Published in final edited form as:

Ear Hear. 2017 ; 38(5): e256–e267. doi:10.1097/AUD.0000000000000432.

Human Frequency Following Responses to Vcoded Speech

Saradha Ananthakrishnan^{1,2}, Xin Luo^{2,3}, and Ananthanarayan Krishnan²

¹Department of Audiology, Speech-Language Pathology, and Deaf studies, Towson University, Towson, MD 21252

²Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN 47906

³Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287

Abstract

Objectives—Vocoders offer an effective platform to simulate the effects of cochlear implant (CI) speech processing strategies in normal-hearing listeners. Several behavioral studies have examined the effects of varying spectral and temporal cues on vocoded speech perception; however, little is known about the neural indices of vocoded speech perception. Here, the scalp-recorded Frequency Following Response (FFR) was used to study the effects of varying spectral and temporal cues on brainstem neural representation of specific acoustic cues, the temporal envelope periodicity related to fundamental frequency (F0) and temporal fine structure (TFS) related to formant and formant-related frequencies, as reflected in the phase-locked neural activity in response to vocoded speech.

Design—In Experiment 1, FFRs were measured in 12 normal-hearing, adult listeners in response to a steady-state English back vowel /u/ presented in an unaltered, unprocessed condition as well as six sine-vocoder conditions with varying numbers of channels (1, 2, 4, 8, 16 and 32), while the temporal envelope cutoff frequency was fixed at 500 Hz. In Experiment 2, FFRs were obtained from 14 normal-hearing, adult listeners in response to the same English vowel /u/, presented in an unprocessed condition and four vocoded conditions where both the temporal envelope cutoff frequency (50 vs. 500 Hz) and carrier type (sine wave vs. noise band) were varied separately with the number of channels fixed at 8. Fast Fourier Transform (FFT) was applied to the time waveforms of FFR to analyze the strength of brainstem neural representation of temporal envelope periodicity (F0) and TFS related peaks (formant structure).

Results—Brainstem neural representation of both temporal envelope and TFS cues improved when the number of channels increased from 1 to 4, followed by a plateau with 8 and 16 channels, and a reduction in phase-locking strength with 32 channels. For the sine-vocoders, peaks in the FFR_{TFS} spectra corresponded with the low-frequency sine-wave carriers and sideband frequencies in the stimulus spectra. When the temporal envelope cutoff frequency increased from 50 to 500 Hz, an improvement was observed in brainstem F0 representation with no change in brainstem

Corresponding Author: Saradha Ananthakrishnan, 8000 York Road, Van Bockelen 103D, Towson University, Towson, MD 21252, USA. sananthakrishnan@towson.edu. Phone: 410-704-6369. Fax: 410-704-4131.

Financial Disclosures/Conflicts of Interest: This research was funded by the National Institutes of Health (NIH) R01 DC008549 (A.K.), the Department of Speech, Language, and Hearing Sciences (Purdue University), the Faculty Development and Research Committee (FDRC) Grant (FDRC # 140-255 (S.A.)) (Towson University) and the Department of Audiology, Speech Language Pathology and Deaf Studies (Towson University)

representation of spectral peaks proximal to the first formant frequency (F1). There was no significant effect of carrier type (sine- vs. noise-vocoder) on brainstem neural representation of F0 cues when the temporal envelope cutoff frequency was 500 Hz.

Conclusions—While the improvement in neural representation of temporal envelope and TFS cues with up to 4 vocoder channels is consistent with the behavioral literature, the reduced neural phase-locking strength noted with even more channels may be due to the narrow bandwidth of each channel as the number of channels increases. Stronger neural representation of temporal envelope cues with higher temporal envelope cutoff frequencies is likely a reflection of brainstem neural phase-locking to F0-related periodicity fluctuations preserved in the 500-Hz temporal envelopes, which are unavailable in the 50-Hz temporal envelopes. No effect of temporal envelope cutoff frequency was seen for neural representation of TFS cues, suggesting that spectral sideband frequencies created by the 500-Hz temporal envelopes did not improve neural representation of F1 cues over the 50-Hz temporal envelopes. Lastly, brainstem F0 representation was not significantly affected by carrier type with a temporal envelope cutoff frequency of 500 Hz, which is inconsistent with previous results of behavioral studies examining pitch-perception of vocoded stimuli.

Keywords

Neural phase-locking; frequency following response; temporal envelope; temporal fine structure; vocoded speech; cochlear implant

INTRODUCTION

Speech perception depends on both spectral and temporal cues. Spectral cues refer to the information available in the frequency spectrum of the stimulus, including fundamental frequency (F0) and harmonics, formant structure and transitions, as well as spectral slope and contrast between peaks and valleys. Temporal cues, on the other hand, refer to the information available in the amplitude fluctuations of the stimulus over time, which can be classified as envelope (2–50 Hz), periodicity (50–500 Hz), and fine structure cues (>500 Hz) (Rosen 1992). Low frequency periodicity cues provide voicing, manner and prosodic information, which are all related to the fundamental frequency of the signal; information about the spectrum of the stimulus, including its formant structure, is provided by fine structure cues.

Current cochlear implants (CIs) partially restore hearing sensation to profoundly deaf people and support generally good speech perception in quiet using limited spectral and temporal cues (Zeng 2004). In CIs, the incoming speech signal is divided into a number of frequency channels to provide spectral cues (although with low resolution). The temporal envelope (rather than fine structure) of speech signal in each channel is extracted to modulate the amplitude of a high-rate pulse train delivered to the corresponding electrode within the cochlea based on the tonotopic organization of human auditory system (Wilson et al. 1991).

As CI speech processing strategies evolved over the years, the relative contributions of spectral and temporal cues to speech perception with CIs have received considerable interest. However, studies of these cues in real CI users showed highly variable results and were often hampered by numerous confounding factors (e.g., electrode placement, neural

survival, etiology and duration of deafness). Instead, researchers tested relatively homogeneous normal-hearing listeners using carefully controlled vocoded speech to simulate the effects of CI processing on speech perception. Vocoders work in a similar manner as CI processors, except that the temporal envelope extracted from each channel is used to modulate a noise band of the same width of the channel (Shannon et al. 1995) or a sine wave at the center of the channel (Dorman et al. 1997) instead of the high-rate pulse train. The temporal envelope cutoff frequency and the number of channels can be independently varied to control the temporal and spectral cues available in the acoustic CI simulations, respectively. Spectral resolution is manipulated by varying the number of channels while temporal resolution is manipulated by varying the temporal envelope cutoff frequency.

