Speech-evoked ABR: Effects of age and simulated neural temporal jitter

Sara K. Mamo[1,i], **John H. Grose**[2], and **Emily Buss**[2]

[1]Division of Speech and Hearing Sciences, University of North Carolina–Chapel Hill

[2]Department of Otolaryngology-Head & Neck Surgery, University of North Carolina–Chapel Hill

1. Introduction

Older adults who have clinically normal audiometric hearing sensitivity nevertheless often experience difficulty with speech recognition, especially in noise. Given the ubiquity of this observation, it is generally understood that speech perception abilities can deteriorate as a function of age independently of hearing loss as determined by an audiogram. Many studies of older adults with normal hearing sensitivity point to auditory temporal processing deficits as a contributing factor to decreased speech perception performance (Bergman, 1980; Strouse et al., 1998; Gordon-Salant & Fitzgibbons, 2001; Dubno et al., 2002; Dubno et al., 2003; Pichora-Fuller et al., 2007; Grose et al., 2009).

Age-related deficits have been measured across a range of temporal processes, including encoding of both transient and sustained complex signals. Among these temporal processing deficits is reduced periodicity coding, which is thought to affect pitch tracking and, consequently, perceptual segregation of simultaneous sounds (Vongpaisal & Pichora-Fuller, 2007). Investigating periodicity coding allows for consideration of both temporal envelope and fine structure cues critical to speech understanding in noise. A speech signal is made up of the slower amplitude modulations of the overall stimulus (i.e., temporal envelope) carried by the more rapid pressure oscillations (i.e., temporal fine structure). While temporal envelope cues are sufficient for speech intelligibility in quiet (Shannon et al., 1995), some argue that it is the fine structure coding that is necessary for speech intelligibility in noisy, complex backgrounds because temporal fine structure supports accurate segregation of the target and masker (Divenyi & Simon, 1999; Schneider & Pichora-Fuller, 2001; Pichora-

Corresponding author: Sara K. Mamo, Johns Hopkins Center on Aging and Health, 2024 E. Monument Street, Suite 2-700, Baltimore, MD, 21205; smamo1@jhmi.edu; 1-443-287-7195.
[i]Present address: Department of Otolaryngology-Head & Neck Surgery, Johns Hopkins School of Medicine

[1]Note that these eight subjects did not need to be excluded from the FFRenv analysis due to the cancelation of stimulus artifact when the alternating polarities are added.

Fuller & Souza, 2003; Hopkins & Moore, 2009). Differentiating these two characteristics of the speech signal in order to assess their separate contributions to speech perception difficulties is challenging because isolating one type of cue tends to disrupt other spectro-temporal cues (Apoux et al., 2013). Given that both temporal envelope and fine structure components likely contribute to speech perception in complex backgrounds, further understanding of the mechanisms underlying the processing of these cues can help shed light on speech perception deficits in older adults.

Evidence of reduced periodicity coding for temporal envelope and fine structure cues in older adults can be found in psychophysical, speech perception, and electrophysiological research. In terms of psychophysical studies, poor periodicity coding in older adults has been found using measures of temporal modulation transfer function (He et al., 2008), low-rate frequency modulation detection (He et al., 2007; Grose & Mamo, 2012), and interaural phase/time differences (Ross et al., 2007; Grose & Mamo, 2010). In addition, Abel et al. (1990) found age-related differences in frequency difference limens at low frequencies for older adults with normal hearing and suggested that a loss of neural synchrony may explain the age-related deficits. Of particular interest to the current investigation, Pichora-Fuller and Schneider (1992) tested a masking level difference (MLD) paradigm and proposed that reduced temporal resolution in the aging auditory system could be modeled as increased neural jitter. That is, diminished neural phase locking or synchrony at the level of the auditory nerve fibers and/or brainstem could lead to a decline in periodicity coding in older adults. Pichora-Fuller and Schneider (1992) investigated this using a psychophysical MLD experiment where the masking release relies, in part, on interaural time differences. When external stimulus delays were applied independently to right and left ears in a dichotic MLD condition, the thresholds of younger adults were affected by the external delay imposed on the stimulus. However, the response pattern of the older adults did not change as a result of the external delays. The interpretation of the finding that external delays did not further degrade the response pattern for the older adult group was that the fixed degree of external delay was small relative to the magnitude of internal jitter in the older listeners but not in the younger listeners.

Pichora-Fuller et al. (2007) developed a simulation of the hypothesized age-related increase in neural jitter and applied it to complex stimuli in an algorithm that manipulated the lower frequency region of a complex signal (where phase-locking is likely to play a role). The manipulation systematically corrupted the periodicity of the stimulus. In effect, the temporal jitter algorithm simulated the internal loss of neural synchrony by externally corrupting the complex periodic stimulus. Motivation for this came in part from the observation by Vongpaisal and Pichora-Fuller (2007) that fundamental frequency (F_0) difference limens (DLs) were elevated in older listeners, accompanied by diminished performance in a concurrent vowel identification task. They suggested that temporal synchrony deficits might explain poor F_0 and harmonicity coding leading to difficulty in vowel segregation. The basis for this is that concurrent vowel segregation relies strongly on pitch tracking of the F_0 of each harmonic complex, and this depends on periodicity coding of the complex speech stimulus. Specifically, pitch tracking allows a listener to use the pitch associated with a speaker's F_0 to segregate target speech from the competing background. Given that the pitch of harmonic complexes like voiced speech is dominated by the lower-numbered, often-

resolved harmonics (e.g., Bernstein and Oxenham, 2003) it is to be expected that reduced neural phase locking should impair precise representation of the F_0 and its associated harmonics. In support of this, Russo et al. (2012) found that some older listeners with normal audiograms were poorer than young listeners at perceiving melodic mistunings for melodies carried by lower-numbered harmonics, and that this could be well described in terms of a model of neural periodicity coding in which the degree of phase locking was reduced leading to reduced pitch strength. Pichora-Fuller et al. (2007) applied the temporal jitter algorithm to speech and demonstrated that when the periodicity of the speech material was disrupted, the performance of younger adults could be degraded to the point that it matched that of older adults listening to uncorrupted speech in a speech-in-noise task. This supports the hypothesis that diminished periodicity coding in older listeners contributes to their speech-in-noise difficulties. Later studies applying the jitter algorithm to speech have further supported this hypothesis (MacDonald et al., 2010; Smith et al., 2012). Given this successful simulation of increased neural jitter in the aging auditory system for speech perception in noise, the purpose of the present study was to incorporate the temporal jitter model into an electrophysiological investigation of neural synchrony coding in the aging auditory system.

The assumption underlying the present study is that the periodicity reflected in electrophysiological responses evoked by complex harmonic stimuli depends both on the periodicity of the stimuli and the degree of neural synchrony inherent in the auditory system. Due to a reliance on precise neural synchrony, auditory brainstem response (ABR) measures lend themselves well to investigations of age-related changes in neural phase locking and neural jitter. Moreover, passive electrophysiological measures are informative in trying to distinguish effects related to sensory encoding from more central and cognitive aspects of auditory processing. Evidence from electrophysiological measures of periodicity coding has supported the hypothesis of reduced fidelity of temporal envelope (Purcell et al., 2004; Leigh-Paffenroth & Fowler, 2006; Grose et al., 2009) and fine structure (Ross et al., 2007; Clinard et al., 2010; Grose & Mamo, 2012) encoding as a function of age. Whereas response characteristics from older adults in brainstem measures evoked by tonal stimuli are quite well established, relatively little is known about the brainstem response to complex stimuli.

