



## Research paper

## Neural encoding in the human brainstem relevant to the pitch of complex tones

Ananthanarayan Krishnan<sup>a,\*</sup>, Christopher J. Plack<sup>b</sup><sup>a</sup>Department of Speech Language Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA<sup>b</sup>Human Communication and Deafness Division, School of Psychological Sciences, The University of Manchester, Manchester M13 9PL, UK

## ARTICLE INFO

## Article history:

Received 19 August 2010

Received in revised form

6 December 2010

Accepted 8 December 2010

Available online 16 December 2010

## ABSTRACT

Psychoacoustic studies have shown that complex tones containing resolved harmonics evoke stronger pitches than complex tones with only unresolved harmonics. Also, unresolved harmonics presented in alternating sine and cosine (ALT) phase produce a doubling of pitch. We examine here whether the temporal pattern of phase-locked neural activity reflected in the scalp recorded human frequency following response (FFR) preserves information relevant to pitch strength, and to the doubling of pitch for ALT stimuli. Results revealed stronger neural periodicity strength for resolved stimuli, although the effect of resolvability was weak compared to the effect observed behaviorally; autocorrelation functions and FFR spectra suggest a different pattern of phase-locked neural activity for ALT stimuli with resolved and unresolved harmonics consistent with the doubling of pitch observed in our behavioral estimates; and the temporal pattern of neural activity underlying pitch encoding appears to be similar at the auditory nerve (auditory nerve model response) and the rostral brainstem level (FFR). These findings suggest that the phase-locked neural activity reflected in the scalp recorded FFR preserves neural information relevant to pitch that could serve as an electrophysiological correlate of the behavioral pitch measure. The scalp recorded FFR may provide for a non-invasive analytic tool to evaluate neural encoding of complex sounds in humans.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Pitch is a fundamental auditory perceptual attribute that plays a crucial role in the perception of speech, language, music, and in the analysis of complex auditory scenes. In speech, voice pitch carries information about both the prosodic features and talker identity. In tonal languages, like Mandarin, variations in pitch signal meaning differences in monosyllabic words. In music, changes in pitch convey melody. In auditory scene analysis, differences in pitch provide a major cue for sound source segregation (Bregman et al., 1990; Darwin et al., 1995).

Studies evaluating neural encoding of the pitch of harmonic complex tones in the auditory nerve (AN) and cochlear nucleus (CN) have shown that pitch cues are available in both the temporal patterns of neural activity (phase locking) and in the spatial distribution of activity along the tonotopic axis. Most studies have focused on temporal encoding schemes because they provide a unified and parsimonious way of explaining a diverse range of pitch phenomena (Cariani and Delgutte, 1996a, b; Meddis and O'Mard, 1997).

The temporal models are based solely on pooling the timing information available in the inter-spike intervals represented in the simulated (Meddis and O'Mard, 1997) or actual (Evans, 1983; Shofner, 1991b; Rhode, 1995; Cariani and Delgutte, 1996a, b; Cedolin and Delgutte, 2005; Larsen et al., 2008) neural activity across nerve fibers without regard to the frequency-to-place mapping. These periodicity cues can be extracted by an autocorrelation-type mechanism (Licklider, 1951; Meddis and Hewitt, 1991; Yost, 1996), which is equivalent to an all-order inter-spikeinterval distribution for neural spike trains. This interval-based representation can predict the pitch of both resolved and unresolved harmonics (Meddis and Hewitt, 1991; Cariani and Delgutte, 1996a; Cedolin and Delgutte, 2005), although basic autocorrelation models do not capture the superior pitch encoding for resolved harmonics (Carlyon, 1998; Bernstein and Oxenham, 2005). Thus, neural phase locking related to fundamental frequency (F0) may play a dominant role in the encoding of the low pitch associated with complex sounds.

Most periodic complex sounds evoke low pitches associated with their F0, sometimes termed periodicity pitch (deBoer, 1976; Evans, 1978; Moore, 1989). Energy may or may not be present at the F0. The salience of the perceived pitch for these complex sounds has been shown to decrease as the lowest present harmonic number is increased (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003b). The difference in pitch salience between complex sounds

\* Corresponding author. Tel.: +1765 494 3793; fax: +1 765 494 0771.

E-mail addresses: [rkrish@purdue.edu](mailto:rkrish@purdue.edu) (A. Krishnan), [chris.plack@manchester.ac.uk](mailto:chris.plack@manchester.ac.uk) (C.J. Plack).

with low- and high-order harmonics has been explained in terms of harmonic resolvability (Shackleton and Carlyon and Shackleton, 1994). That is, the absolute bandwidth of cochlear filters increases with increasing center frequency. Since the frequency spacing between harmonics is constant in Hz, each low-order harmonic of a complex tone dominates the output of a single (relatively narrow) auditory filter. These harmonics are effectively separated out by the cochlea, and are said to be resolved. In contrast, multiple high-order harmonics above about the 10th harmonic (Plomp, 1964; Bernstein and Oxenham, 2003b) fall within the (relatively broad) auditory filters and are therefore unresolved. Thus, unresolved harmonics provide only temporal cues for pitch; the waveform resulting from a combination of unresolved harmonics has a period equal to that of the complex tone. Although both resolved and unresolved harmonics of a common F0 produce the same pitch, complex sounds containing resolved harmonics tend to evoke stronger pitches and have smaller F0 discrimination thresholds than complex sounds with only unresolved harmonics (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994). However, when successive harmonics are presented to alternating ears, increasing the peripheral resolution, the variation in F0 discrimination with harmonic number is largely unaffected (Bernstein and Oxenham, 2003b). This seems to depend in part on the ear that receives the even or odd harmonics being randomly allocated between observation intervals, encouraging listeners to use a central pitch (Bernstein and Oxenham, 2008). Hence, harmonic number, rather than peripheral harmonic resolution, may be crucial in determining pitch salience.

In addition, unresolved harmonics presented in alternating sine and cosine (ALT) phase produce a temporal envelope periodicity of 2F0 and a pitch matched to 2F0 (Shackleton and Carlyon, 1994a). This doubling in pitch is thought to be a behavioral measure of harmonic interaction on the basilar membrane.

We examine here whether these changes in pitch salience, and the doubling of pitch, are reflected in the phase-locked neural activity reflected in the scalp recorded human frequency following response (FFR). The scalp recorded FFR reflects sustained phase-locked neural activity in a population of neural elements within the rostral brainstem (Worden and Marsh, 1968; Marsh et al., 1974; Smith et al., 1975; Glaser et al., 1976), presumably the inferior colliculus (IC). FFRs have been shown to preserve information about formants of speech sounds (Krishnan, 1999, 2002; Krishnan and Parkinson, 2000; Aiken and Picton, 2008) and pitch relevant information of both steady-state (Greenberg et al., 1987) and dynamic complex sounds including speech (Krishnan et al., 2004, 2005) and nonspeech “iterated ripple noise” stimuli (Swaminathan et al., 2008a, b; Krishnan and Gandour, 2009; Krishnan et al., 2009a, b). If indeed neural information relevant to pitch salience and pitch doubling is preserved in the FFR then it may provide for a robust electrophysiologic measure to evaluate the nature of the human brainstem neural activity relevant to pitch.

## 2. Materials and methods

### 2.1. Participants

Ten adult listeners (3 male, 7 female) ranging in age from 18 to 24 years participated in the FFR experiment. Five subjects participated in the behavioral experiment, three of whom also participated in the FFR experiment. All participants exhibited normal hearing sensitivity (i.e., better than 15 dB HL in both ears) at octave frequencies between 500 and 8000 Hz. All participants were students (enrolled at Lancaster University or Purdue University) who were paid for their participation. Each participant gave informed consent in compliance with a protocol approved by the Ethics Committee of Lancaster University and the Institutional Review Board of Purdue University.

