Effects of acoustic periodicity and intelligibility on the neural oscillations in response to speech

Kurt Steinmetzger [a)*] **& Stuart Rosen** [a)]

[a)] Speech, Hearing and Phonetic Sciences, UCL, Chandler House, 2 Wakefield St, London WC1N 1PF, United Kingdom

[*]Author to whom correspondence should be addressed. Electronic mail: kurt.steinmetzger.12@ucl.ac.uk

Abstract

Although several studies have investigated neural oscillations in response to acoustically degraded speech, it is still a matter of debate which neural frequencies reflect speech intelligibility. Part of the problem is that effects of acoustics and intelligibility have so far not been considered independently. In the current electroencephalography (EEG) study the amount of acoustic periodicity (i.e. the amount of time the stimulus sentences were voiced) was manipulated, while using the listeners' spoken responses to control for differences in intelligibility. Firstly, the total EEG power changes in response to completely aperiodic (noise-vocoded) speech and speech with a natural mix of periodicity and aperiodicity were almost identical, while an increase in theta power (5–6.3 Hz) and a trend for less beta power (11–18 Hz) were observed in response to completely periodic speech. These two effects are taken to indicate an information processing conflict caused by the unnatural acoustic properties of the stimuli, and that the subjects may have internally rehearsed the sentences as a result of this. Secondly, we separately investigated effects of intelligibility by sorting the trials in the periodic condition according to the listeners' spoken responses. The comparison of the intelligible trials, largely unintelligible trials, as well as completely unintelligible spectrally rotated speech revealed that the total EEG power in the delta band (1.7–2.7 Hz) was markedly increased during the second half of the intelligible sentences, which suggests that delta oscillations are an indicator of successful speech understanding.

Keywords: EEG; speech; delta; theta; intelligibility; periodicity

1. Introduction

In order to shed light on the underlying neural mechanisms and cognitive processes involved when attempting to understand spoken speech, a growing number of magneto- and electroencephalographic (M/EEG) studies focusses on the time-frequency properties of the neural signals rather than traditional waveform analyses (for reviews see Giraud & Poeppel, 2012; Weisz & Obleser, 2014). The current experiment adds to existing knowledge by investigating effects of acoustics and intelligibility separately, two factors that usually vary together when speech signals are acoustically manipulated. Specifically, we manipulated the amount of acoustic periodicity, while controlling for differences in intelligibility, and vice versa. In the context of speech, periodicity denotes that a sound is produced by the periodic vibrations of the vocal folds, resulting in voiced speech with a pitch corresponding to the vibration rate. Unvoiced speech sounds, in contrast, emanate from constrictions in the vocal tract and have aperiodic fluctuations in energy, leading to a noisy sound quality and the absence of a pitch.

A popular speech processing technique that has been used in the neurosciences (e.g. Davis & Johnsrude, 2003; Obleser & Weisz, 2012; Peelle et al., 2013; Scott et al., 2000) as well as in psychoacoustic studies concerned with the simulation of cochlear implants (e.g. Qin & Oxenham, 2003; Schoof et al., 2013; Shannon et al., 1995), is noise-vocoding (henceforth referred to as the *aperiodic* condition). By filtering the unprocessed input speech into a specified number of frequency bands, it allows the spectral resolution of the synthesised output speech to be varied in a controlled manner, a feature that is closely related to speech intelligibility. At the same time, using noise as speech source results in a loss of the natural mix of voiced and voicelessness, and consequently also any voice pitch information, making it resemble whispered speech.

Nevertheless, our previous behavioural work (Steinmetzger & Rosen, 2015) has shown that preserving periodicity information in a vocoder (henceforth the *mixed* condition) does not lead to improved intelligibility rates. This suggests that periodicity information, despite its salience, is a redundant cue, at least in non-tonal languages and quiet listening conditions. The first question the current study addresses, is thus whether EEG time-frequency responses are similarly unaffected by the absence of any source periodicity.

To enable a more comprehensive investigation of the effects of periodicity, we included a third processing condition in which the same speech materials were synthesised with a completely periodic source (henceforth the *periodic* condition). Acoustically this condition is in fact closer to natural speech, which is mostly voiced (Dellwo, et al., 2007; Fourcin, 2010), than aperiodic noise-vocoded speech. However, because natural speech does not contain periodic sounds with much energy in the frequency region above 4 kHz, it sounds very unnatural. Additionally, periodicity is such a salient cue that it obscures weaker cues such as intensity differences, thereby making the information transmitted contradictory. For unvoiced fricatives like /s/ and /ʃ/, for example, aperiodic high-frequency energy is missing and replaced by periodic energy, which makes it difficult to identify these sounds. Consequently, periodic speech has substantially lower intelligibility rates than the other two conditions (Ardoint, et al., 2014; Steinmetzger & Rosen, 2015).

In order to control for this expected difference in intelligibility, the single trials were sorted according to the spoken responses of the participants. This approach was also chosen to enable the direct comparison of intelligible and unintelligible trials in the periodic condition. Consequently, the current study also provides the opportunity to investigate how the EEG time-frequency responses are affected by the intelligibility of the speech materials in the absence of any acoustic differences. This approach is akin to studies generating a pop-out effect by presenting the same stimulus materials twice, first without any additional information and then again after providing a written transcript (Sohoglu, et al., 2012) or the unprocessed recording (Millman, et al., 2015), but avoids any predictive top-down processing.

In addition to the acoustically matched unintelligible trials, the current study included completely unintelligible spectrally rotated speech as a second control condition (henceforth the *rotated* condition). Rotated speech has a similar spectro-temporal complexity as unrotated speech and has been used in several of the studies mentioned above (e.g. Becker et al., 2013; Peelle et al., 2013; Scott et al., 2000). Yet, apart from not being a precise acoustic match, the obvious meaninglessness casts doubts on whether it is indeed an adequate control condition.

Importantly, in contrast to the event-related potentials (ERPs), the EEG time-frequency analyses in the current paper were not assumed to be affected by the perception of a voice pitch. Direct cortical recordings and MEG experiments have shown that the presence of a pitch coincides with increased high gamma power (>80 Hz; e.g. Griffiths et al., 2010; Sedley et al., 2012). However, due to potential muscle artefacts and the low signal strength when recorder with cortical EEG, we did not these frequencies in the current analysis. Moreover, functional magnetic resonance imaging (fMRI) signals, which fluctuate at rates of less than .5 Hz (He & Raichle, 2009), have been shown to be larger for signals with a pitch (e.g. Norman-Haignere et al., 2013; Patterson et al., 2002). Yet, frequencies below .5 Hz similarly lie outside the possible frequency range of EEG time-frequency analyses, because they would require excessively long baseline and stimulus windows.

Based on the results of Strauß and colleagues (2014a), who have recently reported increased theta activity (here 3–7 Hz) in a left-lateralised fronto-temporal network for so-called ambiguous pseudo-words in an auditory word recognition task, we expected increased theta power in the periodic condition. The pseudo-words used by Strauß, et al. (2014a) were characterised by a wrong core vowel, making them resemble the periodic condition in the current study to some extent. Theta oscillations have also been associated with the storage of sequentially presented verbal information in working memory and the phonological loop (Roux & Uhlhaas, 2014). Based on this idea, Strauß, et al. (2014a) suggested that subjects may have internally rehearsed the unusual pseudo-words in order to classify them as words or non-words. More generally, this effect was taken to indicate an information processing conflict (Botvinick et al., 2001, 2004), although studies eliciting response conflicts in non-speech tasks, for example by using the Stroop paradigm, have usually reported mid-frontal theta power increases (Cohen & Donner, 2013; Hanslmayr, et al., 2008).

