

PAPER • OPEN ACCESS

Computational modeling of the auditory brainstem response to continuous speech

To cite this article: Marina Saiz-Alía and Tobias Reichenbach 2020 *J. Neural Eng.* **17** 036035

View the [article online](#) for updates and enhancements.



The Department of Bioengineering at the University of Pittsburgh Swanson School of Engineering invites applications from accomplished individuals with a PhD or equivalent degree in bioengineering, biomedical engineering, or closely related disciplines for an open-rank, tenured/tenure-stream faculty position. We wish to recruit an individual with strong research accomplishments in Translational Bioengineering (i.e., leveraging basic science and engineering knowledge to develop innovative, translatable solutions impacting clinical practice and healthcare), with preference given to research focus on neuro-technologies, imaging, cardiovascular devices, and biomimetic and biorobotic design. It is expected that this individual will complement our current strengths in biomechanics, bioimaging, molecular, cellular, and systems engineering, medical product engineering, neural engineering, and tissue engineering and regenerative medicine. In addition, candidates must be committed to contributing to high quality education of a diverse student body at both the undergraduate and graduate levels.

[CLICK HERE FOR FURTHER DETAILS](#)

To ensure full consideration, applications must be received by June 30, 2019. However, applications will be reviewed as they are received. Early submission is highly encouraged.



PAPER

OPEN ACCESS

RECEIVED
20 April 2020ACCEPTED FOR PUBLICATION
27 May 2020PUBLISHED
2 July 2020

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Computational modeling of the auditory brainstem response to continuous speech

Marina Saiz-Alía¹ and Tobias Reichenbach^{1,2} ¹ Department of Bioengineering and Centre for Neurotechnology, Imperial College London, South Kensington Campus, London, SW7 2AZ, United Kingdom² Author to whom any correspondence should be addressed.E-mail: reichenbach@imperial.ac.uk**Keywords:** auditory brainstem response, speech, computational model, auditory periphery

Abstract

Objective. The auditory brainstem response can be recorded non-invasively from scalp electrodes and serves as an important clinical measure of hearing function. We have recently shown how the brainstem response at the fundamental frequency of continuous, non-repetitive speech can be measured, and have used this measure to demonstrate that the response is modulated by selective attention. However, different parts of the speech signal as well as several parts of the brainstem contribute to this response. Here we employ a computational model of the brainstem to elucidate the influence of these different factors. **Approach.** We developed a computational model of the auditory brainstem by combining a model of the middle and inner ear with a model of globular bushy cells in the cochlear nuclei and with a phenomenological model of the inferior colliculus. We then employed the model to investigate the neural response to continuous speech at different stages in the brainstem, following the methodology developed recently by ourselves for detecting the brainstem response to running speech from scalp recordings. We compared the simulations with recordings from healthy volunteers. **Main results.** We found that the auditory-nerve fibers, the cochlear nuclei and the inferior colliculus all contributed to the speech-evoked brainstem response, although the dominant contribution came from the inferior colliculus. The delay of the response corresponded to that observed in experiments. We further found that a broad range of harmonics of the fundamental frequency, up to about 8 kHz, contributed to the brainstem response. The response declined with increasing fundamental frequency, although the signal-to-noise ratio was largely unaffected. **Significance.** Our results suggest that the scalp-recorded brainstem response at the fundamental frequency of speech originates predominantly in the inferior colliculus. They further show that the response is shaped by a large number of higher harmonics of the fundamental frequency, reflecting highly nonlinear processing in the auditory periphery and illustrating the complexity of the response.

1. Introduction

The auditory-brainstem response (ABR) is an evoked potential traditionally believed to be generated from different structures below the cortex, in particular from the auditory-nerve fibers and the auditory brainstem nuclei [1, 2]. It confers a convenient tool to non-invasively probe subcortical auditory processing, as well as to assess impairment in the auditory periphery. Typical ABR measurements employ many repeated short clicks [3, 4]. However, the auditory brainstem also exhibits a frequency-following

response (FFR) to the periodicity of pure tones [5, 6], mostly as a result of the phase-locked activity of neurons in the rostral brainstem [1, 2, 7, 8].

Similarly, speech evokes a complex ABR that encodes several aspects of the acoustic stimulus, such as the onset of a syllable as well as the fundamental frequency of the voiced speech parts [9, 10]. The measurement of brainstem responses to speech stimuli may aid audiological assessments of hearing-impaired people who cannot respond behaviourally, and in particular inform on their neural speech coding [11, 12]. Recent studies have accordingly

developed statistical methods for detecting the brainstem response to different features of continuous, non-repetitive speech [13–15].

While brainstem responses to short sounds such as clicks or single syllables require a large number of repetitions of the sound and subsequent averaging of the neural measurement, the brainstem response to continuous speech can be measured without repeating the speech signal. The brainstem response to continuous speech may accordingly be useful in assessing higher-level cognitive processes where repetition of a stimulus could lead to neural adaptation as well as fatigue in subjects. As an example, in a previous study, we have shown how the brainstem response at the fundamental frequency of speech can be recorded (speech-ABR), despite the unavoidable temporal variations of the fundamental frequency in natural speech. We have then employed this method to show that the speech-ABR is modulated by selective attention to one of two competing speakers, demonstrating a subcortical mechanism for listening in noisy backgrounds that involves the extensive feedback loops coming from the central auditory cortex [13].

Major questions remain, however, regarding the precise origin of the speech-ABR, as well as what parts of the speech signal evoke it. First, regarding the neural origin, different parts of the brainstem likely contribute to the speech-ABR, but experimental measurements so far have not been able to resolve those, instead showing only a single aggregate scalp-recorded response. In addition, recent magnetoencephalographic (MEG) and electroencephalographic (EEG) studies have revealed a cortical contribution to the FFR, namely at the fundamental frequency of short speech tokens [16–18]. Such a cortical response likely occurs for the speech-ABR as well, and will add to the different subcortical sources. Second, regarding the parts of the speech signal that cause the speech-ABR, it is well known that the brainstem still responds at the fundamental frequency of speech even when the fundamental frequency itself has been removed from the signal [5, 6]. The response is then driven by higher harmonics and reflects different nonlinearities in the auditory periphery. As an example, the cochlea introduces a pronounced compressive nonlinearity, and this can lead to the phenomenon of ‘formant capture’ by which harmonics adjacent to the formant regions are emphasised [19, 20]. However, it remains unclear to which degree the fundamental frequency itself as well as the many higher harmonics that are present in a speech signal contribute to the speech-ABR.

Computational modelling of the brainstem response can elucidate these questions. A number of computational models of the auditory periphery and subcortical processing indeed exist and have been employed extensively to study click-evoked brainstem responses, FFRs and how they are affected by

sensorineural hearing loss as well as neuronal impairment [8, 21–25]. The auditory brainstem response to short speech tokens has been modelled as well [26]. The response of the auditory brainstem at the fundamental frequency of continuous speech has, however, not yet been assessed in a computational model.

Computational models of the brainstem response have shown that the most important stages include the middle ear that modifies the spectrum of a complex sound, the inner ear that decomposes a complex sound into different frequency components and introduces a compressive nonlinearity regarding the amplitudes, the synapse between the inner hair cells of the inner ear and the attached auditory-nerve fibers, and the subsequent responses in the cochlear nuclei (CN) and the inferior colliculus (IC). The existing computational models of these different stages range from detailed biophysical descriptions to more phenomenological approaches [25, 27, 28].