### 2.2. Stimuli

For both the FFR and the behavioral experiment complex tone burst stimuli with equal-amplitude harmonics added in either sine (SIN) phase or ALT phase were used. SIN-phase complexes had an F0 of 90 Hz (SIN 90) or 180 Hz (SIN 180). ALT phase complexes had an F0 of 90 Hz (ALT 90). For each of the complexes, harmonics were filtered into one of four spectral regions: 360–900 Hz, 720–1260 Hz, 1080–1620 Hz, and 1440–1980 Hz (3-dB cutoffs, 450 dB/oct slopes). The stimulus waveforms and their spectra are shown in Fig. 1. These spectral regions were chosen to include stimuli with low-order harmonics that are completely resolved, and stimuli with higher-order harmonics that are completely unresolved, in the cochlea. A low-pass Gaussian noise was gated with each complex to mask combination tones. The cutoff frequency of the noise was 180 Hz below the start of the complex pass-band in each case (i.e., 180 Hz, 540 Hz, 900 Hz, and 1260 Hz for the four spectral regions). The spectrum level of the noise was chosen so that the noise had the same average spectral density as the complex. The noise was frozen (same sample) for each complex tone condition. All stimuli had an overall duration of 150 ms including the 5 ms onset and offset ramps.

### 2.3. FFR recording protocol

Participants reclined comfortably in an acoustically and electrically shielded booth. They were instructed to relax and refrain from extraneous body movement to minimize myogenic artifacts. Subjects were allowed to sleep through the duration of the FFR experiment. FFRs were recorded from each participant in response to monaural presentation of the twelve stimuli. Stimuli were presented to the right ear at an overall level of 85 dB SPL through a magnetically shielded insert earphone (Bio-logic, ER-3A). Note that this means that the level of the complex tone decreased slightly (by about 3 dB) as overall bandwidth increased from 900 to 1980 Hz. Stimuli were presented at a repetition rate of 3.13/sec. The presentation order of the stimuli was randomized both within and across participants. Control of the experimental protocol was accomplished by a signal generation and data acquisition system (Intelligent Hearing Systems, SmartEP, Advanced Research Module).

FFRs were recorded differentially between a non-inverting (+) electrode placed on the midline of the forehead at the hairline (Fz) and an inverting (–) electrode placed on (i) the ipsilateral mastoid (A2); and (ii) the 7th cervical vertebra (C7). Another electrode placed on the mid-forehead (Fpz) served as the common ground. FFRs were recorded simultaneously from the two electrode configurations, and subsequently averaged for each stimulus condition to yield a response with a higher signal-to-noise ratio (Krishnan et al., 2010). All inter-electrode impedances were maintained at or below 1 k $\Omega$ . The raw EEG inputs were amplified by 200,000 and band-pass filtered from 30 to 3000 Hz (6 dB/octave roll-off, RC response characteristics). In addition, each averaged FFR was digitally band-pass filtered from 60 to 2500 Hz post acquisition. Sweeps containing activity exceeding  $\pm 30$   $\mu$ V were rejected as artifacts. Each FFR response waveform represents the average of 2000 artifact free trials over a 160 ms acquisition window.

### 2.4. FFR data analysis

Since the FFR reflects phase-locked activity in a population of neural elements, we adopted a temporal analysis scheme in which we examined the periodicity information contained in the aggregate distribution of neural activity (Langner, 1983; Rhode, 1995; Cariani and Delgutte, 1996a, b). Short-term autocorrelation functions (ACFs) and running autocorrelograms (ACGs) were computed from the averaged FFRs derived from each participant to index variation in

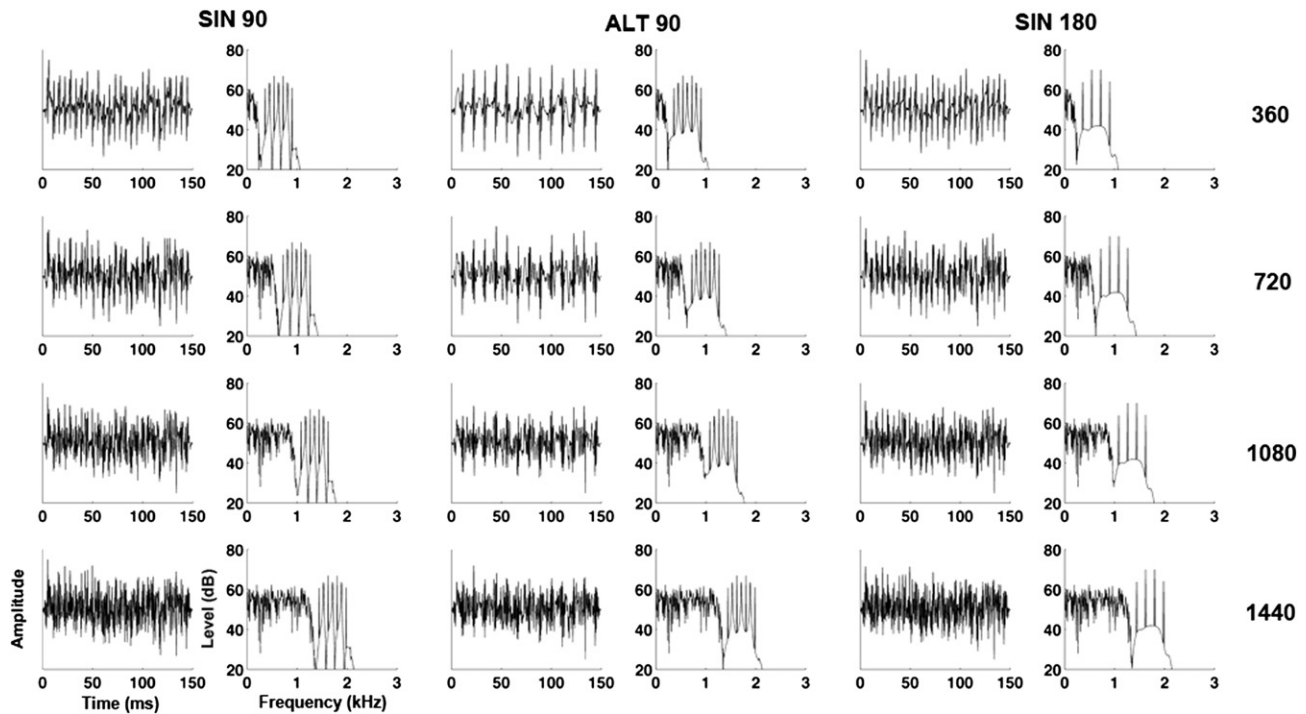


Fig. 1. Stimulus waveforms (left panel) and spectra (right panel) of the complex harmonic sounds used in the study.

FFR periodicities over the duration of the response. The ACF is equivalent to an all-order inter-spike interval histogram (ISIH) and represents the dominant pitch periodicities present in the neural response (Cariani and Delgutte, 1996b; Krishnan et al., 2004). The ACG represents the short-term ACF of windowed frames of a compound signal, i.e.,  $ACG(\tau, t) = X(t) \times X(t - \tau)$  for each time-lag  $\tau$ . It is a three-dimensional plot quantifying the variations in periodicity and neural pitch strength (i.e., degree of phase locking) as a function of time. The horizontal axis represents the time at which single ACF “slices” are computed while the vertical axis represents their corresponding time-lags, i.e., pitch periods. The intensity of each point in the image represents the instantaneous ACF magnitude computed at a given time within the response. Mathematically, the running ACF is the time-domain analog to the frequency-domain spectrogram. In terms of neurophysiology, it represents the running distribution of all-order inter-spike intervals present in the population neural activity (Cariani and Delgutte, 1996a; Sayles and Winter, 2008).