A recent theoretical approach has linked increased alpha power (~7–13 Hz) to the selective inhibition of brain areas that are not currently task relevant (Jensen & Mazaheri, 2010). Applied to speech perception, it has been proposed that alpha oscillations might be actively enhanced in order to cope with a demanding task, particularly listening to speech in the presence of background noise (Obleser, et al., 2012; Strauß, et al., 2014b; Wöstmann, et al., 2015). For words presented in quiet listening conditions, on the other hand, alpha activity was found to be increasingly suppressed with higher intelligibility levels (Becker, et al., 2013; Obleser & Weisz, 2012). However, in these studies the intelligibility of the mostly noise-vocoded stimuli varied along with their acoustic properties (i.e. the number of frequency bands in the vocoder) and hence also the subjective listening effort, which is similarly thought to depend on the degree of acoustic degradation (Obleser & Weisz, 2012; Wöstmann, et al., 2015). Sorting the trials in the periodic condition according to the spoken behavioural responses provided the opportunity to test whether there is indeed a direct relation between alpha suppression and speech intelligibility.

2. Material and methods

2.1 Participants

Eighteen normal-hearing right-handed subjects (8 females, mean age = 21.6 years, SD = 2.3 years) took part in the study. All participants were native speakers of British English and had audiometric thresholds of less than 20 dB HL at frequencies between 125 and 8000 Hz. All subjects gave written consent and the study was approved by the UCL research ethics committee.

2.2 Stimuli

The stimulus materials used in this experiment were recordings of the IEEE sentences (Rothauser, et al., 1969) spoken by an adult male Southern British English talker with a mean F0 of 121.5 Hz that were cut at zero-crossings right before sentence onset and normalised to a common root-mean-square (RMS) level. The IEEE sentence corpus consists of 72 lists with 10 sentences each and is characterized by similar phonetic content and difficulty across the lists, as well as an overall low semantic predictability. Every sentence contains five keywords (nouns, verbs, or adjectives; e.g. *The <u>birch</u> <u>canoe</u> <u>slid</u> on the <u>smooth</u> <u>planks</u>.*).

All stimulus materials were processed prior to the experiment using a channel vocoder implemented in MATLAB (Mathworks, Natick, MA). For all three vocoding conditions (aperiodic, mixed, and periodic) the original recordings of the IEEE sentences were first band-pass filtered into eight bands using zero phase-shift sixth-order Butterworth filters. The filter spacing was based on equal basilar membrane distance (Greenwood, 1990) across a frequency range of .1–8 kHz (upper filter cut-offs in Hz: 242, 460, 794, 1307, 2094, 3302, 5155, 8000; filter centre frequencies in Hz: 163, 339, 609, 1023, 1658, 2633, 4130, 6426). The output of each filter was full-wave rectified and low-pass filtered at 30 Hz (zero phase-shift fourth-order Butterworth) to extract the amplitude envelope. The low cut-off value was chosen in order to ensure that no temporal periodicity cues were present in the aperiodic condition.

In order to synthesise aperiodic speech, the envelope of each individual band was multiplied with a broadband white noise carrier. In the mixed condition, the envelope of each band was also multiplied with a broadband white noise, but only in time windows where the original speech was unvoiced. Sections that were voiced in the original recordings were synthesised by multiplying the envelopes with a pulse train following the natural F0 contour. The individual pulses had a duration of one sample point, i.e. about 23 μs at a sampling rate of 44.1 kHz. The F0 contours of the original sentences were extracted using ProsodyPro version 4.3 (Xu, 2013) implemented in PRAAT (Boersma & Weenink, 2013), with the F0 extraction sampling rate set to 100 Hz. The resulting F0 contours were corrected manually where necessary and then used to determine the distance between the individual pulses of the pulse train sources. Based on the original intermittent F0 contours, we also produced artificial continuous F0 contours by interpolation through unvoiced sections and periods of silence. These continuous F0 contours were used to produce the pulse train sources for the periodic condition.

Finally, in all three vocoding conditions, the eight sub-band signals were again band-pass filtered using the same filters as in the analysis stage of the process. Before the individual bands were summed together, the output of each band was adjusted to the same RMS level as found in the original recordings.

Spectrally rotated speech was produced using a technique introduced by Blesser (1972) and implemented in MATLAB. Here, the waveforms of the mixed condition described above were first multiplied with an 8 kHz sinusoid, resulting in a spectral rotation around the midpoint frequency of 4 kHz. Note, that this procedure also renders the rotated speech inharmonic, since the frequencies of the component tones will not be multiples of a particular F0 anymore. The rotated waveforms were then filtered (FFT-based FIR filter, order 256) to have the average UK long-term speech spectrum (Byrne, et al., 1994) and, finally, scaled to the same RMS level as the original waveforms in the mixed condition.

Figure 1 shows an unprocessed example sentence along with the same sentence processed in the four ways described.

Figure 1 about here

2.3 Procedure

Each participant listened to 80 aperiodic, 80 mixed, 160 periodic, and 80 rotated sentences. There were twice as many trials in the periodic condition because we wanted to ensure a sufficient number of unintelligible trials. All 4 conditions were presented in blocks of 10 sentences (i.e. 1 complete IEEE sentence list) and the order of the conditions and IEEE lists was randomised. Only the first 40 IEEE lists were used in the main experiment and none of the sentences was presented more than once. Participants were asked to repeat as many words as possible after every sentence. The verbal responses were logged by the experimenter before the next sentence was played and no feedback was given following the responses. The presentation of the stimuli and the logging of the responses was carried out using Presentation version 17.0 (Neurobehavioral Systems, Berkeley, USA).

Single trials consisted of a silent pre-stimulus interval with random duration (1.5–2.5 s), a stimulus sentence (average duration = 2.04 s, SD = .24 s) followed by a silent interval of .25 s, a short beep sound signalling the participants to start responding, the spoken responses, and the subsequent logging of the responses by the experimenter.

Before being tested, the subjects were familiarised with the materials by listening to 10 aperiodic, mixed, and periodic examples sentences each (IEEE lists 41–43). During the familiarisation phase every sentence was directly followed by its unprocessed counterpart, and again followed by the processed sentence.

The main part of the experiment took about 70 minutes to complete and subjects were allowed to take breaks whenever they wished to. The experiment took place in a double-walled sound-attenuating and electrically-shielded booth, with the computer signal being fed through the wall onto a separate monitor. Participants sat in a comfortable reclining chair during EEG acquisition and told to not move their eyes during sentence presentation. The stimuli were converted with 16-bit resolution and a sampling rate of 22.05 kHz using a Creative Sound Blaster SB X-Fi sound card (Dublin, Ireland) and presented over Sennheiser HD650 headphones (Wedemark, Germany). The presentation level was about 71 dB SPL over a frequency range of .1–8 kHz as measured on an artificial ear (type 4153, Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark).

