

Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension

Jonathan E. Peelle^{1,4}, Joachim Gross^{2,3} and Matthew H. Davis¹

¹MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK and ²School of Psychology and ³Centre for Cognitive Neuroimaging (CCNi), University of Glasgow, Glasgow G12 8QB, UK

⁴Present address: Center for Cognitive Neuroscience and Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, USA

Address correspondence to Dr Matthew H. Davis, Cognition and Brain Sciences Unit, Medical Research Council, 15 Chaucer Road, Cambridge CB2 7EF, UK. Email: matt.davis@mrc-cbu.cam.ac.uk.

A growing body of evidence shows that ongoing oscillations in auditory cortex modulate their phase to match the rhythm of temporally regular acoustic stimuli, increasing sensitivity to relevant environmental cues and improving detection accuracy. In the current study, we test the hypothesis that nonsensory information provided by linguistic content enhances phase-locked responses to intelligible speech in the human brain. Sixteen adults listened to meaningful sentences while we recorded neural activity using magnetoencephalography. Stimuli were processed using a noise-vocoding technique to vary intelligibility while keeping the temporal acoustic envelope consistent. We show that the acoustic envelopes of sentences contain most power between 4 and 7 Hz and that it is in this frequency band that phase locking between neural activity and envelopes is strongest. Bilateral oscillatory neural activity phase-locked to unintelligible speech, but this cerebro-acoustic phase locking was enhanced when speech was intelligible. This enhanced phase locking was left lateralized and localized to left temporal cortex. Together, our results demonstrate that entrainment to connected speech does not only depend on acoustic characteristics, but is also affected by listeners' ability to extract linguistic information. This suggests a biological framework for speech comprehension in which acoustic and linguistic cues reciprocally aid in stimulus prediction.

Keywords: entrainment, intelligibility, prediction, rhythm, speech comprehension, MEG

Introduction

Oscillatory neural activity is ubiquitous, reflecting the shifting excitability of ensembles of neurons over time (Bishop 1932; Buzsáki and Draguhn 2004). An elegant and growing body of work has demonstrated that oscillations in auditory cortex entrain (phase-lock) to temporally-regular acoustic cues (Lakatos et al. 2005) and that these phase-locked responses are enhanced in the presence of congruent information in other sensory modalities (Lakatos et al. 2007). Synchronizing oscillatory activity with environmental cues provides a mechanism to increase sensitivity to relevant information, and thus aids in the efficiency of sensory processing (Lakatos et al. 2008; Schroeder et al. 2010). The integration of information across sensory modalities supports this process when multi-sensory cues are temporally correlated, as often happens with natural stimuli. In human speech comprehension, linguistic cues (e.g. syllables and words) occur in quasi-regular ordered sequences that parallel acoustic information. In the current study, we therefore test the hypothesis that nonsensory

information provided by linguistic content would enhance phase-locked responses to intelligible speech in human auditory cortex.

Spoken language is inherently temporal (Kotz and Schwartz 2010), and replete with low-frequency acoustic information. Acoustic and kinematic analyses of speech signals show that a dominant component of connected speech is found in slow amplitude modulations (approximately 4–7 Hz) that result from the rhythmic opening and closing of the jaw (MacNeilage 1998; Chandrasekaran et al. 2009), and which are associated with metrical stress and syllable structure in English (Cummins and Port 1998). This low-frequency envelope information helps to convey a number of important segmental and prosodic cues (Rosen 1992). Sensitivity to speech rate—which varies considerably both within and between talkers (Miller, Grosjean et al. 1984)—is also necessary to effectively interpret speech sounds, many of which show rate dependence (Miller, Aibel et al. 1984). It is not surprising, therefore, that accurate processing of low-frequency acoustic information plays a critical role in understanding speech (Drullman et al. 1994; Greenberg et al. 2003; Elliott and Theunissen 2009). However, the mechanisms by which the human auditory system accomplishes this are still unclear.

One promising explanation is that oscillations in human auditory and/or periauditory cortex entrain to speech rhythm. This hypothesis has received considerable support from previous human electrophysiological studies (Ahissar et al. 2001; Luo and Poeppel 2007; Kerlin et al. 2010; Lalor and Foxe 2010). Such phase locking of ongoing activity in auditory processing regions to acoustic information would increase listeners' sensitivity to relevant acoustic cues and aid in the efficiency of spoken language processing. A similar relationship between rhythmic acoustic information and oscillatory neural activity is also found in studies of nonhuman primates (Lakatos et al. 2005, 2007), and thus appears to be an evolutionarily conserved mechanism of sensory processing and attentional selection. What remains unclear is whether these phase-locked responses can be modulated by nonsensory information—in the case of speech comprehension, by the linguistic content available in the speech signal.

In the current study we investigate phase-locked cortical responses to slow amplitude modulations in trial-unique speech samples using magnetoencephalography (MEG). We focus on whether the phase locking of cortical responses benefits from linguistic information, or is solely a response to acoustic information in connected speech. We also use source localization

methods to address outstanding questions concerning the lateralization and neural source of these phase-locked responses. To separate linguistic and acoustic processes we use a noise-vocoding manipulation that progressively reduces the spectral detail present in the speech signal but faithfully preserves the slow amplitude fluctuations responsible for speech rhythm (Shannon et al. 1995). The intelligibility of noise-vocoded speech varies systematically with the amount of spectral detail present (i.e. the number of frequency channels used in the vocoding) and can thus be adjusted to achieve markedly different levels of intelligibility (Fig. 1A). Here, we test fully intelligible speech (16 channel), moderately intelligible speech (4 channel), and 2 unintelligible control conditions (4 channel rotated and 1 channel). Critically, the overall amplitude envelope—and hence the primary acoustic signature of speech rhythm—is preserved under all

conditions, even in vocoded speech that is entirely unintelligible (Fig. 1B). Thus, if neural responses depend solely on rhythmic acoustic cues, they should not differ across intelligibility conditions. However, if oscillatory activity benefits from linguistic information, phase-locked cortical activity should be enhanced when speech is intelligible.

Materials and Methods

Participants

Participants were 16 healthy right-handed native speakers of British English (aged 19–35 years, 8 female) with normal hearing and no history of neurological, psychiatric, or developmental disorders. All gave written informed consent under a process approved by the Cambridge Psychology Research Ethics Committee.