The processes underlying speech recognition rely on both spectral and temporal cues; Qin and Oxenham (2005) demonstrated an improvement in F0 difference limens as the number of channels was increased, leading to presence of both temporal and spectral cues which in turn, improved pitch salience. Similarly, F0-based voice gender identification scores were highest with the highest temporal envelope cutoff frequency (320 Hz) and the greatest number of channels (32) tested in Fu et al. (2004). However, the general consensus from behavioral studies is that effective speech recognition can be achieved with primarily temporal cues, even in the face of reduced spectral cues (Shannon et al. 1995; Fu et al. 1998; Xu et al. 2002; Fu et al. 2004). For example, the number of channels required for effective speech perception is as low as 4–6, for consonant, vowel and sentence recognition in English (Shannon et al. 1995) and Mandarin (Fu et al. 1998; Xu et al. 2002). Further, high speech recognition scores were measured when slow-varying temporal cues < 50 Hz were used (Shannon et al. 1995). In addition, spectral resolution (manipulated by varying the number of channels) appears to play a more important role in English vowel recognition (Shannon et al. 1995) and Mandarin consonant and vowel recognition (Fu et al. 1998; Fu et al. 2004), while temporal resolution (manipulated by varying the temporal envelope cutoff frequency) seems to play a more important role in pitch-related listening tasks such as tone recognition (Fu et al. 1998; Xu et al. 2002) and voice gender identification (Fu et al. 2004; Luo et al. 2007) with CI simulations.

While both noise and sine-vocoders have been widely used for CI simulations, recent studies have indicated an advantage to speech recognition for sine-vocoders over noise-vocoders, specifically with respect to gender identification (Fu et al. 2004; Gonzalez & Oliver 2005) and vowel recognition (Fu et al. 2004). Souza and Rosen (2009) and Stone et al. (2008) further observed that sine-vocoders outperformed noise-vocoders only when the temporal envelope cutoff frequency was high enough (e.g., 300 Hz) to allow transmission of periodicity cues; the reverse was true when the temporal envelope cutoff frequency (e.g., 30 Hz) was lower than the signal F0. The speech recognition advantage of sine-wave carriers over noise-band carriers has been attributed to the lack of random level variations in sine-vocoders that may distort the speech envelope, interfering with amplitude modulation detection, and the presence of spectral “sidebands” around sine-wave carriers that may represent F0-related periodicity cues in the speech envelope, especially with high temporal envelope cutoff frequencies.

While the perception of vocoded speech in normal-hearing listeners is well-documented through behavioral studies, there is presently limited literature examining the neural representation of vocoded signals in normal-hearing listeners. Friesen et al. (2009) successfully recorded the P1-N1-P2 complexes of cortical auditory evoked potentials (CAEPs) in response to consonant-vowel-consonant stimuli that are often used to evaluate CI performance in clinic. As the number of channels increased from 2 to 16, the response amplitude increased while the response latency decreased, along with improved vowel recognition scores in behavioral studies. Peelle et al. (2013) collected magnetoencephalography (MEG) responses in normal-hearing listeners to both unprocessed and noise-vocoded sentences and found greater cerebral phase-locking coherence as the number of channels in the noise-vocoder increased from 1 to 16. Both Friesen et al. (2009) and Peelle et al. (2013) examined neural representation of vocoded signals at the cortical level; to our knowledge, there are no published studies evaluating the neural representation of temporal and spectral cues presented in the vocoded speech at the level of the brainstem in humans. Further, no study has examined the effects of carrier type, temporal envelope cutoff frequency, and their interaction on neural responses to vocoded speech. Understanding how neural representation of key acoustic features responsible for speech perception such as envelope and TFS cues may be altered when vocoded stimuli are presented to listeners with normal hearing may eventually lay the foundation for future research into the neurophysiology underlying speech perception for similar stimuli in clinical populations, such as CI users. In the long term, such knowledge could be useful when devising CI speech processing strategies for better encoding and perception of specific acoustic features in the speech signal.

The human scalp recorded Frequency Following Response (FFR) reflects sustained neural activity in an ensemble of neural elements within the rostral brainstem that is phase locked to both the envelope periodicity and the TFS of a complex stimulus. As such, the FFR provides an objective, non-invasive window to evaluate the fine grained neural representation of both envelope and formant related information of various complex sounds in human listeners. For example, the FFR to the temporal envelope of the stimulus has been shown to preserve pitch and periodicity information for both harmonic complex tones with time-invariant pitch (Krishnan & Plack 2009; Greenberg et al. 1987) and Mandarin tones with time-varying pitch (Krishnan et al. 2004); formant related information of both steady-state and time-varying vowel-like sounds (Krishnan 1999; Krishnan 2002; Krishnan & Parkinson 2000; Plyler & Ananthanarayan 2001; Ananthakrishnan et al. 2016). The aim here is to evaluate the neural representation of both envelope periodicity and TFS cues by systematically varying the number of channels, temporal envelope cutoff frequency, and carrier type of vocoded stimuli. Based on the results of previous behavioral studies summarized above, brainstem neural representation of both envelope periodicity and formant-related spectral peaks is expected to be facilitated with increase in the number of channels, with relatively greater change for representation of formant related information. On the other hand, using sine-wave carriers instead of noise-band carriers and increasing the temporal envelope cutoff frequency from 50 to 500 Hz would facilitate only the neural representation of pitch relevant information.

MATERIALS AND METHODS

Participants

Twelve adults (four males, eight females) aged between 21 and 25 years ($M = 22.72$, $SD = 1.19$) participated in Experiment 1, which evaluated the effects of changing the number of channels in sine-vocoded speech. Fourteen adults (three males, eleven females) aged between 21 and 44 years ($M = 24.25$, $SD = 6.36$) participated in Experiment 2, which evaluated the effect of changing the temporal envelope cutoff frequency in sine and noise-vocoders with channel number fixed to 8. Six of the subjects who participated in Experiment 1 also participated in Experiment 2. All participants exhibited normal hearing sensitivity with thresholds better than 20 dB HL in both ears at octave frequencies from 250 to 8000 Hz. All participants were paid and gave informed consent in compliance with a protocol approved by the Institutional Review Boards of Purdue University and Towson University.

Stimuli

The original unprocessed stimulus was a synthetic steady-state English back vowel /u/ ($F_0 = 120$ Hz; $F_1 = 360$ Hz; $F_2 = 970$ Hz; duration: 250 ms) presented in quiet at 80 dB SPL. This specific stimulus was selected as the first two formants occur at frequencies well below 1500 Hz, which is the upper limit for neural phase-locking in the rostral brainstem. Further, this vowel has been previously used successfully to elicit FFRs in normal hearing listeners (Krishnan 2002; Ananthakrishnan et al. 2016) as well as hearing-impaired listeners (Ananthakrishnan et al. 2016). FFRs were recorded for both the original unprocessed stimulus and various vocoded versions of the stimulus. The noise-vocoders were similar to those in Shannon et al. (1995), while the sine-vocoders were similar to those in Dorman et al. (1997). In vocoder processing, the original signal was first processed through a pre-emphasis filter (1st order high-pass Butterworth filter at 1200 Hz) and then band-pass filtered (4th order Butterworth filters) into a number of channels. The Greenwood function (Greenwood 1990) was used to calculate the cutoff frequencies of individual band-pass filters. The center frequency of each channel was calculated as the geometric mean of the two cutoff frequencies for that channel. The overall input frequency range was from 200 to 7000 Hz. This bandwidth is similar to those used in CIs, which typically exclude the first harmonic or fundamental frequency (Vandali et al. 2000). Amplitude envelopes were extracted from the output of each band-pass filter using half-wave rectification and low-pass filtering. The extracted amplitude envelopes were used to modulate either white noise (for noise-vocoders) or sinusoids at the center of each band-pass filter (for sine-vocoders). The amplitude-modulated noise or sinusoids were band-limited by the same band-pass filters and summed across all channels. Finally, the noise- or sine-vocoded signals were normalized to have the same root mean square (RMS) level as the original signal.