2.4 EEG recording and pre-processing

The continuous EEG was recorded using a Biosemi ActiveTwo system (Amsterdam, Netherlands) with 61 Ag-AgCl scalp electrodes mounted on a cap according to the extended international 10-20 system. Four additional external electrodes were used to record the vertical and horizontal eletrooculogram (EOG) by placing them on the outer canthus of each eye as well as above and below the left eye. Two more external electrodes were used to record the signal from the left and right mastoids. EEG signals were recorded with a sampling rate of 1024 Hz and an analogue anti-aliasing low-pass filter with a cut-off frequency of 200 Hz.

The EEG data were processed offline using EEGLAB 12.0.2.5b (Delorme & Makeig, 2004). The waveforms were first down-sampled to 512 Hz, re-referenced to the mean of the two mastoids, and then filtered (zero-phase shift Hamming-windowed sinc FIR filter, order 3380, using the firfilt plugin version 1.5.3.) with a .1 Hz high-pass filter and a 100 Hz low-pass filter. An independent component analysis (ICA) was used to remove artefacts caused by eye blinks, eye movements, and muscular activity. Epochs ranging from -1000–3000 ms around sentence onset were extracted and rejected if amplitudes exceeded ±120 μV, if linear trends exceeded 120 μV in a 500 ms gliding window, or if the trial was lying outside a ±6 SD range (for single channels) and ±3 SD (for all channels) of the mean voltage probability distribution or the mean distribution of kurtosis values. On average 73.5% (294/400, SD = 13%, range = 57–95%) of the total number of trials passed the rejection procedure.

2.5 EEG time-frequency analysis

In order to ensure the same signal-to-noise ratio in each condition, all analyses in the present paper are based on matched trial numbers across conditions. For each individual participant, the number of trials was determined by the condition with the fewest trials, with excess trials in the other conditions omitted randomly.

The pre-processed EEG data were first sorted according to the spoken responses. For the analysis of periodicity only trials with all five keywords correct were considered in order to control for differences in intelligibility. This resulted in an average of 35.7 trials in each of the 4 processing conditions (SD = 9.7, range = 23–56). For the analysis of intelligibility, trials in the periodic condition with all five keywords correct were compared to trials with maximally two correct keywords (i.e. less than half of the sentence correctly repeated). Three participants with less than 10 unintelligible trials were excluded due to the low signal-to-noise ratio of the data. The remaining 15 participants (8 females) had an average number of 21.7 trials per condition (SD = 7.3, range = 14–35).

The time-frequency decomposition of the pre-processed and sorted data was conducted by computing the event-related spectral perturbation (ERSP) as implemented in EEGLAB. The ERSP is a measure of the relative change in power from baseline to stimulus period (Makeig, 1993). For each time-frequency point in the stimulus window, the spectral power is divided by the average power of the respective frequency bin in the baseline window and transformed into a dB value. The data were analysed from 1–30 Hz in 200 log-spaced frequency steps by convolving them with a set of Morlet wavelets, whose widths increased linearly with frequency from 1–15 cycles. This resulted in an analysis window ranging from -442.4–2442.4 ms around sentence onset. For the sake of simplicity, the rounded values -500–2500 ms will be used henceforth. Within this window, the ERSP for each of the 200 frequency bins was calculated 100 times, resulting in a decomposition step size of about 29 ms. In order to limit the overlap between pre- and post-stimulus activity due to the windowing of the time-frequency analysis, the baseline window lasted from -1000–-100 ms (Shahin et al., 2008, 2009). All analyses in the current study are based on estimates of the total EEG power. In order to obtain the total (i.e. time- but not necessarily phase-locked) EEG power, the ERSP was computed for the single trial data and averaged afterwards (Tallon-Baudry & Bertrand, 1999).

Statistical differences between conditions were examined using non-parametric cluster-based permutation tests (Maris & Oostenveld, 2007). Firstly, it was tested whether there was a linear relationship between the amount of periodicity (aperiodic vs. mixed vs. periodic) or intelligibility (rotated vs. unintelligible periodic vs. intelligible periodic) and the total EEG power by computing separate two-sided regression *t*-tests for dependent samples at each electrode. In both cases, the whole stimulus window (0–2500 ms) and the complete array of analysed neural frequencies (1–30 Hz) was included in the test. Individual time-frequency sample points were considered to belong to a cluster if their *F*-values fell below the alpha level of .05, if the same was true for at least 3 neighbouring channels, and if they were connected to other significant sample points surrounding them. This procedure provides a weak control for false positive findings due to multiple comparisons by only allowing effects that are coherent in time, frequency, and space. Next, the individual *F*-values within a given cluster were summed to obtain the cluster-level statistic. The significance probability of a cluster was then assessed by comparing this cluster-level statistic to the one obtained after randomly re-allocating the individual trials to the conditions. This step was repeated 1000 times and the final cluster *p*-value was then determined by the proportion of these Monte Carlo iterations in which the cluster-level statistic was exceeded. The same statistical technique was also applied when two conditions were compared, but in this case two-sided t-tests were used in order to determine the *p*-values of the individual time-frequency sample points. In all statistical tests reported in this paper, an effect was considered to be significant if the cluster *p*-value was smaller than .05.

3. Results

3.1 Behavioural data

The averaged spoken behavioural responses obtained after each trial (Fig. 2) show that the aperiodic and mixed conditions are equally intelligible (88.8% and 90.0% correct keywords on average; $t(17) = -1.60$, $p = .13$), while the rotated condition is completely unintelligible (0%), and periodic speech is slightly less intelligible (77.4%) than aperiodic ($t(17) = -8.42$, $p < .001$) and mixed speech ($t(17) = -11.60$, $p < .001$). Furthermore, we compared the responses to the first and the second half of the trials in the periodic condition and found no significant differences (77.8% and 77.0%; $t(17) = .70$, $p = .49$), indicating that there were no learning effects over the course of the 160 trials.

Figure 2 about here

3.2 Periodicity

The total EEG power changes in response to the fully intelligible trials (all five keywords correctly repeated) in the aperiodic, mixed, and periodic conditions are shown in Fig. 3. The periodic condition was found to substantially deviate from the other two conditions, which had a very similar response pattern.

Firstly, as shown by the spectrogram plots in the upper part of Fig. 3, there was a power increase in the upper delta and theta region during the middle part of the stimulus window in the periodic condition. A cluster-based regression *t*-test including all three conditions confirmed that there was a linear positive relation between the amount of periodicity in the stimuli and the EEG power in this time-frequency region. This effect was most pronounced at electrode F7, but included 4 more electrodes in the left temporal scalp region (FT7, FC5, T7, and C5). These 5 electrodes all showed a consistent significant effect in time-frequency-electrode space from about 5–6.3 Hz and 845–1320 ms (p = .045). This time-frequency window is indicated by the black boxes in the spectrogram plots and the corresponding scalp distributions are shown in the lower part of Fig. 3. Subsequent pairwise comparisons revealed a significant difference between the aperiodic and periodic conditions in the same time-frequency-electrode region (p = .047), but no additional effects (see supplementary materials).