Recent research has shown reduced temporal envelope and fine structure encoding at the level of the brainstem in older adults when presented with a synthetic /da/ stimulus (Anderson et al., 2012; Ruggles et al., 2012). The use of speech stimuli provides an opportunity to investigate encoding of the F_0 and associated harmonics, which are important components for speech perception (Vongpaisal & Pichora-Fuller, 2007). The electrophysiological response to the temporal envelope or fine structure can be differentially enhanced by adding or subtracting responses to alternating polarities, respectively (Aiken & Picton, 2008). Typically, the temporal envelope response is accentuated by summing responses to stimuli presented with alternating polarities, and the fine structure response is accentuated by taking the difference between responses to stimuli with opposite polarities. In the speech-evoked ABR (sABR) literature, the temporal envelope response highlights the fundamental frequency and the lower harmonics of the stimulus, while the fine structure response highlights mid-to-upper harmonics. While there are many response features of interest in the sABR recording (e.g., onset, transition, sustained features), the current

investigation will be constrained to the sustained periodic portion of the response in order to focus on synchronicity elicited by relatively steady spectral stimulation. For a more comprehensive overview of sABR response components in older adults, the reader is referred to Anderson and colleagues (2012).

The purpose of this study is to assess age-related effects in the strength of periodicity coding, presumed to reflect the degree of neural phase locking. These effects will be examined using the sABR evoked by a /da/ speech token in younger and older adults with normal hearing sensitivity (Experiment 1). A second step of this investigation will apply the temporal jitter model of Pichora-Fuller and colleagues to systematically degrade the response pattern of the younger adults (Experiment 2). For the intact stimulus, the hypothesis is that the evoked response of older adults will exhibit reduced spectral magnitude of the harmonic components associated with the periodic portion of the stimulus. In addition, it is hypothesized that applying the model's temporal jitter algorithm to the sABR stimulus will result in younger listeners exhibiting response characteristics similar to those observed in older listeners obtained using non-jittered sABR stimuli.

2. Experiment 1: Effects of Age on the Encoding of Periodicity in the sABR

The purpose of Experiment 1 was to test the hypothesis that periodicity coding, as characterized by sABR recordings, declines as a function of age. This was tested by recording sABRs in response to a /da/ stimulus in normal hearing younger and older adults.

2.1. Methods

2.1.1. Participants—Two groups of listeners participated: Older adults (n = 22; 15 women) and younger adults (n = 22; 17 women). The mean age of the older adults was 70.5 years (s.d. = 4.8 years) with a range from 65 to 80 years old. Younger adults ranged from 18 to 30 years old with a mean age of 23.2 years (s.d. = 3.2 years). All listeners had thresholds 20 dB HL from 250–4000 Hz in the test ear (Figure 1). There were four exceptions among the older adults: one had a threshold of 25 dB HL at 2000 Hz, two had a threshold of 25 dB HL at 4000 Hz, and one had a threshold of 30 dB HL at 4000 Hz. Test ear was chosen based on best hearing thresholds for the older adult subjects (right ear = 11) and assigned to balance right/left presentation for the younger adults (right ear = 12). All participants signed a consent form approved by the UNC-Chapel Hill Institutional Review Board and were paid for their participation.

2.1.2. Stimulus—A 170-ms synthetic speech stimulus (/da/), provided by the Auditory Neuroscience Laboratory at the Northwestern University School of Communication, was employed in this study. The speech token contains a stop burst, a formant transition, and a 120-ms steady vowel. The F_0 is 100 Hz throughout the stimulus. The first three formants shift during the initial 50-ms portion of the stimulus. Formant 1 rises from 400 to 720 Hz, formant 2 falls from 1700 to 1240 Hz, and formant 3 falls from 2580 to 2500 Hz; subsequently, formants 1–3 remain constant during the steady vowel. Formants 4–6 are fixed at 3300, 3750, and 4900 Hz, respectively, over the entire 170-ms stimulus (Anderson et al., 2012). Presentation level was 80 dB peak equivalent SPL. Stimuli were presented with alternating polarity at a rate of 3.9/second through a shielded insert ER2 earphone (Etymotic

Research, Inc.; Elk Grove Village, IL). Mu-metal casing and electrical shielding tape (3M; Moncure, NC) were used to shield the transducers, as well as the cables within the sound booth (Campbell et al., 2012). The stimulus was controlled via a custom MATLAB (Mathworks; Natick, MA) program and output through a Tucker-Davis-Technologies (TDT; Alachua, FL) digital signal processor, which also sent a time-locked trigger to a SynAmpsRT Neuroscan 4.3 recording system (Compumedics; Charlotte, NC).

2.1.3. Recording—Electrophysiological recordings were collected using a SynAmpsRT Neuroscan 4.3 system. A bipolar, single-channel recording with a vertical/midline electrode montage was used, with the non-inverting electrode at Fz, the inverting electrode at the nape of neck, and the ground electrode at Fpz. Recordings were digitized at a sampling rate of 20,000 Hz. The continuous EEG recordings were bandpass filtered online with cutoff frequencies of 0.5 and 3000 Hz. Offline digital bandpass filters with cutoff frequencies of 100 and 3000 Hz with a 12-dB/octave roll-off were applied before analysis. Artifact rejection was applied to any epoch exceeding ±35 µV. Stimuli were presented with alternating polarity and approximately 3000 sweeps per stimulus polarity were collected; the final number of sweeps included in statistical analyses was determined after offline filtering and artifact rejection (see Section 2.1.4.). A total of 6000 sweeps per stimulus condition were collected in 3 blocks of 2000 sweeps. The conditions were interwoven randomly to avoid order effects related to subject resting state. The recordings that passed artifact rejection in each block of 2000 sweeps were combined after offline analysis using the weighted average transform in the Neuroscan 4.3 editing software, which adjusts the average based on the number of sweeps in each file. The participant rested in a recliner in a sound-attenuated booth and was free to either sleep or watch a silent movie of choice with subtitles. Recordings were made during one or two sessions lasting no more than 2 hours each. Stimulus presentation and recording parameters were largely based on the tutorial from Skoe and Kraus (2010).

2.1.4. Analysis—All analyses were done offline. For this investigation of periodicity coding, only the response to the sustained portion of the 170-ms /da/ stimulus was analyzed by windowing the final 120 ms of the response. The recording associated with each stimulus presentation was first excised from the continuous EEG trace by windowing from –40 ms to 190 ms relative to stimulus onset; this was then truncated to an epoch extending from 60 to 180 ms for analysis of the response to the periodic portion of the stimulus. The responses to this sustained portion of the stimulus were analyzed in two ways: in terms of the frequency following response to the envelope (FFRenv), and in terms of the frequency following response to the fine structure (FFRtfs). To isolate the FFRenv, the responses to the alternating stimulus polarities were added together, thus largely canceling the fine structure components and enhancing the envelope response. To extract the FFRtfs, the average responses associated with each polarity were subtracted. In both cases, the responses were analyzed using Fourier analysis, and the grand mean responses were computed in the frequency domain. Frequency resolution for these analyses was 8.3 Hz/bin. A signal-to-noise ratio (SNR dB) was computed by comparing the amplitudes of the fundamental frequency and each harmonic component to the noise floor surrounding each of these respective components. A noise floor value was calculated for each response component by

averaging the amplitude from four frequency bins over a +/− 25 Hz range, excluding the frequency bins immediately adjacent to that containing the component of interest. A criterion of 3 dB SNR was used to determine whether a response was reliably present. Any SNR value < 3 dB was replaced with a value of 0 dB for all statistical analyses.

Statistical analysis of the data consisted of repeated-measures analyses of variance (ANOVAs) with frequency-component SNRs as the within-subject factors and age group as the between-subjects factor. The spectral magnitudes for the FFRenv were analyzed using the F_0 and harmonic components 2–6 as the within-subject factor and age group as the between subject factor. The response spectra for the FFRtfs were analyzed using harmonic components 2–7 as the within-subject factor and age group as the between-subject factor. Interactions were probed with pairwise comparisons with Bonferroni adjustments for multiple comparisons. Greenhouse-Geisser corrections for violations of sphericity were applied where appropriate.