In Experiment 1, the original vowel stimulus and six sine-vocoded signals with a fixed temporal envelope periodicity of 500 Hz (to preserve F_0 -related periodicity) and six different number of channels (1, 2, 4, 8, 16, and 32-channels), to vary the amount of spectral cues provided, were utilized to record the FFRs.

In Experiment 2, both noise- and sine-vocoded stimuli were utilized. The number of channels was fixed at 8 to simulate the spectral resolution that CI users may effectively have (Friesen et al. 2001). Similar to previous studies (Shannon et al. 1995; Fu et al. 1998; Xu et al. 2002; Luo & Fu 2004), the cutoff frequency of low-pass filter for temporal envelope extraction was 500 or 50 Hz, which varied the amount of temporal envelope cues available in vocoded signals (i.e., with or without F0-related periodicity cues). There were five conditions in total: 8-channel noise-vocoded signals with 50- or 500-Hz temporal envelopes (denoted as Noise-50 and Noise-500, respectively), 8-channel sine-vocoded signals with 50- or 500-Hz temporal envelopes (denoted as Sine-50 and Sine-500, respectively), and the original unprocessed signal.

Data acquisition

FFR recording protocol and data analysis were similar to those described in Ananthakrishnan et al. (2016). Participants were situated in a comfortable recliner in an acoustically and electrically shielded booth. They were instructed to ignore the sounds they heard while relaxing and refraining from extraneous body movements to minimize movement artifacts. Participants were allowed to sleep during FFR recording. FFRs were recorded from each participant in response to monaural stimulation at a fixed presentation level of 80 dB SPL and a repetition rate of 2.76 stimuli/second. Alternating stimulus polarity was used. The presentation order of the stimuli was randomized both within and across participants. The experimental protocol was controlled using a signal generation and data acquisition system (Intelligent Hearing Systems) with a sampling rate of 40 kHz. The stimulus waveforms were routed through a digital to analog module and presented via a magnetically shielded insert earphone (Etymotic, ER-3A) to the right ear of participants.

FFRs were recorded differentially between a non-inverting (positive) electrode placed on the midline of the forehead at the hairline (Fz) and linked inverting (reference) electrodes placed on (1) the right mastoid (A2) and the left mastoid (A1) and (2) the 7th cervical vertebra (C7). Another electrode placed on the mid-forehead (Fpz) served as the common ground. For each stimulus condition, the FFRs from the two electrode configurations, were averaged to yield a response with a higher signal-to-noise ratio (Krishnan et al. 2009). All inter-electrode impedances were maintained below 1 k Ω . The FFR recordings were amplified by 200,000 and band-pass filtered from 50 to 3000 Hz (6 dB/octave roll-off). The final FFRs were the average responses of 4000 stimulus presentations (2000 sweeps each in condensation and rarefaction polarities) over a 250-ms analysis window. It took about 120 minutes to complete the protocol for each experiment.

When alternating stimulus polarity is used (as in this study), the split buffer averaging feature in the data acquisition system stores the average FFRs to the condensation and rarefaction polarities in separate buffers. Stimulus polarity inversion does not affect the neural response to the temporal envelope, but leads to an inversion of the neural response to the fine structure. As a result, addition of the responses to the condensation and rarefaction polarities yields FFRs phase-locked to the temporal envelope of the stimulus (FFR_{ENV}), while subtraction of these responses yields FFRs phase-locked to the fine structure of the stimulus (FFR_{TFS}) (Krishnan, 2002; Aiken and Picton, 2008). Addition and subtraction of

polarity-inverted responses followed in this study have been widely used in previous FFR studies to obtain brainstem neural representations of temporal envelope and fine structure, respectively (Krishnan 2002; Aiken & Picton 2008; Elsisy & Krishnan 2008; Smalt et al. 2012; Anderson et al. 2013; Ananthakrishnan et al. 2016). Further, in order to rule out any polarity-sensitive amplitude effects on FFR_{ENV} and FFR_{TFS} strength, RMS amplitudes were first compared for FFRs obtained in condensation and rarefaction polarity to the 8-channel sine-vocoder condition with a temporal envelope cutoff frequency of 500 Hz (a stimulus common to both experiments). There was no statistically significant difference between the RMS amplitudes of the FFRs obtained in the two polarities (paired t-test: $t_{10} = -0.80$, $p = 0.44$), ruling out any polarity-sensitive amplitude effects on the summation and subtraction of FFRs collected in opposing polarities in this study.

Data analysis

Spectral analysis—Magnitude spectra were computed over the whole duration of the 4000-sweep averaged responses using the Fast Fourier Transform (FFT). Spectra of added FFRs (FFR_{Condensation} + FFR_{Rarefaction}) yield FFRs that reflect neural phase-locking to the envelope periodicity (FFR_{ENV}), characterized by a spectral peak at F0 and its integer multiples. When FFT is applied to the subtraction of the condensation and rarefaction FFRs (FFR_{TFS}), the response spectrum is characterized by robust peaks at formant-related spectral peaks or the fine structure components present in the vocoded stimuli (Krishnan 2002; Aiken and Picton 2008). Peak FFT magnitudes were measured at F0 from the FFR_{ENV} spectra, and at the spectral peak closest to the first formant frequency (F1) from the FFR_{TFS} spectra.

Statistical analysis—Several one-way repeated measures analyses of variance (RM ANOVAs) were conducted on the FFR data to evaluate the effects of vocoding on brainstem neural representation of the F0 and first formant related spectral peak. Specifically, one-way repeated measures ANOVAs were performed to evaluate the effect of varying the number of vocoder channels on 1) F0 and 2) formant-related spectral peak magnitudes in the FFR_{ENV} and FFR_{TFS} spectra, respectively. Additional one-way repeated measures ANOVAs were utilized to determine the effects of varying temporal envelope cutoff frequency and carrier type on 1) the F0 magnitude of the FFR_{ENV}, and 2) formant-related spectral peak magnitude in FFR_{TFS} spectra.

RESULTS

Neural representation of the envelope periodicity (F0)

Experiment 1: Effect of the number of vocoder channels on F0 representation

—Grand average FFR time waveforms representing neural phase-locking to the envelope of the unprocessed vowel /u/, and sine vocoded conditions are shown in Figure 1 (panel B). In general, FFR waveforms show clear periodicity up to the 4-channel condition, with somewhat reduced periodicity in the 8- and 16-channel conditions and no discernable periodicity for the 32-channel condition.

