

1 **The auditory-brainstem response to continuous,**
2 **non-repetitive speech is modulated by the speech**
3 **envelope and reflects speech processing**

4 **Running title: Auditory-brainstem response to continuous, non-repetitive**
5 **speech**

6 **Chagit S. Reichenbach^{1,2,3}, Chananel Braiman¹, Nicholas D. Schiff³, A. J. Hudspeth², and**
7 **Tobias Reichenbach^{4*}**

8 ¹Tri-Institutional Training Program in Computational Biology and Medicine, Weill Cornell Medical
9 College, New York, NY 10065, U.S.A.

10 ²Howard Hughes Medical Institute and Laboratory of Sensory Neuroscience, The Rockefeller
11 University, New York, NY 10065, U.S.A.

12 ³Department of Neuroscience, Brain and Mind Research Institute, Weill Cornell Medical College,
13 New York, NY 10065, U.S.A.

14 ⁴Department of Bioengineering, Imperial College London, South Kensington Campus, SW7 2AZ,
15 London, U.K.

16 *To whom correspondence should be addressed (email: reichenbach@imperial.ac.uk)

17

18 **Abstract**

19 The auditory-brainstem response (ABR) to short and simple acoustical signals is an important
20 clinical tool used to diagnose the integrity of the brainstem. The ABR is also employed to
21 investigate the auditory brainstem in a multitude of tasks related to hearing, such as processing
22 speech or selectively focusing on one speaker in a noisy environment. Such research measures the
23 response of the brainstem to short speech signals such as vowels or words. Because the voltage
24 signal of the ABR has a tiny amplitude, several hundred to a thousand repetitions of the acoustic
25 signal are needed to obtain a reliable response. The large number of repetitions poses a challenge to
26 assessing cognitive functions due to neural adaptation. Here we show that continuous, non-
27 repetitive speech, lasting several minutes, may be employed to measure the ABR. Because the
28 speech is not repeated during the experiment, the precise temporal form of the ABR cannot be
29 determined. We show, however, that important structural features of the ABR can nevertheless be
30 inferred. In particular, the brainstem responds at the fundamental frequency of the speech signal,
31 and this response is modulated by the envelope of the voiced parts of speech. We accordingly
32 introduce a novel measure that assesses the ABR as modulated by the speech envelope, at the
33 fundamental frequency of speech and at the characteristic latency of the response. This measure has
34 a high signal-to-noise ratio and can hence be employed effectively to measure the ABR to
35 continuous speech. We use this novel measure to show that the auditory brainstem response is
36 weaker to intelligible speech than to unintelligible, time-reversed speech. The methods presented
37 here can be employed for further research on speech processing in the auditory brainstem and can
38 lead to the development of future clinical diagnosis of brainstem function.

39

40 **Introduction**

41 The auditory-brainstem response (ABR) is an evoked potential generated from the auditory
 42 brainstem nuclei in response to auditory stimuli. Because the ABR can be measured noninvasively
 43 through scalp electrodes, it is widely used in both research and clinical settings to probe subcortical
 44 acoustic processing (Hood, 1998;Hall, 2007). Since the discovery of the ABR in 1970 by Jewett
 45 and colleagues (Jewett et al., 1970), a wealth of studies have investigated how the auditory
 46 brainstem processes a variety of acoustic signals. Such studies have mostly measured the ABR in
 47 response to simple stimuli such as clicks or pure tones. In particular, the auditory brainstem can
 48 exhibit a frequency-following response to the periodicity of a pure tone (Galbraith, 1994;Galbraith,
 49 1995). The frequency-following response has a striking similarity to the eliciting periodic stimulus
 50 in both the temporal and the spectral domain. It presumably represents the phase-locked activity of
 51 neurons in the rostral brainstem, predominantly in the inferior colliculus, lateral lemniscus, and
 52 cochlear nucleus (Smith et al., 1975;Sohmer et al., 1977;Chandrasekaran and Kraus, 2010;Du et al.,
 53 2011).

54 Speech evokes a complex ABR that encodes many aspects of the complicated acoustic
 55 stimulus. A pioneering study in 1980 showed that formants are encoded in the speech-evoked ABR
 56 (Greenberg, 1980). Since then, a diverse set of speech stimuli, including Mandarin syllables, words
 57 such as “lily,” “apple,” and “piano,” consonant-vowel sounds, and short sentences have been used
 58 to elicit ABRs (Krishnan et al., 2004;Russo et al., 2004;Aiken and Picton, 2006;Parbery-Clark et
 59 al., 2009;Skoe and Kraus, 2010;Choi et al., 2013). Galbraith and colleagues demonstrated that the
 60 speech-evoked ABR resembles the eliciting stimulus so closely that it can be understood quite
 61 accurately by naïve participants when played to them as sound (Galbraith et al., 1995). It has further
 62 been demonstrated that the speech-evoked ABR can be affected significantly by aspects of the
 63 acoustic presentation, such as the level of environmental noise or whether the stimulation is
 64 monaural, dichotic, or diotic (Galbraith et al., 1998;Anderson and Kraus, 2010;Li and Jeng, 2011).

65 Important questions remain, however, regarding the role of the auditory brainstem in speech
 66 processing. Extensive efferent neural pathways project from higher areas of the auditory system
 67 such as the auditory cortex back to different areas of auditory brainstem, including the inferior
 68 colliculus and the cochlear nuclei. These connections suggest that the brainstem can play a role in
 69 high-level aspects of speech processing (Diamond et al., 1969;Weedman and Ryugo, 1996;Mulders

70 and Robertson, 2000; Du et al., 2011; Barbas et al., 2013). Training in languages or in music can
 71 affect the subcortical processing of speech as measured through the speech-evoked ABR
 72 (Musacchia et al., 2007; Hornickel et al., 2012). Furthermore, short, repeated intelligible speech
 73 stimuli elicit a larger ABR than reversed, unintelligible speech (Galbraith et al., 2004). However,
 74 whether and how the ABR is modulated by higher cognitive functions such as attention and
 75 comprehension remains debated. Some researchers have measured a different latency of the peak
 76 response to click stimuli during attention to an auditory stimulus as opposed to other sounds (Brix,
 77 1984; Ikeda et al., 2010), but other studies have found no significant difference (Collet and Duclaux,
 78 1986; Connolly et al., 1989). Although the effect is small, the amplitude of the frequency-following
 79 response may be modulated by attention to pure tones (Galbraith and Doan, 1995; Galbraith et al.,
 80 2003). Attending to single vowels yields differences in the amplitude of the brainstem's response,
 81 but the results are inconsistent between subjects (Lehmann and Schönwiesner, 2014). A reason for
 82 these dissonant findings may be the brevity of the signals employed, on the order of tens of
 83 milliseconds. Measuring the ABR requires several hundred to a thousand repetitions of the same
 84 stimuli, potentially allowing for neural adaptation and reducing the effect of efferent feedback
 85 (Lasky, 1997; Neupane et al., 2014).