2.2. Results

Analysis of the audiometric thresholds of both age groups show significant main effects of frequency ($F(2.50,105.19) = 14.13$, $p < 0.001$) and age group ($F(1,42) = 62.20$, $p < 0.001$), as well as a significant interaction of frequency and age ($F(2.50,105.19) = 18.94$, $p < 0.001$). Further analysis of the interaction via pairwise comparisons with Bonferroni adjustments for multiple comparisons revealed no group differences at the low-frequency thresholds and significant differences between age groups for all test frequencies from 1000–8000 Hz. The differences ranged from < 5 dB at 1000 Hz to about 12 dB at 4000 Hz, close to the upper limit of the spectral content of the steady state portion of the stimulus. Despite age-related differences in thresholds at octave frequencies from 1000–8000 Hz, nearly all thresholds for the older adults (with four exceptions described above) were still within normal audiometric limits (20 dB HL) for 1000–4000 Hz, ensuring audibility of the experimental stimuli.

Approximately 6000 sweeps were collected for each test condition, but the exact number of sweeps submitted to analysis was determined after the application of offline filtering and artifact rejection. For all test conditions there was no group difference in the numbers of sweeps included in the averaging process ($t(42) = −0.83$, $p = 0.41$). Across all subjects, a mean of 5577 sweeps per individual per condition were included in the analyses.

The grand mean time domain waveform responses are shown for the younger and older adult groups in Figure 2. No analysis was undertaken on the time domain waveforms, but the responses are consistent with previous studies. Figure 3 shows the FFRenv response in the frequency domain with F_0 marked by an open circle, each expected harmonic 2–6 marked by an asterisk, the brackets indicate the interquartile range of the data at each component, and the noise floor depicted by gray shading. It can be seen that the younger adults (upper panel) had higher amplitude responses for all expected response components than the older adults (lower panel). All analyses were computed based on the SNR (dB) of the response amplitude compared to the surrounding noise floor. There was a significant within-subjects main effect of frequency ($F(3.73,156.85) = 7.89$, $p < 0.001$) and a significant between-subjects effect of age ($F(3.73,156.85) = 40.96$, $p < 0.001$). The interaction of age and frequency was not significant ($F(1,42) = 1.69$, $p = 0.16$). These results suggest that the

older adults had less robust encoding of the temporal envelope overall, and this age effect did not depend on frequency. Analysis of the noise floor surrounding each harmonic component showed a main effect of frequency region ($F(3.36,141.06) = 107.66$, $p < .001$), which was expected due to the rise in noise floor with decreasing frequency (see Figure 3). There was no significant difference in noise floor between age groups ($F(1,42) = .29$, $p = 0.59$), but a significant interaction between frequency region and age was observed ($F(3.36,141.06) = 2.68$, $p = 0.04$). However, pairwise comparisons with Bonferroni adjustments for multiple comparisons did not reveal significant differences in noise floor between groups at any harmonic frequency region.

For the FFRtfs analysis, eight younger adults were excluded. These subjects were tested prior to appropriate shielding of the transducer and associated cables. Evidence of residual stimulus artifact was present in control traces collected from these listeners when the electrical signal was delivered to the transducer but the acoustic signal was blocked from reaching the subject's ear. Therefore the responses from these subjects were not included in the analyses.[1] For the remaining 14 younger adults and all 22 older adults, the group mean responses in the frequency domain are shown in Figure 4, with the frequencies associated with harmonics 2–7 marked with asterisks and the interquartile ranges. As with the FFRenv, the younger adults have higher amplitude responses for all expected harmonic components. There was a significant within-subjects main effect of frequency ($F(5,170) = 7.25$, $p < 0.001$) and a significant between-subjects effect of age ($F(1,34) = 28.71$, $p < 0.001$) for the FFRtfs. The interaction of age and frequency was not significant ($F(5,170) = 1.90$, $p = 0.10$). These results suggest poorer encoding of the temporal fine structure for older adults compared to younger adults, which was not dependent on frequency.

Analysis of the noise floor in these spectra yielded a significant main effect of frequency ($F(2.59,88.09) = 83.35$, $p < 0.001$). Although the between-subjects effect of age group was not statistically significant ($F(1,34) = 3.70$, $p = 0.06$), there was a significant frequency-by-age interaction ($F(2.59,99.09) = 3.35$, $p = 0.03$). Pairwise comparisons of the noise floor levels in the test frequency regions between age groups showed significant differences around the third ($p = 0.004$) and seventh ($p = 0.02$) harmonic components. The source of the differences at these specific frequencies is not clear. Given the marginal elevation in noise floor for the older adults in the presence of robust amplitude differences between older and younger adults, the difference in noise floor was not considered to be a confound for analysis of the SNR differences between groups.

The results of Experiment 1 support the hypothesis that periodicity coding in older listeners is less robust than in younger listeners even when audiometric thresholds are largely within normal limits. This is consistent with diminished neural synchrony in the senescent auditory system. The next experiment tests this hypothesis further by implementing a simulation of neural jitter.

3. Experiment 2: Effect of Simulated Neural Jitter on the sABR

Experiment 1 found age-related amplitude effects consistent with those observed previously (Anderson et al., 2012). Experiment 2 incorporated the neural jitter simulation of Pichora-

Fuller et al., (2007) to evaluate the extent to which neural jitter might account for the reduced amplitude sABR in older adults.

3.1 Methods

3.1.1. Participants—All younger adults and a subset of older adults (n = 7) who participated in Experiment 1 also participated in Experiment 2. The average age for the subset of older adults was 68.3 years (s.d. = 2.8 years). One younger adult was excluded from analysis for Experiment 2 due to high artifact (50% of sweeps rejected) on the second day of testing. All other participant characteristics are consistent with Experiment 1.

3.1.2. Stimulus—The /da/ stimulus was presented in its original (non-jittered) form and with three levels of externally applied temporal jitter. The applied jitter was based on the algorithm of Pichora-Fuller and colleagues (2007). In this approach, the jitter algorithm is applied only to frequencies below 1200 Hz. The first step was to filter the stimulus into two frequency bands, one above and one below 1200 Hz. This was achieved by converting the stimulus into the frequency domain using a Fast Fourier Transform (FFT), zeroing out magnitude components associated with frequencies above or below 1200 Hz, and then transforming the array back into the time domain using an inverse FFT.

The jitter is implemented by resampling the time-domain waveform, introducing delays (positive or negative). The delays associated with each point in the time domain array are based on a low-pass Gaussian noise sample. The amplitude of the Gaussian noise at the corresponding data point determines the delay value, and the amplitude of the original /da/ stimulus is then replaced by the /da/ amplitude value at the assigned delay value. After applying the jitter algorithm in the time domain, the stimulus is recombined with the unmodified portion of the original stimulus > 1200 Hz to create the full-frequency time-domain stimulus. This manipulation results in a time waveform with a disrupted amplitude pattern that perturbs the periodicity of the original stimulus.

The low-pass filter cutoff and standard deviation of the Gaussian noise determine the degree of jitter (i.e., delay magnitude and rate of change over time) applied to the stimulus (for details see, Miranda and Pichora-Fuller, 2002). A low pass cutoff of 500 Hz was constant for all jitter conditions in the present experiment. Three levels of temporal jitter were tested using three standard deviations of the noise: mild = 0.25, moderate = 0.5, and extreme = 1.0. The jitter algorithm was implemented using a custom MATLAB script, and a new random jitter was applied to the /da/ stimulus prior to each presentation. The frequency spectra of the stimuli with various degrees of jitter are shown in Figure 5 in order to display the disruption of the periodicity of the signal.