Grand average FFT spectra of the FFR_{ENV} in the unprocessed, and sine vocoded conditions are shown in Figure 1 (panel C). Regardless of the number of channels, a spectral peak was

consistently seen at the stimulus F0 of 120 Hz suggesting periodicity information was well preserved for all stimuli. Individual and mean F0 magnitudes calculated from the FFT spectra are shown in Figure 2. As shown in Figures 1 (panel C) and 2, the F0 magnitude increased as the number of channels increased from 1 to 4, plateaued between 8- and 16-channels, and then decreased in the 32-channel condition. A one-way repeated measures ANOVA showed a significant effect of the number of channels on the F0 magnitudes ($F_{6, 53} = 3.31, p = 0.008$). Tukey pair-wise comparisons indicated a significant difference in F0 magnitudes for only the channel 1 vs. 4 comparison ($p = 0.03$), and channel 4 vs. 32 comparison ($p = 0.02$). These changes in the periodicity information preserved in the FFR may reflect temporal changes in stimulus periodicity consequent to changes in the number of channels of the vocoded stimulus.

Experiment 2: Effects of temporal envelope cutoff frequency and carrier type on F0 representation—

Grand average FFR time waveforms representing neural phase-locking to the envelope of the unprocessed vowel /u/, and the four vocoder conditions (i.e., Sine-500, Sine-50, Noise-500, and Noise-50) are shown in Figure 3 (panel B: 500 Hz cutoff in black; and 50 Hz cutoff in red). As expected, FFR_{ENV} time waveforms appeared more robust and periodic when the temporal envelope cutoff frequency was 500 Hz as compared to 50 Hz in both the sine- and noise-vocoder conditions. Further, the grand average waveforms indicated reduced overall amplitude and periodicity in the noise-vocoder condition as compared to the sine-vocoded condition when the temporal envelope cutoff frequency was 500 Hz.

Grand average FFT spectra of the FFR_{ENV} and the F0 magnitudes in the unprocessed and four vocoder conditions (i.e., Sine-500, Sine-50, Noise-500, and Noise-50) are shown in Figure 3 (panel C: 500 Hz cutoff in black; and 50 Hz cutoff in red) and Figure 4, respectively. As observed in Figure 3 (panel C), for the 500 Hz temporal envelope cutoff frequency, both the sine-wave and noise-band carriers showed a robust FFR spectral peak at F0 followed by increasingly smaller magnitude peaks at integer multiples of the F0. In contrast, no or only a very weakly identifiable peak was observed at 120 Hz for both the sine- and noise-vocoders for the 50 Hz envelope frequency cutoff condition. Individual and mean F0 magnitudes are plotted in Figure 4. Irrespective of carrier type, mean F0 magnitudes were stronger in the 500 Hz condition than the 50 Hz condition; further, no differences were observed between either the sine or noise-vocoder conditions, at either low pass filter cutoff frequency. A one-way repeated measures ANOVA confirmed the above observations; a significant effect of the stimulus condition was observed on the F0 magnitudes of the FFR_{ENV} ($F_{4, 13} = 14.38, p < 0.001$). Post-hoc Tukey pairwise comparisons indicated that regardless of carriers, the F0 magnitudes were significantly stronger in the unprocessed condition and the 500-Hz temporal envelope condition as compared to the 50-Hz temporal envelope conditions ($p < 0.01$). Finally, there were no effects of carrier type, i.e. no significant differences in the F0 magnitudes between sine and noise-vocoders with either 500-Hz temporal envelopes ($p = 0.864$) or 50-Hz temporal envelopes ($p = 0.792$). Overall, stronger F0 magnitudes observed in the FFR_{ENV} with 500-Hz temporal envelope cutoff frequency likely reflect brainstem neural phase-locking to the

F0 and low-frequency harmonics of the stimuli at 120, 240, and 360 Hz preserved in the 500-Hz temporal envelopes, which are unavailable in the 50-Hz temporal envelopes.

Brainstem neural representation of temporal fine structure (F1)

Experiment 1: Effect of the number of vocoder channels on F1 representation

—Grand average FFR time waveforms representing neural phase-locking to the temporal fine structure of the unprocessed vowel /u/, and all the sine-vocoder conditions are shown in Figure 5 (panel B). A slight decrease in the overall waveform amplitude was observed as the number of channels increases from 1 to 32.

Grand average FFT spectra of the FFR_{TFS} in the unprocessed and all sine-vocoder conditions are shown in Figure 5 (panel C). In the unprocessed condition, robust spectral peaks of the FFR_{TFS} were observed at the harmonics of stimulus F0 (240, 360, and 480 Hz, etc.), with the highest amplitude observed for the F1-related harmonic (360 Hz). However, in the sine-vocoder conditions, spectral peaks of the FFR_{TFS} were observed at the center frequencies of the vocoder channels (i.e., the sine wave carrier frequencies) and the sideband frequencies around the center frequencies generated by temporal envelope modulations (i.e., center frequency $\pm n \times F0$, where n is an integer and $F0$ is 120 Hz). Note that the brainstem neurons generating the FFR_{TFS} could only phase lock to the center and sideband frequencies below about 1000 Hz. For example, in the 1-channel condition, the single center frequency was 1183 Hz and spectral peaks of the FFR_{TFS} were observed at the lower sideband frequencies of 1183 Hz: 214, 336, 456, 576, and 698 Hz. In this instance, the largest peak was observed at 456 Hz— a F1 related sideband frequency. Similarly, in the 2-channel condition, the center frequency of channel 1 was 534 Hz and spectral peaks of the FFR_{TFS} were observed at 534 Hz as well as the sideband frequencies around 534 Hz: 170, 291, 413, 655, and 776 Hz. The strongest peak was the one closest to F1 at 413 Hz. In the 4-channel condition, the strongest spectral peak in the FFR_{TFS} was observed at 344 Hz, center frequency of channel 1. With 8 channels, the strongest peak in the response occurred at 461 Hz, center frequency of channel 2. In both the 4- and 8-channel conditions, small spectral peaks were also observed at the sideband frequencies around the center frequencies of low-numbered channels. When the number of channels was increased to 16 and 32, spectral peaks of the FFR_{TFS} were observed at the center frequencies of low-numbered channels, but not at the sideband frequencies around the center frequencies, possibly because each channel was too narrow to preserve the sideband frequencies. The strongest spectral peak in the FFR_{TFS} in the 16-channel condition occurred at 408 Hz, center frequency of channel 3. In the 32-channel condition, the most dominant spectral peak in the response was noted at 383 Hz, center frequency of channel 5. In general, the strongest spectral peak of the FFR_{TFS} in each condition was at a channel center frequency or a sideband frequency near F1, reflecting the brainstem neural representation of vowel information in vocoded speech. Individual and mean FFR_{TFS} magnitudes at the channel center frequency or sideband frequency closest to F1 for all stimulus conditions are shown in Figure 6. A one-way repeated measures ANOVA showed a significant effect of stimulus type on neural representation of TFS cues ($F_{6, 11} = 8.761$; $p < 0.001$). Brainstem representation of the TFS or F1 cues was significantly stronger in the unprocessed condition as compared to any of the vocoded conditions ($p < 0.017$). Although TFS representation was stronger when the number of channels was 1, 2 or 4 as

compared to 8, 16 or 32, these differences among the vocoded conditions were not significant ($p > 0.24$).