To analyze the robustness of encoding to stimuli differing in pitch, and in pitch salience, we quantified the neural *periodicity strength* of each response waveform. From each FFR, the normalized ACF (calculated over the entire duration of the response) was computed in order to determine the dominant periodicities contained within the response. The first prominent peak in the ACF away from time-lag zero was taken as the magnitude of neural pitch strength (Krishnan et al., 2005). Since, two ACF peaks (one corresponding to the period of 90 Hz and the other corresponding to the period of 180 Hz) with different magnitudes were observed in the FFR for the ALT stimuli, the ACF peak 1 (P1) to Peak 2 (P2) ratio was used to determine the dominant pitch relevant periodicity. It was reasoned that if the ACF peak associated with the longer delay (P2) was dominant then it would yield a relatively smaller ratio compared to when the ACF peak associated with the shorter delay (P1) was dominant. The magnitude of the ACF peak (the measure of neural pitch strength) has shown a high correspondence with the pitch salience of a stimulus in both behavioral studies with humans and single-unit neurophysiologic studies in

animal models (Cariani and Delgutte, 1996a; Yost et al., 1996). While this autocorrelation based pitch strength measure was utilized to quantify FFR pitch data, spectral analyses (estimate of power spectral density using Welch’s method using the default settings in Matlab) were also performed on the FFRs to determine if the changes in the spectral pattern of the FFR corresponded with pitch changes produced by stimuli used in the study.

### 2.5. Statistical analysis of FFR data

ACF P1/P2 ratio was analyzed using a two-way mixed model ANOVA with subjects as a random factor, and stimulus (SIN 90, ALT 90, SIN 180) and cutoff frequency (360, 720, 1080, 1440) as within-subject factors in order to assess whether pitch encoding differed between stimuli and across cutoff frequencies. In addition, a separate one-way ANOVA was utilized to evaluate if changes in neural pitch strength as a function of cutoff frequency were significant for FFRs in response to SIN 90, and SIN 180 stimuli.

### 2.6. Comparison of FFR data with auditory nerve model predictions

In order to determine if a temporal pitch encoding scheme based on neural phase locking is still available at the level of the inferior colliculus, the presumed site of FFR generation, the pitch relevant temporal pattern of neural activity reflected in the FFR was compared to the temporal pattern of the auditory nerve output for our stimuli predicted by a multichannel autocorrelation model of pitch processing (Meddis and Hewitt, 1991) as implemented by (Patterson et al., 1995). Essentially, the model consists of four sequential processing stages: (1) peripheral frequency selectivity; (2) within-channel half-wave rectification and low-pass filtering; (3) within-channel periodicity extraction using autocorrelation; and (4) cross channel summary autocorrelation which represents the aggregate of all temporal periodicity information contained in the entire ensemble of AN fibers. This summary ACF is similar to an all-order inter-spike interval histogram commonly reported in

single-unit studies (Cariani and Delgutte, 1996b). Similar to our FFR response analysis, the peak 1 to peak 2 ratio measures were obtained from the summary autocorrelation function generated by the auditory nerve model for all stimuli.

### 2.7. Behavioral measures

A three interval forced choice (3IFC) odd-one-out paradigm was used to obtain the behavioral data from five normal hearing listeners. On each trial, listeners were presented with the three complexes (SIN 90, ALT 90, and SIN 180) for one of the four spectral regions consecutively in a random order. The task was to choose which interval contained the tone with a different pitch, via a mouse click in a custom GUI coded in MATLAB® 7.5 (The MathWorks, Inc., Natick, MA). The aim of the procedure was to determine, for each frequency region, if ALT 90 sounded more similar to SIN 90 (same F0) or to SIN 180 (same envelope periodicity). In the former case, listeners tended to choose SIN 180 as the odd-one-out. In the latter case, listeners tended to choose SIN 90 as the odd-one-out.

The spectral and temporal characteristics of the stimuli were the same as those used for the FFR experiment. The inter-stimulus interval on each trial was 500 ms. Stimuli were presented monaurally to the right ear at an overall level of 80 dB SPL through circumaural headphones (Sennheiser HD 580). For each frequency region, two practice blocks and five experimental blocks of 20 trials were presented. The frequency regions were tested in a random order for each participant. The behavioral experiment was conducted after the FFR experiment.

## 3. Results

### 3.1. Encoding of pitch as reflected in the FFR

The grand averaged FFR waveforms for the SIN 90, ALT 90, and SIN 180 stimuli across the four spectral regions are shown in Fig. 2.

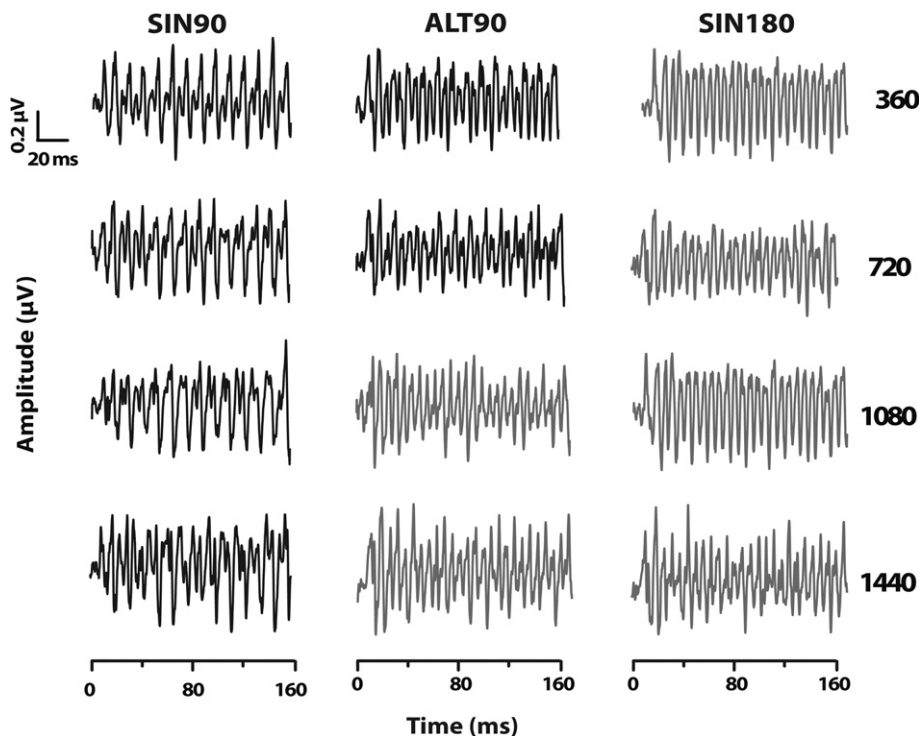


Fig. 2. Grand averaged FFR waveforms across spectral regions (identified on the right) for SIN 90, ALT 90, and SIN 180 stimuli.

The FFR waveforms for SIN 90 and SIN 180 clearly show a lower frequency and a higher frequency periodicity, respectively, across the four spectral regions. Visually the FFR waveforms for the ALT 90 stimuli have periodicity that is closer to the SIN 90 FFRs for the two lower spectral regions and a periodicity that is closer to SIN 180 FFRs at the two higher spectral regions. For all stimuli, the FFR periodicity appears to become less robust for stimuli with harmonics in the higher spectral regions.

Mean FFR spectra for each of the four spectral regions of the three stimuli are shown in Fig. 3. The spectral data reinforce the initial observations on the FFR waveforms. For the SIN 90 and SIN 180 stimuli, FFR spectra show clear peaks at the F0, and at integer multiples corresponding to the 90 and 180 Hz F0s, respectively. Also, the magnitude of the spectral peaks for both stimuli appears to decrease, and the peaks show poorer resolution for the highest spectral region. For the ALT 90 stimulus conditions, the FFR spectral patterns for stimuli with harmonics in the two lower spectral regions are similar to the FFR spectral pattern for the SIN 90 stimulus conditions. That is, spectra are consistent with a 90 Hz F0. In contrast, the spectral pattern for stimuli with harmonics in the two higher spectral regions for this stimulus is similar to the spectral pattern observed for the SIN 180 stimulus conditions. That is, spectra are consistent with a 180 Hz F0.

