012). Further analyses of alpha and gamma power may enable us to account for the missing effect of rhythmicity and even exploit the observed inter-subject variability. Predictions for such an analysis would entail greater alpha suppression as well as higher gamma activity during the non-rhythmic familiarization phase. Moreover, for participants with larger repetition effects in the non-rhythmic condition, alpha suppression and gamma power increase should be maximal as these participants achieve higher performance in the vigilance mode compared to the entrained mode. In contrast, participants with larger repetition effects in the rhythmic conditions are expected to display relatively less alpha suppression and gamma power increase, resulting in a disadvantage for the vigilance mode.

Complex processing of non-rhythmic signals

Bearing in mind the proposed interpretations of our results, an account of entrainment as a core mechanism underlying speech segmentation may be too simplistic after all. While rhythmic patterns seem to be ubiquitous in our environment (e.g. in speech or music) not all languages are rhythmic in an isochronous sense and non-rhythmic signals are still generally perceivable. This raises questions such as: What other neural mechanisms ensure successful speech processing in the absence of neural entrainment? Furthermore, do those mechanisms only substitute neural entrainment for non-rhythmic signals or do they act jointly when processing rhythmic signals? In order to integrate the role of neuronal entrainment into auditory perception, it is hence crucial to understand processing of non-rhythmic signals. Generally, neural phase has been shown to modulate perception not only for entrained oscillations but also in absence of acoustic rhythm (Molly J Henry & Obleser, 2012; Neuling, Rach, Wagner, Wolters, & Herrmann, 2012). A recent study used non-speech stimuli without rhythmic structure to look at instantaneous phase effects of neural activity on auditory target detection. They revealed a first glimpse of the complex interplay among different neural frequencies and the influence of phase-phase combinations on detection (M. J. Henry, Herrmann, & Obleser, 2016). As discussed in the beginning, speech-processing success has been found to correlate with strength of entrainment in a specific frequency band corresponding to stimulus rhythm. Given the recent evidence by Henry et al (2016), it seems that in the absence of rhythmic structure perception relies on complex phase-phase combinations of several frequencies. Thus, rhythmic cues in past studies may have lead to an oversimplified picture of the underlying processes (Obleser, Herrmann, & Henry, 2012). One proposal for future investigation is to analyze instantaneous phase-phase information in a larger range of frequencies during listening to both rhythmic and non-rhythmic continuous speech. Thereby, one could test whether the interplay among several frequencies can predict optimal speech processing beyond strength of entrainment.

Conclusion

Results of the current study clearly demonstrate that foreign listeners, without previous lexical knowledge, are able to segment words repeatedly presented in sentence context. Further, we showed that strong acoustic rhythmic patterns induce low frequency oscillatory entrainment at the syllabic rate. Crucially, our results do not confirm the hypothesis that the presence of and neural entrainment to rhythmic cues in the speech signal play a mechanistic role for word segmentation of foreign speech. By targeting foreign listeners, our study contributes to our knowledge about preferences and importance of speech cues during second language learning. Also, it provides evidence for neural entrainment under very close to natural conditions, using non-degraded sentences. At the same time, due to preserved intonation and co-articulation as well as an infant-orientated speech rate, the stimuli pose certain limitations on our conclusions. For future research, it will be important to improve the experimental design to isolate effects of rhythmic cues and to find a meaningful behavioral measure of word segmentation to better evaluate the modulations in oscillatory entrainment to speech rhythm.