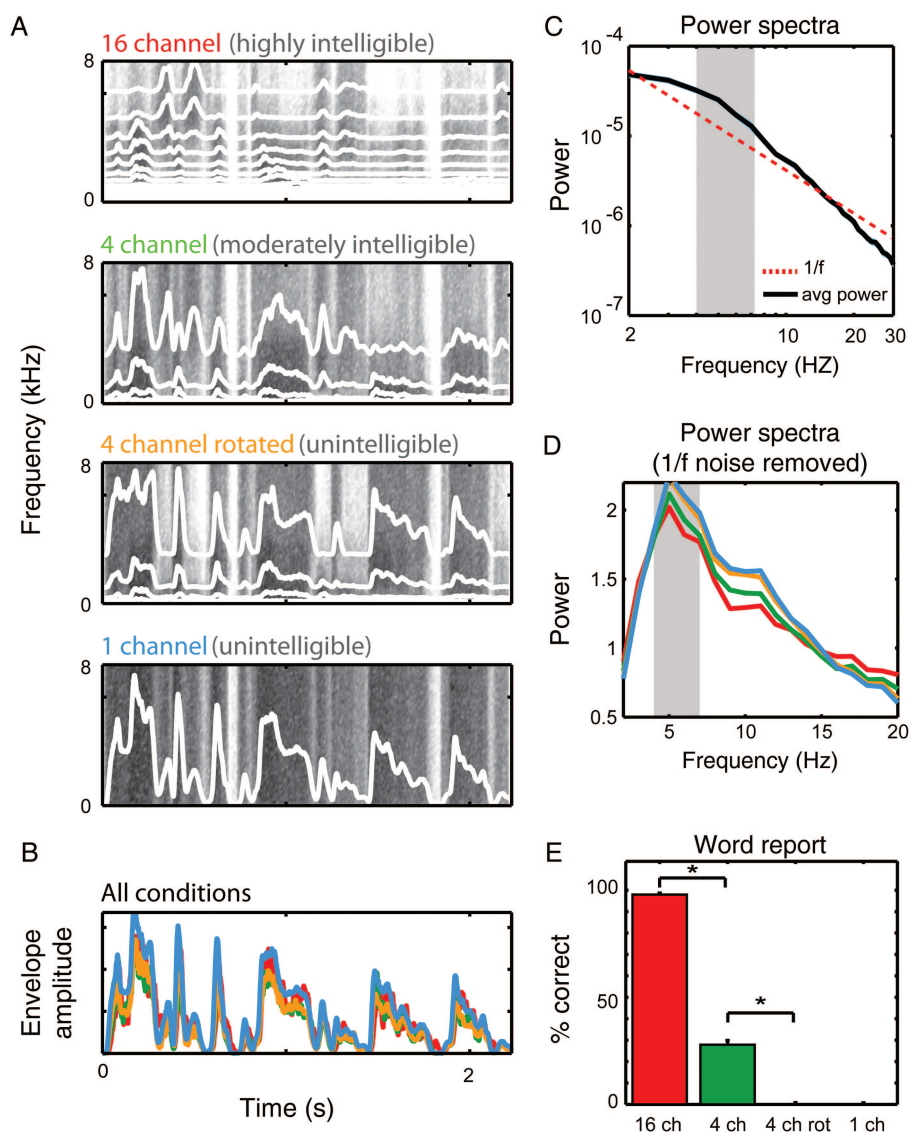


Figure 1. Stimulus characteristics. (A) Spectrograms of a single example sentence in the 4 speech conditions, with the amplitude envelope for each frequency band overlaid. Spectral change for the 16 channel sentence is absent from the 1 channel sentence. This spectral change is created by differences between the amplitude envelopes in multichannel vocoded speech. (B) Despite differences in spectral detail, the overall amplitude envelope contains only minor differences among the 4 conditions. (C) The modulation power spectrum of sentences in each condition shows $1/f$ noise as expected. Shading indicates 4–7 Hz where speech signals are expected to have increased power. (D) Residual modulation power spectra for each of the 4 speech conditions: after $1/f$ noise is subtracted highlights the peak in modulatory power between 4 and 7 Hz. (E) Word report accuracy for sentences presented in each of the 4 speech conditions. Error bars here and elsewhere reflect standard error of the mean with between-subject variability removed (Loftus and Masson 1994).

Materials

We used 200 meaningful sentences ranging in length from 5 to 17 words ($M=10.9$, $SD=2.2$) and in duration from 2.31 to 4.52 s ($M=2.96$, $SD=0.45$) taken from previous experiments (Davis and Johnsrude 2003; Rodd et al. 2005). All were recorded by a male native speaker of British English and digitized at 22 050 Hz. For each participant, each sentence occurred once in an intelligible condition (16 or 4 channel) and once in an unintelligible condition (4 channel rotated or 1 channel).

Noise vocoding was performed using custom Matlab scripts. The frequency range of 50–8000 Hz was divided into 1, 4, or 16 logarithmically spaced channels. For each channel, the amplitude envelope was extracted by full-wave rectifying the signal and applying a lowpass filter with a cutoff of 30 Hz. This envelope was then used to amplitude modulate white noise, which was filtered again before recombining the channels. In the case of the 1, 4, and 16 channel conditions, the output channel frequencies matched the input channel frequencies. In the case of 4 channel rotated speech, the output frequencies were inverted, effectively spectrally rotating the speech information (Scott et al. 2000). Because the selected number of vocoding channels followed a geometric progression, the frequency boundaries were common across conditions, and the corresponding envelopes were nearly equivalent (i.e. the sum of the lowest 4 channels in the 16 channel condition was equivalent to the lowest channel in the 4 channel condition) with only negligible differences due to filtering. Both the 1 channel and 4 channel rotated conditions are unintelligible but, because of their preserved rhythmic properties (and the experimental context), were likely perceived as speech or speech-like by listeners.

We focused our analysis on the low-frequency information in the speech signal based on prior studies and the knowledge that envelope information is critically important for comprehension of vocoded speech (Drullman et al. 1994; Shannon et al. 1995). We extracted the amplitude envelope for each stimulus, using full wave rectification and a lowpass filter at 30 Hz for use in the coherence analysis (Fig. 1B). This envelope served as the acoustic signal for all phase-locking analyses.

Procedure

Prior to the experiment, participants heard several example sentences in each condition, and were instructed to repeat back as many words as possible from each. They were informed that some sentences would be unintelligible and instructed that if they could not guess any of the words presented they should say “pass.” This word report task necessarily resulted in different patterns of motor output following the different intelligibility conditions, but was not expected to affect neural activity during perception. Each trial began with a short auditory tone and a delay of between 800 and 1800 ms before sentence presentation. Following each sentence, participants repeated back as many words as possible and pressed a key to indicate they were finished; they had as much time to respond as they needed. The time between this key press and the next trial was randomly varied between 1500 and 2500 ms. Data collection was broken into 5 blocks (i.e. periods of continuous data collection lasting approximately 10–12 min), with sentences randomly assigned across blocks. (For 5 participants, a programming error resulted in them not hearing any 4 channel rotated sentences, but these were replaced with additional 1 channel sentences. Analyses including the 4 channel rotated condition are performed on only 11 participants hearing this condition.) Stimuli were presented using E-Prime 1.0 software (Psychology Software Tools Inc., Pittsburgh, PA, USA), and participants' word recall was recorded for later analysis. Equipment malfunction resulted in loss of word report data for 5 of the participants, and thus word report scores are reported only for the participants who had behavioral data in all conditions.

MEG and Magnetic Resonance Imaging (MRI) Data Collection

MEG data were acquired with a high-density whole-scalp VectorView MEG system (Elekta-Neuromag, Helsinki, Finland), containing a magnetometer and 2 orthogonal planar gradiometers located at each of 102 positions (306 sensors total), housed in a light magnetically shielded room. Data were sampled at 1 kHz with a bandpass filter from 0.03 to

330 Hz. A 3D digitizer (Fastrak Polhemus Inc., Colchester, VA, USA) was used to record the positions of 4 head position indicator (HPI) coils and 50–100 additional points evenly distributed over the scalp, all relative to the nasion and left and right preauricular points. Head position was continuously monitored using the HPI coils, which allowed for movement compensation across the entire recording session. For each participant, structural MRI images with 1 mm isotropic voxels were obtained using a 3D magnetization-prepared rapid gradient echo sequence (repetition time = 2250 ms, echo time = 2.99 ms, flip angle = 9°, acceleration factor = 2) on a 3 T Tim Trio Siemens scanner (Siemens Medical Systems, Erlangen, Germany).