Mean ACFs and the corresponding mean ACGs are shown in Fig. 4. The ACF for the SIN 90 condition shows a prominent peak at a delay corresponding to the period of 90 Hz (11.1 ms) whose magnitude is smaller for the 1440–1980 spectral region. Consistent with this is the observation of a single band of phase-locked activity at this delay in the ACG. For the SIN 180 condition, the ACF shows two peaks (a shorter delay corresponding to the period of 180 Hz (5.55 ms), and a longer delay corresponding to the period of 90 Hz) with the first peak slightly greater in amplitude across the spectral regions. Also, peak magnitude decreases with increasing low frequency cutoff of the spectral regions. The two bands corresponding to these peaks are clearly discernible in the ACG. Again,

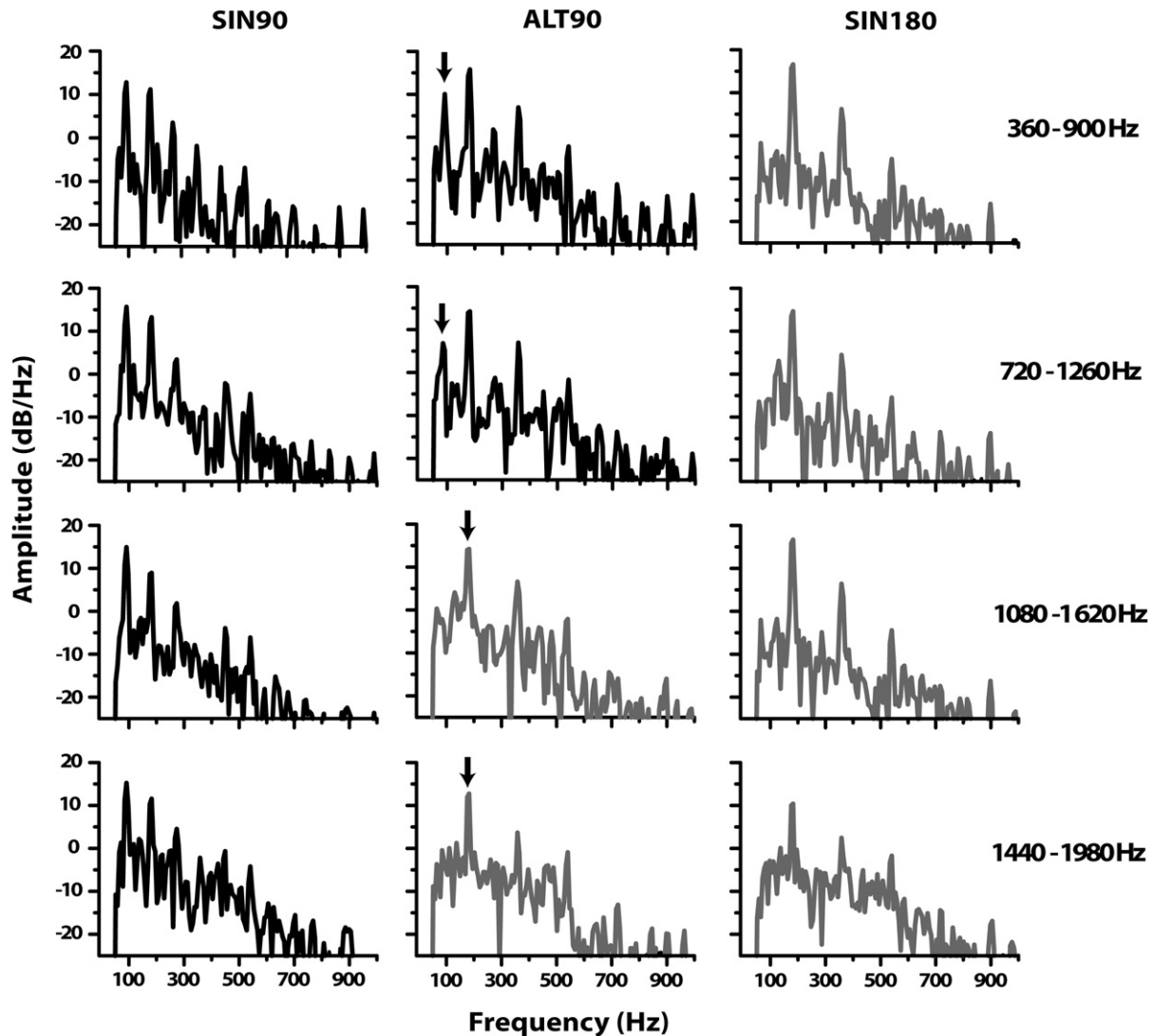


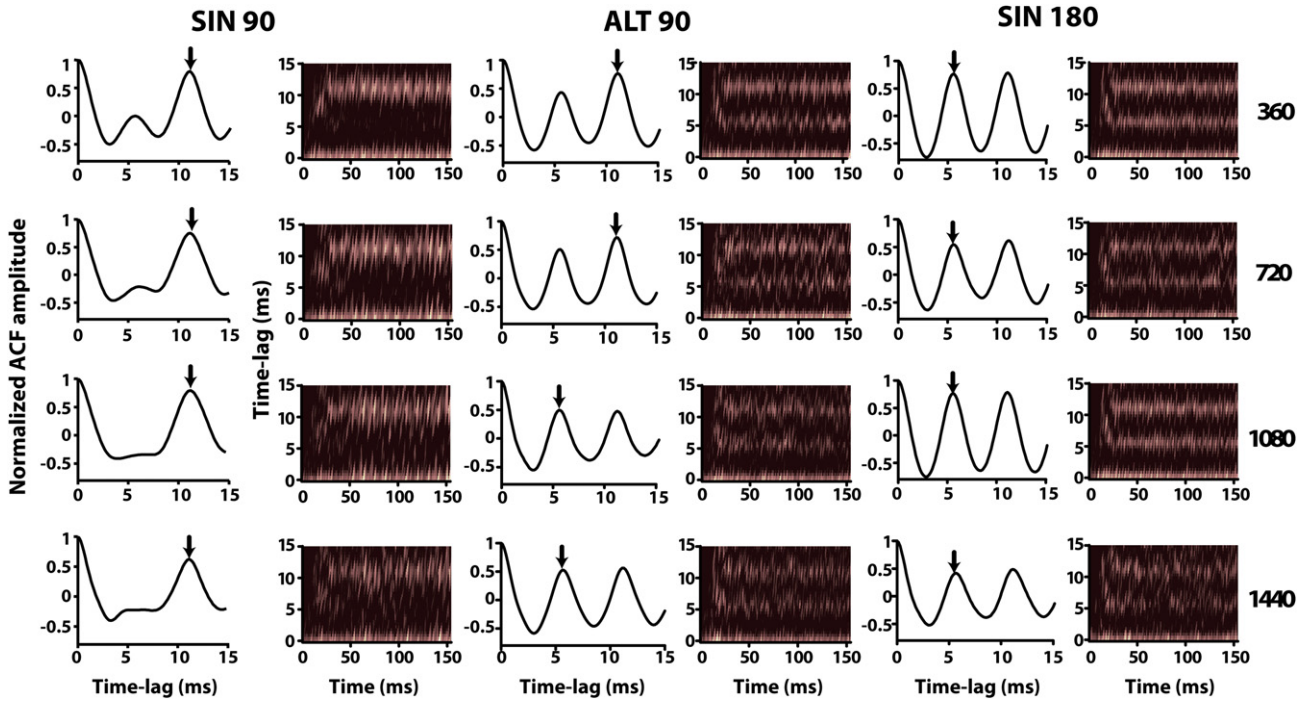
Fig. 3. Mean FFR spectra for each stimuli plotted across the four spectral regions. The spectral regions are identified the right of each row of plots.

note that the bands reflecting phase-locked activity at these periods become less prominent for stimuli with harmonics in the higher spectral regions. ANOVA Omnibus results revealed significant main effects for cutoff frequency ( $F_{3,10} = 8.75$ ,  $p = 0.0003$  for SIN 90, and  $F_{3,10} = 9.11$ ,  $p = 0.0002$  for SIN 180) suggesting that pitch strength decreased with increasing cutoff frequency. Posthoc Tukey–Kramer adjusted multiple comparisons ( $\alpha = 0.001$ ) revealed that, for both SIN 90 and SIN 180 stimuli; the two lowest cutoff frequencies (360 and 720 Hz) together produced significantly higher neural pitch strengths than the two highest cutoff frequencies (1080 and 1440 Hz) (i.e.,  $[360 = 720] > [1080 = 1440]$ ). The ACFs for the ALT 90 also show two peaks. However, unlike the ACFs for SIN 180, the second peak at a delay corresponding to the period of 90 Hz is relatively bigger than the first peak (corresponding to a periodicity of 180 Hz) for the two lower spectral regions. For the two higher spectral regions, the two peaks are either similar in magnitude or the first peak is greater in magnitude than the second peak. The prominence of the band at the longer delay in the ACF for the two lower spectral regions is consistent with these observations.