References

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2008). Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech. J Neurosci, 28 (15), 3958–3965. doi: 10.1523/JNEUROSCI.0187-08.2008

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, a., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Sciences of the United States of America, 98 (23), 13367–13372. doi: 10.1073/pnas.201400998

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer. Computer program, Version 6.0.19 . Retrieved from http://www.praat.org

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. doi: 10.1016/j.neuroimage.2008.09.015

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. Journal of experimental psychology. Learning, memory, and cognition, 16 (6), 1084–1096. doi: 10.1037/0278-7393.16.6.1084

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. Journal of Memory and Language, 31 (2), 218–236. doi: 10.1016/0749-596X(92)90012-M

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. Language and Speech, 40 , 141–201.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience, 1–10. doi: 10.1038/nn.4186

Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2016). Temporal Modulations Reveal Distinct Rhythmic Properties of Speech and Music. bioRxiv. doi: http://dx.doi.org/10.1101/059683

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. Frontiers in Human Neuroscience, 8 (May), 311. doi: 10.3389/fnhum.2014.00311

Doyle, M. C., Rugg, M. D., & Wells, T. (1996). A comparison of the electrophysiological effects of formal and repetition priming. Psychophysiology, 33 (2). doi:
10.1111/j.1469-8986
.1996.tb02117.x

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., & Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. The Journal of neuroscience : the official journal of the Society for Neuroscience, 15 (4), 2748–2755.

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: a matter of signal quality. Perception & psychophysics, 69 (6), 856–864. doi: 10.3758/BF03193922

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science, 291 (5508), 1560–3. doi: 10.1126/science.291.5508.1560

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. Neuron, 56 (6), 1127–1134. doi: 10.1016/ j.neuron.2007.09.038

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nature Neuroscience, 15 (4), 511–517. doi: 10.1038/nn.3063

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech - A syllable-centric perspective. Journal of Phonetics, 31 (3-4), 465–485. doi: 10.1016/j.wocn.2003.09.005

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proceedings of the National Academy of Sciences of the United States of America, 98 (2), 694–9. doi: 10.1073/pnas.98.2.694

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). _-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proceedings of the National Academy of Sciences of the United States of America, 108 (48), 19377–82. doi: 10.1073/pnas.1117190108

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. Reviews of Modern Physics, 65 (2), 413–497. doi: 10.1103/RevModPhys.65.413

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. Proceedings of the National Academy of Sciences of the United States of America, 111 (41), 14935–14940. doi: 10.1073/ pnas.1408741111

Henry, M. J., Herrmann, B., & Obleser, J. (2016). Neural Microstates Govern Perception of Auditory Input without Rhythmic Structure. Journal of Neuroscience, 36 (3), 860–871. doi: 10.1523/JNEUROSCI.2191-15.2016

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proceedings of the National Academy of Sciences of the United States of America, 109 (49), 2009–100. doi: 10.1073/pnas.1213390109

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. Journal of Memory and Language, 44 , 548–567. doi: 10.1006/jmla.2000.2755

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. Cognitive psychology, 39 (3-4), 159–207. doi: 10.1006/cogp.1999.0716

Kayser, S. J., Ince, R. a. a., Gross, J., & Kayser, C. (2015). Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha. Journal of Neuroscience, 35 (44), 14691–14701. doi: 10.1523/JNEUROSCI.2243-15 .2015

Kim, B. J., Chang, S. A., Yang, J., Oh, S. H., & Xu, L. (2015). Relative contributions of spectral and temporal cues to Korean phoneme recognition. PLoS ONE, 10 (7), 3255–3267. doi: 10.1371/journal.pone.0131807

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. Journal of neurophysiology, 94 (3), 1904–1911. doi: 10.1152/jn.00263.2005

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. Behavior research methods, 44 (2), 325–43. doi: 10.3758/s13428-011-0146-0

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. Neuron, 54 (6), 1001–1010. doi: 10.1016/ j.neuron.2007.06.004

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164 (1), 177–190. doi: 10.1016/ j.jneumeth.2007.03.024

Mattys, S. L. (2004). Stress versus coarticulation: toward an integrated approach to explicit speech segmentation. Journal of experimental psychology. Human perception and performance, 30 (2), 397–408. doi: 10.1037/0096-1523.30.2.397

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. Journal of Experimental Psychology: General, 134 (4), 477–500. doi: 10.1037/0096-3445.134.4.477

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. Journal of Neuroscience, 26 (29), 7629–7639. doi: 10.1523/JNEUROSCI.5501-05.2006