MEG Data Analysis

External noise was removed from the MEG data using the temporal extension of Signal-Space Separation (Taulu et al. 2005) implemented in MaxFilter 2.0 (Elekta-Neuromag). The MEG data were continuously compensated for head movement, and bad channels (identified via visual inspection or MaxFilter; ranging from 1 to 6 per participant) were replaced by interpolation. Subsequent analysis of oscillatory activity was performed using FieldTrip (Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands: <http://www.ru.nl/neuroimaging/fieldtrip/>). In order to quantify phase locking between the acoustic signal and neural oscillations we used coherence, a frequency-domain measure that reflects the degree to which the phase relationships of 2 signals are consistent across measurements, normalized to lie between 0 and 1. In the context of the current study, this indicates the consistency of phase locking of the acoustic and neural data across trials, which we refer to as cerebro-acoustic coherence. Importantly, coherence directly quantifies the synchronization of the acoustic envelope and neural oscillations, unlike previous studies that have looked at the consistency of neural response across trials without explicitly examining its relationship with the acoustic envelope (Luo and Poeppel 2007; Howard and Poeppel 2010; Kerlin et al. 2010; Luo et al. 2010).

The data were transformed from the time to frequency domain using a fast Fourier transform (FFT) applied to the whole trial for all MEG signals and acoustic envelopes using a Hanning window, producing spectra with a frequency resolution of approximately 0.3 Hz. The cross-spectral density was computed for all combinations of MEG channels and acoustic signals. We then extracted the mean cross-spectral density of all sensor combinations in the selected frequency band. We used dynamic imaging of coherent sources (DICS) (Gross et al. 2001) to determine the spatial distribution of brain areas coherent to the speech envelope. This avoids making the inaccurate assumption that specific sensors correspond across individuals despite different head shapes and orientations, although results must be interpreted within the limitations of MEG source localization accuracy. It also allows data to be combined over recordings from magnetometer and gradiometer sensors. DICS is based on a linearly constrained minimal variance beamformer (Van Veen et al. 1997) in the frequency domain and allows us to compute coherence between neural activity at each voxel and the acoustic envelope. The beamformer is characterized by a set of coefficients that are the solutions to a constrained minimization problem, ensuring that the beamformer passes activity from a given voxel while maximally suppressing activity from all other brain areas. Coefficients are computed from the cross-spectral density and the solution to the forward problem for each voxel. The solution to the forward problem was based on the single shell model (Nolte 2003). This dominant orientation was computed for each voxel from the first eigenvector of the cross-spectral density matrix between both tangential orientations. The resulting beamformer coefficients were used to compute coherence between acoustic and cortical signals in a large number of voxels covering the entire brain.

Computations were performed separately for 4, 5, 6, and 7 Hz and then averaged before performing group statistics. For each participant, we also conducted coherence analyses on 100 random pairings of acoustic and cerebral data, which we averaged to produce random coherence images. The resulting tomographic maps were spatially normalized to Montreal Neurological Institute (MNI) space, resampled to 4 mm isotropic voxels, and averaged across 4–7 Hz. Voxel-based group analyses were performed using 1-sample *t*-tests and region of

interest (ROI) analyses in SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). Results are displayed using MNI-space templates included with SPM8 and MRICron (Rorden and Brett 2000).

Results

Acoustic Properties of the Sentences

To characterize the acoustic properties of the stimuli, we performed a frequency analysis of all sentence envelopes using a multitaper FFT with Slepian tapers. The spectral power for all sentence envelopes averaged across condition is shown in Figure 1C, along with a $1/f$ line to indicate the expected noise profile (Voss and Clarke 1975). The shaded region indicates the range between 4 and 7 Hz, where we anticipated maximal power in the speech signal. The residual power spectra after removing the $1/f$ trend using linear regression are shown in Figure 1D. This shows a clear peak in the 4–7 Hz range (shaded) that is consistent across condition. These findings, along with previous studies, motivated our focus on cerebro-acoustic coherence between 4 and 7 Hz, which is well matched over all 4 forms of noise vocoding.

Behavioral Results

To confirm that the intelligibility manipulations worked as intended, we analyzed participants' word report data, shown in Figure 1E. As expected, the 1 channel ($M=0.1\%$, $SD=0.1\%$, range = 0.0–0.6%) and 4 channel rotated ($M=0.2\%$, $SD=0.1\%$, range = 0.0–0.4%) conditions were unintelligible with essentially zero word report. Accuracy for these unintelligible conditions did not differ from each other ($P=0.38$), assessed by a nonparametric sign test. The word report for the 16 channel condition was near ceiling ($M=97.9\%$, $SD=1.5\%$, range = 94.4–99.6%) and significantly greater than that for the 4 channel condition ($M=27.9\%$, $SD=8.2\%$, range = 19.1–41.6%) [$t(8)=26.00$, $P<0.001$ (nonparametric sign test $P<0.005$)]. The word report in the 4 channel condition was significantly better than that in the 4 channel rotated condition [$t(8)=10.26$, $P<0.001$ (nonparametric sign test $P<0.005$)]. Thus, connected speech remains intelligible if it is presented with sufficient spectral detail in appropriate frequency ranges (i.e. a multichannel, nonrotated vocoder). These behavioral results also suggest that the largest difference in phase locking will be seen between the fully intelligible 16 channel condition and 1 of the unintelligible control conditions. Because the 4 channel and 4 channel rotated conditions are the most closely matched acoustically but differ in intelligibility, these behavioral results suggest 2 complementary predictions: first, coherence is greater in the 16 channel condition than in the 1 channel condition; secondly, coherence is greater in the 4 channel condition than in the 4 channel rotated condition.

Cerebro-Acoustic Coherence

We first analyzed MEG data in sensor space to examine cerebro-acoustic coherence across a range of frequencies. For each participant, we selected the magnetometer with the highest summed coherence values between 0 and 20 Hz. For that sensor, we then plotted coherence as a function of frequency, as shown in Figure 2A for 2 example participants. For each participant, we also conducted a nonparametric permutation analysis in which we calculated coherence for 5000 random pairings of acoustic envelopes with neural data;

based on the distribution of values obtained through these random pairings, we were able to determine the chance of obtaining coherence values for the true pairing. In both the example participants, we see a coherence peak between 4 and 7 Hz that exceeds the $P<0.005$ threshold based on this permutation analysis. For these 2 participants, greatest coherence in this frequency range is seen in bilateral frontocentral sensors (Fig. 2A). The maximum-magnetometer coherence plot averaged across all 16 participants, shown in Figure 2B, also shows a clear peak between 4 and 7 Hz. This is consistent with both the acoustic characteristics of the stimuli and the previous literature, and therefore supports our decision to focus on this frequency range for further analyses.