Figure 3 about here

Secondly, there was a trend for less low beta power (11–18 Hz) in the periodic condition throughout the stimulus window. This observation was confirmed by a post-hoc analysis in which the whole stimulus window was statistically examined, but the frequencies range was reduced to 10–20 Hz. As indicated by the uncorrected *p*-values, this trend was strongest over the central scalp region, particularly at electrode Cz, and present throughout the stimulus window. As can be told from both the spectrogram plots at electrode Cz and the corresponding scalp maps, there was again hardly any difference between the aperiodic and mixed conditions.

Figure 4 about here

3.3 Intelligibility

The total EEG power changes for completely unintelligible rotated speech, largely unintelligible periodic speech (trials with maximally two out of five keywords correct, henceforth referred to as the *unintelligible periodic* condition), and fully intelligible periodic speech (all five keywords correct, henceforth the *intelligible periodic* condition) are shown in Fig. 5. There was a general trend for greater power changes when the speech was more intelligible. In particular, the neural response in the rotated condition was very small, apart from an initial burst of activity following the acoustic onset of the sentences. Unintelligible and intelligible periodic speech, in contrast, showed sustained activity in the theta band (~4–7 Hz) throughout the stimulus window. Crucially, intelligible periodic speech also led to a substantial increase in delta power (1–4 Hz) during the second half of the stimulus window, which was absent in both the rotated and unintelligible periodic conditions.

A cluster-based regression *t*-test including all three conditions confirmed that there was a linear positive relation between the intelligibility of the stimuli and the EEG power in the delta and theta region during the second half of the stimulus window. This effect was most pronounced at electrode Fz, but included 7 more electrodes in the frontal scalp region (Fpz, Fp2, AF3, AFz, AF4, F3, and F1). These 8 electrodes all showed a more or less consistent significant effect in time-frequency-electrode space from about 2.4–5 Hz and 1400–2500 ms (p = .01). This time-frequency window is indicated by the black boxes in the spectrogram plots and the corresponding scalp distributions are shown in the lower part of Fig. 5.

Figure 5 about here

Subsequent pairwise comparisons showed that this cluster consisted of two overlapping smaller clusters, one in the theta band and another in the delta band (Fig. 6). Firstly, the direct comparison of the rotated and intelligible periodic conditions returned a cluster from about 3–4.9 Hz and 1100–2500 ms with a slightly left-lateralised frontal location (p = .01). This effect was strongest at electrode F3 and in total included 11 electrodes showing a more or less consistent significant difference in time-frequency-electrode space (Fp1, Fpz, Fp2, AF3, AFz, AF4, F3, F1, Fz, FC1, and FCz). Secondly, when comparing the unintelligible and intelligible periodic conditions directly, another cluster with a slightly right-lateralised frontal distribution was obtained. This cluster had a similar temporal extension, but did not overlap in frequency with the previous one (~1.7–2.7 Hz/1200–2300 ms; p = .019). Here, the strongest effect was observed at electrode F6 and in total 7 electrodes showed a consistent significant difference (Fpz, Fp2, AF4, Fz, F2, F4, and F6).

Figure 6 about here

3.4 Acoustic comparison of the unintelligible and intelligible periodic conditions

In order to test whether there were any substantial acoustic differences between the unintelligible and intelligible periodic conditions, we ran a number of additional acoustic analyses. Firstly, both the average duration of the sentences (means = 2.05/2.04 s, SDs = .22/.26, medians = 2.04/2.01, $t(173)$ = .34, p = .73) and the average F0 frequencies of the concatenated sentences in the two conditions (means = 119.33/119.01 Hz, SDs = 25.42/24.20, medians = 115.46/116.18, $t(173)$ = -.42, p = .68) were found to show little difference. The F0 frequencies were initially extracted with a sampling rate of 100 Hz, but down-sampled to .56 Hz for statistical testing to obtain the same degrees of freedom as for the comparison of sentence duration.

Secondly, we compared the power spectra of the two concatenated sets of sentences (computed using Welch's method, FFT size = 1024 samples, sampling rate = 22.05 kHz). The left panel of Fig. 7 shows that the spectra are virtually identical, which is underlined by a very high Pearson's correlation coefficient of the frequency bins closest to the centre frequencies of the eight vocoder bands ($r = .99$, $p < .001$). Lastly, we computed the average modulation spectra of the amplitude envelopes of all the sentence in the two conditions using the front-end of the mr-sEPSM speech intelligibility model (Jørgensen, et al., 2013). The modulation spectra were averaged over all 22 gammatone audio filters of the model, resulting in the simple line plot shown in the right panel of Fig. 7. The high correlation coefficient of the nine modulation filter centre frequencies across conditions ($r = .99$, $p < .001$) again confirms that there is little difference between the two conditions.

Figure 7 about here

4. Discussion

The present study sought to identify effects of acoustic periodicity and intelligibility in the EEG time-frequency responses to acoustically presented sentences. We thereby attempted to overcome the limitation that acoustic factors and intelligibility have not been examined independently in previous studies. Firstly, it was found that despite considerable acoustic differences, the total responses in the aperiodic and mixed conditions were almost identical. In contrast, the total EEG power in the periodic condition differed substantially from the other two conditions, even after controlling for the lower intelligibility. Differences were observed in the theta and low beta bands. Secondly, completely unintelligible rotated speech and largely unintelligible as well as intelligible periodic speech we compared. Here, we observed hardly any power changes in the rotated condition, apart from the acoustic onset response, but a substantial increase in delta power during the second half of the intelligible periodic sentences, when compared to their acoustically similar unintelligible counterparts.

4.1 Periodicity

The increase in theta power in the periodic condition agrees well with the results of Strauß, et al. (2014a), who found significantly more theta activity in a left-lateralised fronto-temporal network for ambiguous pseudo-words. Our results thus corroborate the idea that increased theta activity in this region is an indicator of response conflicts in auditory speech tasks. As suggested by Roux and Uhlhaas (2014), enhanced theta power in the context of speech tasks may indicate that verbal information is kept in the phonological loop, where the materials are sub vocally rehearsed. In line with this idea, the power decrease in the low beta range in the periodic condition seems to be a mu rhythm de-synchronisation, often observed before imagined or real movements (e.g. Cohen & Donner, 2013; Pfurtscheller et al., 1997; Wisniewski et al., 2015). It thus appears that both effects that distinguish the periodic speech from the other two conditions stand in relation to each other. Importantly, the participants correctly repeated every stimulus sentence in each of the three conditions included in the current analysis. Hence, the effect cannot be a result of motor preparation *per se*, but must be due to specific processes associated with the periodic condition.

Previous studies have reported that more intelligible words lead to a greater suppression of alpha activity (Becker, et al., 2013; Obleser & Weisz, 2012). In line with these findings, we have observed no differences in the alpha range between the aperiodic, mixed, and periodic conditions after controlling for differences in intelligibility. On the other hand, we did also not observe a suppression of alpha activity. The absence of this effect is in fact the only notable difference between the EEG spectrogram plot of the 8-channel aperiodic (i.e. noise-vocoded) condition in the study of Obleser and Weisz (2012) and the one in the current study. This might be due to the fact that we have used the relatively long and difficult IEEE sentences, and not single words. Although the alpha power level did not appear to decrease towards the onset of the sentences (see Fig. 3), it may have been lowered throughout the pre-stimulus window, indicating a state of 'anticipatory attention' preceding a demanding task (Klimesch, 2012). However, since we presented the conditions in blocks of ten sentences (i.e. whole IEEE sentence lists), subjects could also form expectancies regarding the upcoming stimulus, which may have caused the alpha power level to remain relatively stable between the baseline and stimulus windows.

