Figures and figure supplements

The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention

Antonio Elia Forte et al
Figure 1. The brainstem response to running speech. (a) Speech (black) contains voiced parts with irregular oscillations at a time-varying fundamental frequency and higher harmonics. We extract a fundamental waveform (red) that oscillates nonlinearly and nonstationary at the fundamental frequency. (b) The autocorrelation of the fundamental waveform (red) peaks when the delay vanishes and oscillates at the average fundamental frequency. The cross-correlation of the fundamental waveform with its Hilbert transform (blue) can be seen as an imaginary part of the autocorrelation. The amplitude of the resulting complex cross-correlation (black) shows a life-time of a few ms. (c) The correlation of the speech-evoked brainstem response, recorded from one subject, to the fundamental waveform of the speech signal (red) as well as to its Hilbert transform (blue) can serve as real and imaginary parts of a complex correlation function. Its amplitude (black) peaks at a latency of 9 ms. The latency of the correlation is not altered by the processing of the speech signal or of the neural recording, and contains neither a stimulus artifact nor the cochlear microphonic (Figure 1—figure supplement 1).

DOI: https://doi.org/10.7554/eLife.27203.002
Figure 1—figure supplement 1. Controls for latencies induced by signal processing as well as for the source of the measured brainstem response to running speech. (a) The cross-correlation between the original speech signal with the fundamental waveform (red) as well as with its Hilbert transform.
(blue) and the resulting amplitude (black) show a peak at 0 ms and no phase shift. The processing of the acoustic signal does accordingly not change the latency or phase of that signal. (b) The computation of the cross-correlation of the fundamental waveform to the neural recording involved processing of the neural signal such as through filtering. However, the cross-correlation between the recorded neural signal and the filtered version shows a peak at vanishing latency. The processing of the neural signal did therefore not alter the latency. (c) When the earphones are placed close to the ears, but not inside the ear canal, preventing a subject from hearing the speech signal, the cross-correlation between the recorded neural signal and the fundamental waveform of speech (red) as well as its Hilbert transform (blue) do not yield a measurable peak. The amplitude of the resulting complex correlation function (black) does not peak either, demonstrating the absence of a stimulus artifact. (d) When a subject listened to a speech signal and then to the same signal with reversed polarity, and when the average over the neural recordings to both stimulus presentations was employed for the analysis, the complex cross-correlation showed the same structure as when it was computed using the neural response to one stimulus only. This shows the absence of a stimulus artifact as well as the absence of the cochlear microphonic in the measured response. (e) Putative cortical contributions to the neural response would occur at latencies above 15 ms and likely between 50–500 ms. The complex correlation at those latencies shows, however, no significant peak. To enable comparison, all recordings were obtained from the same subject for whom we report the exemplary recording in Figure 1c.

DOI: https://doi.org/10.7554/eLife.27203.003
Figure 2. Simplistic model of the auditory brainstem response to continuous speech. (a) The fundamental waveform (red) as well as its Hilbert transform (blue) oscillate with a varying amplitude and frequency. (b) We model a simplistic brainstem response in which bursts of neural spikes occur at each cycle of the fundamental waveform, at a phase of $\frac{1}{4} \pi$ rad (black dots). Furthermore, all neural bursts are shifted by a temporal delay of 8 ms. (c) When adding realistic noise as emerges from scalp recordings, and then computing the complex correlation with the fundamental waveform as performed for the actual brainstem recording, we find a peak at the modelled delay of 8 ms. The modelled phase of $\frac{1}{4} \pi$ rad is obtained as the inverse phase of the complex correlation at that latency.

DOI: https://doi.org/10.7554/eLife.27203.004
Figure 3. Modulation of the brainstem response to speech by selective attention. (a) The brainstem's response to the male speaker is larger for each subject when attending the speaker (dark blue) than when ignoring it (light blue). The average ratio of the brainstem responses to the attended and to the ignored male speaker is significantly larger than 1 (black, mean and standard error of the mean). (b) With the exception of subject 13, the neural response to the female voice is also larger when subjects attend to it (dark red) instead of ignoring it (light red). The average ratio of the brainstem responses to the attended and to the ignored female speaker is significantly larger than one as well (black, mean and standard error of the mean).

DOI: https://doi.org/10.7554/eLife.27203.005
Figure 3—figure supplement 1. Correlation between the amplitude of the brainstem response and the fundamental frequency of the speech signal. (a) The correlation between the amplitude of the brainstem’s response to a single speaker and its fundamental frequency is slightly negative. (b, c) The correlation coefficient is shown in each panel. The data points represent individual measurements, and the trend line indicates the direction of the correlation. R values are given for each panel:

- Panel (a): $R = -0.16988$
- Panel (b): $R = -0.23035$
- Panel (c): $R = -0.16397$

Figure 3—figure supplement 1 continued on next page
amplitude of the brainstem's response to an attended (b) as well as an ignored (c) speaker decreases slightly with increasing fundamental frequency as well. The range of fundamental frequencies in the competing speaker scenarios (b, c) is larger than in the single speaker situation (a) since the latter consists of a single female speaker only whereas the former contain both a male and a female speaker, with the female speaker having a fundamental frequency that is on average larger than that of the male speaker.

DOI: https://doi.org/10.7554/eLife.27203.006