3.3.3. Recording—All recording procedures were the same as Experiment 1.

3.1.4. Analysis—All waveform analysis procedures were the same as Experiment 1. T-tests were performed on the most robust components from the envelope and fine structure responses to compare the jittered responses from the younger adults to the non-jittered responses of the older adults. Components of interest were F_0 and the 2nd harmonic (H2) for the FFRenv analysis (i.e., added polarities), and the 4th harmonic (H4) for the FFRtfs

analysis (i.e., subtracted polarities). These components were chosen based on their being the most robust components in Experiment 1 and in previously reported studies (Anderson et al., 2012; Ruggles et al., 2012). For the subset of older adults included in Experiment 2, paired t-tests were used to compare the non-jittered responses obtained from those listeners in Experiment 1 to the jittered responses obtained from those listeners in Experiment 2.

3.2. Results

Figure 6 shows group mean FFRenv spectra for the mild (middle panel) and extreme (lower panel) jitter conditions for the younger adults; for reference, the response to the non-jittered stimuli from Experiment 1 are also shown (upper panel). The mild jitter reduced the F_0 response amplitude by 32% for the younger adults. The moderate jitter condition is not shown, but the response amplitudes were intermediate to those recorded in the mild and extreme conditions. Visual inspection of the averaged jittered results suggests that for the mild level of jitter, the F_0 component for the younger adults approximates the F_0 component for the older adults in the non-jittered condition (cf. Fig. 2, lower panel). An independent samples, t-test of this comparison yielded no difference between the non-jittered response of older adults and the mild-jittered response for the younger adults for the F_0 component ($t(41) = 0.93$, $p = 0.36$). However at the mild level of jitter, the younger adults still exhibit a more robust response for the H2 component than the H2 component in the non-jittered response of the older adults ($t(37.77) = 5.55$, $p < .001$). Only at the extreme jitter level were there no group differences between younger adults' jittered response and the older adults' non-jittered response for F_0 ($t(41) = -1.10$, $p = 0.28$) and H2 ($t(41) = -0.29$, $p = 0.77$).

Figure 7 shows FFRenv spectra for the older adults for non-jittered and jittered stimuli. Analyses were conducted to compare the non-jittered responses to the jittered responses within the subgroup of older adults who were tested on all conditions. Whereas the younger adults showed a significant decrease in the amplitude at F_0 even at the mild jitter condition, no significant difference in the SNR of F_0 was observed for older adults as analyzed by paired t-tests comparing the responses to non-jittered and any of the jittered stimuli conditions: mild ($t(6) = -0.88$, $p = 0.41$), moderate ($t(6) = 2.13$, $p = 0.06$), and extreme ($t(6) = 1.95$, $p = 0.10$). Again only the mild and extreme conditions are shown, but the responses in the moderate condition were consistent with the other two conditions. Only four of the seven older adults who participated in the jittered conditions had a present response (SNR > 3 dB) for H2 in the non-jittered condition, and so statistical analysis was not completed for the H2 component for the older adults.

Figure 8 shows the FFRtfs spectra for the young listeners for the mild jitter condition (lower panel); for reference, the mean response to the non-jittered stimuli from Experiment 1 is also shown (upper panel). For the FFRtfs, the introduction of mild jitter resulted in a dramatic reduction in the response amplitudes of all harmonic components. Analysis of the most robust component (H4) for the FFRtfs showed that at the mild level of jitter, the younger adults were no different than the older adults in the non-jittered condition ($t(41) = -1.04$, $p = 0.30$). The subgroup of older adults who participated in these conditions did not exhibit a reduction in H4 between the non-jittered and mild-jittered presentations ($t(6) = -1.13$, $p = 0.30$). For both age groups, the FFRtfs response quality for the moderate and extreme jitter

conditions was too poor to permit analysis, with over 70% of responses from the moderate and extreme jitter conditions being within 3 dB of the noise floor.

4. Discussion

This study focused on the neural encoding of the periodic, steady-state vowel portion of a synthetic speech stimulus. The hypothesis that older adults have reduced magnitudes of the spectral components of their responses was supported for both temporal envelope and fine structure analyses. Further, the response components for the younger adults were systematically degraded by applying a temporal jitter to the stimulus; supporting the idea that neural jitter is responsible for the response pattern of results observed for the older adults. Finally, data from a subset of older adults supported the hypothesis that the application of external jitter does not further degrade the recorded responses of these older listeners presumably because internal noise, or neural jitter, is the limiting factor for encoding the response.

4.1. Speech-evoked ABR: FFRenv and FFRtfs

The results reported here using the 170-ms /da/, complemented by others (Anderson et al., 2012; Ruggles et al., 2012), point to age-related deficits for encoding the sustained vocalic portion of the stimulus, which is consistent with previous research using tonal stimuli (Purcell et al., 2004; Leigh-Paffenroth & Fowler, 2006; Grose et al., 2009). For young listeners, the magnitude of the results in the present study was smaller than that previously reported by Anderson et al. (2012). There are a number of stimulus and procedural factors that could contribute to such differences (e.g., monaural vs. binaural presentation).

Other studies using complex stimuli, specifically the 40-ms /da/ stimulus, to examine age-related effects in hearing have found mixed results with regard to deterioration of periodicity encoding at the level of the brainstem (Vander Werff & Burns, 2011; Clinard & Tremblay, 2013). Specifically, while Vander Werff and Burns (2011) found no age-related differences for the sustained portion of the response, Clinard and Tremblay (2013) suggested an age-related difference in the onset of neural phase locking to a periodic stimulus. The lack of consistent age-related deficits for the sustained components of the 40-ms /da/ are surprising due to the F_0 (and therefore temporal modulation rate) of 103 to 120 Hz. One possibility is that the brevity of the stimulus precludes phase locking to the periodicity of the signal; if that is the case, then this stimulus may not provide a good measure of periodic neural phase locking.

In further support of an association between poorer temporal envelope encoding and age, Ruggles and colleagues (2012) reported that the phase locking value (PLV) to the periodic portion of the same 170-ms /da/ token used in the present study was degraded for middle-aged adults. Specifically, in a cohort of listeners aged 20–55 years, encoding of F_0 significantly declined as a function of age. In the Anderson et al. (2012) sABR study, responses were analyzed not only in terms of phase locking and spectral magnitude, but also in terms of waveform latency. In that comparison of younger and older adults, there were no latency differences for the steady-state portion of the response, but there was poorer F_0 encoding by the older adults, reflected in both the phase locking and spectral magnitude

measures. The finding of an age-related difference in the frequency domain but not the time domain suggests that analyses in the frequency domain may be more sensitive to differences in the fidelity of complex stimulus encoding. In the data presented here, there were many cases where response waveform morphology in the time domain was sufficiently poor to render peak identification for latency measures difficult, yet the frequency domain representation of the response yielded interpretable data.

Electrophysiological studies investigating time-locked encoding of fine structure components of complex stimuli as well as pure tones have also suggested age-related differences. Tonal studies have shown a deficit for monaurally presented mid-, but not low-frequency tones (Clinard et al., 2010; Grose & Mamo, 2012). Specifically, Clinard and colleagues (2010) found age-related reductions in response amplitude for pure tone FFR for tones ~1000 Hz, but not for tones that were ~500 Hz. On the other hand, declines in fine structure encoding have been demonstrated for the synthetic speech token, even in low-frequency regions such as 400–700 Hz, (Anderson et al., 2012; Ruggles et al., 2012); such findings are consistent with the results in the current experiment. More parametric FFR studies with complex tonal stimuli may help clarify these disparate findings.