In summary, the very similar EEG responses in the aperiodic and mixed conditions suggests the existence of a default response pattern to speech signals that are relatively easy to understand. A deviation from this pattern, as in the case of the periodic condition, in turn may indicate that a speech signal sounds unnatural and interferes with normal processing.

4.2 Intelligibility

Spectrally rotated speech was introduced to neuroscience in an attempt to provide an adequate non-speech analogue for intelligible speech (Rosen & Iverson, 2007; Scott, et al., 2000). However, despite the speech-like acoustic properties, we did not observe any substantial total or evoked activity in the rotated condition, apart from the initial acoustic onset response. The largest part of the signal thus resembled a recording of silence. This suggests that the attempt to mimic the acoustic properties of speech in an unintelligible control condition may in fact be needless, at least in M/EEG studies. In contrast, the neural response in the unintelligible periodic condition, which included trials with up to two correctly repeated keywords out of five, resembled its intelligible counterpart much more and additionally provides the benefit of being acoustically matched. In particular, only the two periodic conditions showed activity in the theta band throughout the stimulus window.

The main finding when comparing the total EEG power in the three conditions was the pronounced increase in delta power in the intelligible periodic condition. Given that there is no acoustic event that could have triggered this effect, it shows a surprisingly sharp onset at around 1000 ms after sentence onset and was present throughout the remainder of the stimulus window. As the average duration of the sentences was 2.04 s, the onset of this effect coincides with the beginning of the second half of the sentences. In line with the rule that the lower the frequency of a neural oscillation, the wider its distribution (e.g. Buzsáki & Draguhn, 2004), this effect was observed at the vast majority of electrodes, but only reached significance level in the frontal scalp region. Although the phase of delta oscillations has been shown to entrain to tone sequences (Lakatos, et al., 2005), and the detection of small loudness differences of tones has been reported to depend on delta power (Herrmann, et al., 2016), the power of delta oscillations has so far not been associated with speech intelligibility. Importantly, delta power increases towards the end of the stimulus window were observed in all three conditions in the analysis of periodicity (see Fig. 3), which demonstrates that this effect is not confined to the unnatural sounding periodic condition. Clearly, further research is needed to explore the exact relation between delta oscillations and speech intelligibility, particularly its time course. It is noteworthy that the average word duration of clearly articulated English is about 400 ms (i.e. 2.5 Hz; Hazan & Baker, 2011), and thus falls right into the middle of the delta band. However, since the increase in delta power was only observed during the second half of the intelligible sentences, it does not appear to be associated with the intelligibility of the individual words. Instead, one may speculate that this effect reflects the understanding of the meaning of the sentences as a whole.

Finally, we did again not observe any significant differences in the alpha band, although Fig. 5 shows that alpha power after sentence onset is slightly increased in the unintelligible periodic condition. This trend resembles the finding that alpha power is enhanced when speech signals are embedded in background noise, which was suggested to reflect the attempt to cope with demanding listening conditions (e.g. Strauss et al., 2014b; Wilsch et al., 2015). However, it appears that the target speech needs to be both difficult to understand, as is the case for the unintelligible periodic condition, and be presented in noise in order to lead to pronounced alpha power changes. Our data hence suggest that for speech presented in quiet, there is no strong association between alpha power and speech intelligibility, after controlling for acoustic differences.

5. Conclusion

By manipulating the amount of source periodicity in the materials, the present study has shown that total EEG power changes in response to speech do not reflect acoustic stimulus properties as such, but the perceptual effects of these properties. Even after controlling for differences in intelligibility, responses to fully periodic speech, an artificial condition that makes it difficult to identify the individual speech sounds, deviated markedly from the two other conditions with an entirely aperiodic or mixed source excitation. The neural responses in the latter two conditions, on the other hand, were very similar, despite their acoustic differences. In a second analysis, EEG power changes to unintelligible and intelligible speech were compared. Firstly, the very sparse neural response to spectrally rotated speech casts strong doubts on whether it is a suitable unintelligible control condition in M/EEG studies. Secondly, the direct comparison of the unintelligible and intelligible trials in the periodic condition revealed an increase in delta power during the second half of the sentences. The current results thus suggest that delta oscillations are a possible neural correlate of successful speech understanding.

Acknowledgements

This project has been funded with support from the European Commission under Contract FP7-PEOPLE-2011-290000. We thank Johannes Zaar and Torsten Dau for their advice and the code of the mr-sEPSM model.

References

Ardoint, M., Green, T., & Rosen, S. (2014). The intelligibility of interrupted speech depends upon its uninterrupted intelligibility. *J. Acoust. Soc. Am., 136*, EL275–EL280.

Becker, R., Pefkou, M., Michel, C. M., & Hervais-Adelman, A. G. (2013). Left temporal alpha-band activity reflects single word intelligibility. *Front. Syst. Neurosci., 7*, Article 121.

Blesser, B. (1972). Speech perception under conditions of spectral transformation: I. Phonetic characteristics. *J. Speech Hear. Res., 15*, 5–41.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. In (5.3.49 ed.).

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev., 108*, 624–652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci., 8*, 539–546.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304*, 1926–1929.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., & Lui, C. (1994). An international comparison of long-term average speech spectra. *J. Acoust. Soc. Am., 96*, 2108–2120.

Cohen, M. X., & Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *J. Neurophysiol., 110*, 2752–2763.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci., 23*, 3423–3431.

Dellwo, V., Fourcin, A., & Abberton, E. (2007). Rhythmical classification of languages based on voice parameters. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1129–1132). Saarbrücken, Germany.

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

Fourcin, A. (2010). A note on voice timing and the evolution of connected speech. *Logoped. Phoniatr. Vocology, 35*, 74–80.

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

Greenwood, D. D. (1990). A cochlear frequency-position function for several species – 29 years later. *J. Acoust. Soc. Am., 87*, 2592–2605.

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., Brugge, J. F., & Howard, M. A. (2010). Direct recordings of pitch responses from human auditory cortex. *Curr. Biol., 20*, 1128–1132.

Hanslmayr, S., Pastötter, B., Bäuml, K.-H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *J. Cognit. Neurosci., 20*, 215–225.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *J. Acoust. Soc. Am., 130*, 2139–2152.

He, B. J., & Raichle, M. E. (2009). The fMRI signal, slow cortical potential and consciousness. *Trends Cogn. Sci., 13*, 302–309.

Herrmann, B., Henry, M. J., Haegens, S., & Obleser, J. (2016). Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. *Neuroimage, 124*, 487–497.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci., 4*, Article 186.

Jørgensen, S., Ewert, S. D., & Dau, T. (2013). A multi-resolution envelope-power based model for speech intelligibility. *J. Acoust. Soc. Am., 134*, 436–446.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci., 16*, 606–617.