86 Here we endeavored to measure the response of the auditory brainstem to continuous
 87 speech. To avoid potential adaptive effects to the stimulus we presented a non-repeating speech
 88 signal. We thus faced an important technical challenge: the ABR is of the order of microvolts and
 89 thus much smaller than the background electrical activity contributed by the cortex. Because we
 90 sought not to repeat the stimulus, we could not average the ABR over multiple repetitions and
 91 consequently could not establish the precise temporal waveform of the ABR. Instead we employed
 92 advanced data analytics to obtain meaningful features of the ABR from the recordings.

93 Two important structural aspects of speech are promising in order to extract meaningful
 94 features of the ABR to continuous speech. First, every human voice has a distinct spectral structure:
 95 when a person speaks, the vocal folds open and close at a fundamental frequency that typically lies
 96 between 150 Hz and 250 Hz for a woman or between 100 Hz and 200 Hz for a man. Most of the
 97 spectrum of speech accordingly lies within distinct frequency bands, namely at the fundamental
 98 frequency and its more than ten lowest harmonics (Fig. 1a). The ABR to speech tracks this spectral
 99 structure: an amplitude spectrum of the response shows peaks at the fundamental frequency as well

100 as its harmonics (Skoe and Kraus, 2010;Jeng et al., 2011). Although this response shows
101 similarities to the frequency-following response to pure tones, a major difference exists. The
102 speech-evoked ABR at the fundamental frequency is evoked not only by the fundamental frequency
103 itself, but also by the harmonics in the speech signal. Indeed, the auditory brainstem exhibits a
104 response at the fundamental frequency even when that frequency itself has been removed from the
105 stimulus (Galbraith, 1994;Galbraith and Doan, 1995).

106 Second, the envelope of continuous speech traces important building blocks of speech,
107 namely phonemes, syllables, and words. Cortical oscillations, especially in the delta and theta
108 frequency bands, can entrain to the envelope of speech (Ding and Simon, 2012;Power et al.,
109 2012;Horton et al., 2013;Peelle et al., 2013;Zion Golumbic et al., 2013). This neural entrainment is
110 modulated by higher cognitive functions: it is stronger for an attended speech stream than for an
111 unattended one (Horton et al., 2013) and may be larger for intelligible than for unintelligible speech
112 (Peelle et al., 2013;Ding et al., 2014;Ding and Simon, 2014). The entrainment of cortical
113 oscillations to the speech envelope may accordingly represent a neural mechanism for speech
114 processing (Giraud and Poeppel, 2012).

115 Here we show that the fundamental frequency and the envelope of speech can be employed
116 effectively to measure the ABR to continuous, non-repetitive speech. In particular, the brainstem
117 responds to the fundamental frequency of a continuous, non-repetitive speech stream, the response
118 is modulated by the envelope, and the envelope's modulation greatly increases the signal-to-noise
119 ratio of the ABR. Because continuous speech has a fundamental frequency that varies over time and
120 thus hinders an assessment of the ABR (Fig. 1a), we have used the computer-linguistic program
121 Praat to convert natural speech into monotone speech in which the fundamental frequency and its
122 higher harmonics remain constant throughout the speech stream (Fig. 1b) (Boersma, 2002;Deroche
123 and Culling, 2011). This monotone speech is easily intelligible and complex enough to elicit
124 sustained attention from human subjects. We then employ the developed method to investigate how
125 the ABR to continuous speech is modulated by speech intelligibility.

126 **Results**

127 We recorded ABRs from healthy volunteers presented with a monotone speech stream. To avoid
 128 stimulus artifacts we removed the fundamental frequency and the first two harmonics from the
 129 speech (Fig. 1c), thus ensuring that any signal measured by the electrodes at those frequencies did
 130 not result from the electrical activity of the earphones. We found that the brainstem exhibited a
 131 reliable response at the constant fundamental frequency and at higher harmonics of the monotone
 132 speech (Fig. 2a). Because it was the largest and most informative regarding speech processing, we
 133 focused on the amplitude at the fundamental frequency. The absence of a response at the
 134 fundamental frequency in control recordings, in which the earphones were near the ear but not in
 135 the ear canal so that subjects could not hear the speech, confirmed that the measured ABR at the
 136 fundamental frequency was not a stimulus artifact.

137 *Modulation of the ABR with the speech envelope*

138 Speech includes voiced and voiceless components (Fig. 2b). The voiceless part contains a broad
 139 range of frequencies. The voiced elements result from vowels, among other features, and exhibit a
 140 distinct spectral structure with a fundamental frequency and many harmonics. Because the ABR to
 141 speech at the fundamental frequency arises from the voiced parts of the speech, we hypothesized
 142 that the timecourse of the ABR at the fundamental frequency is modulated by the envelope of the
 143 voiced parts of speech.

144 To investigate the envelope modulation of the ABR, we multiplied the measured brainstem
 145 response by the envelope of the voiced components of the speech at different temporal delays. The
 146 amplitude of the resulting signal at the fundamental frequency was then determined through
 147 spectral analysis at each temporal delay; we refer to this signal as the *envelope-modulated ABR at*
 148 *the fundamental frequency*. If the ABR at the fundamental frequency results from the voiced parts
 149 of speech, then the envelope-modulated ABR at that frequency should have a peak at a
 150 characteristic delay that corresponds to the latency between the speech stimulus and the neural
 151 response in the brainstem. We also computed the modulation of the ABR by the envelope of the
 152 entire speech and by the envelope of the voiceless parts. The modulation of the ABR by the
 153 envelope of the whole speech signal should yield a peak at the same latency, albeit with a smaller

154 magnitude. Modulating the ABR by the envelope of the voiceless components alone should not
 155 produce a peak, for these speech components do not yield an ABR at the fundamental frequency.

156 For modulation by the envelope of the entire speech or by the envelope of the voiced parts
 157 we measured a peak in the envelope-modulated ABR at the fundamental frequency at a delay of
 158 10 ms (Fig. 3a). The peak was larger for the correlation with the voiced components of the speech
 159 than for that with the whole speech. Modulation by the voiceless parts of the speech yielded a
 160 negative peak at around 35 ms. These results indicate that the brainstem response to the
 161 fundamental frequency reflects primarily the voiced components of the speech. The characteristic
 162 latency of the response exceeds those of the peaks in the standard click-evoked ABR, but
 163 corresponds to the latency observed in the ABR to vowels (Skoe and Kraus, 2010).

164 *Improving the signal-to-noise ratio through modulating or correlating the ABR with the speech* 165 *envelope*

166 Modulating the ABR with the voiced speech envelope at the characteristic delay of 10 ms largely
 167 eliminates the periods in a recording during which the brainstem does not respond at the
 168 fundamental frequency. We thus expected that the modulation of the ABR with the speech envelope
 169 would reduce the noise in the recordings. To quantify the putative noise reduction, we computed the
 170 signal-to-noise ratio for the amplitude of the peak in the envelope-modulated ABR at the
 171 fundamental frequency. To obtain an estimate of the inter-subject variability of the signal-to-noise
 172 ratio, we calculated the ratio for each individual and determined the population mean and its
 173 standard error. We also calculated the signal-to-noise ratio for the Fourier amplitude of the ABR at
 174 the fundamental frequency, without modulating the signal by the speech envelope, and determined
 175 the population mean and the associated standard error. We found that, by taking the envelope
 176 modulation of the ABR, we obtained a signal-to-noise ratio of the amplitude at the fundamental
 177 frequency that was more than threefold as large as when we computed the Fourier amplitude of the
 178 ABR alone (Fig. 3c). The difference was highly significant ($p < 0.001$).