4.3. Temporal Jitter Model

This experiment showed a reduction in neural encoding of the temporal characteristics of the complex stimulus in the younger adult listeners as the degree of stimulus jitter increased. The F_0 and harmonic components of the response were systematically reduced in amplitude. For the younger adults, mild jitter was sufficient to reduce the most robust components in both the FFRenv (F_0) and FFRtfs (H4) analyses to magnitudes similar to those observed in older adults for non-jittered stimuli. Interestingly, the parameters associated with the mild jitter condition were the same as those employed by Pichora-Fuller and colleagues to reduce the performance of young adults in a speech-in-noise task to a level equivalent to that of older adults. Thus, both behaviorally and physiologically the younger adults with increased stimulus jitter showed reduced performance that was comparable to that of the older adults. This parallel finding suggests that the temporal jitter model may accurately simulate an age-related reduction of neural synchrony for coding the temporal envelope and fine structure of a complex sound. One limitation of the temporal jitter algorithm is that, because it is applied to a single frequency band < 1200 Hz, it likely produces correlated jitter across the nerve fibers tuned to these lower frequencies, while natural internal jitter is presumably independent across nerve fibers. Nevertheless, applying the jitter to all frequencies < 1200 Hz resulted in a systematic decline in the amplitude of the sABR in younger adults.

Although the number of older adults who participated in the jittered conditions of Experiment 2 was relatively small, the most prominent components of the FFRenv and the FFRtfs response analyses were not degraded in this subgroup compared to the response to the non-jittered stimulus. However, the results for the FFRtfs analysis among this small subset of older adults must be interpreted with caution due to the high variability and noise in the response components in both the non-jittered and jittered conditions. Nevertheless, this pattern supports the hypothesis that the poorer responses from the older adults for the

non-jittered stimulus are limited by internal neural jitter and thus are not further degraded by externally applied jitter.

Future research should consider additional models of reduced neural synchrony in the aging auditory system. For example, factors other than phase locking, such as neural survival, may contribute to the magnitude of complex brainstem responses. Lopez-Poveda and Barrios (2013) successfully manipulated stimuli to simulate in young adults the speech-in-noise deficits observed in older listeners. Their model is based on deafferentation, or loss of auditory nerve fibers, and provides an alternative approach to characterizing the neural response patterns associated with poor speech-in-noise performance in older adults. Whereas the temporal jitter model employed in the current experiment targets encoding of frequencies below 1200 Hz, where neural phase locking to the temporal fine structure is most robust, the Lopez-Poveda and Barrios model focuses on neural responses to high-frequency sounds. As such, that model targets a different spectral range of speech information.

5. Conclusion

Findings reported here are consistent with age-related reductions in neural synchrony as measured at the level of the brainstem. This study also provides an objective measure to support the model of temporal jitter in the aging auditory system put forward by Pichora-Fuller and colleagues. In fact, a mild level of jitter reduced the most prominent components in both temporal envelope and fine structure response spectra of the younger adults to the levels observed in the older adult group in response to non-jittered stimuli. Moving forward, an approach that applies the jitter algorithm independently across multiple frequency bands might contribute to a further understanding of the sABR differences between younger and older adults across different response components. In addition, while scalp-recorded evoked potentials have limitations in terms of determining specific response generators, parametric adjustment of the spectral content of the stimulus presented to the listener would provide informative data in terms of understanding what nerve fibers are dominating the recorded response. While the body of sABR literature related to the aging auditory system has increased rapidly in recent years, there are relative few studies that have investigated complex tonal stimuli in the aging brainstem and parametrically adjusted, for example, the harmonic structure of the complex (Marmel et al., 2013; Gockel et al., 2011). In this study, we focused on the response from the sustained portion of the stimulus in order to relate our findings generated with a complex stimulus to other findings generated by on-going tonal stimuli; this focus precluded consideration of how earlier segments of the stimulus (e.g., noise burst, formant transition) may have affected the later sustained response. More parametric manipulations of complex stimuli and investigation of the subsequent changes in the brainstem recorded response would improve our understanding of temporal processing in older adults.

In conclusion, the current investigation contributes to the growing body of sABR literature which finds reduced encoding of complex sounds at the level of the brainstem in the senescent system. Two key factors that are important to the broader understanding of speech-in-noise difficulties among older adults are: 1) the participants in this experiment had

clinically normal audiometric thresholds in the region of the spectral content of the stimulus; and 2) the passive electrophysiological recording was largely free from any cognitive aging effects, which likely contribute to speech-in-noise difficulties. As such, in the clinical management of older adults with hearing complaints, it is important to recognize that even among adults without audiometric evidence of hearing loss, the encoded signal may lack the fidelity needed for listening in complex backgrounds.

Acknowledgments

References

Abel SM, Krever EM, Alberti PW. Auditory detection, discrimination and speech processing in ageing noise-sensitive and hearing impaired listeners. Scandinavian Audiology. 1990; 19:49–54.

Aiken SJ, Picton TW. Envelope and spectral frequency-following responses to vowel sounds. Hearing Research. 2008; 245(1–2):35–47. [PubMed: 18765275]

Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N. Aging affects neural precision of speech encoding. Journal of Neuroscience. 2012; 32(41):14156–14164. [PubMed: 23055485]

Apoux F, Yoho SE, Youngdahl CL, Healy EW. Role and relative contribution of temporal envelope and fine structure cues in sentence recognition by normal-hearing listeners. Journal of the Acoustical Society of America. 2013; 134(3):2205–2212. [PubMed: 23967950]

Bergman, M. Aging and the Perception of Speech. Baltimore: University Park Press; 1980.

Berstein JG, Oxenham AJ. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? Journal of the Acoustical Society of America. 2003; 113(6): 3323–3334. [PubMed: 12822804]

Campbell T, Kerlin JR, Bishop CW, Miller LM. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear and Hearing. 2012; 33(1):144–150. [PubMed: 21760513]

Clinard CG, Tremblay KL. Aging degrades the neural encoding of simple and complex sounds in the human brainstem. Journal of the Academy of Audiology. 2013; 24(7):590–599.

Clinard CG, Tremblay KL, Krishnan AR. Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. Hearing Research. 2010; 264(1–2):48–55. [PubMed: 19944140]

Divenyi P, Simon H. Hearing in aging: Issues old and young. Current Opinion Otolaryngology Head Neck Surgery. 1999; 7:282–289.

Dubno JR, Horwitz AR, Ahlstrom JB. Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. Journal of the Acoustical Society of America. 2002; 111(6):2897–2907. [PubMed: 12083223]

Dubno JR, Horwitz AR, Ahlstrom JB. Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. Journal of the Acoustical Society of America. 2003; 113:2084–2094. [PubMed: 12703719]

Gockel HE, Carlyon RP, Mehta A, Plack CJ. The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. Journal of the Association for Research in Otolaryngology. 2011; 12(6):767–782. [PubMed: 21826534]

Gordon-Salant S, Fitzgibbons PJ. Sources of age-related recognition difficulty for time-compressed speech. Journal of Speech, Language and Hearing Research. 2001; 44(4):709–719.