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature, 485 (7397), 233–6. doi: 10.1038/

nature11020

Neuling, T., Rach, S., Wagner, S., Wolters, C. H., & Herrmann, C. S. (2012). Good vibrations: Oscillatory phase shapes perception. NeuroImage, 63 (2), 771–778. doi: 10.1016/j.neuroimage.2012.07.024

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. Psychological review, 115 (2), 357–395. doi: 10.1037/0033-295X.115.2 .357

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Sensitivity to Spectral and Temporal Features. The Journal of Neuroscience, 28 (32), 8116–8123. doi: 10.1523/JNEUROSCI.1290 -08.2008

Obleser, J., Herrmann, B., & Henry, M. J. (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. Frontiers in Human Neuroscience, 6(August), 2008–2011. doi: 10.3389/fnhum.2012.00250

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cerebral Cortex, 22 (11), 2466–2477. doi: 10.1093/ cercor/bhr325

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience, 2011 . doi: 10.1155/2011/156869

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. Frontiers in Psychology, 3 (SEP), 1–17. doi: 10.3389/fpsyg.2012 .00320

Pellegrino, F., Coupe, C., & Marsico, E. (2011). Across-language perspective on speech information rate. Language, 87 , 539-558.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. Science, 298 (October), 604–607.

Peña, M., & Melloni, L. (2012). Brain Oscillations during Spoken Sentence Processing. Journal of cognitive neuroscience, 24 (5), 1149–1164. doi: 10.1162/jocn_a_00144

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. Speech Communication, 41 (1), 245–255. doi: 10.1016/S0167-6393(02)00107-3

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: a study based on speech resynthesis. The Journal of the Acoustical Society of America, 105 (1), 512–21. doi: 10.1121/1.424522

Rosenberg, J., Amjad, A., Breeze, P., & Brillinger, D. M., D. R.and Halliday. (1989). The Fourier Approach to the Identification of Functional Coupling between Neuronal Spike Trains. Progress in Biophysics and Molecular Biology, 53 , 1-31.

Rugg, M., Furda, J., & Lorist, M. (1988). The Effects of Task on the Modulation of Event-Related Potentials by Word Repetition (Vol. 25) (No. 1). doi: 10.1111/j.1469-8986.1988.tb00958.x

Rugg, M. D., Doyle, M. C., & Wells, T. (1995). Word and nonword repetition within- and across-modality: an event-related potential study. Journal of cognitive neuroscience, 7 (2), 209–27. doi: 10.1162/jocn.1995.7.2.209

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. Journal of Memory and Language, 35 (4), 606–621. doi: 10.1006/jmla.1996.0032

Saksida, A., Langus, A., & Nespor, M. (2016). Co-occurrence statistics as a language dependent cue for speech segmentation. Developmental Science, 1–11. doi: 10.1111/desc.12390

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. Trends in Neurosciences, 32 (1), 9–18. doi: 10.1016/j.tins.2008.09.012

Snijders, T. M., Kooijman, V., Cutler, A., & Hagoort, P. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. Brain Research, 1178 (1), 106–113. doi: 10.1016/j.brainres.2007.07.080

Spaak, E., de Lange, F. P., & Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. The Journal of neuroscience : the official journal of the Society for Neuroscience, 34 (10), 3536–44. doi: 10.1523/JNEUROSCI.4385-13.2014

Stefanics, G., Hangya, B., Hernadi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase Entrainment of Human Delta Oscillations Can Mediate the Effects of Expectation on Reaction Speed. Journal of Neuroscience, 30 (41), 13578–13585. doi: 10.1523/JNEUROSCI.0703-10.2010

Steinhauer, K., Alter, K., & Friederici, a. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nature neuroscience, 2 (2), 191–196. doi: 10.1038/5757

Stolk, A., Todorovic, A., Schoffelen, J. M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. Neuroimage, 68 , 39–48.

VanRullen, R., Busch, N. A., Drewes, J., & Dubois, J. (2011). Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. Frontiers in Psychology, 2 (April), 1–9. doi: 10.3389/fpsyg.2011.00060