Experiment 2: effects of temporal envelope cutoff frequency and carrier type on F1 representation

Grand average FFR time waveforms representing neural phase-locking to the temporal fine structure of the unprocessed vowel /u/, and four vocoder conditions (i.e., Sine-500, Sine-50, Noise-500, and Noise-50) are shown in Figure 7 (panel B). There were no major observable differences in grand average waveform morphology or amplitude between the 500-Hz and 50-Hz low-pass cutoff frequencies within either the sine- or noise-vocoder conditions.

Grand average FFT spectra of the FFR_{TFS} in the unprocessed and four vocoder conditions (i.e., Sine-500, Sine-50, Noise-500, and Noise-50) are shown in Figure 7 (panel C). As in Experiment 1, spectral peaks in the unprocessed condition were noted at the harmonics of stimulus F0 (240, 360, and 480 Hz, etc.), with the greatest amplitude observed for the F1-related harmonic (360 Hz). In both the Sine-500 and Sine-50 conditions, spectral peaks of the FFR_{TFS} were observed at 268 and 461 Hz, the center frequencies of channels 1 and 2, respectively. Unlike the Sine-500 condition, the Sine-50 condition did not show small spectral peaks at the sideband frequencies around these center frequencies. This was because the 50-Hz temporal envelopes were flat for the steady-state vowel /u/ and did not have amplitude modulations at F0. In the noise-vocoder conditions, as expected, no spectral peaks of the FFR_{TFS} were identified at all. Individual and mean FFR_{TFS} magnitudes at the spectral peak closest to F1 for the clean and sine-vocoder conditions are shown in Figure 8. A one-way repeated measures ANOVA indicated a significant effect of stimulus condition on FFR_{TFS} peak magnitudes ($F_{2, 13} = 19.232$, $p < 0.001$). Pairwise Tukey comparisons further indicated that brainstem representation of the TFS or F1 cues was significantly stronger in the unprocessed condition as compared to either the 500-Hz or 50-Hz temporal envelope condition for the sine-vocoded signals ($p < 0.001$). However, no significant differences were observed in the FFR_{TFS} peak magnitudes around F1 between the 500-Hz and 50-Hz temporal envelopes in the sine-vocoder conditions ($p = 0.167$). The stimulus energy within each channel was distributed mostly at the center frequency in the Sine-50 condition, but across the center and sideband frequencies in the Sine-500 condition. This may explain the slightly lower FFR_{TFS} peak magnitudes at the center frequencies in the Sine-500 condition than in the Sine-50 condition. Statistical analysis could not be performed on the neural responses to the noise-vocoder conditions as there were no spectral peaks to pick in these responses.

DISCUSSION

The major findings of this study were that (i) in general, brainstem neural representation of F0 as reflected in the FFR_{ENV} improved as the number of channels increased from 1 to 4, plateaued with 8 and 16 channels, and then degraded with 32 channels; (ii) brainstem neural phase-locking to the F0 as reflected in the FFR_{ENV} was stronger when the temporal envelope cutoff frequency was 500 Hz as compared to 50 Hz for both sine and noise-vocoders; however, no significant differences were noted between sine-wave and noise-band carriers; (iii) with the sine-vocoders, brainstem neural representation of F1-related spectral

peaks, i.e., spectral components at a channel center frequency or a spectral sideband frequency close to F1 were not significantly different across the tested channel numbers or temporal envelope cutoff frequencies. To our knowledge, these are the only data that describe the effects of channel number, carrier type and temporal envelope cutoff frequency on brainstem neural responses to vocoded speech.

Influence of number channels on neural representation of the fundamental periodicity

In the present study, FFT analyses indicated an improvement in the F0 magnitudes of FFR_{ENV} as the number of channels increased from 1 to 4, followed by a plateau with 8 and 16 channels and then a degradation with 32 channels. The improvement in brainstem F0 representation from 1 to 4 channels appears consistent with the performance improvement in pitch-related listening tasks from 1 to 4 channels in previous behavioral studies (Xu et al. 2002; Fu et al. 2004; Qin & Oxenham 2005). For instance, Xu et al. (2002) reported improved Mandarin tone recognition as the number of channels increased from 1 to 3, followed by a plateau as the number of channels approached 12. However, an improvement in voice gender identification (which relies on pitch perception) as the number of channels increased from 4 to 32 was noted by Fu et al. (2004). Also, Qin and Oxenham (2005) found better F0 difference limens as the number of channels increased from 1 to 40. These improvements in pitch perception were not consistent with the significant *degradation* in brainstem F0 representation with 32 channels in this study. Neural phase-locking to F0 may be degraded in the 32-channel condition due to the narrow channel bandwidth associated with a high number of channels. When there were less than 4 channels, each channel was wide enough for the extraction and modulation of the 500-Hz temporal envelopes. In such a case, adding channels with consistent amplitude modulation rates may strengthen the brainstem F0 representation and help pitch perception. As the number of channels further increased, the channels became narrower than 500 Hz, which would limit the cutoff frequencies of the extracted temporal envelopes in the sine-vocoded signal. This effect may eventually significantly impair the neural phase-locking to the temporal periodicity cues. On the other hand, improved spectral resolution and place coding of F0 with more channels may overcome the degraded neural phase-locking to temporal periodicity cues, and thus lead to better pitch perception in behavioral studies.

Effects of temporal envelope cutoff frequency and carrier type on brainstem neural representation of the F0

Our observation of improved brainstem F0 representation with the 500-Hz rather than 50-Hz temporal envelopes was consistent with the better Mandarin tone recognition, voice gender identification, and vocal emotion recognition with higher temporal envelope cutoff frequencies (as reviewed in the Introduction). As suggested by Rosen (1992), slow variations in the overall amplitude at rates between 2–50 Hz provide information for vowel identification, manner of articulation, and voicing. On the other hand, F0-related periodicity fluctuations at rates between 50 and 500 Hz provide listeners with voicing, manner, and prosodic cues. In this study, the 500-Hz temporal envelopes preserved the F0 and low-frequency harmonic cues of the stimulus at 120, 240, and 360 Hz that were not available in the 50-Hz temporal envelopes. The brainstem neurons generating the FFR were found to phase lock to the F0-related periodicity fluctuations in the 500-Hz temporal envelopes. This

reveals the neural mechanisms of pitch perception in simulated electric hearing and may explain the improved performance in pitch-related listening tasks with 500-Hz rather than 50-Hz temporal envelopes (e.g., Fu et al., 1998; Xu et al., 2002; Fu et al., 2004; Luo et al., 2007). The subcortical neural responses also suggest that temporal periodicity cues may be useful in enhancing pitch perception with real CIs. Studies have shown promising results in this direction (Vandali & van Hoesel 2012; Milczynski et al. 2009).