We next conducted a whole-brain analysis on source-localized data to see whether the unintelligible 1 channel condition showed significantly greater coherence between the neural and acoustic data than that seen in random pairings of acoustic envelopes and neural data. These results are shown in Figure 3A using a voxel-wise threshold of $P<0.001$ and a $P<0.05$ whole-brain cluster extent correction for multiple comparisons using random field theory (Worsley et al. 1992). This analysis revealed a number of regions that show significant phase locking to the acoustic envelope in the absence of linguistic information, including bilateral superior and middle temporal gyri, inferior frontal gyri, and motor cortex.

Previous electrophysiological studies in nonhuman primates have focused on phase locking to rhythmic stimuli in primary auditory cortex. In humans, primary auditory cortex is the first cortical region in a hierarchical speech-processing network (Rauschecker and Scott 2009), and is thus a sensible place to look for neural responses that are phase locked to acoustic input. To assess the existence and laterality of cerebro-acoustic coherence in primary auditory cortex, we used the SPM Anatomy toolbox (Eickhoff et al. 2005) to delineate bilateral auditory cortex ROIs, which comprised regions TE1.0, TE1.1, and TE1.2 (Morosan et al. 2001): regions were identified using maximum probability maps derived from cytoarchitectonic analysis of postmortem samples. We extracted coherence values from these ROIs for the actual and random pairings of acoustic and neural data for both 16 channel and 1 channel stimuli, shown in Figure 3B. Given the limited accuracy of MEG source localization, and the smoothness of the source estimates, measures of phase locking considered in this analysis may also originate from surrounding regions of superior temporal gyrus (e.g. auditory belt or parabelt). However, by using this pair of anatomical ROIs, we can ensure that the lateralization of auditory oscillations is assessed in an unbiased fashion. We submitted the extracted data to a 3-way hemisphere (left/right) \times number of channels (16/1) \times pairing (normal/random) repeated-measures analysis of variance (ANOVA). This analysis showed no main effect of hemisphere ($F_{1,15}<1$, n.s.), but a main effect of the number of channels ($F_{1,15}=6.4$, $P<0.05$) and pairing ($F_{1,15}=24.7$, $P<0.001$). These results reflect greater coherence for the 16 channel speech than for the 1 channel speech and greater coherence for the true pairing than for the random pairing. Most relevant for the current investigation was the significant 3-way hemisphere \times number of channels \times pairing interaction ($F_{1,15}=4.5$, $P<0.001$), indicating that the phase-locked response was enhanced in the left auditory cortex during the more intelligible 16 channel condition (number of channels \times pairing interaction: $F_{1,15}=10.53$, $P=0.005$), but not in the right

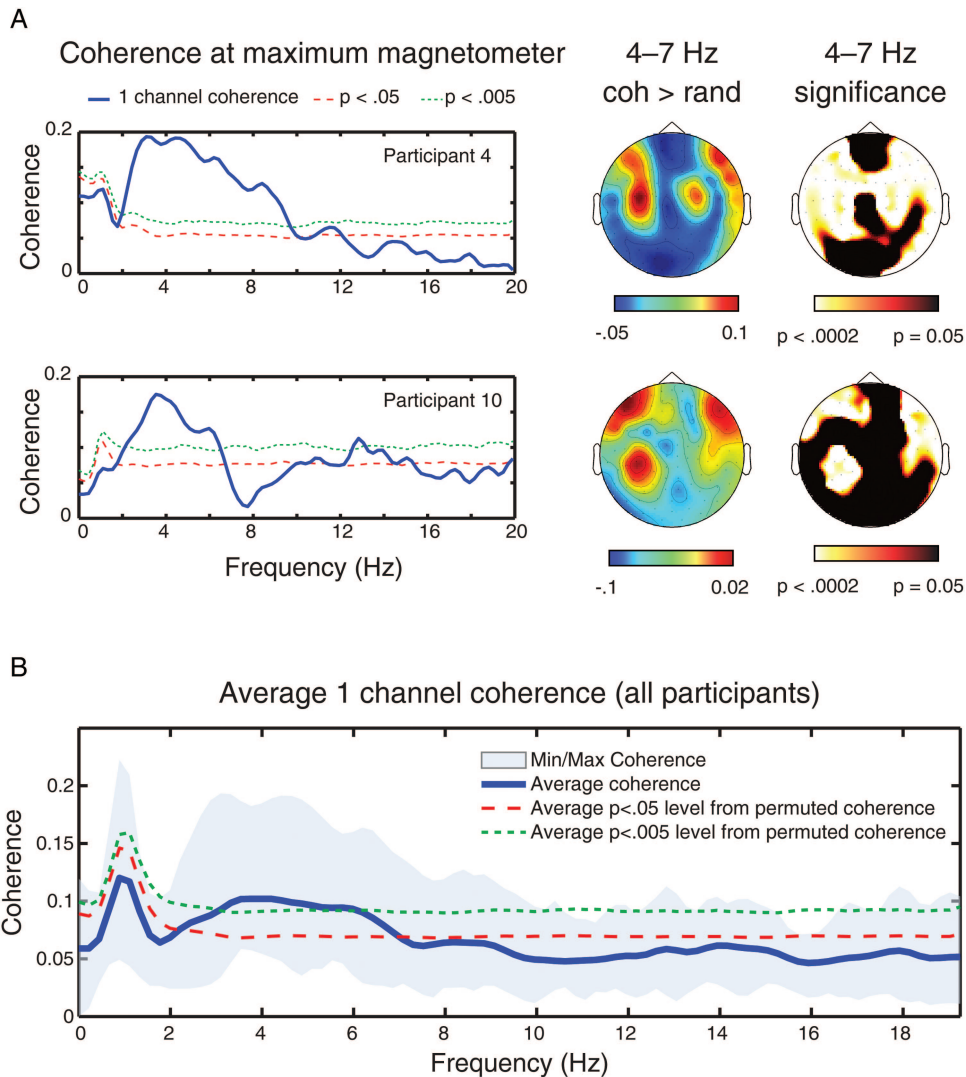


Figure 2. Sensor level cerebro-acoustic coherence for magnetometer sensors. (A) For 2 example participants, the magnetometer with the maximum coherence values (across all frequencies) was selected. Coherence values were then plotted at this sensor as a function of frequency, along with significance levels based on permutation analyses (see text). Topographic plots of coherence values for all magnetometers, as well as a topographic plot showing significance values, are also displayed. (B) Coherence values as a function of frequency computed as above, but averaged for the maximum coherence magnetometer in all 16 listeners. Minimum and maximum values across subjects are also shown in the shaded portion. Coherence values show a clear peak in the 4–7 Hz range.

auditory cortex (number of channels \times pairing interaction: $F_{1,15} < 1$, n.s.). This confirms that cerebro-acoustic coherence in left auditory cortex, but not in right auditory cortex, is significantly increased for intelligible speech.