To further quantify these observed qualitative differences in FFR pitch representation across stimuli and spectral bands, the auto-correlation magnitude at a delay corresponding to 90 Hz and at

a delay corresponding to 180 Hz were compared by expressing the ratio of ACF peak magnitude of the first peak with that of the second peak (ACF P1/P2 ratio). The results of this comparison are shown in Fig. 5. It is clear from this figure that the larger second ACF peak for the SIN 90 stimuli yielded a smaller ratio, which was essentially the same across the spectral bands. Similarly, the nearly equal-magnitude peaks for the SIN 180 stimuli yielded a bigger ratio, which remained the same across spectral bands. In contrast, P1/P2 ACF ratios for the ALT 90 are similar to the ratios observed for SIN 90 for the two lower spectral bands, and similar to SIN 180 for the two higher spectral bands. ANOVA Omnibus results revealed significant main effects of stimulus ( $F_{2,72} = 631.63$ ,  $p < 0.0001$ ) and cutoff frequency ( $F_{3,27} = 7.75$ ,  $p = 0.0007$ ) as well as their interaction ( $F_{6,72} = 21.39$ ,  $p < 0.0001$ ). By stimulus, posthoc Tukey–Kramer adjusted multiple comparisons ( $\alpha = 0.001$ ) revealed that the cutoff frequency only had a significant effect on the ACF P1/P2 ratio in the case of the ALT 90 stimulus. For this stimulus, the two lowest cutoff frequencies (360 and 720 Hz) together produced significantly lower ACF P1/P2 ratios than the two highest cutoff frequencies (1080 and 1440 Hz) (i.e.,  $[360 = 720] < [1080 = 1440]$ ).

Overall, these results suggest that the temporal pattern of neural activity relevant to pitch, as reflected in the FFRs, is different for ALT



**Fig. 4.** Mean autocorrelation functions and their corresponding autocorrelograms for each stimuli plotted across the four spectral regions. The spectral regions are identified the right of each row of plots. Arrows identify the prominent peak in the autocorrelation function. While the ACF y-axis represents normalized amplitude ranging between  $-1$  and  $1$ , ACF y-axis represents lag time (ms)-i.e., pitch periods.

stimuli with resolved and unresolved harmonics. Specifically, the temporal pattern of phase-locked neural activity shifted from a dominant 90 Hz periodicity to a dominant 180 Hz periodicity when the lowest harmonic in the complex was shifted from a completely resolved spectral region to a completely unresolved spectral region. Also, the pitch representation was robust for stimuli with harmonics in the resolved spectral region and less robust when stimulus spectrum contained only unresolved harmonics.

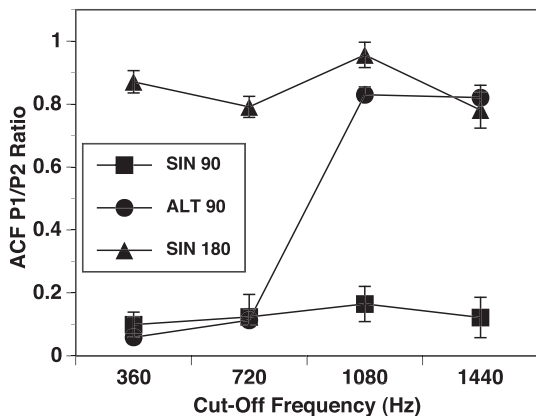
**3.2. Comparison with auditory nerve model responses**

The ACFs representing the auditory nerve model output (in gray) overlaid on the ACFs obtained from the FFR data (in black) are shown in Fig. 6. Comparison of these ACFs for each stimulus and

across spectral bands reveal a striking qualitative similarity both in terms of the location of the peaks and their relative magnitudes. Of particular interest is the close correspondence between the ACFs for the ALT 90 stimulus, including the shift in prominence from peak 2 (corresponding to a fundamental of 90 Hz) for the resolved spectral regions to peak 1 (corresponding to a fundamental of 180 Hz) for the unresolved spectral regions. In addition, P1/P2 ratios obtained from the model ACFs (shown in Fig. 7) also show essentially the same pattern observed for the FFR P1/P2 ratio measures. The similarity of the temporal pattern of neural activity at the auditory nerve level (model response) and presumably at the level of the IC, as reflected in the FFR, suggests that the temporal pattern of neural activity relevant to pitch observed at the auditory nerve level may still be available at the level of the IC.

**3.3. Behavioral measures of pitch**

The mean number of responses that the ALT 90 was judged most similar to the SIN 90 (squares showing SIN 180 odd-one-out responses) and the mean number of responses that the ALT 90 was judged most similar to the SIN 180 (circles showing SIN 90 odd-one-out responses) are plotted in Fig. 8. The ALT 90 stimuli were judged to be similar in pitch to the SIN 90 stimuli when harmonics fell in the lowest spectral region (resolved harmonics). In contrast, the ALT 90 stimuli were judged to be similar in pitch to the SIN 180 stimuli when harmonics were restricted to two higher unresolved spectral regions. In other words, the pitch of the ALT 90 switches from 90-Hz equivalent for the low-resolved spectral regions to 180-Hz equivalent for the high-unresolved spectral regions. While our FFR results are similar to these behavioral results suggesting that pitch relevant information preserved in the FFR could serve as an electrophysiological correlate of the behavioral pitch measure, it should be noted that the behavioral results show ambiguity in pitch judgments for the 720 Hz region.



**Fig. 5.** Mean FFR peak 1 to peak 2 ratio (P1/P2 ratio) plotted as a function of spectral regions for the SIN 90, ALT 90, and SIN 180 stimuli. The error bars correspond to 1 SEM.