179 Another method to determine how the voiced parts of speech shape the ABR at the
 180 fundamental frequency is to investigate the cross-correlation between the timecourse of the ABR
 181 and the envelope of the voiced parts of speech. This correlation might also improve the signal-to-
 182 noise ratio. We computed the timecourse of the ABR at the fundamental frequency by dividing the

time series into short time windows and analyzing the Fourier amplitude at the fundamental frequency in each window (short-time Fourier transformation). As another method, we computed the timecourse of the ABR at the fundamental frequency through a wavelet transform using the morlet wavelet as the mother wavelet. Each of these two resulting timecourses was then correlated with the envelope of the voiced parts of speech. As for the envelope-modulated ABR at the fundamental frequency, the cross-correlation exhibited a peak at a delay of 10 ms (Fig. 3b). We then computed the signal-to-noise ratio of the correlation around the characteristic latency for each subject, and calculated the corresponding population mean and its standard error. We found that the differences in the signal-to-noise ratios that were obtained from the Fourier amplitude as well as from the correlation values as computed from the short-time Fourier transformation and the wavelet transformation were all statistically insignificant ($p > 0.05$). The signal-to-noise ratio of the envelope-modulated ABR at the fundamental frequency and at the characteristic delay was, however, severalfold larger than the signal-to-noise ratios obtained using the other methods, and the differences were highly significant ($p < 0.001$; Fig. 3c). The best signal-to-noise ratio thus resulted not from cross-correlating the timecourse of the ABR with the speech envelope, but rather from modulating the ABR by the speech envelope at the characteristic delay and then extracting the amplitude at the fundamental frequency.

Motivated by the substantial increase in the signal-to-noise ratio of the envelope-modulated ABR at the fundamental frequency as opposed to a simple Fourier transform of the ABR, we employed this measure—the envelope-modulated ABR at the fundamental frequency—to investigate how the ABR to continuous speech is modulated by speech intelligibility.

Modulation of the ABR by speech intelligibility

We investigated the influence of the intelligibility of continuous, non-repetitive speech on brainstem activity by presenting subjects with forward and time-reversed monotone speech. Although the two stimuli have an identical spectral composition, only the forward speech is intelligible. Comparison of the neural responses to the two stimuli has previously been used to investigate speech comprehension, to diagnose brain function, and to identify the role of the auditory brainstem in speech processing (Schiff et al., 2005; Deng and Srinivasan, 2010; Howard and Poeppel, 2010; Sunami et al., 2013).

212 We computed the envelope-modulated ABR at the fundamental frequency for each subject,
 213 and from that the population mean and its standard error (Fig. 4a). We found that the envelope-
 214 modulated ABR to reversed speech exceeded that to forward speech by about half for every
 215 latency. By computing the mean amplitude at the peak of the envelope-modulated ABR, we found
 216 that the difference was statistically significant ($p < 0.05$).

217 We then investigated whether the responses to forward and to reverse speech also differed
 218 significantly and consistently at the level of individual subjects. For each individual we computed
 219 the envelope-modulated ABR to forward and time-reversed speech at the fundamental frequency
 220 and at the characteristic latency. Nine out of ten subjects had a larger envelope-modulated ABR at
 221 the fundamental frequency for reverse than for forward speech (Fig. 4b). These differences were
 222 statistically significant in seven of the nine subjects ($p < 0.05$). Only one subject showed a larger
 223 response to forward than to reverse speech, but the difference was insignificant ($p > 0.5$).

224 Discussion

225 Our results demonstrate that important structural features of the brainstem's response to continuous
 226 speech can be detected reliably by electrophysiological means. Although we cannot measure the
 227 precise temporal form of the ABR, as is feasible through repetitive measurements with short
 228 acoustic stimuli such as clicks or vowels, we can extract and quantify structural features of the ABR
 229 that emerge in response to characteristics of continuous speech. We have found specifically that the
 230 brainstem responds at the fundamental frequency of monotone speech even when that frequency is
 231 absent from the speech stimulus. Our results additionally demonstrate that the ABR at the
 232 fundamental frequency is modulated by the envelope of the voiced part of speech and that the
 233 timecourse of the ABR is correlated to the envelope.

234 The envelope modulation of the ABR can be employed to reduce significantly the noise in
 235 the response at the fundamental frequency. Whereas a Fourier transform of a three-minute
 236 recording of the ABR to continuous speech yields a signal-to-noise ratio of only 0.4 for the
 237 amplitude at the fundamental frequency, modulation of the ABR by the envelope of the voiced
 238 parts of speech, at the characteristic delay and at the fundamental frequency, achieves a signal-to-
 239 noise ratio of 1.3. Because the response of the auditory brainstem at the fundamental frequency

240 results from the voiced parts of speech alone, focusing on those components reduces the noise in a
241 recording. Although we likewise expect the correlation of the timecourse of the ABR to the speech
242 envelope to reduce the noise, the resulting signal-to-noise ratio is below that obtained when
243 modulating the ABR with the envelope, and even below the signal-to-noise ratio of the Fourier
244 amplitude. This deficiency likely stems from the short-time Fourier transformation that is required
245 to extract the timecourse of the ABR, which can then be correlated to the envelope. Short-time
246 Fourier transformation has a poor frequency resolution that varies inversely to the duration of the
247 time window. The poor frequency resolution renders the timecourse of the ABR at the fundamental
248 frequency much noisier than the signal obtained by a Fourier transform over a longer recording, as
249 we can employ for the envelope-modulated ABR at the fundamental frequency. The same
250 reasoning applies to the wavelet transform and can explain why this method also yields a small
251 signal-to-noise ratio.

252 The increase of the signal-to-noise ratio by more than a factor of three through modulation
253 of the brainstem response by the envelope of the voiced parts of speech can accelerate auditory-
254 brainstem recordings significantly. According to the central limit theorem, a longer recording
255 improves the signal-to-noise ratio in proportion to the square root of the duration of the recording.
256 Raising the signal-to-noise ratio by a factor of three therefore requires a ninefold longer recording.
257 Conversely, because the method proposed here increases the signal-to-noise ratio by more than a
258 factor of three through computational means, we can reduce the recording time by more than
259 ninefold and still obtain a signal-to-noise ratio similar to that for the longer recording with a simple
260 Fourier analysis. Although the additional numerical analysis requires several layers of computation,
261 all of them can run in real time.