There were no significant differences between the sine-wave and noise-band carriers in brainstem neural representation of the F0, even when the temporal envelope cutoff frequency was 500 Hz. The lack of a carrier effect in the current study was not consistent with the findings from previous behavioral studies that indicated perceptual advantages of sine-wave carriers over noise-band carriers, especially in pitch-related listening tasks such as gender and talker identification (Fu et al. 2004; Gonzalez & Oliver 2005; Stone et al. 2008; Souza & Rosen 2009). The sine-vocoder advantage observed in these studies has been attributed to the lack of random level variations in sine-vocoders that may distort the speech envelope, and the presence of spectral sidebands that may represent F0-related periodicity cues in the speech envelope, especially with high temporal envelope cutoff frequencies. Our results did not provide neurophysiological evidence at the subcortical level that sine-wave carriers are better than noise-band carriers in preserving temporal pitch cues. Such inconsistency may be because the effect of carrier type on brainstem neural representation of the F0 was only tested with 8 channels in the present study. The results in Experiment 1 showed that brainstem neural representation of the F0 was stronger with less than 8 channels. It is thus possible that a stronger effect of carrier type on brainstem neural representation of the F0 may also be observed with a smaller number of channels (e.g., 4). Grand-average waveform periodicity of FFR_{ENV} appeared to be somewhat reduced in the noise- as compared to the sine-vocoder condition when the temporal envelope cutoff frequency was 500 Hz. An additional one-way repeated measures ANOVA was used to compare the periodicity strengths of FFR_{ENV} extracted from the autocorrelation functions in different stimulus conditions (i.e., unprocessed, Sine-500, Sine-50, Noise-500, and Noise-50). Pairwise Tukey comparisons did not reveal significant differences in the periodicity strengths of FFR_{ENV} between the Sine-500 and Noise-500 conditions ($p > 0.34$), similar to the results of F0 magnitudes.

Effect of the number of vocoder channels on brainstem neural representation of the stimulus temporal fine structure

In general, spectra of the FFR_{TFS} to the vocoded signals were different from those of the FFR_{TFS} to the unprocessed signal. In the sine-vocoder conditions, spectral peaks of the FFR_{TFS} were observed at the channel center frequencies and spectral sideband frequencies of the sine-vocoders rather than the harmonic frequencies of the original unprocessed signal. To our knowledge, the phase-locking of brainstem neurons to the sine-wave carriers and sideband frequencies (i.e., artifacts of the sine-vocoder processing) has not been previously reported. The FFR_{TFS} accurately reflected the different temporal fine structures of the carriers in the sine-vocoded stimuli with varying number of channels. This is consistent with previous FFR studies which demonstrated precise representation of TFS cues contained in

various stimuli (Krishnan 1999; Krishnan & Parkinson 2000; Plyler & Ananthanarayan 2001; Ananthakrishnan et al. 2016).

Further, the strongest spectral peak of the FFR_{TFS} in each sine-vocoded condition was at a channel center frequency or a spectral sideband frequency near F1. This may be reflective of a neural phenomenon known as “synchrony capture”, wherein brainstem neurons exhibit enhanced phase-locking to spectral peaks near formant frequencies and diminished phase-locking to other spectral peaks. It is thought that synchrony capture is responsible for formant encoding, which is critical for vowel recognition. This phenomenon has been demonstrated in the single unit response in animal models (Miller et al. 1997; Palmer & Moorjani 1993), as well as in the ensemble response of brainstem FFR to steady-state speech stimuli in normal hearing listeners (Krishnan 2002, Ananthakrishnan et al. 2016). Increased phase-locking strength at channel center frequencies or spectral sideband frequencies near F1 in the FFR_{TFS} to sine-vocoded signals indicates the preservation of synchrony capture, even when the stimulus is greatly degraded. Such a selective boost in the brainstem neural responses may provide rough information about the spectral shape of the original unprocessed signal for normal hearing listeners to identify vowel sounds. These data may explain why speech recognition is possible even with the distorted neural responses to the sine- and noise-vocoded signals (Shannon et al. 1995; Dorman et al. 1997).

Similar to the effect of channel number on brainstem F0 representation, the strength of brainstem neural representation of TFS or F1 cues was greater with 1–4 channels than with 8–32 channels, although the difference was not significant. This finding seems inconsistent with the behavioral results which suggest that in general, speech perception with vocoded signals improves with the number of channels. However, studies of auditory chimeras (Smith et al. 2002) did show that perceptual salience of TFS cues decreased while that of temporal envelope cues increased as the number of channels increased. Similar mechanisms may underlie the similar patterns of degraded brainstem representation for both F0 and formant-related spectral peaks with more channels. It is possible that when there is a large number of channels, the narrow bandwidth of each channel would adversely affect the neural representation of TFS cues.

Effects of temporal envelope cutoff frequency and carrier type on brainstem neural representation of the stimulus temporal fine structure

Spectral peaks of the FFR_{TFS} in the 8-channel sine-vocoder conditions with either 50- or 500-Hz temporal envelopes were observed at the channel center frequencies and spectral sideband frequencies of the sine-vocoders, consistent with the discussion above. With 8 channels, formant information seems to be represented mainly by the channel center frequencies rather than the spectral sideband frequencies. Synchrony capture or greater response magnitudes close to F1 was slightly but not significantly stronger in the Sine-50 condition than in the Sine-500 condition, possibly because 500-Hz temporal envelopes had part of the energy distributed at the spectral sideband frequencies while 50-Hz temporal envelopes did not. This agrees with the lack of effect of temporal envelope cutoff frequency on vowel recognition that mainly used formant cues (Shannon et al. 1995). On the other hand, no spectral peaks were observed in the FFR_{TFS} to noise-vocoded signals, consistent

with the nature of the noise-band carriers (i.e., energy was evenly distributed within a channel rather than focused on the center frequency). With the noise-vocoded signals, listeners may have determined the spectral shape of the original unprocessed signal from the energy distribution across channels. Such a spectral profile analysis may happen at higher levels along the auditory pathway (Wang & Shamma 1995) and was thus not observed in the brainstem neural responses.