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. *J. Neurophysiol., 94*, 1904–1911.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol., 86*, 283–293.

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

Millman, R. E., Johnson, S. R., & Prendergast, G. (2015). The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. *J. Cognit. Neurosci.*, 533–545.

Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J. Neurosci., 33*, 19451–19469.

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic Details. *Cereb. Cortex, 22*, 2466–2477.

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci., 32*, 12376–12383.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*, 767–776.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex, 23*, 1378–1387.

Pfurtscheller, G., Stancak, A., & Edlinger, G. (1997). On the existence of different types of central beta rhythms below 30 Hz. *Electroencephalogr. Clin. Neurophysiol., 102*, 316–325.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J. Acoust. Soc. Am., 114*, 446–454.

Rosen, S., & Iverson, P. (2007). Constructing adequate non-speech analogues: what is special about speech anyway? *Developmental Science, 10*, 165–168.

Rothauser, E. H., Chapman, N. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., & Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans. on Audio and Electroacoustics, 17*, 225–246.

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct WM information? *Trends Cogn. Sci., 18*, 16–25.

Schoof, T., Green, T., Faulkner, A., & Rosen, S. (2013). Advantages from bilateral hearing in speech perception in noise with simulated cochlear implants and residual acoustic hearing. *J. Acoust. Soc. Am., 133*, 1017–1030.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*, 2400–2406.

Sedley, W., Teki, S., Kumar, S., Overath, T., Barnes, G. R., & Griffiths, T. D. (2012). Gamma band pitch responses in human auditory cortex measured with magnetoencephalography. *Neuroimage, 59*, 1904–1911.

Shahin, A. J., Picton, T. W., & Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain Cogn., 70*, 259–266.

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., & Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *Neuroimage, 41*, 113–122.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303–304.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci., 32*, 8443–8453.

Steinmetzger, K., & Rosen, S. (2015). The role of periodicity in perceiving speech in quiet and in background noise. *J. Acoust. Soc. Am., 138*, 3586–3599.

Strauß, A., Kotz, S. A., Scharinger, M., & Obleser, J. (2014a). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage, 97*, 387–395.

Strauß, A., Wöstmann, M., & Obleser, J. (2014b). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci., 8*, Article 350.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci., 3*, 151–162.

Weisz, N., & Obleser, J. (2014). Synchronisation signatures in the listening brain: a perspective from non-invasive neuroelectrophysiology. *Hearing Res., 307*, 16–28.

Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., & Obleser, J. (2015). Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. *Cereb. Cortex, 25*, 1938–1946.

Wisniewski, M. G., Thompson, E. R., Iyer, N., Estepp, J. R., Goder-Reiser, M. N., & Sullivan, S. C. (2015). Frontal midline θ power as an index of listening effort. *Neuroreport, 26*, 94–99.

Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *J. Neurosci., 35*, 1458–1467.

Xu, Y. (2013). ProsodyPro – A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (pp. 7–10). Aix-en-Provence, France.

Figure 1. Stimuli. Waveforms, wide-band spectrograms, and F0 contours for one example sentence (*Say it slowly but make it ring clear*.). A) The unprocessed version of the sentence and the same sentence processed to have B) an aperiodic source, C) a mixed source, D) a periodic source, or E) a mixed source and spectrally rotated. The four processed conditions (B–E) all have eight frequency bands, i.e. the same spectral resolution. The unprocessed version of the sentence in panel A) is shown for the purpose of comparison only.

(*single-column fitting image, colour online only*)

Figure 2. Behavioural data. Boxplots showing the average proportion of correctly repeated keywords in each of the four processing conditions. The black horizontal lines in the boxplots indicate the median value. *n.s.* stands for non-significant and *** indicates a *p*-value < .001.

(*single-column fitting image, colour online only*)

Figure 3. Periodicity. Total EEG power changes relative to baseline for the fully intelligible trials in the aperiodic, mixed, and periodic conditions. The upper part of the figure shows spectrograms of EEG activity recorded at electrode F7. In the panel on the far right, time-frequency sample points with cluster-corrected *p*-values < .05 are shown in red. The scalp distributions of the significant time-frequency window indicated by the black boxes (~5–6.3 Hz/845–1320 ms) are plotted in the lower part of the figure. Electrodes that are part of the significant cluster are shown as black dots and electrode F7, which showed the strongest effect, is indicated by a red dot.

(*two-column fitting image, colour online only*)

Figure 4. Periodicity post-hoc. Total EEG power changes relative to baseline for the fully intelligible trials in the aperiodic, mixed, and periodic conditions. The upper part of the figure

shows spectrograms of EEG activity recorded at electrode Cz. In the panel on the far right, time-frequency sample points with uncorrected p-values < .05 are shown in red. The scalp distributions of the time-frequency window in which significant differences were observed (11–18 Hz/0–2500 ms) are plotted in the lower part of the figure. Electrode Cz, which showed the strongest effect, is indicated by a red dot.

(*two-column fitting image, colour online only*)

Figure 5. Intelligibility. Total EEG power changes relative to baseline for the completely unintelligible rotated condition, largely unintelligible trials in the periodic condition (maximally two out of five correctly repeated keywords), and intelligible trials in the periodic condition (all five keywords correctly repeated). The upper part of the figure shows spectrograms of EEG activity recorded at electrode Fz. In the panel on the far right, time-frequency sample points with cluster-corrected p-values < .05 are shown in red. The scalp distributions of the significant time-frequency window indicated by the black boxes (~2.4–5 Hz/1400–2500 ms) are plotted in the lower part of the figure. Electrodes that are part of the significant cluster are shown as black dots and electrode Fz, which showed the strongest effect, is indicated by a red dot.

(*two-column fitting image, colour online only*)

Figure 6. Intelligibility pairwise comparisons. Pairwise statistical comparisons of the completely unintelligible rotated condition, largely unintelligible trials in the periodic condition (maximally two out of five correctly repeated keywords), and intelligible trials in the periodic condition (all five keywords correctly repeated). Time-frequency sample points with cluster-corrected p-values < .05 are shown in red. The scalp distributions of the two significant time-frequency windows indicated by the black boxes (~3–4.9 Hz/1100–2500 ms

and ~1.7–2.7 Hz/1200–2300 ms) are plotted in the lower part of the figure. Electrodes that are part of the significant clusters are shown as black dots and electrodes F3 and F6, which showed the strongest effects, are indicated by red dots.

(*two-column fitting image, colour online only*)

Figure 7. Acoustic characteristics of the largely unintelligible (maximally two out of five correctly repeated keywords) and intelligible (all five keywords correctly repeated) trials in the periodic condition. The left panel shows the average power spectra, i.e. the stimulus power plotted as a function of audio frequency, and the right panel the average envelope modulation spectra, i.e. the stimulus power plotted as a function of envelope modulation frequency.