262 Our results on the influence of speech intelligibility on the ABR differ from previous
263 findings. An earlier study addressed differences in the ABR to short, repetitive speech signals and
264 their time-reversed versions and found that forward speech elicited a stronger response at the
265 fundamental frequency than did reverse speech (Galbraith et al., 2004). We have observed the
266 opposite result in response to continuous non-repetitive speech: forward, intelligible speech yields a
267 smaller ABR at the fundamental frequency than does reverse, unintelligible speech. The
268 discrepancy between the two studies may reflect differences in how the brain responds to many
269 repetitions of the same short speech signal rather than to a long, non-repetitive, continuous stream

270 of speech. In particular, the brain may adapt to many repeated presentations of the same speech
271 stimulus, an effect that can be avoided through the non-repetitive speech signal that we have
272 employed here.

273 Although forward and reverse speech have the same Fourier spectrum, the two signals differ
274 phonetically. The manner in which voiceless consonants transition into voiced components can
275 differ between the stimuli; owing to nonlinear processing by the cochlea, this can lead to
276 differences in the brainstem response (Dau, 2003). However, in our study we have shown that the
277 ABR at the fundamental frequency results from the voiced parts of the monotone speech,
278 components that we expect to be comparable between forward and reverse speech. Further studies
279 are needed to clarify whether the differences in phonetical structure between forward and reverse
280 monotone speech, associated with the transition of consonants to vowels, cause a difference in the
281 brainstem's response.

282 In this study we have focused on the response at the fundamental frequency only. However,
283 the brainstem also responds at higher harmonics, and these neural signals likely contain information
284 about speech processing as well. Moreover, empirical mode decomposition such as through the
285 Hilbert-Huang transform can extract nonlinear oscillations from a time series, which may be
286 employed to identify nonlinear responses at the fundamental frequency as well as at higher
287 harmonics (Huang and Shen, 2005). Investigating these issues will further clarify the role of the
288 auditory brainstem in speech processing.

289 ABRs are used routinely to evaluate hearing, specifically to assess the integrity of the ear
290 and the brainstem. The results presented here suggest that the ABR to continuous monotone speech
291 can provide valuable information both about the integrity of the brainstem and about auditory
292 processing. Moreover, forward *versus* time-reversed speech stimuli have been used clinically to
293 assess patients in a minimally conscious state (Schiff et al., 2005). Measuring the envelope-
294 modulated ABR to forward and to reverse speech may likewise provide a valuable tool in assessing
295 auditory processing in patients suffering from disorders of consciousness (Giacino et al.,
296 2002;Laureys et al., 2004;Schiff, 2010;Goldfine et al., 2011).

297 **Methods**

298 *Participants*

299 Ten adult volunteers between 19 and 33 years of age participated in the experiments. All subjects
300 had normal hearing and normal or corrected-to-normal vision and had no history of hearing or
301 neurological impairments. All experimental methods were approved by the Imperial College
302 Research Ethics Committee. All experiments were performed in accordance with relevant
303 guidelines and regulations and every subject provided written informed consent prior to the
304 experimental session.

305 *Monotone speech stimuli*

306 Speech samples were obtained from publicly available audiobooks and were converted to monotone
307 speech through the pitch-synchronous overlap-add (PSOLA) approach (Moulines and Charpentier,
308 1990) with the computer-linguistic software Praat (Boersma, 2002). The fundamental frequency of
309 the speaker was set to 89 Hz. To prevent stimulus artifacts, every speech stimulus was high-pass
310 filtered at three times the fundamental frequency of the speaker and thus did not contain the
311 fundamental frequency and the first two harmonics. Reversed speech was created by temporally
312 inverting a speech stimulus. Each speech stimulus lasted three minutes.

313 *Experimental design*

314 The experiment assessed whether the ABR provides information about speech processing by
315 comparing responses to forward and to reversed speech. Each subject listened to both a forward and
316 a reversed three-minute continuous speech stream. The order of the two speech streams was chosen
317 randomly for every subject.

318 *Auditory-brainstem recordings*

319 All recordings for the study were completed during a three-week period. During each session
320 subjects sat in a comfortable chair in a quiet room. Speech stimuli were presented to the subjects
321 through custom electrically-shielded earphones (hf5, Etymotic, U.S.A.) at a comfortable level of
322 70 dB SPL. Sound intensity was calibrated with a microphone (ECM8000, Behringer, Germany).

323 We measured responses from the auditory brainstem through active Ag/AgCl electrodes and
 324 a passive ground electrode (g.LADYbird and g.LADYbirdGND, Guger Technologies, Austria). The
 325 active electrodes were positioned at the cranial vertex (Cz) and on both mastoid processes. The
 326 passive ground electrode was placed on the central forehead (Lehmann and Schönwiesner, 2014).
 327 The impedance between each electrode and the scalp was measured (g.Zcheck, Guger
 328 Technologies, Austria) and confirmed to be below 5 k Ω . A bipolar amplifier (g.BSamp, Guger
 329 Technologies, Austria) enhanced the differences between the voltage signals at the mastoids and
 330 that at the vertex by a factor of 10,000 and band-pass filtered them between 0.1 Hz and 1 kHz. The
 331 analogue voltage signals were digitized at a sampling frequency of 8 kHz with a data-acquisition
 332 card NI PCI 6221 (National Instruments, U.S.A.) and a custom-written Matlab program
 333 (MathWorks, U.S.A). The Matlab program also presented speech signals to the subjects at a
 334 sampling frequency of 44.1 kHz through the computer's internal sound card. The speech signals
 335 were time-locked to the electroencephalographic recordings and the voltage signals were saved for
 336 offline analysis.

337 *Analysis of auditory-brainstem-response signals*

338 We first determined the significance of the brainstem's response at the fundamental frequency by
 339 comparing it to the signal at neighboring frequencies. To quantify the latter signal, which
 340 constitutes the noise floor, we used Matlab to compute the average and the standard deviation of the
 341 Fourier amplitudes from 2 Hz below the fundamental frequency to 2 Hz above it, excluding the
 342 response at the fundamental frequency. We then considered the response at the fundamental
 343 frequency to be significant if its amplitude was at least three standard deviations above the mean
 344 response at the neighboring frequencies. We found that all responses were significant, and verified
 345 that the width of the frequency interval that was used to compute the noise floor did not impact this
 346 result.

347 We were then interested in the response of the auditory brainstem to continuous speech at
 348 the fundamental frequency, as well as in the modulation and correlation with the speech envelope.
 349 We thus employed four different methods to analyze the speech-evoked ABR. The methods were
 350 implemented using custom-written Matlab programs.

In the first method, we assessed the Fourier amplitude of the ABR at the fundamental frequency. The ABR is measured through surface electrodes that record a voltage signal $V(t)$ from a starting time $t = 0$ to a final time t_e . The analogue signal is then sampled at a sampling frequency F_s and thus transformed to a discrete signal $\{V_n\}_{n=0}^{N-1}$ with $N = t_e F_s$. The discrete Fourier transform decomposes the discrete time signal into its frequency components:

$$\tilde{V}_k = \sum_{n=0}^{N-1} V_n e^{-2\pi i k n / N}, \quad k = 0, 1, \dots, N-1. \quad (1)$$

Because the temporal voltage signal is real, the complex Fourier coefficients \tilde{V}_k fulfill the relation $\tilde{V}_k = \tilde{V}_{N-k}^*$ and the magnitude $|\tilde{V}_k| + |\tilde{V}_{N-k}| = 2|\tilde{V}_k|$ is the Fourier amplitude of the periodic component at frequency $f = k / t_e$.