To assess effects of intelligibility on cerebro-acoustic coherence more broadly we conducted a whole-brain search for regions in which coherence was higher for the intelligible 16 channel speech than for the unintelligible 1 channel speech, using a voxel-wise threshold of $P < 0.001$, corrected for multiple comparisons ($P < 0.05$) using cluster extent. As shown in Figure 4A, this analysis revealed a significant cluster of greater coherence centered on the left middle temporal gyrus [13 824 μ L: peak at $(-60, -16, -8)$, $Z = 4.11$], extending into both inferior and superior temporal gyri. A second cluster extended from the medial to the lateral surface of left ventral inferior frontal cortex [17 920 μ L: peak at $(-8, 40, -20)$, $Z = 3.56$]. A third cluster was also observed in the left inferior frontal gyrus [1344 μ L: peak at $(-60, 36, -16)$, $Z = 3.28$], although

this was too small to pass whole-brain cluster extent correction (and thus not shown in Fig. 4). (We conducted an additional analysis in which the source reconstructions were calculated on a single frequency range of 4–7 Hz, as opposed to averaging separate source localizations, as described in Materials and Methods. This analysis resulted in the same 2 significant clusters of increased coherence in nearly identical locations.)

We conducted ROI analyses to assess which of these areas respond differentially to 4 channel vocoded sentences that are moderately intelligible or made unintelligible by spectral rotation. This comparison is of special interest because these 2 conditions are matched for spectral complexity (i.e. contain the same number of frequency bands), but differ markedly in intelligibility. We extracted coherence values for each condition from a sphere (5 mm radius) centered on the middle temporal gyrus peak identified in the 16 channel $>$ 1 channel comparison, shown in Figure 4B. In addition to the expected

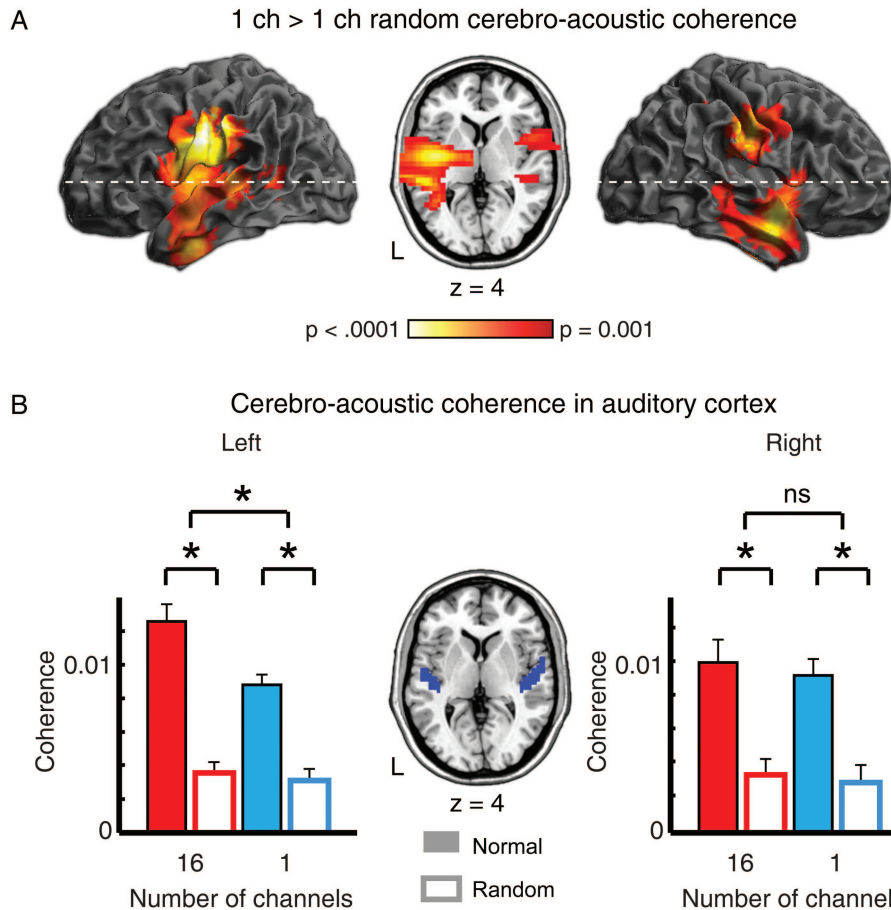


Figure 3. Source-localized cerebro-acoustic coherence results. (A) Source localization showing significant cerebro-acoustic coherence in the unintelligible 1 channel condition compared to a permutation-derived null baseline derived from random pairings of acoustic envelopes to MEG data across all participants. Effects shown are whole-brain corrected ($P < 0.05$). (B) ROI analysis on coherence values extracted from probabilistically defined primary auditory cortex regions relative to coherence for random pairings of acoustic and cerebral trials. Data showed a significant hemisphere \times number of channels \times normal/random interaction ($P < 0.001$).

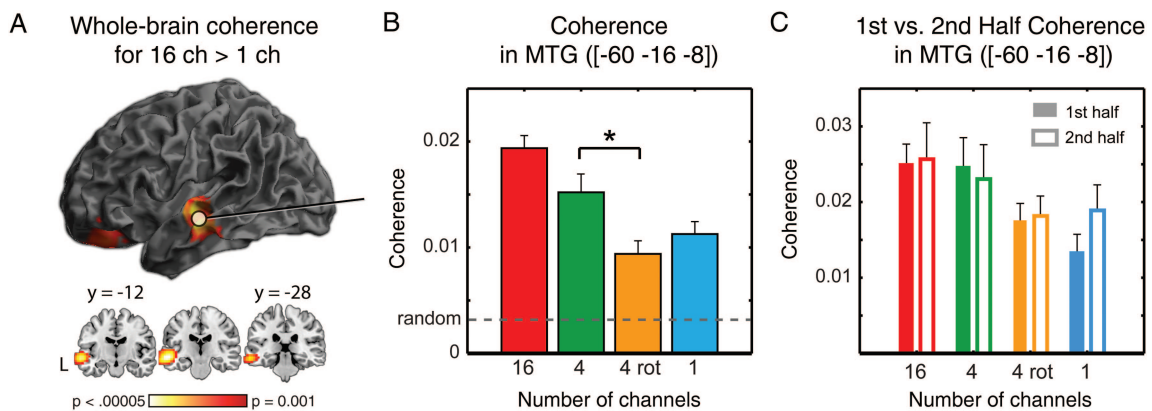


Figure 4. Linguistic influences on cerebro-acoustic coherence. (A) Group analysis showing neural sources in which intelligible 16 channel vocoded speech led to significantly greater coherence with the acoustic envelope than the 1 channel vocoded speech. Effects shown are whole-brain corrected ($P < 0.05$). Coronal slices shown from an MNI standard brain at 8 mm intervals. (B) For a 5 mm radius sphere around the middle temporal gyrus peak ($-60, -16, -8$), the 4 channel vocoded speech also showed significantly greater coherence than the 4 channel rotated vocoded speech, despite being equated for spectral detail. (C) Analysis of the first and second halves of each sentence confirms that results were not driven by sentence onset effects: there was no main effect of sentence half nor an interaction with condition.

difference between 16 and 1 channel sentences [$t(10) = 3.8$, $P < 0.005$ (one-sided)], we found increased coherence for moderately intelligible 4 channel speech compared with unintelligible 4 channel rotated speech [$t(10) = 2.1$, $P < 0.05$]. We also

conducted an exploratory whole-brain analysis to identify any additional regions in which coherence was higher for the 4 channel condition than for the 4 channel rotated condition; however, no regions reached whole-brain significance.