(*single-column fitting image, colour online only*)

Figure 1

Click here to download Figure: 1. Stimuli.pdf

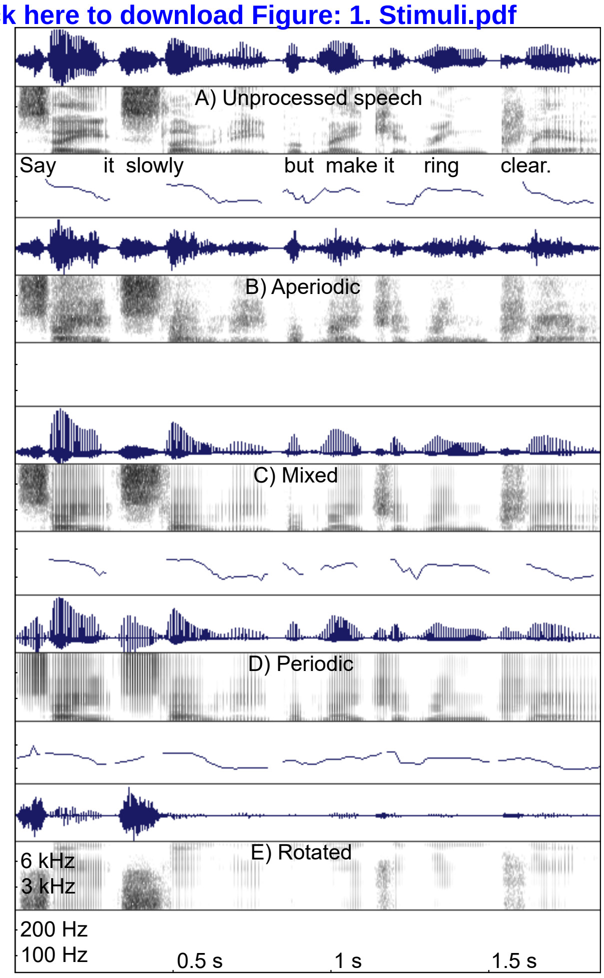


Figure 2

[Click here to download Figure: 2. Behavioural Data.pdf]