To determine the signal-to-noise ratio of the amplitude at the fundamental frequency, the voltage time series was divided into segments of 3 s duration. For each segment the Fourier amplitude at the fundamental frequency was determined, and the amplitudes of the responses from the left and the right brainstem were averaged. From the amplitudes of the segments we then computed the mean and the standard deviation. The signal-to-noise ratio followed as the ratio of the mean amplitude to the standard deviation, that is, as the reciprocal of the coefficient of variation (Bushberg and Boone, 2011). We computed the signal-to-noise ratio for each subject's ABR and from that obtained the population mean and its standard error for the signal-to-noise ratio.

As a second method, we determined the envelope-modulated ABR at the fundamental frequency. We first extracted the envelope of the speech signal $s(t)$ through the Hilbert transformation

$$H[s](t) = -\frac{1}{\pi} \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^{\infty} \frac{s(t+\tau) - s(t-\tau)}{\tau} d\tau. \quad (2)$$

The speech envelope was obtained from the Hilbert transform by low-pass filtering at 30 Hz.

We then determined the voiced and voiceless components of the speech stream and their envelopes. The speech signal was divided into segments of 40 ms duration using Hann windows, and we computed the average speech envelope for each segment. We computed the power cepstrum for each segment,

$$\text{power cepstrum} = \left| \mathcal{F}^{-1} \left(\log \left\{ \left| \mathcal{F}[s(t)] \right|^2 \right\} \right) \right|^2, \quad (3)$$

in which \mathcal{F} denotes the Fourier transform (Benesty et al., 2008). We determined the amplitude at the quefrency that corresponded to the fundamental frequency of the speaker. If this amplitude was

significantly higher than the average of the ten neighbouring frequencies, or if the amplitude of the segment's speech envelope exceeded a minimum threshold level, we considered that segment to correspond to the voiced part of speech, and otherwise to voiceless speech. For constructing the envelope of the voiced parts of speech, we kept the envelopes of all voiced segments but ignored those of voiceless segments. The envelope of the voiceless parts of speech was obtained analogously.

To compute the envelope-modulated ABR at the fundamental frequency, we downsampled the speech envelopes to the same sampling frequency as had been employed for the ABR. Denote by $\{e_n\}_{n=0}^{N-1}$ the resulting discrete time series of the envelope for either the whole speech, its voiced parts, or its voiceless parts. Amplitude modulation of the ABR is then obtained by shifting the envelope by a temporal delay τ , and hence by an index $l = \tau F_s$, with respect to the ABR and by multiplying both signals. We consider the envelope before the speech starts, that is, before time $t = 0$, to be zero. The envelope-modulated ABR at the fundamental frequency, which we denote by $\tilde{V}_{\text{env.mod.}}(\tau)$, follows from the Fourier amplitude at the fundamental frequency:

$$\tilde{V}_k^{(\text{env.mod.})}(\tau) = \sum_{n=0}^{N-1} V_n e_{n-l} e^{-2\pi i k n / N}. \quad (4)$$

The index k is chosen such that it corresponds to the fundamental frequency f_0 , that is, $f_0 = k / t_e$.

We computed the envelope-modulated ABR at the fundamental frequency by first dividing the ABR signal and the corresponding speech envelope into 3 s segments. We then computed the envelope-modulated ABR at the fundamental frequency for each segment and for temporal delays from $\tau = -300$ to $\tau = 500$ ms, and found a peak at the characteristic delay of about $\tau = 10$ ms. For every segment we thus computed the peak amplitude as the mean of the envelope-modulated ABR at delays between $\tau = 0$ and $\tau = 20$ ms. The envelope-modulated ABR at the fundamental frequency and at the characteristic latency for an individual recording followed as the mean of these peak amplitudes across all segments; the standard deviation across the different segments yielded the variation in the envelope-modulated ABR. The signal-to-noise ratio followed as the ratio of the mean to the standard deviation. We computed this ratio for each subject individually, and then used the obtained data to determine the population mean and its standard error.

As a third method, we employed a short-time Fourier transformation to extract the timecourse of the ABR at the fundamental frequency, and then determined the correlation of this

timecourse with the speech envelope. For short-time Fourier transformation, we partitioned the voltage time series into segments of 80 ms duration through Hann windows and computed the Fourier transform for each segment. For every segment we extracted the Fourier amplitude at the fundamental frequency and thus found the discrete timecourse of the ABR at that frequency. We denote this discrete time series, consisting of a number of M data points, by $\{A_n\}_{n=0}^{M-1}$. We downsampled the speech envelope to have the same sampling frequency F_{ABR} as the timecourse of the ABR at the fundamental frequency, resulting in the discrete envelope time series $\{E_n\}_{n=0}^{M-1}$. We computed the cross-correlation of the speech envelope, shifted by various delays τ corresponding to an index $m = \tau F_{ABR}$, with the discrete timecourse of the ABR at the fundamental frequency:

$$(S * E)(\tau) = \frac{1}{\sigma_A \sigma_E} \sum_{n=0}^{M-1} A_n E_{n-m} \quad (5)$$

in which σ_A denotes the standard deviation of the ABR timecourse $\{A_n\}_{n=0}^{M-1}$, and σ_E denotes the standard deviation of the envelope $\{E_n\}_{n=0}^{M-1}$. As for the envelope-modulated signal, the correlation exhibited a peak at a delay of $\tau = 10$ ms. We extracted the signal-to-noise ratio around the peak from the temporal correlation from $\tau = 0$ to $\tau = 20$ ms analogously to the signal-to-noise ratio of the envelope-modulated ABR.

As a fourth method, we used a wavelet transformation to extract the timecourse of the ABR at the fundamental frequency. To this end we employed a morlet wavelet as the mother wavelet at a temporal standard deviation of 16 ms and at the fundamental frequency. The morlet wavelet was chosen to capture the oscillatory behavior of the ABR. We then correlated the obtained timecourse of the ABR to the speech envelope as in the third method.

Statistical analysis

To determine the statistical significance of the differences in the signal-to-noise ratios between the four different methods for extracting the ABR response—the simple Fourier transformation, the cross-correlation of the speech envelope with the timecourse of the ABR at the fundamental frequency determined either through a short-time Fourier transformation or through a wavelet transformation, and the envelope-modulated ABR at the fundamental frequency—we performed two-sample Student's t -tests for pairwise comparisons of the signal-to-noise ratios obtained by these four methods. Although we employed the Bonferroni correction to account for the six

437 pairwise comparisons, the level of statistical significance for each statistical test did not depend on
438 whether this correction was used.