Because we sought to shed light on the relative contributions of the different parts of the brainstem to the speech-ABR, as well as to elucidate how the speech-ABR is shaped by different parts of the speech signal, but not to investigate the biophysical mechanisms of the individual brainstem parts, we employed models of the different stages that were realistic enough to capture the relevant processing, but that omitted further detail. Regarding sound processing in the inner ear, we considered an established model, that could describe typical adaptation properties, realistic phase-locking, as well as the cochlear compressive nonlinearity [29]. We further described the neural responses of the CN through models of spiking globular bushy cells (GBCs), one of the populations of neurons in the ventral cochlear nucleus that exhibit pronounced synchronization to complex sounds [27, 30]. For the IC we employed a phenomenological description that simulated the envelope-tuned behaviour of the neurons in this part of the brainstem [31].

Because the cortical contributions to the FFR remain poorly characterized, for instance regarding their latency, amplitude and frequency characteristics, they have not yet been modelled computationally. We have therefore not included these responses in our model, but restricted the latter to subcortical contributions only. Likewise, we employed a bottom-up approach to model the speech-ABR that did not assess the modulation of the response by selective attention, which would require a description of top-down neural modulation.

Our modelling approach allowed us to quantify the relative contribution of different subcortical stages to the speech-ABR, as well as to quantify how different aspects of a speech signal shape this neural response. To reinforce our findings, we compared the simulations with experimental data of brainstem responses to clicks in noise as well as of brainstem

responses to continuous speech that we published recently [32].

2. Materials and methods

2.1. Computational model

We implemented a computational model that built on existing models of the middle and inner ear, the cochlear nuclei (CN) and the inferior colliculus (IC) (figure 1). First, regarding the middle and inner ear, we employed a humanized phenomenological model of the auditory periphery available in the python package *cochlea* [29, 33]. This modelled the middle ear filtering, the cochlear transmission (through a parallel filter bank), the inner hair cells, the power-law dynamics at the inner hair cell synapses and spike generators. The description of the synapses between the inner hair cells and the auditory-nerve fibers contained a spontaneous rate (SR) as well as rate-level properties that account for different types of auditory nerve fibers. We considered 300 characteristic frequencies ranging from 125 Hz to 20 kHz, distributed using the Greenwood function [34]. For each characteristic frequency we modelled twelve auditory-nerve fibers with a high SR, four fibers with a medium SR, and four fibers with a low SR. This mimicked the reported distribution of SRs of auditory-nerve fibers [35]. The model included a power-law adaptation of the synapses; this incorporated a fractional Gaussian noise [28].

Second, the neural responses of the cochlear nuclei were modelled through a set of globular bushy cells, available in the *cochlear_nucleous* python package [27]. Auditory-nerve fibers converge onto the soma of a globular bushy cell and excite it through large synapses, the endbulbs of Held, the synaptic weights of which we considered as uniform across the different cells. The synaptic connection had a delay of 4 ms. Each globular bushy cell was modelled as a single compartment with Hodgkin-Huxley-type ion channels [36–38]. The auditory-nerve fibers were connected to a population of 200 globular bushy cells, each of which received input from 20 fibers. The model was implemented in the NEURON simulation environment through the Brian simulator for spiking neurons. Globular bushy cells were connected to auditory-nerve fibers at random, namely at a small probability so that on average 20 auditory-nerve fibers were connected to each globular bushy cell (convergence of 20) [27]. The synaptic weight was 8×10^{-9} Siemens. The spiking activity of the auditory nerve and the globular bushy cells were then transformed into a time-varying population activity rate using a boxcar kernel density estimation with a bandwidth of 30 samples.

Third, a phenomenological model of amplitude-modulation processing in the inferior colliculus was used to simulate the neural response at this mid-brain nucleus [31]. Computationally, the neural

response of the CN was convolved with the postsynaptic potentials in the IC. Following recent modeling work, the postsynaptic potentials were approximated by alpha functions [31] whose time constants were based on intracellular recordings of bushy cells [39]. We thereby employed the same model parameters, with the exception of the time constant for the inhibition. For the latter we employed a smaller value of 1 ms to ensure that the speech information around the fundamental frequency was within the band-pass range of the modelled neurons.

The neural response at each of the three model stages, middle and inner ear, cochlear nuclei, and inferior colliculus, was scaled by a corresponding factor, $W_{an} = 3.7 \times 10^{-5}$, $W_{cn} = 0.03$ and $W_{ic} = 1.1$ respectively, to yield electrical amplitudes that were comparable to human scalp recordings (Table VIII-1 in [40] and figure 2). The three scaled waveforms were then added together to simulate the scalp recorded auditory-brainstem activity. Because we sought to keep model parameters to a minimum, we did not introduce additional delays that could have accounted for intermediate processing stages. As a result, the latencies of the individual stages and their interpeak distances may slightly differ from normative data, although the simulated speech-ABR and click-ABR latencies match reference values.

Simulations were performed for continuous speech inputs, sampled at 100 kHz and at an intensity of 85 dB SPL. We used 20 different speech samples, each of which had a duration of 30 s. Speech samples were obtained from the publicly available audiobook ‘*The children of Odin*’ read by Elizabeth Klett (<https://librivox.org>). Her voice had a fundamental frequency of $175 \text{ Hz} \pm 39 \text{ Hz}$ (mean and standard deviation) which exceeded the lowest characteristic frequency of the inner-ear model that we used. We employed the same female voice in our previous experimental measurements of the speech-ABR which allowed us to compare the experimental data to the modelling results [32].

2.2. Computation of the click-ABR

We simulated auditory brainstem responses to clicks (click-ABR) in quiet and in different levels of background noise. For each level of background noise we computed the brainstem response from 20 trials, each of which lasted 30 s. Clicks were presented at a rate of 10 Hz and had a sound intensity of 82 peSPL. We employed white noise at levels of 42, 52, 62 and 72 dB SPL for the background noise.

2.3. Computation of the speech-ABR

Following the methodology developed recently by ourselves for detecting the scalp-recorded brainstem response to running speech, we computed a fundamental waveform of each voiced part of a speech