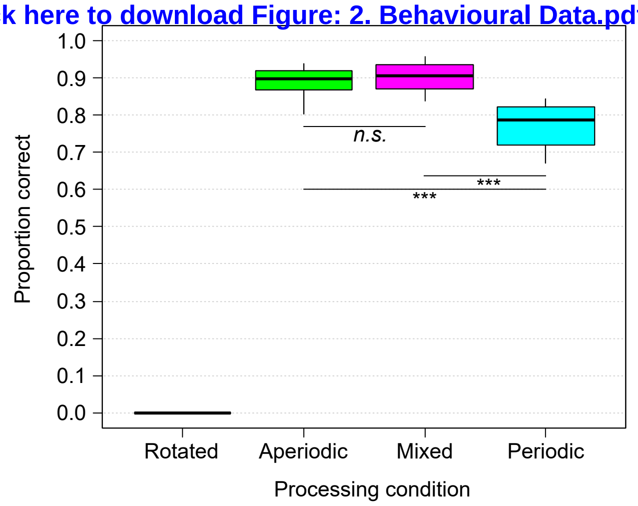


Figure 3

Click here to download Figure: 3. EEG Periodicity.pdf

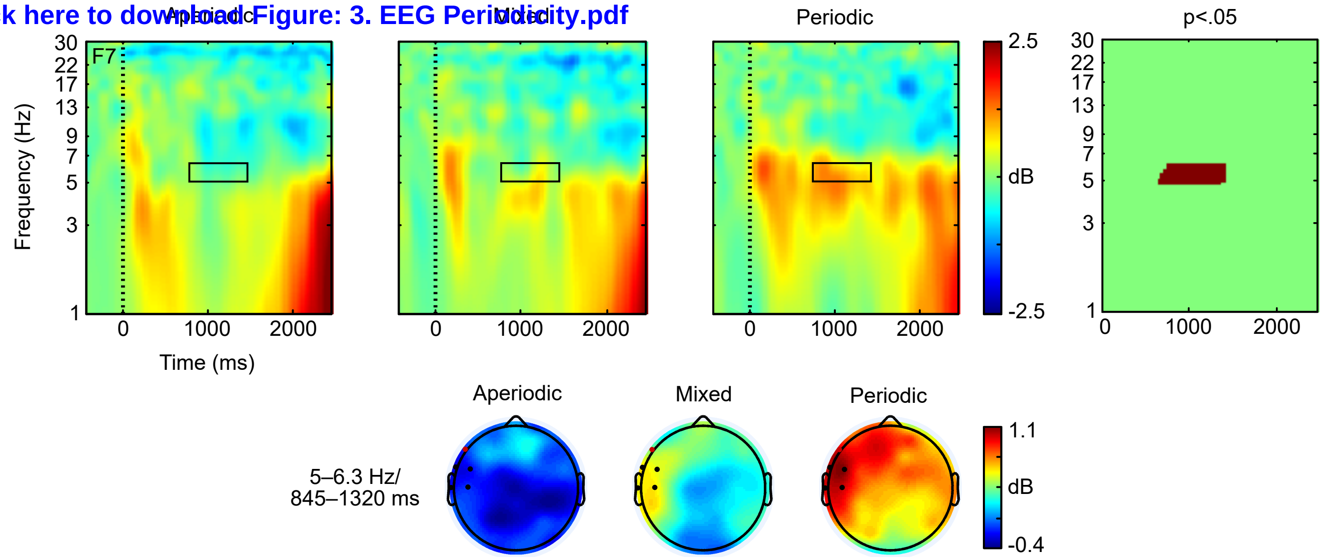


Figure 4

Click here to download Figure: 4. EEG Periodicity Post-hoc Beta.pdf

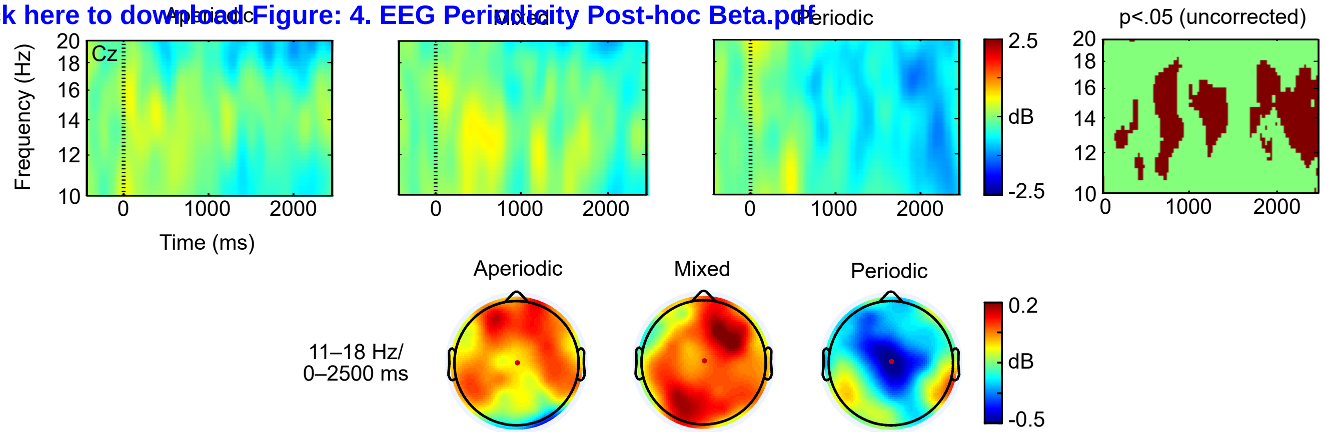


Figure 5

Click here to download Figure: 5. EEG Intelligibility.pdf

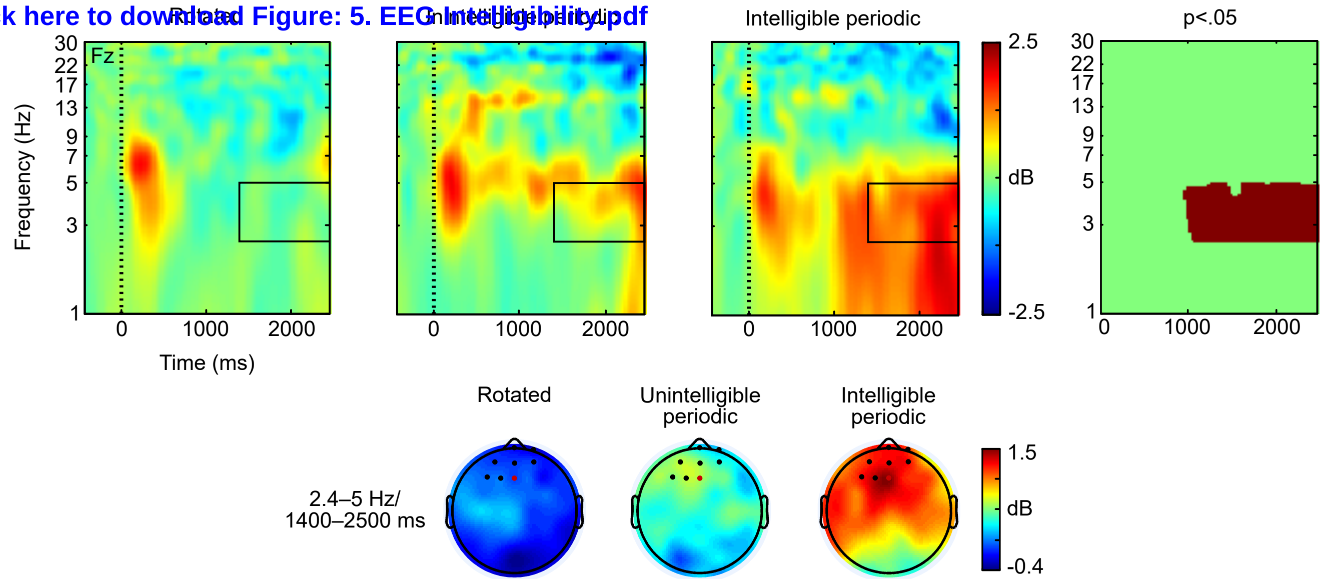
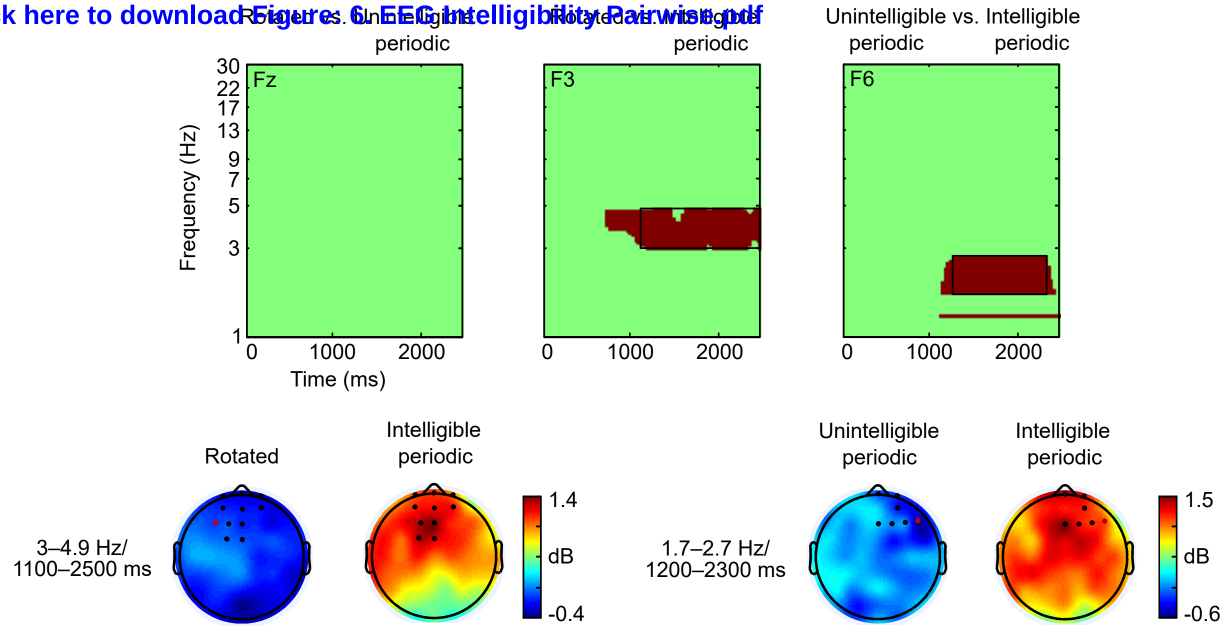


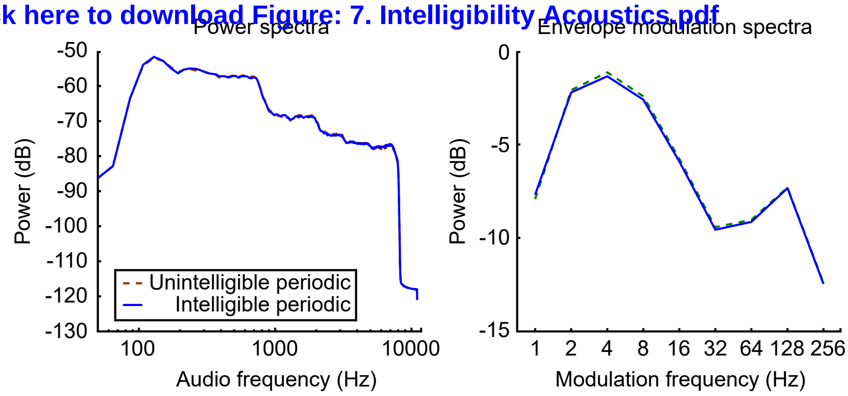
Figure 6



Rotated vs. Unintelligible periodic
Rotated vs. Intelligible periodic
Unintelligible periodic vs. Intelligible periodic
Fz
F3
F6
Frequency (Hz)
30
22
17
13
9
7
5
3
1
0
1000
2000
Time (ms)
Rotated
Intelligible periodic
Unintelligible periodic
Intelligible periodic
3–4.9 Hz/
1100–2500 ms
1.7–2.7 Hz/
1200–2300 ms
1.4
dB
-0.4
1.5
dB
-0.6

Figure 7

Click here to download Figure: 7. Intelligibility Acoustics.pdf



Supplementary Material

Click here to download Supplementary Material: Supp. 3b. EEG Periodicity Pairwise.pdf