439 To assess differences in the ABR to forward and to the time-reversed speech on the
440 population level, we computed the envelope-modulated ABR for forward and reverse speech, for
441 each subject and at a range of delays. We then computed the mean and standard error of the mean
442 of these responses across all individuals. To investigate statistical significance, we analyzed the
443 responses around the peak latency of 10 ms by averaging the responses at latencies between 0 ms
444 and 20 ms for each individual subject, and from that computed the mean and standard error of the
445 mean across all subjects. We performed a two-sample Student's *t*-test to assess whether the
446 difference in the mean amplitudes for the forward and the time-reversed conditions was statistically
447 significant.

448 To investigate differences in the ABR to forward and reverse speech on the level of
449 individual subjects, we computed the envelope-modulated ABR at the fundamental frequency and
450 at the characteristic latency as described above. We obtained the mean and the standard error of the
451 mean across all segments. This gave us an average response as well as a measure of the variability
452 of the envelope-modulated ABR in an individual subject. To assess the statistical significance of
453 the difference in the responses to forward and to reverse speech in an individual subject, we
454 performed a paired, two-sample Student's *t*-test.

455 The results of the statistical tests are indicated in the figures through asterisks: no asterisk is
456 given when results are not significant ($p > 0.05$), one asterisk when results are significant (*,
457 $0.01 < p < 0.05$), two asterisks when significance is high (**, $0.001 < p < 0.01$), and three asterisks
458 when significance is very high (***, $p < 0.001$).

459 **Acknowledgements**

460 This research was supported by a National Science Foundation Graduate Research Fellowship to
461 C.S.R., by a National Institutes of Health T32 pre-doctoral training grant to C.B, and by EPSRC
462 grant EP/M026728/1 to T.R.. A.J.H. is an Investigator of Howard Hughes Medical Institute.

463 **Competing financial interests**

464 The authors declare no competing financial interests.

465 **Author Contributions**

466 C. S. R., C. B., N. D. S., A. J. H. and T. R. designed the research, performed the experiments,
467 analyzed the data, and prepared the manuscript.

468 **References**

- 469 Aiken, S.J., and Picton, T.W. (2006). Envelope following responses to natural vowels. *Audiol.*
470 *Neurotol.* 11, 213-232.
- 471 Anderson, S., and Kraus, N. (2010). Objective neural indices of speech-in-noise perception. *Trends*
472 *Ampl.* 14, 73-83.
- 473 Barbas, H., Bunce, J., Medalla, M., Stuss, D., and Knight, R. (2013). Prefrontal pathways that
474 control attention. *Princ. Front. Lobe Funct.*, 31-48.
- 475 Benesty, J., Sondhi, M.M., and Huang, Y. (2008). *Springer handbook of speech processing*.
476 Springer.
- 477 Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott Int.* 5, 341-345.
- 478 Brix, R. (1984). The Influence of Attention on the Auditory Brain Stem Evoked Responses
479 Preliminary Report. *Acta Oto-lar.* 98, 89-92.
- 480 Bushberg, J.T., and Boone, J.M. (2011). *The essential physics of medical imaging*. Lippincott
481 Williams & Wilkins.
- 482 Chandrasekaran, B., and Kraus, N. (2010). The scalp - recorded brainstem response to speech:
483 Neural origins and plasticity. *Psychophysiol.* 47, 236-246.
- 484 Choi, J.M., Purcell, D.W., Coyne, J.-a.M., and Aiken, S.J. (2013). Envelope Following Responses
485 Elicited by English Sentences. *Ear Hear.* 34, 637-650.
- 486 Collet, L., and Duclaux, R. (1986). Auditory brainstem evoked responses and attention:
487 Contribution to a controversial subject. *Acta Oto-Lar.* 101, 439-441.
- 488 Connolly, J.F., Aubry, K., McGillivray, N., and Scott, D.W. (1989). Human brainstem auditory
489 evoked potentials fail to provide evidence of efferent modulation of auditory input during
490 attentional tasks. *Psychophysiol.* 26, 292-303.
- 491 Dau, T. (2003). The importance of cochlear processing for the formation of auditory brainstem and
492 frequency following responses. *J. Acoust. Soc. Am.* 113, 936-950.
- 493 Deng, S., and Srinivasan, R. (2010). Semantic and acoustic analysis of speech by functional
494 networks with distinct time scales. *Brain Res.* 1346, 132-144.
- 495 Deroche, M.L., and Culling, J.F. (2011). Voice segregation by difference in fundamental frequency:
496 Evidence for harmonic cancellation. *J. Acoust. Soc. Am.* 130, 2855-2865.
- 497 Diamond, I., Jones, E., and Powell, T. (1969). The projection of the auditory cortex upon the
498 diencephalon and brain stem in the cat. *Brain Res.* 15, 305-340.
- 499 Ding, N., Chatterjee, M., and Simon, J.Z. (2014). Robust cortical entrainment to the speech
500 envelope relies on the spectro-temporal fine structure. *Neuroim.* 88, 41-46.
- 501 Ding, N., and Simon, J.Z. (2012). Emergence of neural encoding of auditory objects while listening
502 to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854-11859.
- 503 Ding, N., and Simon, J.Z. (2014). Cortical entrainment to continuous speech: functional roles and
504 interpretations. *Front. Hum. Neurosci.* 8.