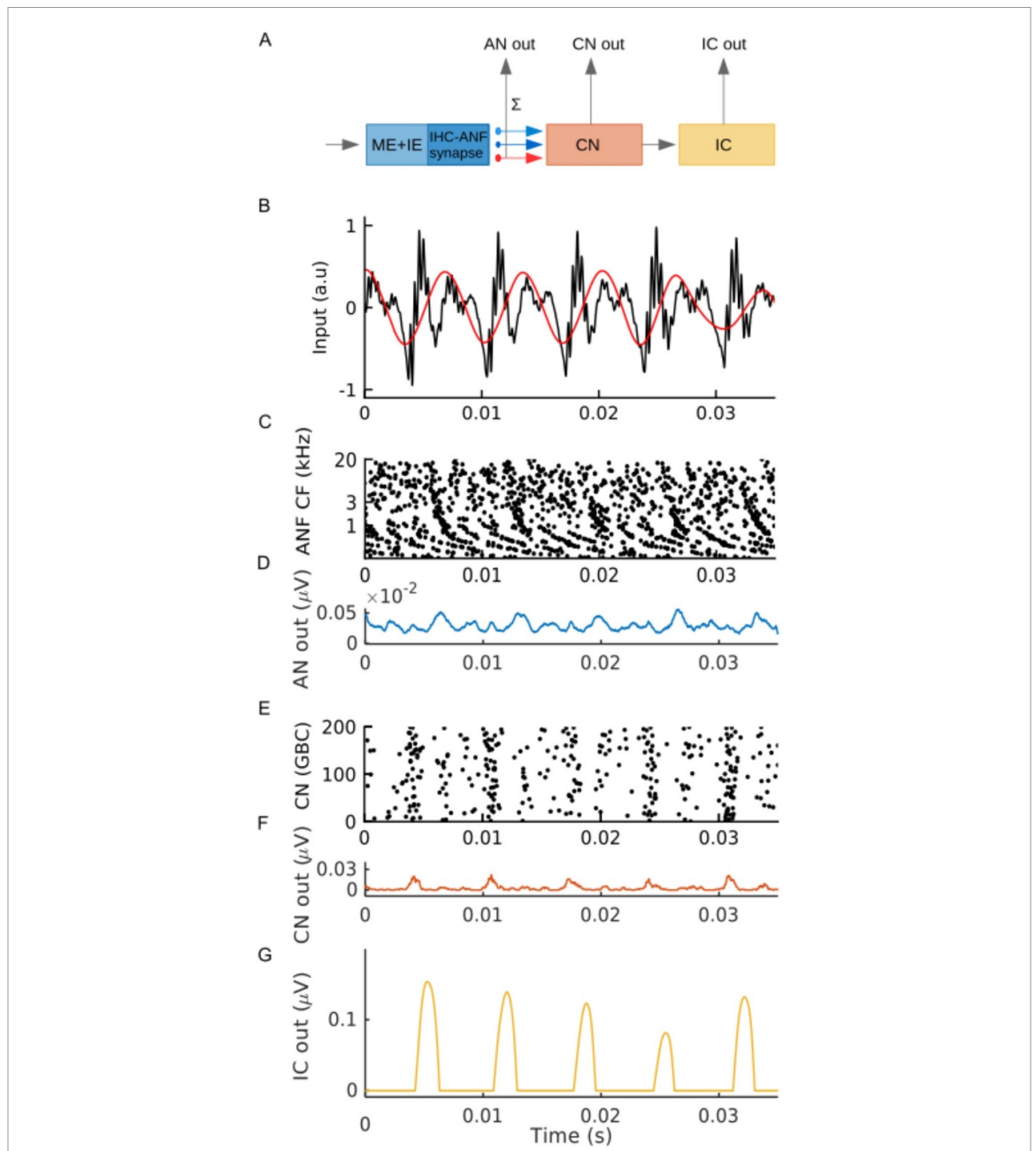
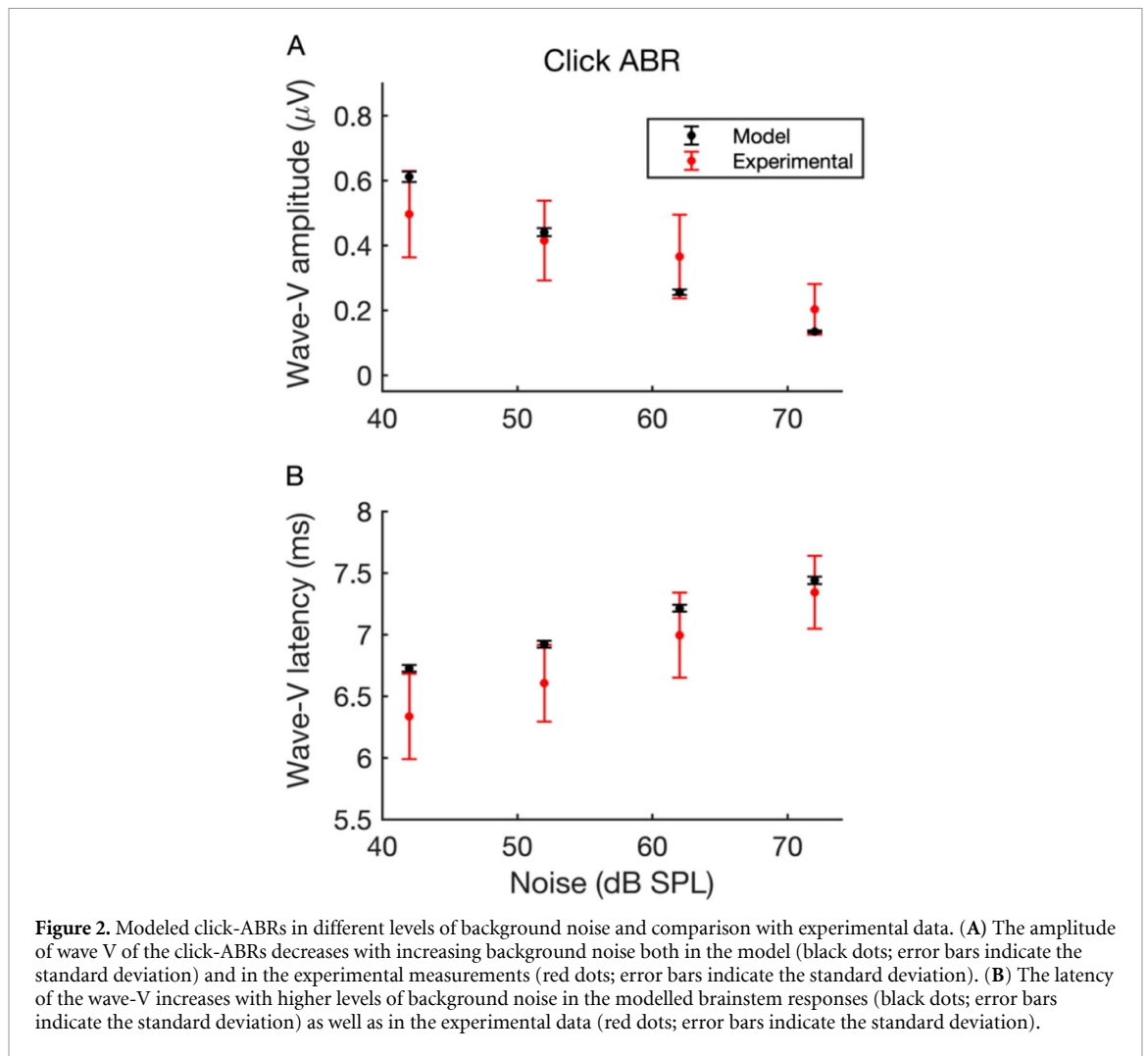


Figure 1. The computational model. (A) The model consists of three main parts. First, a model of the auditory periphery describes the middle ear (ME), inner ear (IE), inner hair cells (IHC) and the attached auditory-nerve fibers (ANF). The output from the auditory-nerve fibers is then fed into the second part, a model of the cochlear nuclei (CN), followed by a model of the inferior colliculus (IC). (B) A short sample of a speech signal, namely the vowel 'e' (black), shows oscillations at the fundamental frequency as well as at higher harmonics. The corresponding fundamental waveform (red) oscillates at the time-varying fundamental frequency. (C), (D) The spike trains of the auditory-nerve fibres (C) are summed over the different frequency channels and temporally smoothed (D). For illustration purposes we display only the response of every thirtieth auditory-nerve fibre. (E), (F) The model of the globular bushy cells in the cochlear nuclei produces spike trains as well that are then summed over the different neurons and smoothed (F). (G) The output from the model of the inferior colliculus yields a firing rate.

signal, using empirical mode decomposition (EMD) (figure 1(B)) [13, 41]. We further determined the Hilbert transform of the fundamental waveform, and obtained a complex waveform that had the fundamental waveform as its real part and the Hilbert transform as its imaginary part. This complex waveform was then correlated with the simulated brainstem responses, at each of the three main levels of the brainstem that we modelled, as well as with the combined

neuronal response, at a range of negative and positive temporal delays. The magnitude and delay of the response were obtained from the peak of the correlation's magnitude. The noise level was computed as the 95th percentile of the complex cross-correlation's amplitude, for latencies from -60 ms to -20 ms and 20 ms to 100 ms. The ratio of the amplitude to the noise level defined the signal-to-noise ratio (SNR) of the response.