Future directions

Presently, our understanding of vocoded speech perception is largely from behavioral results. Only two studies (to our knowledge) have examined neural representation of vocoded signals, and both focused on cortical evoked potentials. The current study describes brainstem neural representation of acoustic information in vocoded signals, adding to our knowledge of vocoded speech processing and perception. While our data establish the feasibility of FFR as a possible electrophysiological tool to study brainstem neural representation of vocoded speech, we cannot make better-informed predictions of vocoded speech perception until we have access to both behavioral data and corresponding neural metrics from multiple levels in the auditory system (e.g., brainstem and cortex) in the same group of subjects. Note that while there is some degree of agreement between the current FFR data and the behavioral literature on vocoded speech perception (in terms of the effect of the temporal envelope cutoff frequency), there are also a few key inconsistencies (e.g., the degraded neural representation of F0 with 32 channels and the lack of carrier effect). The disconnect between behavioral and electrophysiological measurements is not unexpected; the “brain-behavior” inconsistencies, as termed by Friesen et al. (2009), may be attributed to the different stages of auditory processing involved in behavioral responses and electrophysiological potentials. It is likely that vocoded speech perception is also influenced by higher level processing capacities such as cognition and memory (Friesen et al. 2009). FFR recordings in the current study were measured in response to novel vocoded signals in a group of fairly homogeneous subjects in terms of age, education level, and hearing ability, who had no prior listening training with these stimuli. Therefore, the recorded FFRs are likely reflective of early, pre-attentive, sensory-level processing, and thus may have limited power to predict vocoded speech perception. However, it should be noted that phase locked neural activity in the brainstem, as reflected by the FFR, is indeed preserving information about certain acoustic features that may contribute to the development of speech perception, presumably at a later stage in the processing hierarchy. At the current time, there is no neural index to determine how acoustic features important for speech perception are reflected in CI-stimulated auditory system. Given its ability to capture both envelope and fine structure information, the FFR may serve as an efficient platform to study neural representation of CI-speech perception. This, in turn, may facilitate development of improved signal processing strategies to optimize neural representation of acoustic features that are important for speech perception. Future research is needed to confirm if FFR is indeed a useful tool to understand the neural underpinnings of speech perception in the CI population. Further, given the susceptibility of FFR to long-term plasticity effects from training or experience, it is reasonable to hypothesize that brainstem neural representation may vary with significant long-term hearing loss and subsequent use of amplification/prosthetic devices. However, in the present study, we have limited ourselves to discussing only measurable results, such as

envelope and TFS strength in the brainstem FFR in response to vocoded stimuli in normal hearing individuals, rather than making extrapolations to CI users with subject and processing variables that could potentially alter the FFR.

CONCLUSIONS

The present study examined the effects of varying spectral and temporal cues on brainstem neural representation of speech. Specifically, we examined the effects of varying the number of vocoder channels, temporal envelope cutoff frequency, and carrier type on neural representation of temporal envelope (F0) and TFS cues (F1-related spectral peaks) as reflected in the brainstem FFR to vocoded signals. To our knowledge, this is the first study examining the human FFR to vocoded speech stimuli. Both temporal envelope and TFS representation improved when the number of channels increased from 1 to 4; however, a decrement was noted for both as the number of channels increased to 32. We hypothesize that this decrement in neural phase-locking is due to the narrow bandwidth of each channel as the number of channels increases. When the temporal envelope cutoff frequency increased, we noted an improvement in brainstem F0 representation and no change in neural representation of F1-related spectral peaks. It is likely that the brainstem neurons phase lock to the F0-related periodicity fluctuations preserved in the 500-Hz temporal envelopes, which are unavailable in the 50-Hz temporal envelopes. Further, spectral sideband frequencies of the 500-Hz temporal envelopes did not improve neural representation of F1 cues over the 50-Hz temporal envelopes. Finally, inconsistent with previous behavioral studies on pitch-related listening tasks, the brainstem F0 representation was slightly but not significantly different between the sine-wave and noise-band carriers when the temporal envelope cutoff frequency was 500 Hz.

Acknowledgments

S.A. designed and performed experiments, analyzed data and wrote the paper; A.K. and X.L. provided critical revisions; X.L. assisted with statistical analysis. The authors wish to thank Laura Grinstead, B.S. (Towson University) for assistance with data collection, and Dr. Gavin M. Bidelman (University of Memphis), Dr. Christopher J. Smalt (MIT Lincoln Laboratory) and Mr. Venkatakrishnan Vijayaraghavan for assistance with MATLAB coding. Research supported by NIH R01 DC008549 (A.K.), the Department of Speech, Language and Hearing Sciences (Purdue University), FDRC # 140255 (S.A), and the Department of Audiology, Speech Language Pathology and Deaf Studies (Towson University). Portions of this article were presented at the *Conference on Implantable Prostheses (CIAP) 2013* between July 14–19, 2013 at Lake Tahoe, CA.

References

- Aiken SJ, Picton TW. Envelope and spectral frequency-following responses to vowel sounds. *Hearing research*. 2008; 245(1–2):35–47. DOI: 10.1016/j.heares.2008.08.004 [PubMed: 18765275]
- Ananthakrishnan S, Krishnan A, Bartlett E. Human Frequency Following Response: Neural Representation of Envelope and Temporal Fine Structure in Listeners with Normal Hearing and Sensorineural Hearing Loss. *Ear and Hearing*. 2016
- Anderson S, Parbery-Clark A, White-Schwoch T, et al. Effects of hearing loss on the subcortical representation of speech cues. *The Journal of the Acoustical Society of America*. 2013; 133(5): 3030–8. <http://doi.org/10.1121/1.4799804>. [PubMed: 23654406]
- Elsisy H, Krishnan A. Comparison of response characteristics of acoustical and neural distortion product at 2f1-f2 in normal hearing adults. *International Journal of Audiology*. 2008; 47:431–438. [PubMed: 18574781]