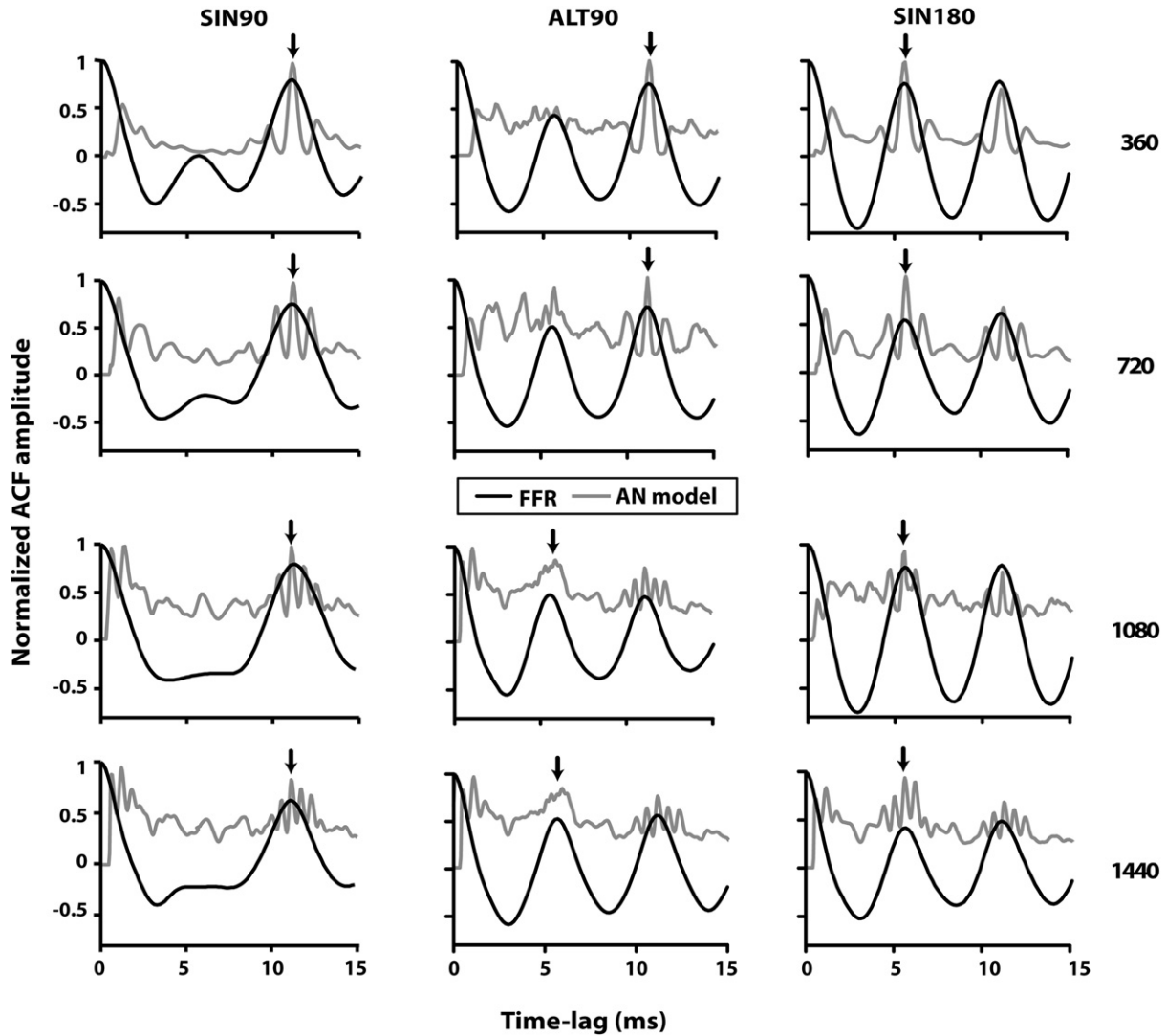


Fig. 6. Comparison of the autocorrelation functions for the FFR (black trace) and the auditory nerve model (grey) for each stimuli across the four spectral regions.

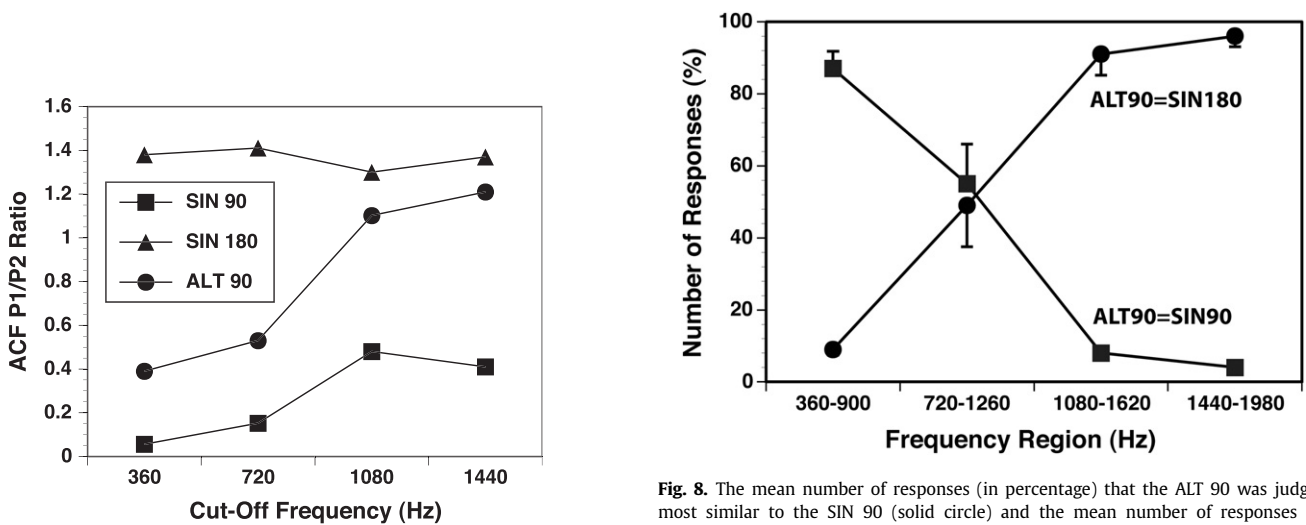


Fig. 7. Mean auditory nerve model peak 1 to peak 2 ratio (P1/P2 ratio) plotted as a function of spectral regions for the SIN 90, ALT 90, and SIN 180 stimuli. The error bars correspond to 1 SEM.

Fig. 8. The mean number of responses (in percentage) that the ALT 90 was judged most similar to the SIN 90 (solid circle) and the mean number of responses (in percentage) that the ALT 90 was judged most similar to the SIN 180 (solid square) plotted as a function of spectral region of the stimulus components. The error bars correspond to 1 SEM. Note that the curves are not exactly complementary, as there were a few responses for which SIN 90 was judged most similar to SIN 180.

#### 4. Discussion

Three main observations can be made from the results of this study: (1) FFR periodicity strength decreases as harmonic resolution decreases, consistent *qualitatively* with previous behavioral measures. However, the reduction in the FFR response is small in comparison to the large change in the F0 difference limen (FODL) in the behavioral results. For example, Houtsma and Smurzynski (1990) reported an increase in the FODL from about 0.25% to 2.5% as the lowest harmonic number increased from 7 to 13. (2) FFR neural pitch strength ratio (derived from the ACF peak 1 and peak 2 magnitude) and FFR spectral patterns suggest different patterns of phase-locked neural activity for ALT 90 stimuli with resolved and unresolved harmonics, broadly consistent with the doubling of pitch observed behaviorally; (3) the temporal pattern of neural activity underlying pitch encoding appears to be similar at the auditory nerve (auditory nerve model response) and the rostral brainstem level (FFR).

##### 4.1. Neural encoding of the pitch of harmonic complexes

Most physiological studies on pitch encoding of harmonic complexes in the auditory nerve (AN), cochlear nucleus (CN), and IC have focused on temporal pitch cues available in the pooled inter-spike interval distributions (ISIDs) (Evans, 1983; Palmer, 1990, 1993; Shofner, 1991b; Rhode, 1995; Cariani and Delgutte, 1996a, b; Winter et al., 2003; Cedolin and Delgutte, 2005; Shackleton et al., 2009). The pooled interval distribution is the sum of the all-order inter-spike interval distributions and is closely related to summary autocorrelation used in some models of pitch (Meddis and Hewitt, 1991). This interval-based pitch representation works with both resolved and unresolved harmonics (Meddis and Hewitt, 1991; Cariani and Delgutte, 1996a, b; Cedolin and Delgutte, 2005). The observation of robust peaks at delays corresponding to the F0 in the autocorrelation functions for the SIN 90 and SIN 180 stimuli in our FFR data suggest that pitch relevant information is preserved in the phase-locked activity in the neural elements generating the FFR. In addition, to the extent that the magnitude of the ACF peak represents neural pitch strength, and therefore perceived pitch salience (Cariani and Delgutte, 1996a, b), the smaller ACF peak magnitudes for stimuli with unresolved components compared to stimuli with resolved components in our FFR data suggests a reduction in pitch salience for stimuli with unresolved components. While this latter observation is consistent with psychophysical and model data, it is generally thought that the autocorrelation model does not sufficiently account for the greater pitch salience for stimuli with resolved harmonics compared to stimuli with unresolved harmonics (Carlyon and Shackleton, 1994, 1998; Meddis and O'Mard, 1997; Bernstein and Oxenham, 2003a, 2005). This is also seen in our data, in that the reduction in response is relatively small compared to the large change in salience as measured by the FODL.