We next investigated whether coherence varied within a condition as a function of intelligibility, as indexed by word report scores. Coherence values for the 4 channel condition, which showed the most behavioral variability, were not correlated with single-subject word report scores across participants or with differences between high- and low-intelligibility sentences within each participant. Similar comparisons of coherence in an ROI centered on the peak of the significant frontal cluster for 4 channel and 4 channel rotated speech and between-subject correlations were nonsignificant (all $P_s > 0.53$). An exploratory whole-brain analysis also failed to reveal any regions in which coherence was significantly correlated with word report scores.

Finally, we conducted an additional analysis to verify that coherence in the middle temporal gyrus was not driven by differential responses to the acoustic onset of intelligible sentences. We therefore performed the same coherence analysis as before on the first and second halves of each sentence separately, as shown in Figure 4C. If acoustic onset responses were responsible for our coherence results, we would expect coherence to be higher at the beginning than at the end of the sentence. We submitted the data from the middle temporal gyrus ROI to a condition \times first/second half repeated-measures ANOVA. There was no effect of half ($F_{10,30} < 1$) nor an interaction between condition and half ($F_{10,30} < 1$). Thus, we conclude that the effects of speech intelligibility on cerebro-acoustic coherence in the left middle temporal gyrus are equally present throughout the duration of a sentence.

Discussion

Entraining to rhythmic environmental cues is a fundamental ability of sensory systems in the brain. This oscillatory tracking of ongoing physical signals aids temporal prediction of future events and facilitates efficient processing of rapid sensory input by modulating baseline neural excitability (Arieli et al. 1996; Busch et al. 2009; Romei et al. 2010). In humans, rhythmic entrainment is also evident in the perception and social coordination of movement, music, and speech (Gross et al. 2002; Peelle and Wingfield 2005; Shockey et al. 2007; Cummins 2009; Grahn and Rowe 2009). Here, we show that cortical oscillations become more closely phase locked to slow fluctuations in the speech signal when linguistic information is available. This is consistent with our hypothesis that rhythmic entrainment relies on the integration of multiple sources of knowledge, and not just sensory cues.

There is growing consensus concerning the network of brain regions that support the comprehension of connected speech, which minimally include bilateral superior temporal cortex, more extensive left superior and middle temporal gyri, and left inferior frontal cortex (Bates et al. 2003; Davis and Johnsrude 2003, 2007; Scott and Johnsrude 2003; Peelle et al. 2010). Despite agreement on the localization of the brain regions involved, far less is known about their function. Our current results demonstrate that a portion of left temporal cortex, commonly identified in positron emission tomography (PET) and functional MRI (fMRI) studies of spoken language (Davis and Johnsrude 2003; Scott et al. 2006; Davis et al. 2007; Friederici et al. 2010; Rodd et al. 2010), shows increased phase locking with the speech signal when speech is intelligible. These findings suggest that the distributed speech comprehension network expresses predictions that aid the

processing of incoming acoustic information by enhancing phase-locked activity. Extraction of the linguistic content generates expectations for upcoming speech rhythm through prediction of specific lexical items (DeLong et al. 2005) or by anticipating clause boundaries (Grosjean 1983), as well as other prosodic elements that have rhythmic correlates apparent in the amplitude envelope (Rosen 1992). Thus, speech intelligibility is enhanced by rhythmic knowledge, which in turn provides the linguistic information necessary for the reciprocal prediction of upcoming acoustic signals. We propose that this positive feedback cycle is neurally instantiated by cerebro-acoustic phase locking.

We note that the effects of intelligibility on phase-locked responses are seen in relatively low-level auditory regions of temporal cortex. Although this finding must be interpreted within the limits of MEG source localization, it is consistent with electrophysiological studies in nonhuman primates in which source localization is straightforward (Lakatos et al. 2005, 2007), as well as with interpretations of previous electrophysiological studies in humans (Luo and Poeppel 2007; Luo et al. 2010). The sensitivity of phase locking in auditory areas to speech intelligibility suggests that regions that are anatomically early in the hierarchy of speech processing show sensitivity to linguistic information. One interpretation of this finding is that primary auditory regions—either in primary auditory cortex proper, or in neighboring regions that are synchronously active—are directly sensitive to linguistic content in intelligible speech. However, there is consensus that during speech comprehension, these early auditory regions do not function in isolation, but as part of an anatomical-functional hierarchy (Davis and Johnsrude 2003; Scott and Johnsrude 2003; Hickok and Poeppel 2007; Rauschecker and Scott 2009; Peelle et al. 2010). In the context of such a hierarchical model of speech comprehension, a more plausible explanation is that increased phase locking of oscillations in auditory cortex to intelligible speech reflects the numerous efferent auditory connections that provide input to auditory cortex from secondary auditory areas and beyond (Hackett et al. 1999, 2007; de la Mothe et al. 2006). The latter interpretation is also consistent with proposals of top-down or predictive influences of higher-level content on low-level acoustic processes that contribute to the comprehension of spoken language (Davis and Johnsrude 2007; Gagnepain et al. 2012; Wild et al. 2012).

An important aspect of the current study is that we manipulated intelligibility by varying the number and spectral ordering of channels in vocoded speech. Increasing the number of channels increases the complexity of the spectral information in speech, but does not change its overall amplitude envelope. Greater spectral detail—which aids intelligibility—is created by having different amplitude envelopes in different frequency bands. That is, in the case of 1 channel vocoded speech, there is a single amplitude envelope applied across all frequency bands and therefore no conflicting information; in the case of 16 channel vocoded speech, there are 16 non-identical amplitude envelopes, each presented in a narrow spectral band. If coherence is driven solely by acoustic fluctuations, then we might expect that presentation of a mixture of different amplitude envelopes would reduce cerebro-acoustic coherence. Conversely, if rhythmic entrainment reflects neural processes that track intelligible speech signals, we would expect the reverse, namely increased coherence for speech

signals with multiple envelopes. The latter result is precisely what we observed.

In noise-vocoded speech, using more channels results in greater spectral detail and concomitant increases in intelligibility. One might thus argue that the observed increases in cerebro-acoustic coherence in the intelligible 16 channel condition were not due to the availability of linguistic information, but to the different spectral profiles associated with these stimuli. However, this confound is not present in the 4 channel and 4 channel rotated conditions, which differ in intelligibility but are well matched for spectral complexity. Our comparison of responses with 4 channel and spectrally rotated 4 channel vocoded sentences thus demonstrates that it is intelligibility, rather than dynamic spectral change created by multiple amplitude envelopes (Roberts et al. 2011), that is critical for enhancing cerebro-acoustic coherence. Our results show significantly increased cerebro-acoustic coherence for the more-intelligible, nonrotated 4 channel sentences in the left temporal cortex. Again, this anatomical locus is in agreement with PET and fMRI studies comparing similar stimuli (Scott et al. 2000; Obleser et al. 2007; Okada et al. 2010).

We note with interest that both our oscillatory responses and fMRI responses to intelligible sentences are largely left lateralized. In our study, both left and right auditory cortices show above-chance coherence with the amplitude envelope of vocoded speech, but it is only in the left hemisphere that coherence is enhanced for intelligible speech conditions. This finding stands in contrast to previous observations of right lateralized oscillatory responses in similar frequency ranges shown with electroencephalography and fMRI during rest (Giraud et al. 2007) or in fMRI responses to nonspeech sounds (Boemio et al. 2005). Our findings, therefore, challenge the proposal that neural lateralization for speech processing is due solely to asymmetric temporal sampling of acoustic features (Poeppel 2003). Instead, we support the view that it is the presence of linguistic content, rather than specific acoustic features, that is critical in changing the lateralization of observed neural responses (Rosen et al. 2011; McGettigan et al. 2012). Some of these apparently contradictory previous findings may be explained by the fact that the salience and influence of linguistic content are markedly different during full attention to trial-unique sentences—as is the case in both the current study and natural speech comprehension—than in listening situations in which a limited set of sentences is repeated often (Luo and Poeppel 2007) or unattended (Abrams et al. 2008).