- Friesen LM, Tremblay KL, Rohila N, et al. Evoked cortical activity and speech recognition as a function of the number of simulated cochlear implant channels. *Clinical Neurophysiology*. 2009; 120(4):776–782. [PubMed: 19250865]
- Fu QJ, Chinchilla S, Galvin JJ. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology*. 2004; 5(3):253–260. [PubMed: 15492884]
- Fu QJ, Zeng FG, Shannon RV, et al. Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*. 1998; 104(1):505–510. [PubMed: 9670541]
- Gonzalez J, Oliver JC. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *The Journal of the Acoustical Society of America*. 2005; 118(1):461–470. [PubMed: 16119365]
- Greenberg S, Marsh JT, Brown WS, et al. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research*. 1987; 25(2–3):91–114. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3558136>. [PubMed: 3558136]
- Greenwood DD. A cochlear frequency-position function for several species-29 years later. *The Journal of the Acoustical Society of America*. 1990; 87(6):2592–2605. [PubMed: 2373794]
- Krishnan A. Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiology and Neurotology*. 1999; 4(2):95–103. [PubMed: 9892760]
- Krishnan A. Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research*. 2002; 166(1–2):192–201. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12062771>. [PubMed: 12062771]
- Krishnan A, Gandour JT, Bidelman GM, et al. Experience dependent neural representation of dynamic pitch in the brainstem. *Neuroreport*. 2009; 20(4):408. [PubMed: 19223793]
- Krishnan A, Parkinson J. Human frequency-following response: representation of tonal sweeps. *Audiology and Neurotology*. 2000; 5(6):312–321. [PubMed: 11025331]
- Krishnan A, Plack C. Auditory brainstem correlates of basilar membrane nonlinearity. *Journal of Audiology and Neuro-Otology*. 2009; 14:88–97. [PubMed: 18827479]
- Krishnan A, Xu Y, Gandour JT, et al. Human frequency-following response: Representation of pitch contours in Chinese tones. *Hearing research*. 2004; 189(1–2):1–12. DOI: 10.1016/S0378-5955(03)00402-7 [PubMed: 14987747]
- Luo, X., Fu, QJ. Importance of pitch and periodicity to Chinese-speaking cochlear implant patients. *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on; IEEE; 2004. p. iv-1.*
- Luo X, Fu QJ, Galvin JJ. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*. 2007; 11(4):301–315. [PubMed: 18003871]
- Miller RL, Schilling JR, Franck KR, et al. Effects of acoustic trauma on the representation of the vowel “eh” in cat auditory nerve fibers. *The Journal of the Acoustical Society of America*. 1997; 101(6):3602–16. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9193048>. [PubMed: 9193048]
- Milczynski M, Wouters J, Van Wieringen A. Improved fundamental frequency coding in cochlear implant signal processing. *The Journal of the Acoustical Society of America*. 2009; 125(4):2260–2271. [PubMed: 19354401]
- Palmer AR, Moorjani PA. Responses to speech signals in the normal and pathological peripheral auditory system. *Progress in brain research*. 1992; 97:107–115.
- Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*. 2013; 23(6):1378–1387. [PubMed: 22610394]
- Plyler PN, Ananthanarayan AK. Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*. 2001; 12(10):523–33. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11791939>. [PubMed: 11791939]
- Qin MK, Oxenham AJ. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and hearing*. 2005; 26(5):451–460. [PubMed: 16230895]

- Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions: Biological Sciences*. 1992; 336(1278):367–373. Retrieved from <http://www.jstor.org/stable/55906>. [PubMed: 1354376]
- Shannon RV, Feng FG, Kamath V, et al. Speech recognition with primarily temporal cues. *Science*. 1995; 270(5234):303–304. [PubMed: 7569981]
- Smalt CJ, Krishnan A, Bidelman GM, et al. Distortion products and their influence on representation of pitch-relevant information in the human brainstem for unresolved harmonic complex tones. *Hearing Research*. 2012; 292(1–2):26–34. <http://doi.org/10.1016/j.heares.2012.08.001>. [PubMed: 22910032]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*. 2002; 416:87–90. [PubMed: 11882898]
- Souza P, Rosen S. Effects of envelope bandwidth on the intelligibility of sine-and noise-vocoded speech. *The Journal of the Acoustical Society of America*. 2009; 126(2):792–805. [PubMed: 19640044]
- Stone MA, Füllgrabe C, Moore BC. Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region. *The Journal of the Acoustical Society of America*. 2008; 124(4):2272–2282. [PubMed: 19062865]
- Vandali AE, van Hoesel RJ. Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking. *The Journal of the Acoustical Society of America*. 2012; 132(1):392–402. [PubMed: 22779486]
- Vandali AE, Whitford LA, Plant KL, Clark GM. Speech perception as a function of electrical stimulation rate using the Nucleus 24 cochlear implant system. *Ear and Hearing*. 2000; 21(6):608–624. [PubMed: 11132787]
- Wang K, Shamma SA. Spectral shape analysis in the central auditory system. *IEEE Transactions on Speech and Audio Processing*. 1995; 3(5):382–395.
- Xu L, Tsai Y, Pfingst BE. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *The Journal of the Acoustical Society of America*. 2002; 112(1):247–258. [PubMed: 12141350]
- Zeng FG. Trends in cochlear implants. *Trends in Amplification*. 2004; 8(1):1–34. [PubMed: 15247993]

Figure 1.

Stimulus Fast Fourier Transforms (FFT) (panel A), grand average FFR_{ENV} response waveforms (panel B), and response FFTs (panel C) for the unprocessed (blue), 1- (black), 2- (orange), 4- (green), 8- (turquoise), 16- (grey) and 32- (red) channel stimuli. Vertical arrows in response FFTs indicate the spectral peaks (F₀ in FFR_{ENV}) that were chosen for analysis in each condition.

Figure 2.

Individual (black) and mean (orange) F0 magnitudes measured in the FFR_{ENV} condition for unprocessed, 1-, 2-, 4-, 8-, 16- and 32-channel stimuli. The asterisk (*) indicates a statistically significant difference ($p < 0.05$).

Figure 3. Stimulus Fast Fourier Transforms (FFT) (panel A), grand average FFR_{ENV} response waveforms (panel B), and response FFTs (panel C) for the unprocessed (blue), and sine- and noise-vocoded stimuli with temporal envelope cutoff frequencies at 500 (black) and 50 (red) Hz. Vertical arrows in response FFTs indicate the spectral peaks (F₀ in FFR_{ENV}) that were chosen for analysis in each condition.

Figure 4.

Individual (black) and mean (orange) F0 magnitudes measured in the FFR_{ENV} condition for unprocessed, and sine- and noise-vocoded stimuli with temporal envelope cutoff frequencies at 500 and 50 Hz. The asterisk (*) indicates a statistically significant difference ($p < 0.05$).

Figure 5. Stimulus Fast Fourier Transforms (FFT) (panel A), grand average FFR_{TFS} response waveforms (panel B), and response FFTs (panel C) for the unprocessed (blue), 1- (black), 2- (orange), 4- (green), 8- (turquoise), 16- (grey) and 32- (red) channel stimuli. Vertical arrows in response FFTs indicate the spectral peaks (closest to F1 in FFR_{TFS}) that were chosen for analysis in each condition.

Figure 6.

Individual (black) and mean (orange) F1-related spectral peak magnitudes measured in the FFR_{TFS} condition for unprocessed, 1-, 2-, 4-, 8-, 16- and 32-channel stimuli. The asterisk (*) indicates a statistically significant difference ($p < 0.05$).

Figure 7. Stimulus Fast Fourier Transforms (FFT) (panel A), grand average FFR_{TFS} response waveforms (panel B), and response FFTs (panel C) for the unprocessed (blue), and sine- and noise-vocoded stimuli with temporal envelope cutoff frequencies at 500 (black) and 50 (red) Hz. Vertical arrows in response FFTs indicate the spectral peaks (closest to F1 in FFR_{TFS}) that were chosen for analysis in each condition.

Figure 8. Individual (black) and mean (orange) F1-related spectral peak magnitudes measured in the FFR_{TFS} condition for unprocessed, and sine- and noise-vocoded stimuli with temporal envelope cutoff frequencies at 500 and 50 Hz. The asterisk (*) indicates a statistically significant difference ($p < 0.05$).