Grose JH, Mamo SK. Processing of temporal fine structure as a function of age. Ear and Hearing. 2010; 31(6):755–760. [PubMed: 20592614]

Grose JH, Mamo SK. Electrophysiological measurement of binaural beats: effects of primary tone frequency and observer age. Ear and Hearing. 2012; 33(2):187–194. [PubMed: 21926628]

Grose JH, Mamo SK. Frequency modulation detection as a measure of temporal processing: age-related monaural and binaural effects. 2012

Grose JH, Mamo SK, Hall JW III. Age effects in temporal envelope processing: speech unmasking and auditory steady state responses. Ear and Hearing. 2009; 30(5):568–575. [PubMed: 19633565]

He NJ, Mills JH, Ahlstrom JB, Dubno JR. Age-related differences in the temporal modulation transfer function with pure-tone carriers. Journal of the Acoustical Society of America. 2008; 124:3841–3849. [PubMed: 19206810]

He NJ, Mills JH, Dubno JR. Frequency modulation detection: Effects of age, psychophysical method, and modulation waveform. Journal of the Acoustical Society of America. 2007; 122(1):467–477. [PubMed: 17614504]

Hopkins K, Moore BC. The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. Journal of the Acoustical Society of America. 2009; 125(1):442–446. [PubMed: 19173429]

Leigh-Paffenroth ED, Fowler CG. Amplitude-modulated auditory steady-state responses in younger and older listeners. Journal of the American Academy of Audiology. 2006; 17:582–597. [PubMed: 16999253]

Lopez-Poveda EA, Barrios P. Perception of stochastically undersampled sound waveforms: a model of auditory deafferentation. Frontiers in Neuroscience. 2013; 7:124.10.3389/fnins.2013.00124 [PubMed: 23882176]

MacDonald EN, Pichora-Fuller MK, Schneider BA. Effects on speech intelligibility of temporal jittering and spectral smearing of the high-frequency components of speech. Hearing Research. 2010; 261(1–2):63–66. [PubMed: 20093174]

Marmel F, Linley D, Carlyon RP, Gockel HE, Hopkins K, Plack CJ. Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. Journal of the Association for Research in Otolaryngology. 2013; 14:757–766. [PubMed: 23760984]

Miranda T, Pichora-Fuller MK. Temporally jittered NU-6 stimuli produce PI-PB rollover in young normal-hearing listeners. Journal of the American Academy of Audiology. 2002; 13:50–58. [PubMed: 11833992]

Pichora-Fuller MK, Schneider BA. The effect of interaural delay of the masker on masking-level differences in young and old adults. Journal of the Acoustical Society of America. 1992; 91(4 Pt 1):2129–2135. [PubMed: 1597604]

Pichora-Fuller MK, Schneider BA, Macdonald E, Pass HE, Brown S. Temporal jitter disrupts speech intelligibility: a simulation of auditory aging. Hearing Research. 2007; 223(1–2):114–121. [PubMed: 17157462]

Pichora-Fuller MK, Souza P. Effects of aging on auditory processing of speech. International Journal of Audiology. 2003; 42(Suppl 2):S11–S16.

Purcell DW, John SM, Schneider BA. Human temporal auditory acuity as assessed by envelope following responses. Journal of the Acoustical Society of America. 2004; 116:3581–3593. [PubMed: 15658709]

Ross B, Tremblay KL, Picton TW. Physiological detection of interaural phase differences. Journal of the Acoustical Society of America. 2007; 121(2):1017–1027. [PubMed: 17348524]

Ruggles D, Bharadwaj H, Shinn-Cunningham BG. Why middle-aged listeners have trouble hearing in everyday settings. Current Biology. 2012; 22(15):1417–1422. [PubMed: 22727697]

Russo F, Ives T, Goy H, Pichora-Fuller MK, Patterson R. Age-related differences in melodic perception are probably mediated by temporal processing: Empirical and computational evidence. Ear and Hearing. 2012; 33:177–186. [PubMed: 22367092]

Schneider BA, Pichora-Fuller MK. Age-related changes in temporal processing: Implications for speech perception. Seminars in Hearing. 2001; 22:227–239.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270(5234):303–304. [PubMed: 7569981]

Skoe E, Kraus N. Auditory brain stem response to complex sounds: A tutorial. Ear and Hearing. 2010; 31:302–324. [PubMed: 20084007]

Smith SL, Pichora-Fuller MK, Wilson RH, Macdonald EN. Word recognition for temporally and spectrally distorted materials: the effects of age and hearing loss. Ear and Hearing. 2012; 33(3): 349–366. [PubMed: 22343546]

Strouse A, Ashmead DH, Ohde RN, Grantham DW. Temporal processing in the aging auditory system. Journal of the Acoustical Society of America. 1998; 104(4):2385–2399. [PubMed: 10491702]

Vander Werff KR, Burns KS. Brain stem responses to speech in younger and older adults. Ear and Hearing. 2011; 32(2):168–180. [PubMed: 21052004]

Vongpaisal T, Pichora-Fuller MK. Effect of age on F_0 difference limen and concurrent vowel identification. Journal of Speech, Language and Hearing Research. 2007; 50(5):1139–1156.

Highlights

sABRs of normal-hearing older adults showed reduced harmonic content.

Age-related sABR deficits were simulated in young ears using a temporal jitter model.