The lack of a correlation between behavioral word report and coherence across participants in the 4 channel condition is slightly puzzling. However, we note that there was only a range of approximately 20% accuracy across all participants' word report scores. Our prediction is that if we were to use a slightly more intelligible manipulation (e.g. 6 or 8 channel vocoding) or other conditions that produce a broader range of behavioral scores, such a correlation would indeed be apparent. Further research along these lines would be valuable in testing for more direct links between intelligibility and phase locking (cf. Ahissar et al. 2001).

Other studies have shown time-locked neural responses to auditory stimuli at multiple levels of the human auditory system, including auditory brainstem responses (Skoe and Kraus 2010) and auditory steady-state responses in cortex (Picton et al. 2003). These findings reflect replicable neural

responses to predictable acoustic stimuli that have high temporal resolution and (for the auditory steady-state response) are extended in time. To date, there has been no convincing evidence that cortical phase-locked activity in response to connected speech reflects anything more than an acoustic-following response for more complex stimuli. For example, Howard and Poeppel (2010) conclude that cortical phase locking to speech is based on acoustic information because theta-phase responses can discriminate both normal and temporally reversed sentences with equal accuracy, despite the latter being incomprehensible. Our current results similarly confirm that neural oscillations can entrain to unintelligible stimuli and would therefore discriminate different temporal acoustic profiles, irrespective of linguistic content. However, the fact that these entrained responses are significantly enhanced when linguistic information is available indicates that it is not solely acoustic factors that drive phase locking during natural speech comprehension.

Although we contend that phase locking of neural oscillations to sensory information can increase the efficiency of perception, rhythmic entrainment is clearly not a prerequisite for successful perceptual processing. Intelligibility depends on the ability to extract linguistic content from speech: this is more difficult, but not impossible, when rhythm is perturbed. For example, in everyday life we may encounter foreign-accented or dysarthric speakers that produce disrupted speech rhythms but are nonetheless intelligible with additional listener effort (Tajima et al. 1997; Liss et al. 2009). Similarly, short fragments of connected speech presented in the absence of a rhythmic context (including single monosyllabic words) are often significantly less intelligible than connected speech, but can still be correctly perceived (Pickett and Pollack 1963). Indeed, from a broader perspective, organisms are perfectly capable of processing stimuli that do not occur as part of a rhythmic pattern. Thus, although adaptive and often present in natural language processing, rhythmic structure and cerebro-acoustic coupling are not necessary for successful speech comprehension.

Previous research has focussed on the integration of multisensory cues in “unisensory” cortex (Schroeder and Foxe 2005). Complementing these studies, here we have shown that human listeners are able to additionally integrate nonsensory information to enhance the phase locking of oscillations in auditory cortex to acoustic cues. Our results thus support the hypothesis that organisms are able to integrate multiple forms of nonsensory information to aid stimulus prediction. Although in humans this clearly includes linguistic information, it may also include constraints such as probabilistic relationships between stimuli or contextual associations which can be tested in other species. This integration would be facilitated, for example, by the extensive reciprocal connections among multisensory, prefrontal, and parietal regions and auditory cortex in nonhuman primates (Hackett et al. 1999, 2007; Romanski et al. 1999; Petrides and Pandya 2006, 2007).

Taken together, our results demonstrate that the phase of ongoing neural oscillations is impacted not only by sensory input, but also by the integration of nonsensory—in this case, linguistic—information. Cerebro-acoustic coherence thus provides a neural mechanism that allows the brain of a listener to respond to incoming speech information at the optimal rate for comprehension, enhancing sensitivity to relevant dynamic

spectral change (Summerfield 1981; Dilley and Pitt 2010). We propose that during natural comprehension, acoustic and linguistic information act in a reciprocally supportive manner to aid in the prediction of ongoing speech stimuli.

Authors' Contribution

J.E.P., J.G., and M.H.D. designed the research, analyzed the data, and wrote the paper. J.E.P. performed the research.

Funding

The research was supported by the UK Medical Research Council (MC-A060-5PQ80). Funding to pay the Open Access publication charges for this article was provided by the UK Medical Research Council.

Notes

We are grateful to Clare Cook, Oleg Korzyukov, Marie Smith, and Maarten van Casteren for assistance with data collection, Jason Taylor and Rik Henson for helpful discussions regarding data processing, and our volunteers for their participation. We thank Michael Bonner, Bob Carlyon, Jessica Grahn, Olaf Hauk, and Yury Shtyrov for helpful comments on earlier drafts of this manuscript. *Conflict of Interest:* None declared.

References

Abrams DA, Nicol T, Zecker S, Kraus N. 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci.* 28:3958–3965.

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

Arieli A, Sterkin A, Grinvald A, Aertsen A. 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science.* 273:1868–1871.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. 2003. Voxel-based lesion–symptom mapping. *Nat Neurosci.* 6:448–450.

Bishop GH. 1932. Cyclic changes in excitability of the optic pathway of the rabbit. *Am J Physiol.* 103:213–224.

Boemio A, Fromm S, Braun A, Poeppel D. 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci.* 8:389–395.

Busch NA, Dubois J, VanRullen R. 2009. The phase of ongoing EEG oscillations predicts visual perception. *J Neurosci.* 29:7869–7876.

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

Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA. 2009. The natural statistics of audiovisual speech. *PLoS Comput Biol.* 5:e1000436.

Cummins F. 2009. Rhythm as an affordance for the entrainment of movement. *Phonetica.* 66:15–28.

Cummins F, Port R. 1998. Rhythmic constraints on stress timing in English. *J Phonetics.* 26:145–171.

Davis MH, Coleman MR, Absalom AR, Rodd JM, Johnsrude IS, Matta BF, Owen AM, Menon DK. 2007. Dissociating speech perception and comprehension at reduced levels of awareness. *Proc Natl Acad Sci USA.* 104:16032–16037.

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

Davis MH, Johnsrude IS. 2003. Hierarchical processing in spoken language comprehension. *J Neurosci.* 23:3423–3431.

de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA. 2006. Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J Comp Neurol.* 496:27–71.

DeLong KA, Urbach TP, Kutas M. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat Neurosci.* 8:1117–1121.

Dilley LC, Pitt MA. 2010. Altering context speech rate can cause words to appear and disappear. *Psychol Sci.* 21:1664–1670.

Drullman R, Festen JM, Plomp R. 1994. Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am.* 95:2670–2680.

Eickhoff S, Stephan K, Mohlberg H, Grefkes C, Fink G, Amunts K, Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage.* 25:1325–1335.

Elliott TM, Theunissen FE. 2009. The modulation transfer function for speech intelligibility. *PLoS Comput Biol.* 5:e1000302.

Friederici AD, Kotz SA, Scott SK, Obleser J. 2010. Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp.* 31:448–457.