Physiological studies show a consistent correlate of the pitch doubling for ALT stimuli as harmonic resolution decreases. That is, the ISIDs not only show peaks at F0 and 2F0 but also, the interval-based measure of pitch strength is almost as large at the envelope frequency 2F0 as at the F0 for alternating phase stimuli with unresolved harmonics based on both period histograms and ACGs (Horst et al., 1990; Palmer and Winter, 1992, 1993; Cedolin and Delgutte, 2005; Shackleton et al., 2009). However, since physiological studies generally use the first peak in the ACF to estimate pitch, the results for ALT stimuli have been interpreted to suggest that pitch extraction based on autocorrelation is not sensitive to the doubling in pitch observed in behavioral experiments because the first ACF peak occurs at the same delay for both resolved and unresolved stimuli. The P1/P2 ratio metric applied to our FFR data clearly showed a different pattern of neural activity for ALT 90 stimuli with

resolved and unresolved components consistent with the doubling of pitch observed in behavioral studies (Ritsma and Engel, 1964; Lundeen and Small, 1984; Shackleton and Carlyon, 1994a). Given these observations it is plausible that the shift in the relative prominence of the autocorrelation peaks in our FFR data for ALT 90 stimuli serves as a neural correlate of the doubling of perceived pitch observed for ALT 90 stimuli in our behavioral estimates and in previous psychophysical studies (Ritsma and Engel, 1964; Lundeen and Small, 1984; Shackleton and Carlyon, 1994a).

In addition to the autocorrelation analysis showing different pattern of neural activity for the ALT resolved and unresolved stimuli, Fourier analysis of the FFR data showed a clear shift in spectral pattern from one consistent with a 90 Hz F0 for the resolved ALT 90 stimuli to one more consistent with a 180 Hz F0 for unresolved ALT 90 stimuli, again strongly suggesting the presence of a neural correlate for the doubling of pitch in the FFR data. Shackleton et al. (2009) observed that neural activity from clusters of neurons in the central nucleus of IC showed peaks at  $1/F_0$  and  $1/2F_0$  in the ACFs in response to alternating phase stimuli with unresolved components. Since the autocorrelation of stimulus with period  $1/f$ , is expected to produce intervals at all integer multiples of the period ( $n/f$ ); the peak at  $1/F_0$  in alternating phase conditions is the second order response to a period of  $1/2F_0$  (i.e.,  $2/2F_0$ ). Thus, if the response is predominantly at F0, then we should expect components at both F0 and 2F0 in the Fourier analysis but only at  $1/F_0$  in the SACF, whereas if the response is predominantly at 2F0, then we expect a component only at 2F0 in the Fourier analysis but at both  $1/F_0$  and  $1/2F_0$  in the SACF. Thus, the results of the autocorrelation analysis using the P1/P2 ratio and the Fourier analysis of the FFR data taken together, clearly suggest that the temporal pattern of neural activity relevant to perceptual doubling of pitch is indeed preserved in the neural activity underlying the FFR. Furthermore, the similarity between the autocorrelation analyses performed at the level of the auditory nerve (model response) and at the level of the IC (FFR) appears to suggest that a temporal representation of pitch based on pooled neural ISIDs is still potentially available at the level of the midbrain.

There is a caveat regarding the physiological mechanisms underlying the FFR, however. The FFRs measured in the experiment might reflect phase locking to both the envelope and the fine structure of the waveforms at different locations on the basilar membrane. It is thought that the FFR for unresolved harmonics is related to phase locking to the envelope. However, phase locking to envelope might also contribute to the FFR for resolved harmonics, since the output of an auditory filter positioned between two harmonics will beat at the frequency difference between them (i.e., F0). If only two harmonics are driving the response, then the modulation rate will not be affected by the relative phase of the harmonics. Only when three or more harmonics interact will the modulation rate double for ALT phase complexes. In other words, the pattern of shifts in P1/P2 ratio could potentially be explained, at least qualitatively, based on a response to the envelope alone. This differs from the usual assumption of pitch perception models that pitch is derived from combining fine-structure information about the individual frequencies of resolved harmonics. Evidence against an interpretation purely in terms of envelope is that the FFR at F0 was greater for the resolved harmonics in the present study, while the output of an auditory filter between two resolved harmonics would be less than the output of a filter responding to several unresolved harmonics. However, it remains unclear as to what extent the FFR reflects a neural representation that is used by the auditory system to derive pitch.

##### 4.2. Comparison with behavioral measures of pitch

The behavioral pitch estimates for our SIN 90, SIN 180, ALT 90 (resolved components) corresponding to F0 and the doubling of



pitch for our ALT 90 stimuli with unresolved components are consistent with results of previous psychophysical studies (Ritsma and Engel, 1964; Lundeen and Small, 1984; Shackleton and Carlyon, 1994a). In general, these psychophysical results have shown that the pitch of unresolved harmonics is weaker and more dependent on phase relationships than the pitch based on resolved harmonics (Houtsma and Smurzynski, 1990; Carlyon, 1998; Bernstein and Oxenham, 2003b). While stimuli with unresolved harmonics do not provide spectral cues to pitch, they do produce direct temporal cues to pitch because the waveform of a combination of unresolved harmonics has a period equal to that of the complex tone. The greater periodicity strength observed in our FFR data for resolved compared to unresolved stimuli is consistent with these psychophysical results. However, as noted earlier, the decrease in pitch strength (with increasing cutoff frequency) in our neural data is appreciably smaller than the decrease in salience observed in behavioral experiments suggesting that the FFR may not provide a robust correlate of pitch salience (with the caveat that the units of ACF correlation strength and percent F0 difference may not be directly comparable).

For unresolved stimuli, several components will interact to produce amplitude modulation at F0 for SIN stimuli and at 2F0 for ALT stimuli. Our FFR data are consistent with this in that the FFR showed a prominent peak at F0 for SIN stimuli and at 2F0 for ALT unresolved stimuli. Thus, the temporal pattern of neural activity reflected in the FFR appears to be sensitive to both pitch and pitch salience.

## 5. Conclusion

Autocorrelation, and spectral analysis of FFR to complex harmonic tones taken together suggest that the temporal pattern of phase-locked neural activity among a population of neural elements is able to account qualitatively for the greater salience of pitch for stimuli with resolved harmonics compared to unresolved harmonics, although the effect size appears to be small compared to the change in perception. Also, the variation of P1/P2 ratio applied to the ACF is clearly consistent with doubling of pitch for ALT stimuli with unresolved harmonics. In addition, the similarity in the ACFs for the FFR and the auditory nerve model response suggests that the temporal pitch encoding scheme observed at the auditory nerve is still available at the midbrain level to extract pitch relevant information for complex sounds producing low pitch. Finally, the correlation between the FFR data and the behavioral estimates of pitch suggests that the phase-locked neural activity reflected in the scalp recorded FFR preserves sensory level pitch information that may contribute to pitch perception. It can be concluded that the scalp recorded FFR may provide for a non-invasive analytic tool to evaluate neural encoding of complex sounds in humans.

## Acknowledgments

Research supported by a grant from the University of Lancaster, UK; BBSRC (UK) grant BB/D012953/1 (C.P.) and by NIH R01, DC 008549 (A.K.). We would also like to thank Gavin Bidelman and Chris Smalt for their invaluable help in data analysis, and an anonymous reviewer for constructive comments on an earlier version of the manuscript. Reprint requests should be addressed to Ananthanarayan Krishnan, Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, IN, USA 47907-2038, or via email: rkrish@purdue.edu

## References

Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47.