- 505 Du, Y., Kong, L., Wang, Q., Wu, X., and Li, L. (2011). Auditory frequency-following response: A
 506 neurophysiological measure for studying the “cocktail-party problem”. *Neurosci. Biobehav.*
 507 *Rev.* 35, 2046-2057.
- 508 Galbraith, G.C. (1994). Two-channel brain-stem frequency-following responses to pure tone and
 509 missing fundamental stimuli. *Electroencephal. Clin. Neurophysiol.* 92, 321-330.
- 510 Galbraith, G.C., Amaya, E.M., De Rivera, J.M.D., Donan, N.M., Duong, M.T., Hsu, J.N., Tran, K.,
 511 and Tsang, L.P. (2004). Brain stem evoked response to forward and reversed speech in
 512 humans. *Neurorep.* 15, 2057-2060.
- 513 Galbraith, G.C., Arbagey, P.W., Branski, R., Comerci, N., and Rector, P.M. (1995). Intelligible
 514 speech encoded in the human brain stem frequency-following response. *Neuroreport* 6,
 515 2363-2367.
- 516 Galbraith, G.C., Bhuta, S.M., Choate, A.K., Kitahara, J.M., and Mullen Jr, T.A. (1998). Brain stem
 517 frequency - following response to dichotic vowels during attention. *Neurorep.* 9, 1889-
 518 1893.
- 519 Galbraith, G.C., and Doan, B.Q. (1995). Brainstem frequency-following and behavioral responses
 520 during selective attention to pure tone and missing fundamental stimuli. *Int. J. Psychophys.*
 521 19, 203-214.
- 522 Galbraith, G.C., Olfman, D.M., and Huffman, T.M. (2003). Selective attention affects human brain
 523 stem frequency-following response. *Neurorep.* 14, 735-738.
- 524 Giacino, J.T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D.I., Kelly, J.P., Rosenberg,
 525 J.H., Whyte, J., and Zafonte, R. (2002). The minimally conscious state definition and
 526 diagnostic criteria. *Neurol.* 58, 349-353.
- 527 Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging
 528 computational principles and operations. *Nat. Neurosci.* 15, 511-517.
- 529 Goldfine, A.M., Victor, J.D., Conte, M.M., Bardin, J.C., and Schiff, N.D. (2011). Determination of
 530 awareness in patients with severe brain injury using EEG power spectral analysis. *Clin.*
 531 *Neurophysiol.* 122, 2157-2168.
- 532 Greenberg, S. (1980). Wpp, no. 52: Temporal neural coding of pitch and vowel quality. *Working*
 533 *Papers in Phonetics, Department of Linguistics, UCLA.*
- 534 Hall, J.W. (2007). *New handbook of auditory evoked responses.* ASHA.
- 535 Hood, L.J. (1998). *Clinical applications of the auditory brainstem response.* Singular publishing
 536 group San Diego, CA.
- 537 Hornickel, J., Knowles, E., and Kraus, N. (2012). Reliability of the auditory brainstem responses to
 538 speech over one year in school-age children: A reply to Drs. McFarland and Cacace. *Hear.*
 539 *Res.* 287, 3-5.
- 540 Horton, C., D'zmura, M., and Srinivasan, R. (2013). Suppression of competing speech through
 541 entrainment of cortical oscillations. *J. Neurophysiol.* 109, 3082-3093.
- 542 Howard, M.F., and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal
 543 response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104,
 544 2500-2511.
- 545 Huang, N.E., and Shen, S.S. (2005). *Hilbert-Huang transform and its applications.* World
 546 Scientific.
- 547 Ikeda, K., Sekiguchi, T., and Hayashi, A. (2010). Concentrated pitch discrimination modulates
 548 auditory brainstem responses during contralateral noise exposure. *Neurorep.* 21, 359-366.
- 549 Jeng, F.-C., Costilow, C.E., Stangherlin, D.P., and Lin, C.-D. (2011). Relative power of harmonics
 550 in human frequency-following responses associated with voice pitch in american and
 551 chinese adults. *Percept. Mot. skills* 113, 67-86.
- 552 Jewett, D.L., Romano, M.N., and Williston, J.S. (1970). Human auditory evoked potentials:
 553 possible brain stem components detected on the scalp. *Science* 167, 1517-1518.
- 554 Krishnan, A., Xu, Y., Gandour, J.T., and Cariani, P.A. (2004). Human frequency-following
 555 response: representation of pitch contours in Chinese tones. *Hear. Res.* 189, 1-12.

- 556 Lasky, R.E. (1997). Rate and adaptation effects on the auditory evoked brainstem response in
 557 human newborns and adults. *Hear. Res.* 111, 165-176.
- 558 Laureys, S., Owen, A.M., and Schiff, N.D. (2004). Brain function in coma, vegetative state, and
 559 related disorders. *Lancet Neurol.* 3, 537-546.
- 560 Lehmann, A., and Schönwiesner, M. (2014). Selective attention modulates human auditory
 561 brainstem responses: relative contributions of frequency and spatial cues. *PloS One* 9,
 562 e85442.
- 563 Li, X., and Jeng, F.-C. (2011). Noise tolerance in human frequency-following responses to voice
 564 pitch. *J. Acoust. Soc. Am.* 129, EL21-EL26.
- 565 Moulines, E., and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for
 566 text-to-speech synthesis using diphones. *Speech Comm.* 9, 453-467.
- 567 Mulders, W., and Robertson, D. (2000). Evidence for direct cortical innervation of medial
 568 olivocochlear neurones in rats. *Hear. Res.* 144, 65-72.
- 569 Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical
 570 auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104,
 571 15894-15898.
- 572 Neupane, A.K., Gururaj, K., Mehta, G., and Sinha, S.K. (2014). Effect of repetition rate on speech
 573 evoked auditory brainstem response in younger and middle aged individuals. *Audiol. Res.* 4.
- 574 Parbery-Clark, A., Skoe, E., and Kraus, N. (2009). Musical experience limits the degradative
 575 effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100-
 576 14107.
- 577 Peelle, J.E., Gross, J., and Davis, M.H. (2013). Phase-locked responses to speech in human auditory
 578 cortex are enhanced during comprehension. *Cerebr. Cort.* 23, 1378-1387.
- 579 Power, A.J., Foxe, J.J., Forde, E.J., Reilly, R.B., and Lalor, E.C. (2012). At what time is the
 580 cocktail party? A late locus of selective attention to natural speech. *Eur. J. Neurosci.* 35,
 581 1497-1503.
- 582 Russo, N., Nicol, T., Musacchia, G., and Kraus, N. (2004). Brainstem responses to speech syllables.
 583 *Clin. Neurophysiol.* 115, 2021-2030.
- 584 Schiff, N.D. (2010). Recovery of consciousness after brain injury: a mesocircuit hypothesis. *Trends*
 585 *Neurosci.* 33, 1-9.
- 586 Schiff, N.D., Rodriguez-Moreno, D., Kamal, A., Kim, K., Giacino, J.T., Plum, F., and Hirsch, J.
 587 (2005). fMRI reveals large-scale network activation in minimally conscious patients.
 588 *Neurol.* 64, 514-523.
- 589 Skoe, E., and Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear*
 590 *Hear.* 31, 302.
- 591 Smith, J.C., Marsh, J.T., and Brown, W.S. (1975). Far-field recorded frequency-following
 592 responses: evidence for the locus of brainstem sources. *Electroencephal. Clin.*
 593 *Neurophysiol.* 39, 465-472.
- 594 Sohmer, H., Pratt, H., and Kinarti, R. (1977). Sources of frequency following responses (FFR) in
 595 man. *Electroencephal. Clin. Neurophysiol.* 42, 656-664.
- 596 Sunami, K., Ishii, A., Takano, S., Yamamoto, H., Sakashita, T., Tanaka, M., Watanabe, Y., and
 597 Yamane, H. (2013). Neural mechanisms of phonemic restoration for speech comprehension
 598 revealed by magnetoencephalography. *Brain Res.* 1537, 164-173.
- 599 Weedman, D.L., and Ryugo, D.K. (1996). Projections from auditory cortex to the cochlear nucleus
 600 in rats: synapses on granule cell dendrites. *J. Comp. Neurol.* 371, 311-324.
- 601 Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., Mckhann, G.M., Goodman,
 602 R.R., Emerson, R., Mehta, A.D., and Simon, J.Z. (2013). Mechanisms underlying selective
 603 neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77, 980-991.

605 **Figure Legends**

606 **Figure 1: Properties of natural and monotone speech.** (a) The spectrogram of a sample of
 607 natural speech shows that the energy concentrates at the fundamental frequency that typically lies
 608 between 100 Hz and 300 Hz and the corresponding harmonics. The fundamental frequency of
 609 speech and the harmonics vary in time. (b) Monotone speech has been modified to maintain
 610 constant frequencies of the fundamental and its harmonics. (c) The experiments employed
 611 monotone speech that was high-pass filtered at 300 Hz. The power spectrum reveals that the
 612 fundamental frequency, $f_0 = 89$ Hz in this example, and its first two harmonics were absent from the
 613 speech sample.