2.4. Contribution of different frequency bands of speech to the speech-ABR

We sought to investigate the frequency contribution of the different harmonics in the speech signal to the modelled speech-ABR. We therefore filtered the speech signal into different frequency bands (zero-phase FIR bandpass filters, filter order 512). The lowest frequency band was 0–300 Hz and contained the fundamental frequency. We divided the higher frequencies of the speech signal into frequency bands that had equal width on a logarithmic frequency scale, mimicking the cochlear frequency mapping (figure 4). Bands of higher frequency therefore contained more harmonics of the fundamental frequency than those of lower frequency. We normalized the amplitude of the speech signal in each frequency band so that they all had the same intensity of 85 dB SPL. For each frequency band we then simulated the resulting brainstem response, using 20 different speech samples of 30 s in duration, and determined the resulting speech-ABR as outlined above.

2.5. Effect of the fundamental frequency on the speech-ABR

We assessed the effects of the fundamental frequency of the speech on the modelled speech-ABR. We used the software Audacity to modify the fundamental frequency and the corresponding harmonics of the original speech, from a 20% reduction in frequency to a 50% increase, in 10% steps, leading to eight different stimuli. We computed the modelled brainstem responses to each by using 20 different speech samples, each of which had a duration of 30 s.

2.6. Experimental data

We previously recorded both brainstem responses to clicks in background noise as well as speech-ABRs from 42 young normal-hearing adults [32]. The inclusion and exclusion criteria as well as the technical details of the recordings are described in our corresponding previous publication [32].

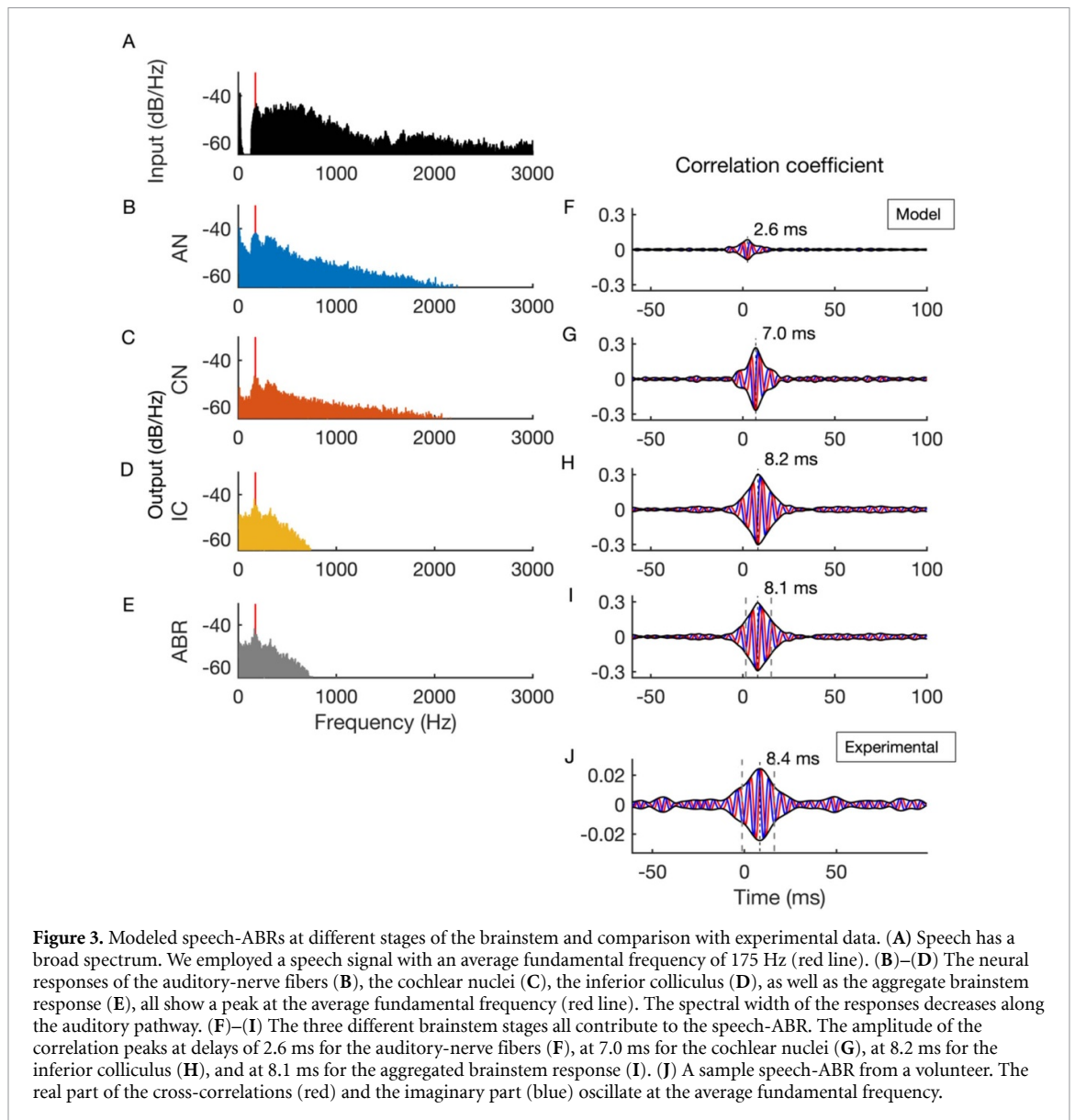


Figure 3. Modeled speech-ABRs at different stages of the brainstem and comparison with experimental data. (A) Speech has a broad spectrum. We employed a speech signal with an average fundamental frequency of 175 Hz (red line). (B)–(D) The neural responses of the auditory-nerve fibers (B), the cochlear nuclei (C), the inferior colliculus (D), as well as the aggregate brainstem response (E), all show a peak at the average fundamental frequency (red line). The spectral width of the responses decreases along the auditory pathway. (F)–(I) The three different brainstem stages all contribute to the speech-ABR. The amplitude of the correlation peaks at delays of 2.6 ms for the auditory-nerve fibers (F), at 7.0 ms for the cochlear nuclei (G), at 8.2 ms for the inferior colliculus (H), and at 8.1 ms for the aggregated brainstem response (I). (J) A sample speech-ABR from a volunteer. The real part of the cross-correlations (red) and the imaginary part (blue) oscillate at the average fundamental frequency.

2.7. Statistical analysis

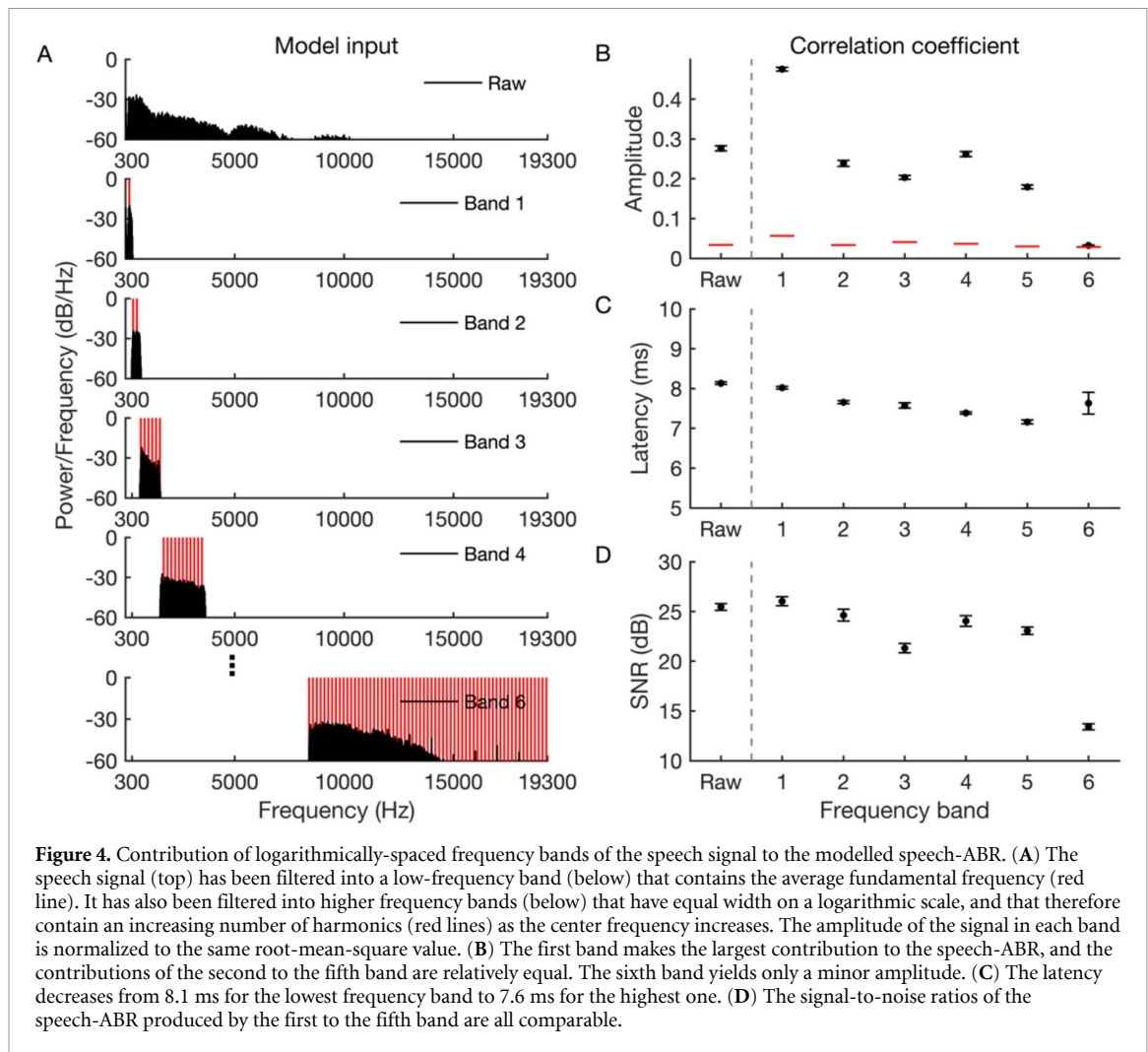
The modelled speech-ABR features, such as amplitude, latency, SNR, and full width at half maximum (FWHM) were averaged across trials. The features followed a normal distribution as assessed through the Shapiro-Wilk test. We accordingly employed parametric tests when assessing the statistical significance of model predictions. Unbalanced ANOVA tests and two-sample Student's *t*-test were used when comparing model results to experimental data.

3. Results

We modelled the brainstem response to speech at three main processing stages in the brainstem, the middle and inner ear, resulting in activity in the auditory-nerve fibers, in the cochlear nuclei and in the inferior colliculus (figure 1). To verify that the models of these different stages gave satisfactory brainstem responses, we first used them to model

the neural responses to clicks. We obtained simulated responses with peaks at 1.090 ± 0.001 ms (mean and SEM) from the auditory-nerve fibers, at 5.461 ± 0.003 ms from the cochlear nuclei, and at 6.687 ± 0.004 ms from the inferior colliculus, corresponding to the main generators of ABR wave I, wave III and wave V, respectively.

We furthermore simulated the brainstem response to clicks in different levels of background noise, and compared the modelled results to the corresponding experimental data that we recorded previously [32] (figure 2). The amplitude of the modelled brainstem responses did not differ significantly from that obtained in the experiments, and for both datasets the amplitude decreased with increasing background noise ($p = 1 \times 10^{-46}$ and $p = 0.5$; two-way ANOVA across noise level and type of dataset, respectively, figure 2(A)). The latencies of the click-ABR wave-V in the simulated brainstem response were roughly comparable to the one seen in



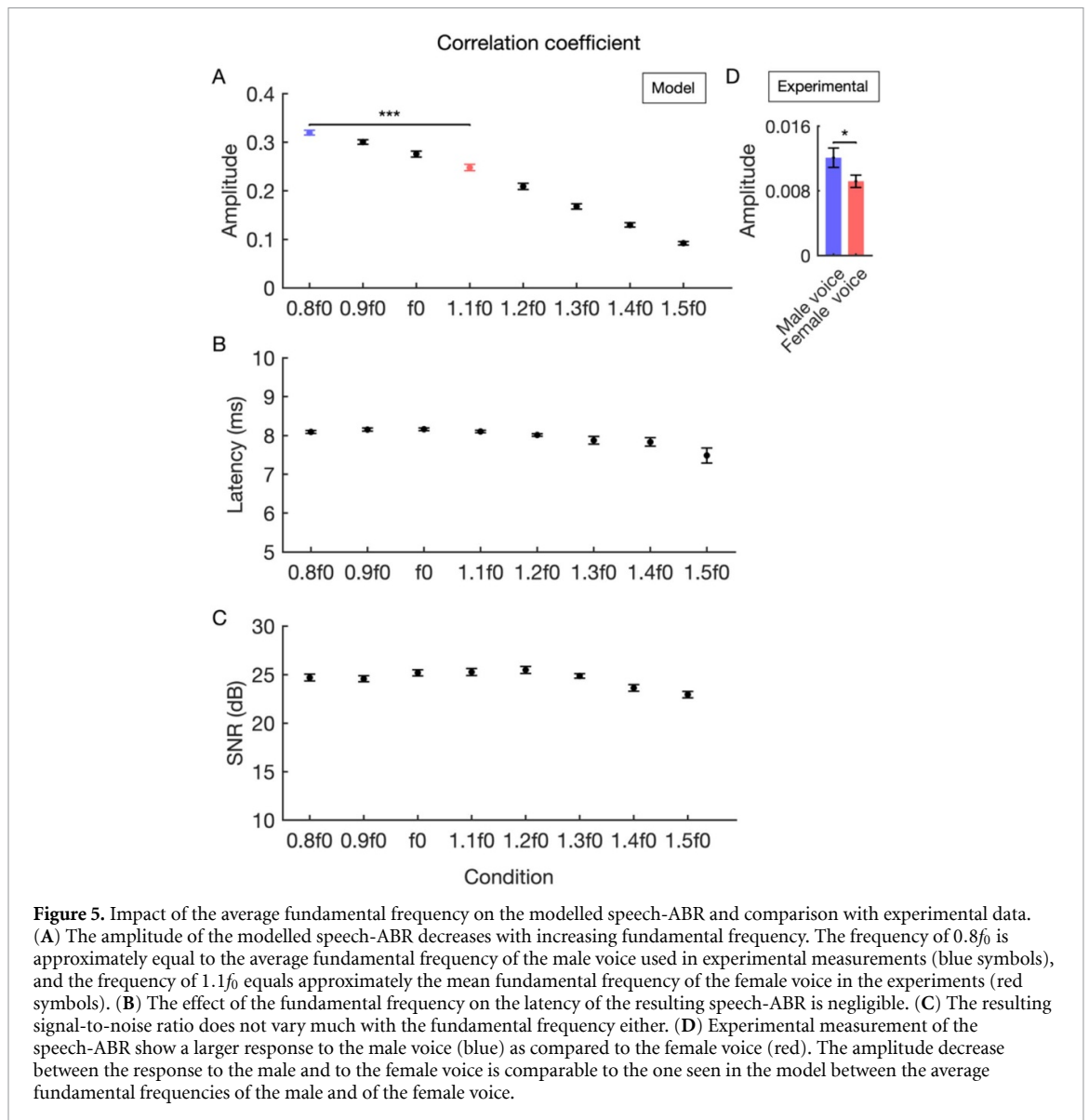
the experimental recordings (figure 2(B)). Both sets of latencies increased significantly as the background noise became louder, and the latencies in the simulated brainstem responses systematically exceeded the experimental measurements ($p = 6 \times 10^{-47}$ and $p = 5 \times 10^{-11}$, two-way ANOVA across noise level and type of dataset, respectively). The increase in latency with the background noise probably reflects a larger contribution of the low-SR auditory-nerve fibers in higher levels of background noise, since the low-SR auditory-nerve fibers have a slower onset than the medium- or high-SR auditory-nerve fibers [42].

Regarding the brainstem's response at the fundamental frequency of a speech signal, we found that all three stages contributed (figures 3(B)–(I)). The auditory-nerve fiber responses contained frequency contributions at a large part of the speech spectrum, whereas the response of the cochlear nuclei and the inferior colliculus occurred predominantly at lower frequencies below 800 Hz. The auditory-nerve fibers contributed to the speech-ABR at an early latency of 2.9 ± 0.1 ms (mean and SEM), while the two subsequent brainstem centers showed longer latencies of 7.23 ± 0.07 and 8.32 ± 0.04 ms. The modelled speech-ABR was largest at the inferior colliculus, with

a cross-correlation amplitude of 0.288 ± 0.006 , compared to 0.079 ± 0.001 at the auditory-nerve fibers and 0.259 ± 0.003 at the cochlear nuclei. The resulting model of the scalp-recorded speech-ABR was dominated by the response at the inferior colliculus, with a cross-correlation amplitude of 0.276 ± 0.007 , a latency of 8.13 ± 0.03 ms and a FWHM of 14.7 ± 0.6 ms. Importantly, the contributions from the three different modelled brainstem stages could no longer be distinguished in the aggregated speech-ABR.

The simulated scalp-recorded speech-ABR resembled our recent experimental observations (figure 3(J)). In particular, the simulated and the experimentally-recorded responses did not differ significantly in either latency or FWHM ($p = 0.9$ and $p = 0.1$, respectively; two-tailed two-sample Student's *t*-tests).

The simulated brainstem responses to clicks and to speech differed in their latencies. The speech-ABR occurred significantly later than wave V of the click-ABR, namely 1.45 ± 0.03 ms later ($p = 9 \times 10^{-21}$, two-tailed one-sample Student's *t*-test), and its latency had a significantly larger variability ($p = 3 \times 10^{-13}$, one-sided two-sample *F*-test for



equal variances). The longer latency of the modelled speech-ABR as compared to the modelled click-ABR mirrored a similar behaviour in the experimental data. The longer latency of the speech-ABR compared to the click-ABR did in fact not differ significantly between the modelled data and the experimental recordings ($p = 0.6$ and $p = 5 \times 10^{-6}$; two-way ANOVA across the type of dataset and stimulus signal, respectively).

We then explored the relative contributions of different frequency bands in speech to the speech-ABR. In particular, due to the extensive nonlinearities in the inner ear as well as in the subsequent neural responses, combinations of higher harmonics of the fundamental frequency can lead to a response at the fundamental frequency. We therefore wondered how much the fundamental frequency in speech itself as well as higher harmonics contributed to the speech-ABR.

We divided the speech signal into a low-frequency band that contained the fundamental frequency, as

well as into higher frequency bands that were equally broad on a logarithmic scale (figure 4). This choice of frequency bands reflects the approximately logarithmic tonotopic map of the inner ear. Furthermore, we scaled the amplitude of the speech stimulus in each frequency band so that all bandpass-filtered signals had the same intensity. Any differences in the resulting brainstem response were consequently due to the neural processing, and not due to amplitude variation in the speech input. We found that the low frequency band that contained the fundamental frequency yielded the highest contribution to the speech-ABR, while the higher frequency bands up to about 8.3 kHz yielded major contributions as well. The amplitudes and signal-to-noise ratios that resulted from the higher frequency bands were approximately equal, up to a frequency of 8.3 kHz. The latency of the brainstem response tended to shorten with increasing frequency.

We further explored the impact of the fundamental frequency itself on the resulting speech-ABR.

To this end we employed speech segments in which the fundamental frequency and the higher harmonics were shifted from lower to higher values. The resulting amplitude of the speech-ABR decreased continuously with increasing fundamental frequency (figure 5). However, both the latency as well as the signal-to-noise ratio remained relatively unchanged. To compare the modelled amplitude's dependence on the fundamental frequency to experimental data, we considered two fundamental frequencies that were 50 Hz apart and that approximately corresponded to those of a male and a female voice for which we had recorded brainstem responses previously [32]. We found that the drop in amplitude was comparable between the modelled responses and the experimental measurements ($p = 0.9$ and $p = 0.003$; two-way ANOVA across the type of dataset and the fundamental frequency, respectively).

4. Discussion

We modelled the human brainstem response at the fundamental frequency of continuous speech (speech-ABR). The simulated brainstem response matched our recent experimental findings regarding the shape of the response, the width and the latency of about 8 ms ([13, 15, 32]; figure 3). In contrast to the modelled speech-ABR, however, the amplitude of the experimentally-measured speech-ABR is affected by other neuronal activities such as cortical ones, by neuronal noise and by recording artefacts. The amplitudes of the modelled and of the experimentally-measured speech-ABR can therefore not be directly compared.

Our model showed that, although we modelled three main stages of the brainstem separately, the net response did not allow to resolve the individual neural responses at the different stages. This agreed with our recent experimental observations that showed a single peak of the speech-ABR at a latency of about 8 ms. The comparison with our simulated data suggests that this response originates predominantly in the inferior colliculus.

Our modelling further revealed that the tracking of the fundamental frequency increased along the auditory pathway. This parallels previous findings that showed increased synchrony of neurons in the cochlear nuclei relative to the auditory-nerve fibers [27, 30], and neurons in the inferior colliculus whose rate was tuned to specific modulation frequencies, further increasing phase locking around the fundamental frequency of the speech [19, 30].

At the same time, the frequency range of the response decreased along the brainstem, corresponding to the well-known low-pass nature of the brainstem response and in agreement with the phase-locking limit up to which FFR can be recorded [43, 44]. Indeed, the brainstem can respond particularly well around the fundamental

frequency of speech and its lower harmonics, a process described as 'formant capture' by which harmonics adjacent to the formant regions are emphasized [19].

The observed latency of the speech-ABR also agrees with the experimental findings on the brainstem response of the frequency-following response to a pure tone as well as shorter speech tokens [9, 45]. Although recent MEG and EEG investigations have found that the frequency-following response as well as the neural response to the fine structure of speech can also have cortical contributions, these presumably occur at latencies beyond 12 ms [18]. Moreover, their sensitivity is presumably reduced for frequencies larger than 100 Hz [18, 46, 47]. Our model did not include cortical generators, and the resulting model data, together with our recent experimental findings, therefore suggest that the EEG-measured response at the fundamental frequency of speech is dominated by subcortical contributions [18].

The modelled speech-ABR occurred later than wave V of the modelled click-ABR, a finding that is in agreement with experimental observations [48–50]. Click-ABRs are elicited by very short stimuli that have a broad frequency spectrum. These stimuli are very different from the continuous speech stimuli that cause the speech-ABR. The higher frequency content of clicks leads to more activation of the high-frequency auditory-nerve fibers that originate from the cochlear base and that are activated earlier than the lower-frequency fibers from more apical cochlear locations, presumably causing a shorter latency of the response. Moreover, natural speech has a higher variability and less regularity than clicks which may introduce interaction and contrast across the neural population, increasing the variability in the latency of the speech-ABR. Our modelling results therefore add further evidence to the growing literature that indicates that onset responses of the brainstem differ from responses to sustained stimuli such as speech [48, 51].

Although we have focussed on the response of the brainstem at the fundamental frequency itself, our modelling work showed that this response does not only come from the speech's fundamental frequency itself, but also from the many higher harmonics. The contribution of these higher harmonics reflects the highly nonlinear processing in the auditory periphery, starting from the cochlea's compressive nonlinearity [42, 43]. It underlies the previous finding that the brainstem responds at the fundamental frequency of the speech even if this frequency has been filtered out of the stimulus [5, 6].

Specifically, our model revealed that harmonics up to 8.3 kHz still yielded a contribution to the speech-ABR. Interestingly, frequency bands that had equal width on a logarithmic scale yielded approximately equally large speech-ABRs. This accords with different studies that point out that the auditory brainstem response is predominantly generated from

auditory-nerve fibers with high characteristic frequencies [52]. At high stimulation levels, as employed here, there is indeed a significant spread of excitation to the basal cochlear region, even for low-frequency stimulation. This is in agreement with previous computational work that highlighted the importance of off-frequency contributions and the importance of the basal locations for low-frequency tone processing as a result of cochlear dispersion and nonlinearities [22]. The contribution of higher harmonics also accords with a number of studies that evaluated the relative contribution of resolved and unresolved harmonics in pitch tracking [53, 54] and demonstrates the different mechanisms of pitch coding. Indeed, it is believed that resolved harmonics may contribute to pitch tracking mostly due to the phase locking to the temporal fine structure in the auditory nerve [26, 55]. Interaction between nearby unresolved harmonics, on the other hand, may lead to beating at the fundamental frequency and to a corresponding response in the auditory nerve fibers [56, 57].

Although a large number of harmonics contributed to the response, the evoked brainstem response decreased above about 8 kHz. Moreover, the contributions of the different frequency bands below 8 kHz did not add up linearly to the overall neural response. Instead, the brainstem response to the whole speech had a magnitude that was approximately similar to the response to the different individual frequency bands below 8 kHz. This behaviour presumably reflects the different nonlinearities and non-monotonic responses in the auditory-nerve fibers that can lead to negative interference and cancellation, in particular at higher stimulation intensities [58, 59].

It is well known that higher frequencies, either of a pure tone, in speech tokens or in a musical note, cause a smaller brainstem response, presumably caused by less phase locking in neurons in response to higher frequencies [2, 45, 60, 61]. Our own previous experimental work as well as that of others similarly found smaller brainstem responses to a female voice than to a male voice, presumably due to the higher fundamental frequency of the female voice [14, 32]. Our computational results reported here recapitulated this effect: we found a degradation of the speech-ABR as the temporal fine structure of the speech shifted towards higher frequencies. In particular, our model simulations predicted a 24% reduction of the response, relative to the male voice, when the input speech is shifted by 50 Hz. This reduction coincides with our previously reported brainstem responses to a male and a female voice, whose fundamental frequencies differed by about 50 Hz on average [32]. The agreement between our model results and previous experimental data suggests that the model results depend on the particular voice only through the fundamental frequency, but that they are not voice-dependent otherwise.

In summary, we have explored a computational model of the brainstem response at the fundamental frequency of continuous speech. We have shown that the model successfully produces many experimental observations of the speech-ABR, in particular regarding its latency, its shape, its origin from many higher harmonics of the speech signal, and its decrease with increasing temporal fine structure of the speech signal. In return, the model showed that the speech-ABR is dominated by the neural responses in the inferior colliculus as opposed to the cochlear nuclei or the auditory-nerve fibers. Furthermore, our results highlight how the speech-ABR is shaped by nonlinearities, by interaction between different harmonics, and by contributions from a large number of frequency bands.

In the future our computational model may allow for an improved experimental detection of the brainstem response to speech. As an example, our model may allow to compute the brainstem response to further speech features, and the modelled response may then be used for the experimental detection. Moreover, future work may employ the model to explore effects of different types of hearing impairment on the speech-ABR, which may lead to better diagnostic tests of hearing loss and speech-in-noise deficits.

Acknowledgments

This research was supported by the Royal British Legion Centre for Blast Injury Studies, from the 'la Caixa' Foundation (LCF/BQ/EU15/10350044) as well as by EPSRC grants EP/M026728/1 and EP/R032602/1.

ORCID iD

Tobias Reichenbach  <https://orcid.org/0000-0003-3367-3511>

References

- [1] Smith J C, Marsh J T and Brown W S 1975 Far-field recorded frequency-following responses: evidence for the locus of brainstem sources *Electroencephalography Clin. Neurophysiol.* **39** 465–72
- [2] Chandrasekaran B and Kraus N 2010 The scalp-recorded brainstem response to speech: neural origins and plasticity *Psychophysiology* **47** 236–46
- [3] Hood L J 1998 *Clinical Applications of the Auditory Brainstem Response* (San Diego, CA: Singular Publishing) pp 49–63
- [4] Hall J W 2006 Introduction to auditory evoked response measurement *New Handbook for Auditory Evoked Responses* (Boston: Pearson) pp 58–108
- [5] Galbraith G C and Doan B Q 1995 Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli *Int. J. Psychophysiol.* **19** 203–14
- [6] Galbraith G C 1994 Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials* **92** 321–30