- Bernstein, J.G.W., Oxenham, A.J., 2003a. Effects of relative frequency, absolute frequency, and phase on fundamental frequency discrimination: data and an autocorrelation model. *J. Acoust. Soc. Am.* 113, 2290.
- Bernstein, J.G.W., Oxenham, A.J., 2003b. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* 113, 3323–3334.
- Bernstein, J.G.W., Oxenham, A.J., 2005. An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *J. Acoust. Soc. Am.* 117, 3816.
- Bernstein, J.G.W., Oxenham, A.J., 2008. Harmonic segregation through mistuning can improve fundamental frequency discrimination. *J. Acoust. Soc. Am.* 124, 1653.
- Bregman, A.S., Liao, C., Levitan, R., 1990. Auditory grouping based on fundamental frequency and formant peak frequency. *Can. J. Psychol.* 44, 400–413.
- Cariani, P.A., Delgutte, B., 1996a. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716.
- Cariani, P.A., Delgutte, B., 1996b. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J. Neurophysiol.* 76, 1717–1734.
- Carlyon, R.P., 1998. Comments on “a unitary model of pitch perception” *J. Acoust. Soc. Am.* 102: 1811–1820 (1997). *J. Acoust. Soc. Am.* 104, 1118–1121.
- Carlyon, R.P., Shackleton, T.M., 1994. Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J. Acoust. Soc. Am.* 95, 3541–3554.
- Cedolin, L., Delgutte, B., 2005. Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *J. Neurophysiol.* 94, 347–362.
- Darwin, C.J., Hukin, R.W., al-Khatib, B.Y., 1995. Grouping in pitch perception: evidence for sequential constraints. *J. Acoust. Soc. Am.* 98, 880–885.
- deBoer, E., 1976. On the residue and auditory pitch perception. In: Keidel, W.D., Neff, W.D. (Eds.), *Handbook of Sensors Physiology*. Springer-Verlag, Berlin, pp. 479–583.
- Evans, E.F., 1978. Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology* 17, 369–420.
- Evans, E.F., 1983. Pitch and cochlear nerve fibre temporal discharge patterns. In: Klinke, R., Hartmann, R. (Eds.), *Hearing: Physiological Bases and Psychophysics*. Springer-Verlag, Berlin, pp. 140–146.
- Glaser, E.M., Suter, C.M., Dasheiff, R., Goldberg, A., 1976. The human frequency-following response: its behavior during continuous tone and tone burst stimulation. *Electroencephalogr. Clin. Neurophysiol.* 40, 25–32.
- Greenberg, S., Marsh, J.T., Brown, W.S., Smith, J.C., 1987. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear. Res.* 25, 91–114.
- Horst, J.W., Javel, E., Farley, G.R., 1990. Coding of spectral fine structure in the auditory nerve. II. Level-dependent nonlinear responses. *J. Acoust. Soc. Am.* 88, 2656–2681.
- Houtsma, A.J., Smurzynski, J., 1990. Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87, 304–310.
- Krishnan, A., 1999. Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiol. Neurootol.* 4, 95–103.
- Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. *Hear. Res.* 166, 192–201.
- Krishnan, A., Gandour, J.T., 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain Lang.* 110, 135–148.
- Krishnan, A., Parkinson, J., 2000. Human frequency-following response: representation of tonal sweeps. *Audiol. Neurootol.* 5, 312–321.
- Krishnan, A., Xu, Y., Gandour, J.T., Cariani, P.A., 2004. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear. Res.* 189, 1–12.
- Krishnan, A., Xu, Y., Gandour, J.T., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* 25, 161–168.
- Krishnan, A., Swaminathan, J., Gandour, J.T., 2009a. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J. Cogn. Neurosci.* 21, 1092–1105.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009b. Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport* 20, 408–413.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010. The effects of tone language experience on pitch processing in the brainstem. *J. Neurol.* 23, 81–95.
- Langner, G., 1983. Evidence for neuronal periodicity detection in the auditory system of the guinea fowl: implications for pitch analysis in the time domain. *Exp. Brain Res.* 52, 333–355.
- Larsen, E., Cedolin, L., Delgutte, B., 2008. Pitch representations in the auditory nerve: two concurrent complex tones. *J. Neurophysiol.* 100, 1301–1319.
- Licklider, J.C., 1951. A duplex theory of pitch perception. *Experientia* 7, 128–134.
- Lundeen, C., Small, A.M.J., 1984. The influence of temporal cues on the strength of periodicity pitches. *J. Acoust. Soc. Am.* 75, 1578–1587.
- Marsh, J.T., Brown, W.S., Smith, J.C., 1974. Differential brainstem pathways for the conduction of auditory frequency-following responses. *Electroen. Clin. Neuro.* 36, 415–424.
- Meddis, R., Hewitt, T., 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery I: pitch identification. *J. Acoust. Soc. Am.* 1862–1882.
- Meddis, R., O'Mard, L., 1997. A unitary model of pitch perception. *J. Acoust. Soc. Am.* 102, 1811–1820.
- Moore, B.C., 1989. *Introduction to the Psychology of Hearing*, third ed.. Academic Press, London.
- Palmer, A.R., 1990. The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea-pig cochlear-nerve fibers. *J. Acoust. Soc. Am.* 88, 1412–1426.

- Palmer, A.R., Winter, I.M., 1992. Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones. In: Horner, K. (Ed.), *Auditory Physiology and Perception*. Pergamon, Oxford, pp. 231–240.
- Palmer, A.R., Winter, I.M., 1993. Coding of the fundamental frequency of voiced speech sounds and harmonic complex tones in the ventral cochlear nucleus. In: Mugnaini, E. (Ed.), *Mammalian Cochlear Nuclei: Organization and Function*. Plenum, New York, pp. 373–384.
- Patterson, R.D., Allerhand, M.H., Giguere, C., 1995. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am.* 98, 1890–1894.
- Plomp, R., 1964. The ear as a frequency analyzer. *J. Acoust. Soc. Am.* 36, 1628–1636.
- Rhode, W.S., 1995. Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus. *J. Acoust. Soc. Am.* 97, 2414–2429.
- Ritsma, R.J., Engel, F.L., 1964. Pitch of frequency-modulated signals. *J. Acoust. Soc. Am.* 36, 1637–1644.
- Sayles, M., Winter, I.M., 2008. Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron* 58, 789–801.
- Shackleton, T.M., Carlyon, R.P., 1994a. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 95, 3529–3540.
- Shackleton, T.M., Liu, L.F., Palmer, A.R., 2009. Responses to diotic, dichotic, and alternating phase harmonic stimuli in the inferior colliculus of guinea pigs. *J. Assoc. Res. Otolaryngol.* 10, 76–90.
- Shofner, W.P., 1991b. Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla. *J. Acoust. Soc. Am.* 90, 2450–2466.
- Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroen. Clin. Neuro.* 39, 465–472.
- Swaminathan, J., Krishnan, A., Gandour, J.T., 2008a. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 19, 1163–1167.
- Swaminathan, J., Krishnan, A., Gandour, J.T., 2008b. Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Trans. Biomed. Eng.* 55, 281–287.
- Winter, I.M., Palmer, A.R., Wiegand, L., Patterson, R.D., 2003. Temporal coding of the pitch of complex sounds by presumed multipolar cells in the ventral cochlear nucleus. *Space Comm.* 41, 135–139.
- Worden, F.G., Marsh, J.T., 1968. Frequency-following (microphonic-like) neural responses evoked by sound. *Electroen. Clin. Neuro.* 25, 42–52.
- Yost, W.A., 1996. Pitch of iterated rippled noise. *J. Acoust. Soc. Am.* 100, 511–518.
- Yost, W.A., Patterson, R., Sheft, S., 1996. A time domain description for the pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.* 99, 1066–1078.