sABR fidelity of older adults seems limited by neural jitter, or internal noise.
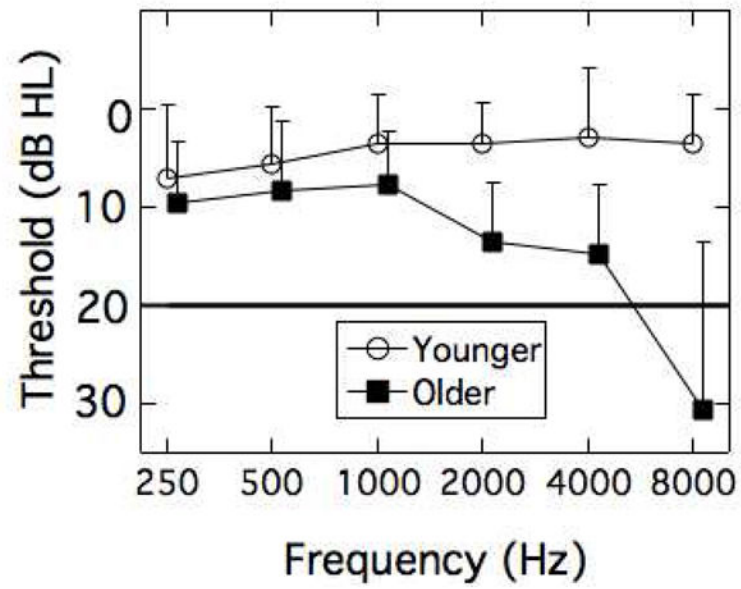
Figure 1.
Mean audiograms. Audiograms for younger (open circles) and older (filled squares) adults.
Error bars show +1 standard deviation. Older adult data points are offset for visualization.
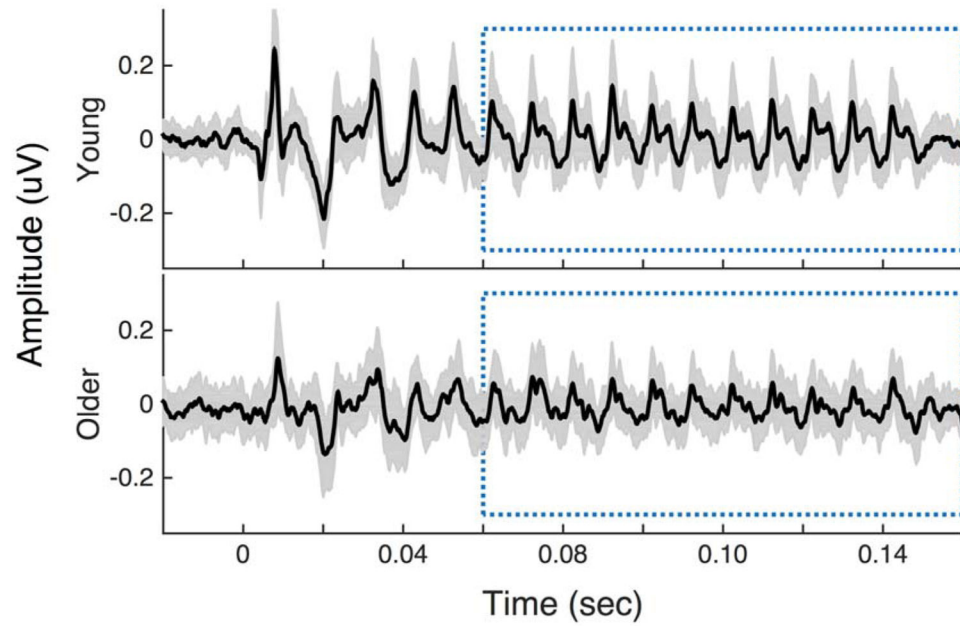
Figure 2.
Grand mean time domain waveforms. The responses from young (top panel) and older (bottom panel) adults are shown with the gray shading displaying +/−1 standard deviation. The portion of the time domain response submitted to FFT analysis is demarcated by the box encompassing 60–180 ms.
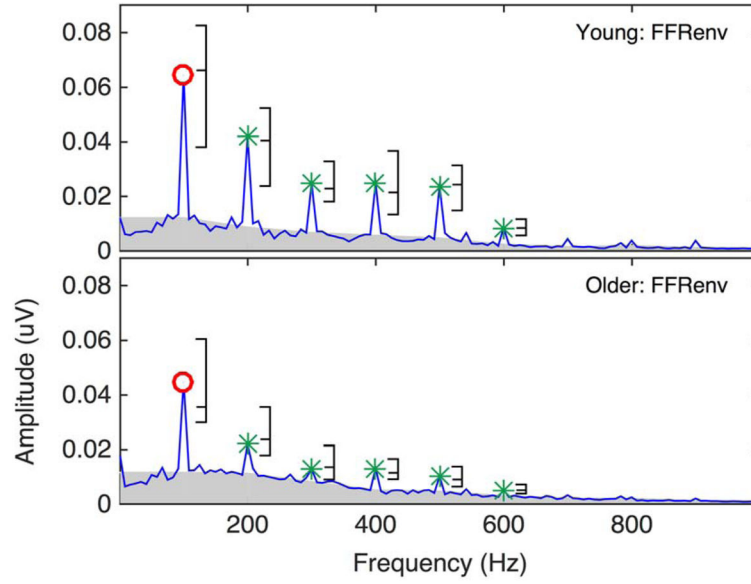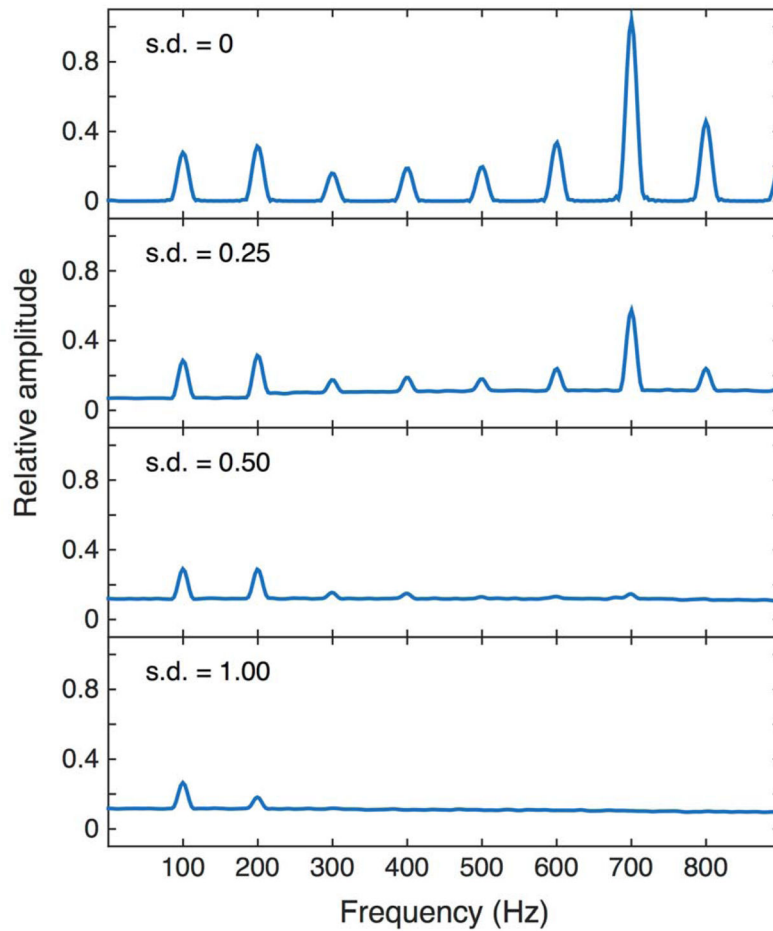
Figure 3.
Grand mean magnitude spectra for the FFRenv for the steady-state portion of the stimulus for young (top panel) and older (bottom panel) adults. The open circles mark F_0 (100 Hz) and the asterisks mark the harmonic components included in analyses. Brackets beside each component display the distribution of the data, with horizontal markers indicating the 25th, 50th and 75th percentiles. The grey shading provides an illustration of the noise floor, with linear extrapolation between values computed near the frequency components of interest.
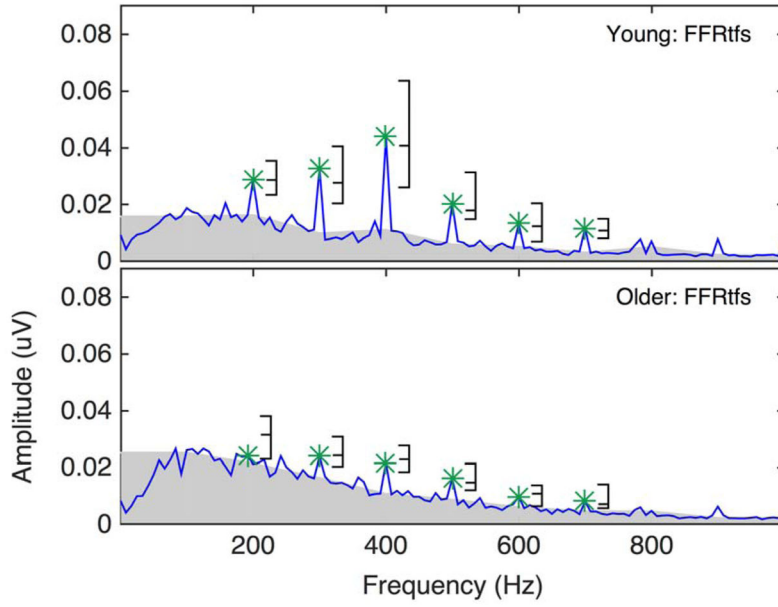
Figure 4.
Grand mean magnitude spectra for the FFRtfs for the steady-state portion of the stimulus for young (top panel) and older (bottom panel) adults. The asterisks mark the harmonic components included in analyses. Brackets beside each component display the distribution of the data in which the bottom and top dashes indicate the 25th and 75th percentiles, respectively, and the middle dash indicates the median. The grey shading provides an illustration of the noise floor, with linear extrapolation between values computed near the frequency components of interest.
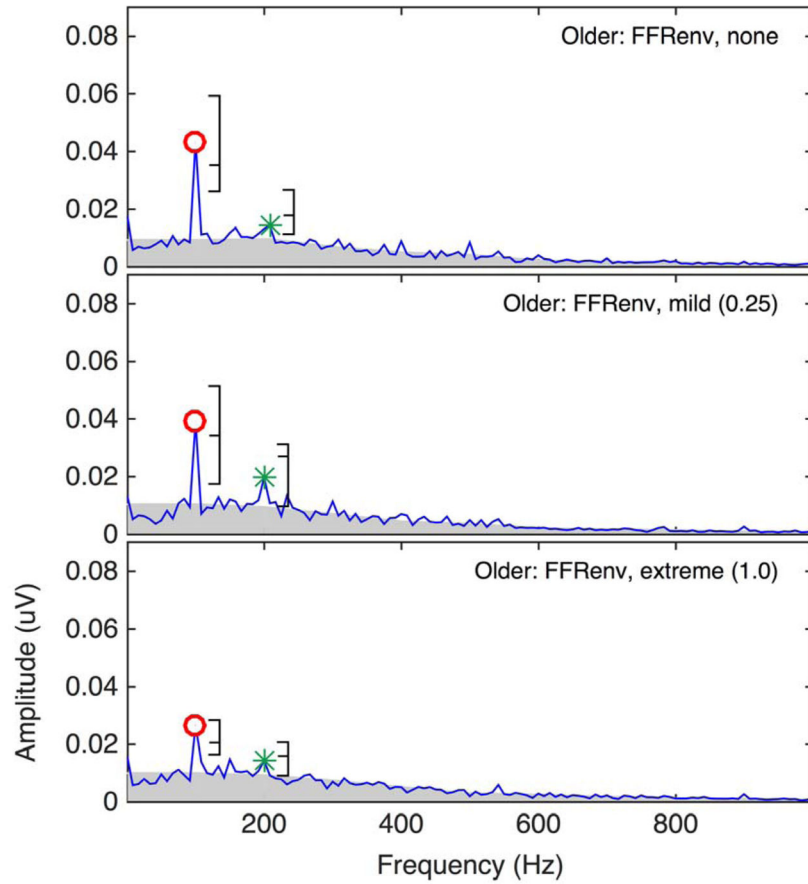
Figure 5.
The mean spectrum for the steady-state portion of the /da/ stimulus, based on 1000 samples averaged in power. Panels show stimuli associated with different degrees of jitter: the non-jittered (s.d. = 0), mild (s.d. = 0.25), moderate (s.d. = 0.50), and extreme (s.d. = 1.0) jitter.