- [7] Sohmer H, Pratt H, Kinarti R, Sohmer H, Pratt H and Kinarti R 1977 Sources of frequency following responses (FFR) in man *Electroencephalogr Clin. Neurophysiol.* **42** 656–64
- [8] Roenne F M, Harte J M, Dau T, Ronne F M, Harte J M and Dau T 2013 Modelling human auditory evoked brainstem responses to speech syllables *Proc. Meet Acoust.* **19** 050120
- [9] Skoe E and Kraus N 2010 Auditory brainstem response to complex sounds: a tutorial *Ear. Hear.* **31** 302–24
- [10] Kraus N and Chandrasekaran B 2010 Music training for the development of auditory skills *Nat. Rev. Neurosci.*
- [11] Parbery-Clark A, Marmel F, Bair J and Kraus N 2011 What subcortical-cortical relationships tell us about processing speech in noise *Eur. J. Neurosci.* **33** 549–57
- [12] BinKhamis G, Elia Forte A, Reichenbach T, O'Driscoll M and Kluk K 2019 Speech auditory brainstem responses in adult hearing aid users: effects of aiding and background noise, and prediction of behavioral measures *Trends Hear.* **23** 2331216519848297
- [13] Forte A E, Etard O and Reichenbach T 2017 The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention *Elife* **6** e27203
- [14] Maddox R K and Lee A K C 2018 Auditory brainstem responses to continuous natural speech in human listeners *Eneuro* **5** 0441-17
- [15] Etard O, Kessler M, Braiman C, Forte A E and Reichenbach T 2019 Decoding of selective attention to continuous speech from the human auditory brainstem response *Neuroimage* **200** 1–11
- [16] Coffey E B J, Herholz S C, Chepesiuk A M P, Baillet S and Zatorre R J 2016 Cortical contributions to the auditory frequency-following response revealed by MEG *Nat. Commun.* **7** 1–11
- [17] Coffey E B J, Musacchia G and Zatorre R J 2017 Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence *J. Neurosci.* **34** 830–8
- [18] Bidelman G M 2018 Subcortical sources dominate the neuroelectric auditory frequency-following response to speech *Neuroimage* **175** 56–69
- [19] Krishnan A 2002 Human frequency-following responses: representation of steady-state synthetic vowels *Hear. Res.* **166** 192–201
- [20] Young E D and Sachs M B 1979 Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers *J. Acoust. Soc. Am.* **66** 1381–403
- [21] Verhulst S, Altoè A and Vasilkov V 2018 Computational modeling of the human auditory periphery: auditory-nerve responses, evoked potentials and hearing loss *Hear. Res.* **360** 55–75
- [22] Encina-Llamas G, Harte J M, Dau T, Shinn-Cunningham B and Epp B 2019 Investigating the effect of cochlear synaptopathy on envelope following responses using a model of the auditory nerve *J. Assoc. Res. Otolaryngol.* **20** 363–82
- [23] Dau T 2003 The importance of cochlear processing for the formation of auditory brainstem and frequency following responses *J. Acoust. Soc. Am.* **113** 936–50
- [24] Schaette R and McAlpine D 2011 Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model *J. Neurosci.* **31** 13452–7
- [25] Verhulst S, Bharadwaj H M, Mehraei G, Shera C A and Shinn-Cunningham B G 2015 Functional modeling of the human auditory brainstem response to broadband stimulations *J. Acoust. Soc. Am.* **138** 1637–59
- [26] Carney L H, Li T and McDonough J M 2015 Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations *eNeuro* **2** 1–12
- [27] Rudnicki M and Hemmert W 2017 High entrainment constrains synaptic depression levels of an in vivo globular bushy cell model *Front Comput. Neurosci.* **11** 16
- [28] Zilany M S A, Bruce I C, Nelson P C and Carney L H 2009 A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics *J. Acoust. Soc. Am.* **2014** 2390
- [29] Zilany M S A, Bruce I C and Carney L H 2014 Updated parameters and expanded simulation options for a model of the auditory periphery *J. Acoust. Soc. Am.* **135** 283–6
- [30] Rhode W S and Greenberg S 1994 Encoding of amplitude modulation in the cochlear nucleus of the cat *J. Neurophysiol.* **71** 1797–825
- [31] Nelson P C and Carney L H 2004 A phenomenological model of peripheral and central neural responses to amplitude-modulated tones *J. Acoust. Soc. Am.* **116** 2173–86
- [32] Saiz-Alía M, Forte A E and Reichenbach T 2019 Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits *Sci. Rep.* **9**
- [33] Rudnicki M, Schoppe O, Isik M, Völkl F and Hemmert W 2015 Modeling auditory coding: from sound to spikes *Cell Tissue Res.* **361** 159–75
- [34] Greenwood D D 1990 A cochlear frequency-position function for several species—29 years later *J. Acoust. Soc. Am.* **87** 2592–605
- [35] Liberman M C 1978 Auditory-nerve response from cats raised in a low-noise chamber *J. Acoust. Soc. Am.* **63** 442–55
- [36] Rothman J S and Manis P B 2003 Differential expression of three distinct potassium currents in the ventral cochlear nucleus *J. Neurophysiol.* **89** 3070–82
- [37] Rothman J S and Manis P B 2003 Kinetic analyses of three distinct potassium conductances in ventral cochlear nucleus neurons *J. Neurophysiol.* **89** 3083–96
- [38] Rothman J S and Manis P B 2003 The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons *J. Neurophysiol.* **89** 3097–113
- [39] Oertel D 1983 Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus *J. Neurosci.* **3** 2043–53
- [40] Picton T 2011 Chapter 8: Auditory brain- stem responses: peaks along the way *Human Auditory Evoked Potentials* (San Diego, CA: Plural Publishing) pp 213–245
- [41] Huang H and Pan J 2006 Speech pitch determination based on Hilbert-Huang transform *Signal Process.* **86** 792–803
- [42] Mehraei G, Hickox A E, Bharadwaj H M, Goldberg H, Verhulst S, Liberman M C and Shinn-Cunningham B G 2016 Auditory brainstem response latency in noise as a marker of cochlear synaptopathy *J. Neurosci.* **36** 3755–64
- [43] Moushegian G, Rupert A L and Stillman R D 1973 Scalp-recorded early responses in man to frequencies in the speech range *Electroencephalogr Clin. Neurophysiol.* **35** 665–7
- [44] Aiken S J and Picton T W 2008 Envelope and spectral frequency-following responses to vowel sounds *Hear. Res.* **245** 35–47
- [45] Batra R, Kuwada S and Maher V L 1986 The frequency-following response to continuous tones in humans *Hear. Res.* **21** 167–77
- [46] Tichko P and Skoe E 2017 Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators *Hear. Res.* **348** 1–15
- [47] Bidelman G M and Momtaz S 2020 Subcortical sources drive the relation between frequency-following responses (FFRs) and speech-in-noise perception (submitted) (<https://doi.org/10.1101/2020.03.29.014233>)
- [48] Song J H, Banai K, Russo N M and Kraus N 2006 On the relationship between speech- and nonspeech-evoked auditory brainstem responses *Audiol Neurotol* **11** 233–41
- [49] Karawani H and Banai K 2010 Speech-evoked brainstem responses in Arabic and Hebrew speakers *Int. J. Audiol.* **49** 844–9
- [50] Johnson K L, Nicol T, Zecker S G and Kraus N 2008 Developmental plasticity in the human auditory brainstem *J. Neurosci.* **28** 4000–7
- [51] Bidelman G M 2015 Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR *Hear. Res.* **323** 68–80

- [52] Abdala C and Folsom R C 1995 The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking *J. Acoust. Soc. Am.* **98** 921–30
- [53] Jeng F C, Costilow C E, Stangherlin D P and Lin C D 2011 Relative power of harmonics in human frequency following responses associated with voice pitch in American and Chinese adults *Percept Mot. Skills* **113** 67–86
- [54] Peng F, McKay C M, Mao D, Hou W and Innes-Brown H 2018 Auditory brainstem representation of the voice pitch contours in the resolved and unresolved components of Mandarin tones *Front Neurosci.* **12** 820
- [55] Carney L H and McDonough J M 2019 Nonlinear auditory models yield new insights into representations of vowels *Attention, Perception, and Psychophysics* **81** 1034–46
- [56] Plack C J and Plack C J 2018 Pitch and periodicity coding *The Sense of Hearing* pp 128–48
- [57] Oxenham A J 2018 How we hear: the perception and neural coding of sound *Annu. Rev. Psychol.* **69** 27–50
- [58] Bruce I C, Sachs M B and Young E D 2003 An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses *J. Acoust. Soc. Am.* **113** 369
- [59] Zilany M S A and Bruce I C 2006 Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery *J. Acoust. Soc. Am.* **120** 1446–66
- [60] Palmer A R and Russell I J 1986 Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells *Hear. Res.* **24** 1–15
- [61] Musacchia G, Sams M, Skoe E and Kraus N 2007 Musicians have enhanced subcortical auditory and audiovisual processing of speech and music *Proc. Natl. Acad. Sci. USA* **104** 15894–8