Gagnepain P, Henson RN, Davis MH. 2012. Temporal predictive codes for spoken words in auditory cortex. *Curr Biol.* 22:615–621.

Giraud A-L, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H. 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron.* 56:1127–1134.

Grahn JA, Rowe JB. 2009. Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J Neurosci.* 29:7540–7548.

Greenberg S, Carvey H, Hitchcock L, Chang S. 2003. Temporal properties of spontaneous speech—a syllable-centric perspective. *J Phonetics.* 31:465–485.

Grosjean F. 1983. How long is the sentence? Prediction and prosody in the online processing of language. *Linguistics.* 21:501–530.

Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmelin R. 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proc Natl Acad Sci USA.* 98:694–699.

Gross J, Timmermann L, Kujala J, Dirks M, Schmitz F, Salmelin R, Schnitzler A. 2002. The neural basis of intermittent motor control in humans. *Proc Natl Acad Sci USA.* 99:2299–2302.

Hackett TA, Smiley JF, Ubert I, Karmos G, Lakatos P, de la Mothe LA, Schroeder CE. 2007. Sources of somatosensory input to the caudal belt areas of auditory cortex. *Perception.* 36:1419–1430.

Hackett TA, Stepniewska I, Kaas JH. 1999. Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 817:45–58.

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

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

Kerlin JR, Shahin AJ, Miller LM. 2010. Attentional gain control of ongoing cortical speech representations in a “cocktail party.” *J Neurosci.* 30:620–628.

Kotz SA, Schwartze M. 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn Sci.* 14:392–399.

Lakatos P, Chen C-M, O’Connell MN, Mills A, Schroeder CE. 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron.* 53:279–292.

Lakatos P, Karmos G, Mehta AD, Ubert I, Schroeder CE. 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science.* 320:110–113.

Lakatos P, Shah AS, Knuth KH, Ubert I, Karmos G, Schroeder CE. 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol.* 94:1904–1911.

Lalor EC, Foxe JJ. 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur J Neurosci.* 31:189–193.

- Liss JM, White L, Mattys SL, Lansford K, Lotto AJ, Spitzer SM, Caviness JN. 2009. Quantifying speech rhythm abnormalities in the dysarthrias. *J Speech Lang Hear Res.* 52:1334–1352.
- Loftus GR, Masson MEJ. 1994. Using confidence intervals in within-subject designs. *Psycho Bull Rev.* 1:476–490.
- Luo H, Liu Z, Poeppel D. 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 8:e1000445.
- Luo H, Poeppel D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron.* 54:1001–1010.
- MacNeilage PF. 1998. The frame/content theory of evolution of speech production. *Behav Brain Sci.* 21:499–546.
- McGettigan C, Evans S, Agnew Z, Shah P, Scott SK. 2012. An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *J Cogn Neurosci.* 24:636–652.
- Miller JL, Aibel IL, Green K. 1984. On the nature of rate-dependent processing during phonetic perception. *Percept Psychophys.* 35:5–15.
- Miller JL, Grosjean F, Lomanto C. 1984. Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica.* 41:215–225.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage.* 13:684–701.
- Nolte G. 2003. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys Med Biol.* 48:3637–3652.
- Obleser J, Wise RJS, Dresner MA, Scott SK. 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci.* 27:2283–2289.
- Okada K, Rong F, Venezia J, Matchin W, Hsich I-H, Saberi K, Serrenes JT, Hickok G. 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex.* 20:2486–2495.
- Peelle JE, Johnsrude IS, Davis MH. 2010. Hierarchical processing for speech in human auditory cortex and beyond. *Front Hum Neurosci.* 4:51.
- Peelle JE, Wingfield A. 2005. Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *J Exp Psychol Hum Percept Perform.* 31:1315–1330.
- Petrides M, Pandya DN. 2007. Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *J Neurosci.* 27:11573–11586.
- Petrides M, Pandya DN. 2006. Efferent association pathways originating in the caudal prefrontal cortex in the macaque monkey. *J Comp Neurol.* 498:227–251.
- Pickett J, Pollack I. 1963. The intelligibility of excerpts from fluent speech: effects of rate of utterance and duration of excerpt. *Lang Speech.* 6:151–164.
- Picton TW, John MS, Dimitrijevic A, Purcell D. 2003. Human auditory steady-state potentials. *Int J Audiol.* 42:177–219.
- Poeppel D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time.” *Speech Commun.* 41:245–255.
- Rauschecker JP, Scott SK. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci.* 12:718–724.
- Roberts B, Summers RJ, Bailey PJ. 2011. The intelligibility of noise-vocoded speech: spectral information available from across-channel comparison of amplitude envelopes. *Proc R Soc Lond Biol Sci.* 278:1595–1600.
- Rodd JM, Davis MH, Johnsrude IS. 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex.* 15:1261–1269.
- Rodd JM, Longe OA, Randall B, Tyler LK. 2010. The functional organization of the fronto-temporal language system: evidence from syntactic and semantic ambiguity. *Neuropsychologia.* 48:1324–1335.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci.* 2:1131–1136.
- Romei V, Gross J, Thut G. 2010. On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? *J Neurosci.* 30:8692–8697.
- Rorden C, Brett M. 2000. Stereotaxic display of brain lesions. *Behav Neurol.* 12:191–200.
- Rosen S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. *Phil Trans R Soc Lond B.* 336:367–373.
- Rosen S, Wise RJS, Chadha S, Conway E-J, Scott SK. 2011. Hemispheric asymmetries in speech perception: sense, nonsense and modulations. *PLoS One.* 6:e24672.
- Schroeder CE, Foxe J. 2005. Multisensory contributions to low-level, “unisensory” processing. *Curr Opin Neurobiol.* 15:454–458.
- Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. 2010. Dynamics of active sensing and perceptual selection. *Curr Opin Neurobiol.* 20:172–176.
- Scott SK, Blank CC, Rosen S, Wise RJS. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain.* 123:2400–2406.
- Scott SK, Johnsrude IS. 2003. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26:100–105.
- Scott SK, Rosen S, Lang H, Wise RJS. 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *J Acoust Soc Am.* 120:1075–1083.
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. *Science.* 270:303–304.
- Shockley K, Baker AA, Richardson MJ, Fowler CA. 2007. Articulatory constraints on interpersonal postural coordination. *J Exp Psychol Hum Percept Perform.* 33:201–208.
- Skoe E, Kraus N. 2010. Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31:302–324.
- Summerfield Q. 1981. Articulatory rate and perceptual constancy in phonetic perception. *J Exp Psychol Hum Percept Perform.* 7:1074–1095.
- Tajima K, Port R, Dalby J. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *J Phonetics.* 25:1–24.
- Taulu S, Simola J, Kajola M. 2005. Applications of the signal space separation method. *IEEE Trans Sign Process.* 53:3359–3372.
- Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A. 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Bio-Med Eng.* 44:867–880.
- Voss RF, Clarke J. 1975. “1/f noise” in music and speech. *Nature.* 258:317–318.
- Wild CJ, Davis MH, Johnsrude IS. 2012. Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage.* 60:1490–1502.
- Worsley KJ, Evans AC, Marrett S, Neelin P. 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. *J Cereb Blood Flow Metab.* 12:900–918.