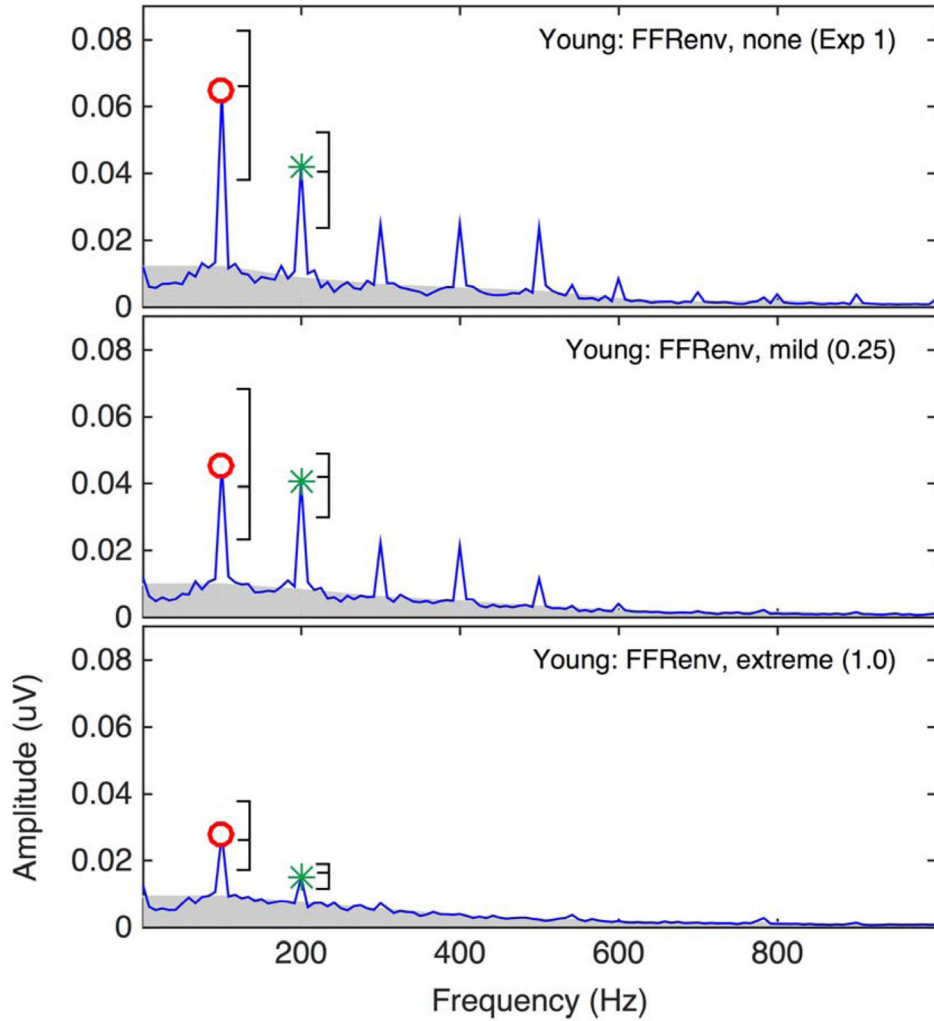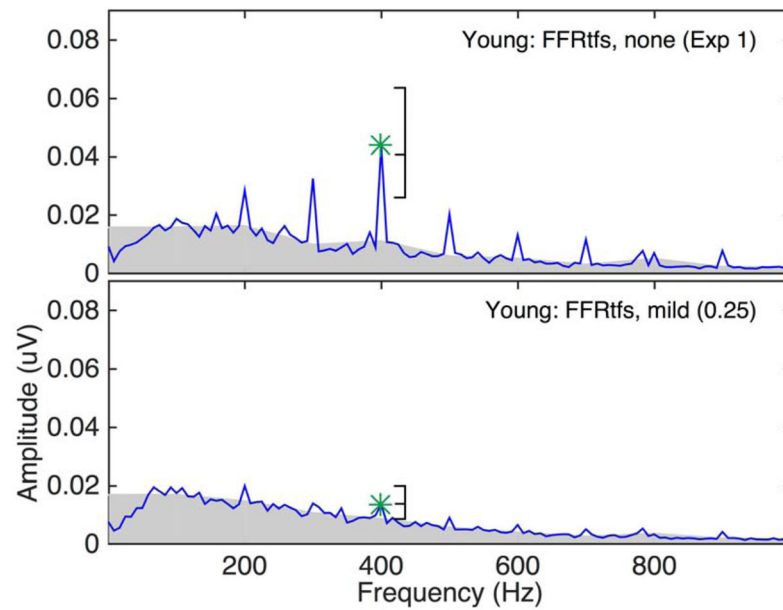
Figure 6.
Jittered FFRenv spectra for younger adults to the mild (middle panel) and extreme (bottom panel) jitter conditions. The original non-jittered FFRenv (top panel) is repeated from Exp. 1 for reference. The open circle marks F_0 (100 Hz) and the asterisk marks the harmonic component included in the jitter analyses (H2; 200 Hz). Brackets beside each component display the distribution of the data in which the bottom and top dashes indicate the 25th and 75th percentiles, respectively, and the middle dash indicates the median. The grey shading provides an illustration of the noise floor, with linear extrapolation between values computed near the frequency components of interest.

Figure 7.
Jittered FFRenv spectra for older adults to the mild (middle panel) and extreme (bottom panel) jitter conditions. The non-jittered FFRenv (top panel) is the grand mean spectra for the older adults from Exp. 2 only (n=7). The open circle marks F_0 (100 Hz) and the asterisk marks the harmonic component included in the jitter analyses (H2; 200 Hz). Brackets beside each component display the distribution of the data in which the bottom and top dashes indicate the 25th and 75th percentiles, respectively, and the middle dash indicates the median. The grey shading provides an illustration of the noise floor, with linear extrapolation between values computed near the frequency components of interest.

Figure 8.
Jittered FFRtfs spectra for younger adults to the mild (bottom panel) jitter condition. The original non-jittered FFRtfs (top panel) is repeated from Exp. 1 for reference. The asterisk marks the harmonic component included in the jitter analyses (H4; 400 Hz). Brackets beside each component display the distribution of the data in which the bottom and top dashes indicate the 25th and 75th percentiles, respectively, and the middle dash indicates the median. The grey shading provides an illustration of the noise floor, with linear extrapolation between values computed near the frequency components of interest.