614 **Figure 2: Response of the auditory brainstem to continuous monotone speech.** (a) The power
 615 spectrum of the ABR to three minutes of monotone speech shows a strong response at the
 616 fundamental frequency (f_0) and at its second harmonic ($2f_0$). (b) A speech waveform (grey) is
 617 characterized by variations on fast and slow time-scales. Slow variations, on the order of hundreds
 618 of milliseconds and above, define the speech envelope (black) that traces distinct syllables and
 619 words. Voiced parts of speech are characterized by a periodicity at the fundamental frequency;
 620 voiceless parts lack this periodic structure.

621 **Figure 3: Envelope-modulated ABR and cross-correlation of the ABR timecourse to the**
 622 **speech envelope.** (a) Modulation of the ABR with the envelope of the voiced parts of speech (red)
 623 as well as with the envelope of the whole speech signal (black) yields a peak at a delay of 10 ms.
 624 Envelope modulation of the ABR with the voiceless parts of speech, however, produces a minimum
 625 value at a delay of about 35 ms. (b) The cross-correlation of the timecourse of the ABR at the
 626 fundamental frequency with the voiced parts of speech (red) and the envelope of the entire speech
 627 stimulus (black) exhibit likewise a maximum at the delay of 10 ms, whereas the cross-correlation
 628 with the envelope of the voiceless speech components yields a minimum at a delay of 15 ms.
 629 (c) The signal-to-noise ratio of the envelope-modulated ABR at the fundamental frequency, and at
 630 the delay of 10 ms, is several fold larger than that obtained by simple Fourier transformation. It also
 631 significantly exceeds the cross-correlation of the speech envelope with the timecourse of the ABR,

632 at the delay of 10 ms, both when the ABR timecourse is extracted through short-time Fourier
633 transformation (STFT) and when it is identified by a wavelet transform.

634 **Figure 4: Modulation of the ABR by speech intelligibility.** (a) The envelope-modulated ABR to
635 unintelligible reverse speech (red) exceeds that to intelligible forward speech (black) when
636 averaged over all subjects. The envelope-modulated ABR to both speech stimuli is largest around
637 the characteristic delay of 10 ms. The standard errors of the mean (shaded) that follow from the
638 variability between the subjects are smaller than the difference between the mean responses, and
639 this difference is statistically significant. (b) For every subject, the response to time-reversed
640 monotone speech (red) exceeds the neural response to forward monotone speech (black). The
641 difference is statistically significant in the majority of the study participants.

Figure 1:

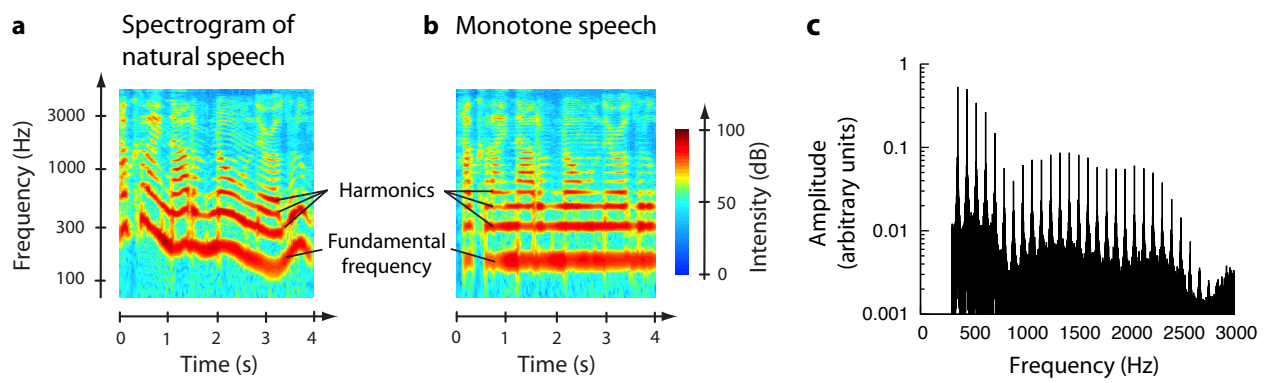


Figure 2:

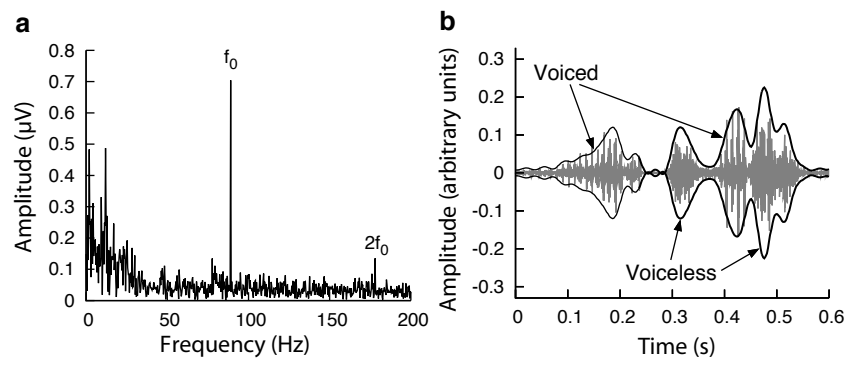


Figure 3:

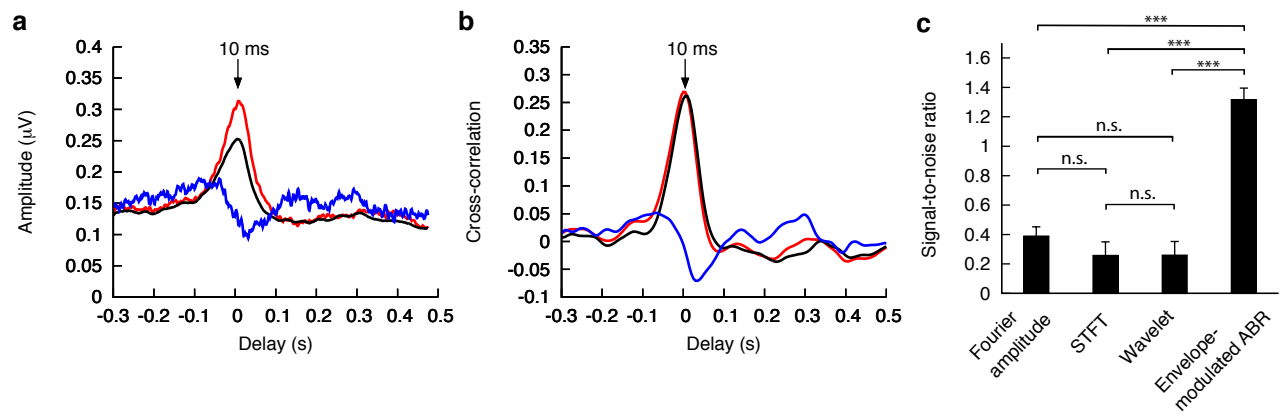


Figure 4:

