The Frequency Following Response (FFR) May Reflect Pitch-Bearing Information But is Not a Direct Representation of Pitch

HEDWIG E. GOCKEL[1], ROBERT P. CARLYON[1], ANAHITA MEHTA[1], AND CHRISTOPHER J. PLACK[2]

[1]MRC-Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK
[2]Division of Human Communication and Deafness, The University of Manchester, Manchester M13 9PL, UK

ABSTRACT

The frequency following response (FFR), a scalp-recorded measure of phase-locked brainstem activity, is often assumed to reflect the pitch of sounds as perceived by humans. In two experiments, we investigated the characteristics of the FFR evoked by complex tones. FFR waveforms to alternating-polarity stimuli were averaged for each polarity and added, to enhance envelope, or subtracted, to enhance temporal fine structure information. In experiment 1, frequency-shifted complex tones, with all harmonics shifted by the same amount in Hertz, were presented diotically. Only the autocorrelation functions (ACFs) of the subtraction-FFR waveforms showed a peak at a delay shifted in the direction of the expected pitch shifts. This expected pitch shift was also present in the ACFs of the output of an auditory nerve model. In experiment 2, the components of a harmonic complex with harmonic numbers 2, 3, and 4 were presented either to the same ear ("mono") or the third harmonic was presented contralaterally to the ear receiving the even harmonics ("dichotic"). In the latter case, a pitch corresponding to the missing fundamental was still perceived. Monaural control conditions presenting only the even harmonics ("2+4") or only the third harmonic ("3") were also tested. Both the subtraction and the addition waveforms showed that (1) the FFR magnitude spectra for "dichotic" were similar to the sum of the spectra for the two monaural control conditions and lacked peaks at the fundamental frequency and other distortion products visible for "mono" and (2) ACFs for "dichotic" were similar to those for "2+4" and dissimilar to those for "mono." The results indicate that the neural responses reflected in the FFR preserve monaural temporal information that may be important for pitch, but provide no evidence for any additional processing over and above that already present in the auditory periphery, and do not directly represent the pitch of dichotic stimuli.

Keywords: complex tones, dichotic presentation, monaural temporal information

INTRODUCTION

The extraction of the pitch of a complex sound is one of the most important functions performed by the auditory system. Pitch conveys melody in music and prosodic information in speech and provides one of the most powerful cues to the perceptual segregation of competing sounds (Darwin and Carlyon 1995). It is perhaps not surprising, therefore, that there is ongoing interest not only in the stimulus parameters that dominate pitch judgments but also in the neural processes that are important for pitch perception. A number of recent studies have provided evidence for a neural representation of pitch in the auditory cortex (Patterson et al. 2002; Penagos et al. 2004; Bendor and Wang 2005; but see Hall and Plack 2007, 2009; Garcia et al. 2010). However, these cortical measures may simply reflect processing that occurs at earlier stages, and several researchers have investigated the

Correspondence to: Hedwig E. Gockel · MRC-Cognition and Brain Sciences Unit · 15 Chaucer Road, Cambridge CB2 7EF, UK. Telephone: 44-1223-273697; fax: 44-1223-359062; e-mail: hedwig.gockel@mrc-cbu.cam.ac.uk

representation of pitch in the auditory brainstem (Smith et al. 1978; Greenberg et al. 1987; Galbraith 1994; Galbraith and Doan 1995; Krishnan 2002, 2006; Krishnan et al. 2004, 2005; Musacchia et al. 2007; Wile and Balaban 2007; Wong et al. 2007; Song et al. 2008; Skoe and Kraus 2010). Of particular interest here is the noninvasive measure known as the frequency following response (FFR), which can be obtained using electrodes attached to the scalp of human participants. The FFR reflects sustained synchronous phase-locked activity in a population of neurons that phase-lock to stimulus-related periodicities (Marsh et al. 1975; Smith et al. 1975; Glaser et al. 1976). The anatomical generators contributing to the FFR are often determined from the latency of the FFR, and the contribution of possible anatomical generators to the observed FFR seems to depend on the electrode configuration, i.e., it is very likely affected by the strength and orientation of the generated electrical field relative to the electrode configuration (e.g., Galbraith 1994; for an overview, see Krishnan 2006). Here, we are concerned with the FFR of a latency between 6 and 10 ms (Smith et al. 1975; Glaser et al. 1976; Skoe and Kraus 2010), suggesting a generation site at the level of the inferior colliculus (IC) or lateral lemniscus (LL).

Greenberg and colleagues (Greenberg et al. 1978; Smith et al. 1978; Greenberg et al. 1987) were among the first to argue that the FFR reflects the neural representation of the pitch of complex sounds in the upper auditory brainstem. They recorded the FFR to a complex tone in which the fundamental frequency (F_0) was absent and showed that the frequency spectrum of the response contained a component at this missing F_0, which persisted even in the presence of a low-pass noise. This mirrored the well-known behavioral finding that listeners hear a missing F_0 (residue pitch or low pitch) even in the presence of a noise that would have masked any distortion product corresponding to that frequency (Licklider 1956). They also showed that the phase sensitivity of the FFR depended on the harmonic numbers of the component frequencies present, in a manner similar to that observed with pitch judgments. Furthermore, they recorded the FFR in response to frequency-shifted complexes, where all components have been shifted by a fixed amount in Hertz re their harmonic frequency values, and calculated the mean of 7–12 estimates of the time interval between negative peaks in the FFR. They argued that this short-term characteristic of the FFR reflected the ambiguous pitches that listeners report for such stimuli. Subsequent experiments on the FFR evoked by tones with a missing fundamental (Galbraith 1994) and by frequency-shifted complexes (Wile and Balaban 2007) led to broadly similar conclusions.

Interest in brainstem responses to complex sounds, such as the FFR and the more generic "complex ABR" ("cABR"; for a review, see Skoe and Kraus 2010) has been reignited by evidence that the responses can be modified by experience and may be degraded in some clinical populations. For example, Krishnan et al. (2005) showed that the representation of the pitch contours that code lexical contrasts in Mandarin Chinese is more accurate in native speakers of that language than in monolingual English speakers. More recently, Carcagno and Plack (2011) showed that F_0 discrimination training could enhance FFR strength in response to band-pass-filtered harmonic complexes; phase locking to the envelope was enhanced by training. Clinical applications of the FFR (or cABR) are suggested by findings by Kraus and colleagues that it can be degraded for children with language impairments (e.g., Russo et al. 2008).

Although the FFR clearly reflects temporal information that the auditory system could use to estimate pitch, an important question remains. Specifically, it is not known whether the FFR reflects neural processes that are involved in the extraction of pitch or whether it simply reflects the neural representation of sounds in the auditory periphery. This issue is important not only for theories of pitch perception but also for accounts of neural plasticity and of language impairments; if the FFR is enhanced by training and degraded by a clinical condition, does this reflect an influence on pitch processing or on more general (and possibly peripheral) temporal representations of sound? Despite the importance of this distinction, it is rarely addressed, with most studies referring to effects on pitch "encoding" at the brainstem level, although sometimes stronger statements are made to the effect that the brainstem "extracts" pitch (Russo et al. 2008) or that the FFR reflects "voice pitch processing" (Krishnan et al. 2005). On the other hand, a recent study of the effects of harmonic number on the periodicity strength of the FFR (quantified as the height of the first prominent peak in the autocorrelation function) in response to complex tones (Krishnan and Plack 2011) showed that although the dependence of FFR periodicity strength on harmonic number and component phase was similar to that reported for perceptual pitch strength, a similar pattern of results was observed using a model of the auditory nerve (AN) response to these sounds. The authors interpreted their results as evidence that the FFR "preserved" sensory-level pitch information.

Here, we addressed this issue using two approaches. In the first, we measured the FFR to frequency-shifted harmonic complexes and compared the results both to the behavioral pitch-matching responses reported in the literature and to the output of a model of the AN response to the same stimuli. We

conclude that temporal information that could account for the reported perceptual pitch shifts is indeed preserved in the FFR, but that a similar representation is likely to be present at the level of the AN. Second, we measured the FFR to three-component complexes consisting of the second, third, and fourth harmonics of a missing F_0. When all components were presented to the same ear, the magnitude spectrum of the FFR showed a peak corresponding to the "missing" F_0. However, when the second and fourth harmonics were presented to one ear and the third harmonic was presented to the other, the same pitch was perceived but no such peak was observed. A similar distinction was found in the ACFs of the FFR in these two conditions; the ACF of the FFR obtained for dichotic presentation of the harmonics was dissimilar to that obtained for monaural presentation, and it did not reflect the pitch. Hence, our results revealed no evidence that the FFR reflects the extraction of a pitch from components presented to opposite ears. Overall, we conclude that the FFR reflects a low-pass-filtered "preservation" of neural responses occurring at earlier stages of processing, but does not reflect pitch processing per se.

EXPERIMENT 1: FREQUENCY-SHIFTED COMPLEX TONES

Rationale

Greenberg et al. (1987) pointed out that the relation between FFR and residue pitch would be cast into doubt if the FFR were simply synchronized to the envelope modulation pattern of tone complexes. Frequency-shifted complex tones, where all components have been shifted by a constant amount in Hertz up or down from their harmonic frequency value, provide an opportunity to test the relation between FFR and pitch. For a frequency-shifted complex, the envelope repetition rate is identical to that of the harmonic complex, but the pitch of the two complexes differs; the pitch change is proportional to the frequency shift and depends on the lowest harmonic present in the complex (de Boer 1956; Schouten et al. 1962; Patterson 1973; Patterson and Wightman 1976; Moore and Moore 2003). Human pitch matches to frequency-shifted complex tones are relatively well described by a slightly modified version of de Boer's (1956) rule, or what has been called by Schouten et al. (1962) the "first effect of pitch shift" (Patterson 1973; Wile and Balaban 2007):

$$\Delta p = \Delta f / n \qquad (1)$$

where Δp corresponds to the pitch shift, Δf corresponds to the frequency shift of each component, and

n corresponds to the second lowest harmonic number present in the complex. If the FFR for complex tones is determined only/mainly by the envelope of the complex, then the FFRs for a harmonic complex and for a frequency-shifted complex would be the same, and thus, FFR would not bear information related to pitch. In spite of the fact that frequency-shifted complex tones provide an opportunity to disentangle envelope-related and pitch-related aspects of the FFR, there seem to be only two published studies actually measuring the FFR for frequency-shifted complexes.

The first, by Greenberg et al. (1987), used three-component complex tones that were either harmonic or shifted up/down by 50% of the 244-Hz F_0. Stimuli were presented in alternating polarity, i.e., successive stimuli were presented alternately in the original waveform polarity and in the inverted waveform polarity. Analysis of time intervals between peaks in the FFR waveform showed a difference between interval values for harmonic and frequency-shifted stimuli; the values for the latter were shifted in the directions of periods corresponding to expected (ambiguous) pitch shifts for both subjects tested. Spectral analysis of the FFR made use of a technique for isolating activity phase-locked to particular stimulus frequencies, originally developed by Goblick and Pfeiffer (1969). First, the FFR waveforms were averaged separately across those trials where the stimulus was played in its original polarity and those where it was played in inverted polarity. Subtraction of the averaged inverted waveform from the averaged original waveform results in a "subtraction waveform" in which envelope-related components of the FFR are minimized (canceled) while the component of the FFR that is phase-locked to the signal frequencies (the temporal fine structure) is enhanced. Conversely, adding the two averaged FFR waveforms, measured for original and inverted polarity stimuli, results in an "addition waveform" in which contributions from neural activity that is related to the stimulus envelope (that is not inverted in phase when the stimulus waveform is inverted) are enhanced while contributions from neural activity phase-locked to the stimulus frequencies are minimized. The results of the spectral analysis showed that the addition waveform was dominated by a large peak at 244 Hz, as expected. Importantly, the subtraction waveform showed a spectral peak at 280 Hz (see Fig. 8 of Greenberg et al. 1987) in addition to peaks corresponding to component frequencies. This frequency (280 Hz) is close to one of the pitches often matched to that stimulus (de Boer 1956; Patterson 1973). The presence of a spectral peak in the FFR at a matched pitch is potentially very important as it would indicate pitch extraction at (or before) the site of FFR generation. Unfortunately, Greenberg et al. (1987) tested only two

subjects in that study and showed the spectral analysis of the FFR subtraction waveform for only one subject and for only one out of the three frequency-shifted complexes that were used.

The second study measuring the FFR for frequency-shifted complexes was conducted by Wile and Balaban (2007). They did not find a peak in the FFR subtraction waveform corresponding to a matched pitch. However, they used a 300-Hz-wide narrowband noise masker centered at F_0 (300 Hz) in order to mask distortion products (the objective of their study was different from the one here; see below). This noise, which had a spectrum level 10 dB below the level of individual primary tones, might also have masked a spectral component in the FFR corresponding to the pitch of the frequency-shifted complex (the frequency shifts corresponded to a maximum of 16.7% of the F_0). In contrast to Greenberg et al. (1987), Wile and Balaban (2007) did not analyze individual intervals in the FFR waveform, nor calculate the ACF for the FFR subtraction waveform.

The main objective of the first experiment was to test whether we could replicate the finding of Greenberg et al. (1987) of a spectral peak in the FFR subtraction waveform corresponding to a possible pitch match for the frequency-shifted complex tone. In order to maximize the chances of observing such a peak, we did not use a masking noise to mask the distortion product as this could also mask a (potential) spectral peak at a possible pitch match. The second objective was to analyze the temporal characteristics of the FFR (via the ACF), and to compare pitch predictions based on this with predictions based on the output of an AN model (Meddis and Hewitt 1991a), in order to assess potential additional processing at the level of the FFR.

Methods

The stimuli were nearly identical to the ones used by Greenberg et al. (1987). They were three-component complexes, all "derived" from a harmonic complex with an F_0 of 244 Hz.

Two of the complex tones were harmonic: the first contained harmonics 2–4 and the second contained harmonics 3–5. Three of the complex tones were frequency-shifted complexes; all components (either harmonics 2–4 or harmonics 3–5) were shifted by the same amount in Hertz away from their nominal (harmonic) frequency values. The amounts of frequency shift applied to each harmonic were 122 and 61 Hz in experiments 1A and 1B, respectively. All conditions were tested for all subjects. Stimulus duration was 100 ms, including 5-ms raised-cosine onset and offset ramps. The relative starting phases of the components were 0°, 120°, and 240° for the

bottom, middle, and top components, respectively. Note, however, that for harmonics with low harmonic number (rank) that are resolved by the peripheral auditory system, the starting phases do not affect the salience of pitch (Moore 2003; Moore and Gockel 2011) and do not correlate with the size of the FFR (Greenberg et al. 1987). Stimuli were presented in quiet, at a level of 70 dB SPL per component, with alternating polarity. The stimuli were generated with 16-bit resolution and a sampling rate of 40 kHz. They were played out through the digital-to-analog converter included in the evoked potentials acquisition system (Intelligent Hearing Systems–SmartEP) and presented binaurally through mu-metal-shielded Etymotic Research ER2 insert earphones, which have a flat frequency response at the human eardrum.

Five subjects (three females) participated. They ranged in age between 18 and 35 years. They all had normal hearing for both ears with pure tone absolute thresholds below 20 dB HL at octave frequencies from 250 to 4,000 Hz. Two of them had some musical training and the others did not. The five subjects were selected from a pool of ten subjects on the basis of initial FFR measurements for pure tones and other complex tones, where they were found to have robust FFR responses, i.e., clear peaks were observed in the magnitude spectrum of the FFR at stimulus frequencies for moderate sound levels. Informed consent was obtained from all subjects. This study was carried out in accordance with the UK regulations governing biomedical research and was approved by the Cambridge Psychology Research Ethics Committee.

Subjects reclined comfortably (in a reclining chair) in a double-walled electrically shielded sound-attenuating booth. They were instructed to relax and to refrain from moving as much as possible during sound presentation and recording. They were allowed to fall asleep. The FFR was recorded differentially between gold-plated scalp electrodes positioned at the midline of the forehead at the hairline (+, Fz) and the seventh cervical vertebra (−, C7). A third electrode placed on the mid-forehead (Fpz) served as the common ground. For this "vertical" electrode montage, the FFR is assumed to reflect sustained phase-locked neural activity from rostral generators in the brainstem (IC, LL; Marsh et al. 1975; Smith et al. 1975; Glaser et al. 1976; Galbraith 1994; Krishnan 2006). Electrode impedances were <1 kΩ for all recordings. The FFR signal was recorded with a sampling period of 0.075 ms, band-pass-filtered from 50 to 3,000 Hz, and amplified by a factor of 100,000. Epochs with voltage changes exceeding 31 μV were automatically discarded and the trial repeated. Stimulus polarity alternated for each presentation, and alternate polarity sweeps were recorded and averaged in separate data buffers

by the SmartEP system. The stimuli were played with a repetition rate of 3.57 per second, in blocks of 2,500 (valid) trials. Two blocks were run for each stimulus, in randomized order across stimuli. The overall duration of a session, including electrode placement and breaks, was about 3 h. Control recordings in which all of the same procedures were followed but with the tubes of the insert earphones blocked resulted in no signal energy above the noise floor at stimulus component, envelope, or distortion product frequencies in the subtraction waveform of the FFR.

Off-line processing was done using MATLAB (The Mathworks, Natick, MA). First, the averaged FFR response for original polarity and for inverted polarity stimuli were either added or subtracted and the result divided by 2, for each subject and condition. The resulting waveform was high-pass- and low-pass-filtered at 150 and 2,000 Hz (eighth-order digital Butterworth filter; 3-dB down cutoff frequencies), respectively. Further analysis was restricted to the time range from 12 to 100 ms after stimulus onset. For spectral analysis, the 88-ms waveform was zero-padded symmetrically to make up a 1-s signal, and the magnitude spectrum was calculated via a discrete Fourier transform. The magnitude spectrum is specified in decibels re 0.01 μV. The averaged magnitude spectrum across subjects was calculated for each condition by averaging across subjects' spectra. ACFs were calculated for each subject and condition for the 12–100 ms section of the FFR waveforms using the MATLAB function "xcorr," with normalization such that the maximum autocorrelation value obtained at lag zero equaled 1, and was then averaged across subjects. Averages across subjects' individual magnitude spectra and ACFs were calculated rather than averages across subjects' FFR waveforms to avoid jitter issues arising from possible differences in onset delay of the FFR between subjects.

Experiment 1A: frequency shifts of 50% of the F_0

Stimuli. The two harmonic complexes contained either harmonics 2–4 or harmonics 3–5 of a 244-Hz F_0. The three frequency-shifted tones contained (1) harmonics 2–4 (of a 244-Hz F_0) shifted down by 122 Hz, resulting in partials at 366, 610, and 854 Hz; (2) harmonics 2–4 shifted up by 122 Hz, resulting in partials at 610, 854, and 1,098 Hz; and (3) harmonics 3–5 shifted up by 122 Hz, resulting in partials at 854, 1,098, and 1,342 Hz.

Results and discussion. The latency of the unprocessed FFRs was about 9 ms, estimated visually as the time point relative to stimulus onset of the first occurrence of a major amplitude excursion followed by a regular pattern in the FFR traces. This is in good agreement with the range of latencies reported in the literature

for FFRs or cABRs (Glaser et al. 1976; Skoe and Kraus 2010) and is consistent with a generation site at the level of the IC or LL.

The averaged magnitude spectra of the FFRs for the five conditions are shown in Figure 1. The blue dashed line and the red solid line indicate the spectra for the addition and the subtraction waveform, respectively. The addition spectra show peaks at/close to the envelope rate (244 Hz) and its integer multiples for all conditions, irrespective of whether the complex is harmonic (panels A and B, top) or shifted in frequency (panels C–E). This would be expected if the FFR partly reflects neural phase locking related to the envelope of the stimulus as the period of the envelope of all tone complexes was constant. The finding also agrees
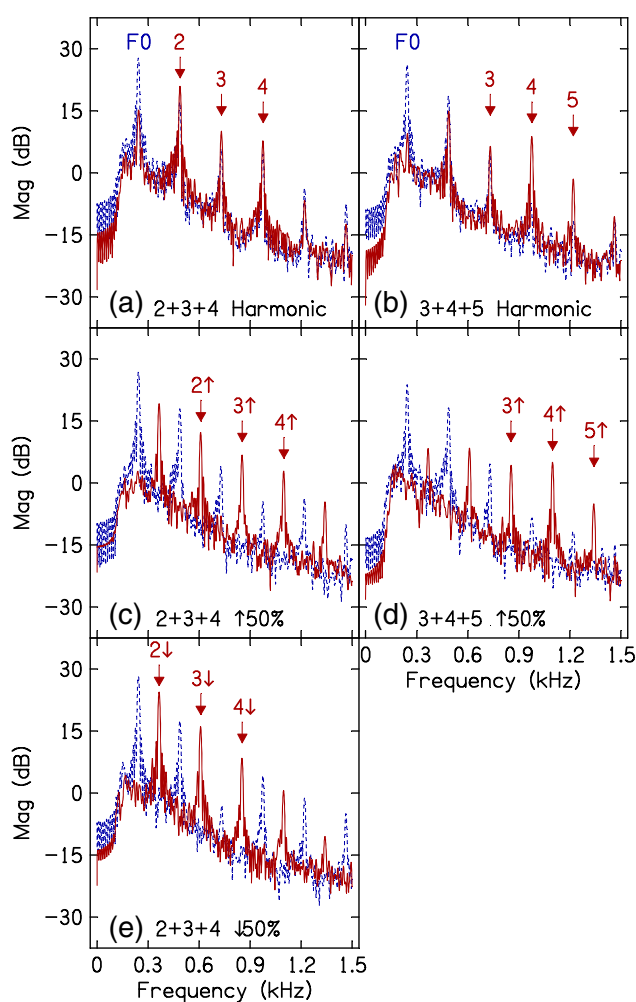

FIG. 1. Magnitude spectra of FFRs for all conditions of experiment 1A, averaged across five subjects, for FFR traces with the two polarities added (*blue dashed line*) or subtracted (*red solid line*). **A** Harmonic complex tone with F_0 of 244 Hz, containing harmonics 2+3+4. **B** As **A**, but containing harmonics 3+4+5. **C** Frequency-shifted complex tone where harmonics 2+3+4 of a 244-Hz F_0 have been shifted up by 122 Hz. **D** As **C**, but with harmonics 3+4+5 shifted up. **E** As **C**, but with harmonics shifted down by 122 Hz.

with that of Greenberg et al. (1987) for a frequency-shifted complex, and with many other studies using harmonic (non-shifted) tone complexes, where a major spectral peak at F_0 (corresponding to the pitch) was observed.

The major interest in this study was whether the spectra of the subtraction waveforms would show peaks at possible pitch matches for the frequency-shifted complexes. According to Eq. 1, pitch matches for our frequency-shifted complexes would be close to either 203 or 305 Hz for the condition plotted in Figure 1E as this stimulus can be regarded either as harmonics 2–4 shifted down by 122 Hz or as harmonics 1–3 shifted up by 122 Hz. Similarly, for the condition plotted in Figure 1C, pitch matches would be close to 214 or 285 Hz as this complex corresponds to harmonics 3–5 shifted down and to harmonics 2–4 shifted up by 122 Hz. Lastly, for the complex containing harmonics 3–5 shifted upwards (panel D), pitch matches close to 220 or 275 Hz would be expected. The spectra for the subtraction waveforms show peaks at/close to the individual frequencies of partials present in the stimulus (indicated by the red arrows) and to cubic distortion products ($2F_1-F_2$, $2F_3-F_2$, $2F_1-F_3$, $2F_3-F_1$). However, a spectral peak corresponding to a possible pitch match was not observed for any of the frequency-shifted tones. This was also true when subtraction waveforms from individual subjects were inspected. Our data show no evidence for a spectral peak in the subtraction waveform of the FFR corresponding to a possible pitch match for harmonics 3–5 shifted up by 122 Hz (panel D), which was the condition for which Greenberg et al. (1987) did report such a peak for the data of one subject. The reason for the discrepancy across studies is unclear. It may be that the subject tested by Greenberg et al. (1987) was unusual; however, it seems more likely that the spectral peak at 280 Hz observed by these authors was just a coincidence as the noise level of the FFR in this frequency region was generally high (see their Fig. 8).

It should be noted that while the absence of a peak at a given frequency in the magnitude spectrum of the FFR does imply that there is no substantial phase locking to that frequency, the opposite does not necessarily hold. For example, even if an assembly of neurons phase lock to a pure tone, one would expect peaks at higher harmonic frequencies in the magnitude spectrum of the FFR due to nonlinearities, such as half-wave rectification, in the response generation. Similarly, although the presence of a peak at $2F_1-F_2$ in the spectrum of the subtraction waveform *can* be the consequence of phase locking to that frequency due to the propagation of this distortion product to its characteristic frequency place on the basilar membrane, this is not necessarily the case. Instead, a single

population of neurons may respond to the combined stimulus, with any nonlinearities in the response generation leading to a peak at a distortion product frequency in the spectrum of the FFR.

Figure 2 shows the averaged ACFs of the FFRs for the five conditions. The ACFs of the addition waveforms (blue dashed line) are similar for all conditions, except for a slight reduction in peak height for complexes containing harmonics 3–5 (panels B and D) in comparison to those containing partials with lower frequencies. All conditions result in a maximum peak at/close to the lag that corresponds to the period of the envelope (4.1 ms). For the subtraction waveform (red solid line), the peak with maximum height (indicated by red arrows) is at/close to the lag corresponding to the period of a 244-Hz F_0 for the harmonic complexes, but is at twice this lag (corresponding to 122 Hz) for the frequency-shifted complexes. Thus, for the shifted complexes, the highest peak in the ACF of the subtraction waveform corre-
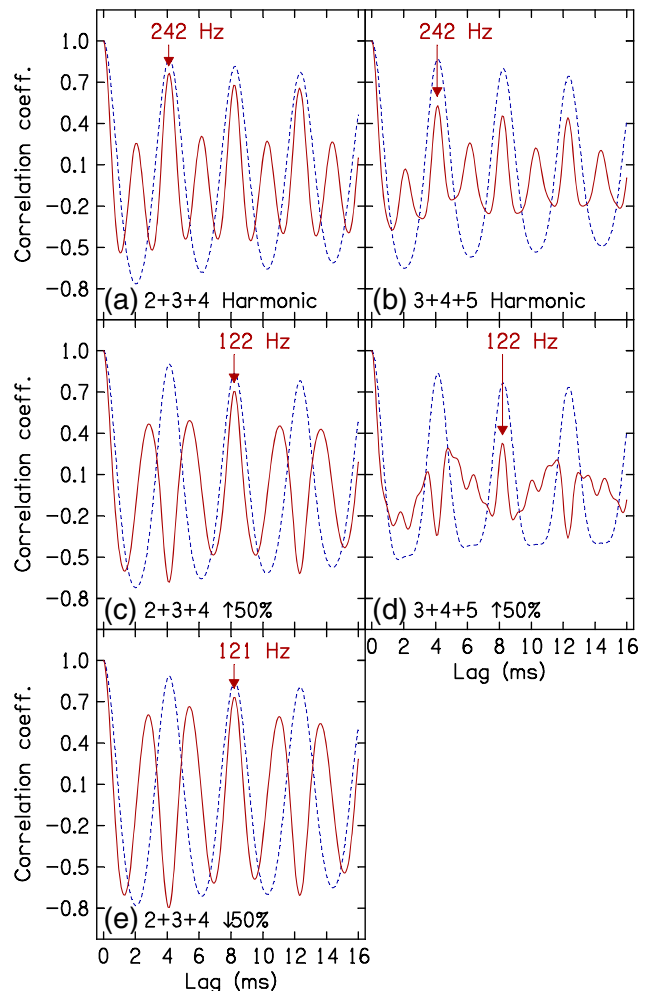

FIG. 2. ACFs of FFRs for all conditions of experiment 1A, averaged across five subjects, for FFR traces with the two polarities added (*blue dashed line*) or subtracted (*red solid line*). **A–E** as **A–E** in Figure 1.

sponds to the true F_0 rather than to a possible pitch match. It should be noted, however, that smaller peaks closer to possible pitch matches are visible near the "244-Hz peak" (in the addition waveform), especially for the upward-shifted harmonics in the 3–5 complex (panel D), indicating some but less strong periodicities in the FFR at these smaller lags (higher frequencies).

Because the true F_0 of 122 Hz of the frequency-shifted complexes is on average only about an octave below the expected pitch matches, it could be argued that (some) listeners might perceive a pitch of 122 Hz. That is, a frequency shift of 50% of the 244-Hz F_0 leads to another "plausible" pitch corresponding to the F_0 of a complex tone containing only odd harmonics of a 122-Hz F_0. For smaller frequency shifts, the true F_0 of the shifted complex is reduced and the partials correspond to higher harmonic numbers of the true F_0, making it much less likely that listeners perceive a pitch corresponding to the true F_0. Because of this, it was decided to repeat the experiment, but for smaller frequency shifts. Another reason to repeat the experiment, but with lower frequency values for the highest shifted harmonics, was that the ACF for the subtraction waveform of the upward-shifted harmonics in the 3–5 complex (panel D) did not show very strong periodicities, with a maximum correlation coefficient of about 0.33. This was most likely due to the low-pass characteristic of the FFR, generally observed in the FFR literature (Krishnan 2006).

Experiment 1B: frequency shifts of 25% of F_0

Methods. In experiment 1A, the frequency shift was 50% of the F_0. In experiment 1B, the frequency shift was reduced to 61 Hz (25% of the F_0). Besides the two harmonic complexes, the following three shifted tone complexes were employed: (1) harmonics 2–4 shifted down by 61 Hz with component frequencies of 427, 671, and 915 Hz; (2) harmonics 2–4 shifted up by 61 Hz with component frequencies of 549, 793, and 1,037 Hz; and (3) harmonics 3–5 shifted down by 61 Hz with component frequencies of 671, 915, and 1,159 Hz. Four of the five subjects who participated in experiment 1A participated in experiment 1B; the fifth subject was not available.

Results and discussion. Figure 3 shows the averaged magnitude spectra of the FFRs for the five conditions. As for experiment 1A, the addition spectra (in blue) show the largest peaks at/close to the envelope rate of 244 Hz and its integer multiples, irrespective of whether the complex is harmonic or frequency-shifted. Following Eq. 1, psychophysical pitch matches for the complexes shifted by 61 Hz are expected to be at about 224 Hz for harmonics 2–4
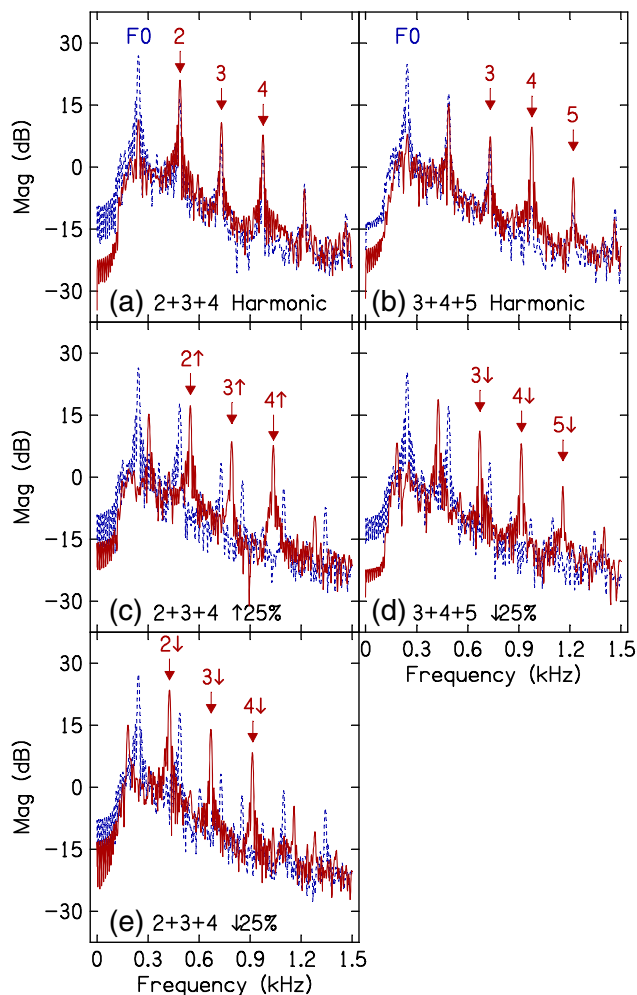

FIG. 3. Magnitude spectra of FFRs for all conditions of experiment 1B, averaged across subjects, for FFR traces with the two polarities added (*blue dashed line*) or subtracted (*red solid line*). **A** Harmonic complex tone with F_0 of 244 Hz, containing harmonics 2+3+4. **B** As **A**, but containing harmonics 3+4+5. **C** Frequency-shifted complex tone where harmonics 2+3+4 of a 244-Hz F_0 have been shifted up by 61 Hz. **D** As **C**, but with harmonics 3+4+5 shifted down by 61 Hz. **E** As **D**, but with harmonics 2+3+4 shifted down.

shifted down (panel E), 264 Hz for harmonics 2–4 shifted up (panel C), and 229 Hz for harmonics 3–5 shifted down (panel D). The peaks in the magnitude spectra of the subtraction waveform of the FFR (in red) are at/close to the frequencies of individual partials present in the stimulus (indicated by the red arrows) and to cubic distortion products ($2F_1-F_2$, $2F_3-F_2$, $2F_1-F_3$, $2F_3-F_1$). For the frequency-shifted stimuli, no spectral peaks corresponding to possible pitch matches were observed (either in the averaged spectra or spectra for individuals).

The averaged ACFs of the FFRs are shown in Figure 4. For the addition waveforms (blue dashed line), the ACFs are similar to those for experiment 1A, except for the frequency-shifted stimulus with harmonics 3–5 (compare Fig. 4D with Fig. 2D), for which the ACF has somewhat more distinct peaks, perhaps
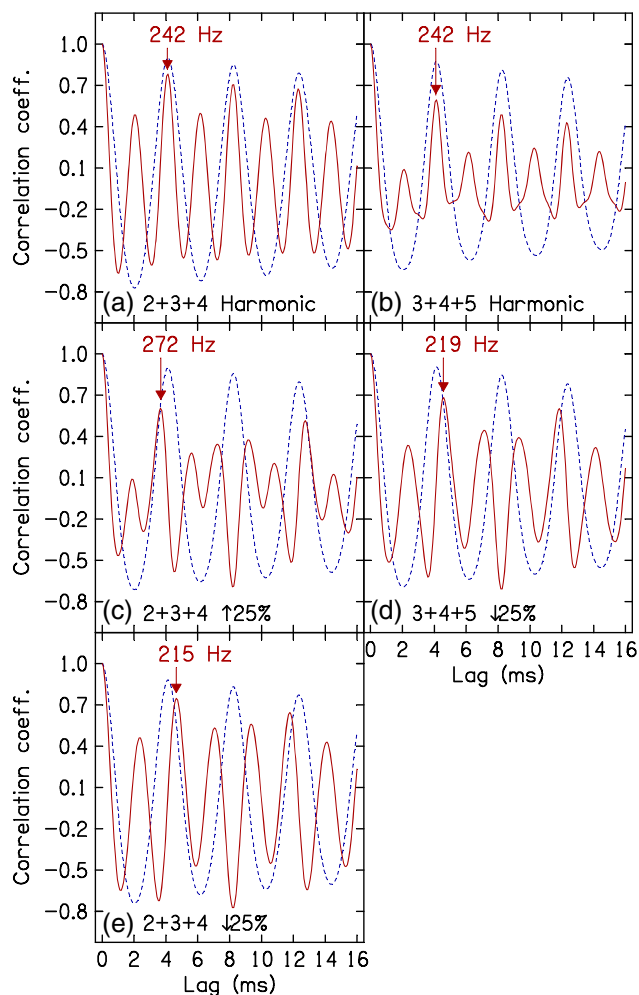
FIG. 4. ACFs of FFRs for all conditions of experiment 1B, averaged across subjects, for FFR traces with the two polarities added (*blue dashed line*) or subtracted (*red solid line*). **A–E** as **A–E** in Figure 3.

because the partials had lower frequencies than in experiment 1A. This tendency is even more obvious in the subtraction waveform (red solid line) for which the ACF has much more distinct and regular peaks than for experiment 1A, indicating stronger phase-locked periodicity in the FFR for the frequency-shifted complex containing harmonics 3–5. Importantly, for all conditions with frequency shifts, the subtraction waveforms now show peaks with maximum heights (indicated by red arrows) that are shifted in the direction of the expected pitch shifts. The differences between the expected pitch and the periodicity corresponding to the lag at the position of the maximum peak are 9, −8, and 10 Hz for harmonics 2–4 shifted down, harmonics 2–4 shifted up, and harmonics 3–5 shifted down, respectively. This indicates that the pitch shifts predicted on the basis of the ACF of the subtraction waveform of the FFR are somewhat larger than the expected perceptual pitch shifts.

Pitch predictions derived from an auditory nerve model

No evidence was found in experiments 1A and 1B for the existence of a spectral component in the subtraction waveform of the FFR at a frequency corresponding to a possible pitch match. However, the ACF of the subtraction waveform clearly indicated neural phase locking at the level of the brainstem with dominant periodicity in the vicinity of possible pitch matches, at least for the smaller frequency shifts. Does this simply reflect information being transmitted from the AN or does it reflect additional processing related to the extraction of pitch? In order to assess this question, pitch predictions based on the ACF of the subtraction waveform of the FFR were compared with those derived from a model of the auditory periphery. The Auditory Modeling System (AMS, version 1.3.0, available at http://dsam.org.uk/) developed by Meddis and O'Mard was used. The model implementation followed the example given for autocorrelation in the AMS tutorial (version 2.4). It is broadly similar to that described in Meddis and Hewitt (1991a, b) and in Meddis and O'Mard (1997), but has been updated to include a more recent nonlinear basilar membrane model and a more recent hair cell model. The inputs to the model were our experimental stimuli saved in wav files and scaled with a routine provided by AMS (the "Ana_Intensity" routine) such that the overall rms level corresponded to 75 dB SPL, the same as in the experiments. The model includes the following stages: (1) Simulation of the operation of the outer and middle ear, as described in Glasberg and Moore (2002); (2) A dual-resonance nonlinear filter bank comprising 60 channels evenly spaced according to Greenwood's function (Greenwood 1990) with center frequencies between 40 and 10,000 Hz, and filter parameters based on Lopez-Poveda and Meddis (2001), simulating the nonlinear response of the human basilar membrane; (3) Simulation of the mechanical to neural transduction at the hair cell using the parameters specified in Table II of Sumner et al. (2002) for the high spontaneous rate AN fiber—this gave the probability of occurrence of a spike in the auditory nerve fibers for each of the 60 channels as a function of time; (4) A running autocorrelation function with time constants according to Wiegrebe (2001), calculated separately within each channel, which provides an estimate of the distribution of intervals between all spikes originating from fibers within a given channel, measured 82 ms (20 cycles of the 244-Hz F_0) after stimulus onset; and (5) A summary ACF (SACF), derived by summing the ACFs across all channels.

As for the pitch predictions based on the ACF of the subtraction trace of the FFR, the predicted pitch was chosen to correspond to $1/\tau$, where τ (with $\tau > 0$)

corresponds to the time lag at which the largest positive correlation was observed. These pitch predictions are shown in Figure 5 (red downward-pointing triangles), together with the psychophysically established pitch estimates according to Eq. 1 (black solid lines) and the predictions based on the ACF of the subtraction waveform of the FFR (yellow circles) as indicated by the red arrows in Figures 2 and 4, discussed above. The pitch predictions based on the SACF of the AN model are generally similar to those based on the ACF of the FFR. The implication of this is that, for the stimuli used here, the periodicity of the temporal information present at the level of the AN seems to be roughly preserved at the level of the brainstem, or IC, the presumed generating site of the FFR. It also means that the FFR reflects as much or as little pitch processing as is present in the AN. This general conclusion is also supported by Cariani and Delgutte's (1996b) report that, in cats, pitches estimated from the pooled interval distributions of the temporal discharge patterns of AN fibers showed shifts corresponding to "the first effect of pitch shift" measured psychophysically in humans; the all-order interspike interval histogram is equivalent to an ACF (Cariani and Delgutte 1996a).
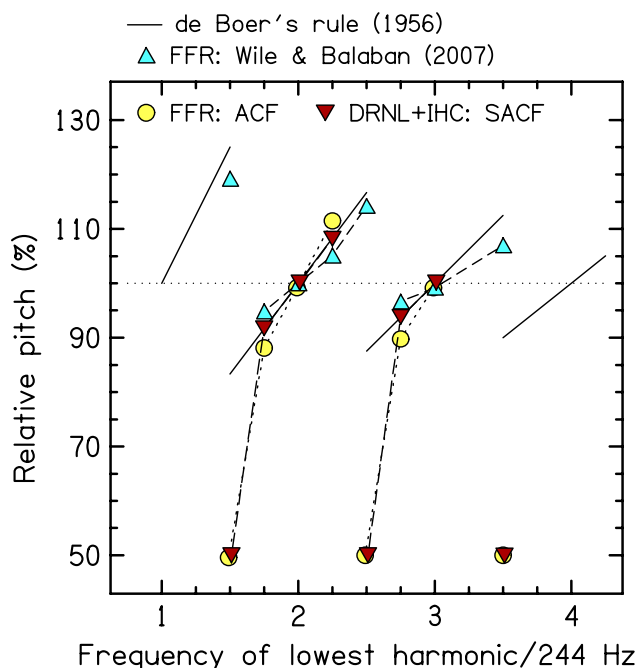

FIG. 5. Pitch predictions for the frequency-shifted complexes used in experiments 1A and 1B based on: (1) psychophysical pitch-matching experiments described by de Boer's rule (*solid line*, see Eq. 1); (2) FFRs measured in experiments 1A and 1B and calculated following Wile and Balaban's (2007) model (*turquoise upward-pointing triangles*); (3) the highest peak in the ACF of the subtraction waveform of the measured FFR (*yellow circles*); and (4) the highest peak in the SACF of the output of an auditory nerve model (Meddis and Hewitt 1991a; Meddis and O'Mard 1997; *red downward-pointing triangles*). See text for details.

Pitch predictions based on the SACF of the AN activity, shown here, deviate markedly from behavioral pitch matches when the frequency shift is 50% of the F_0. In contrast, Meddis and Hewitt (1991a) argued that there was good agreement between the (original) model predictions and psychophysical data even for frequency shifts of 50% of the F_0. While it cannot be excluded that the discrepancy arises from our use of different model parameters (e.g., DNRL vs. gamma-tone filter bank), two points should be noted. Firstly, the SACF of the simulated AN activity did show major peaks close to 244 Hz, but they were not the largest ones. If only peaks close to 244 Hz had been picked as the pitch estimate, predictions would have markedly improved. Secondly, Figure 9 of Meddis and Hewitt shows the SACF for a 100-Hz F_0 complex that has been frequency-shifted by various amounts, but the maximum period plotted is 16 ms (62 Hz). Thus, it is not obvious whether a larger peak was present in the SACF at a period of 20 ms (50 Hz), which is the true F_0 of the harmonic complex shifted by 50% of the F_0.

In summary, pitch predictions for frequency-shifted tones that were based on the ACF of the subtraction waveform of the FFR were similar to those based on the summary ACF of the output of an AN model. This indicates that temporal information, present in the AN and relevant for pitch, is preserved at the level of the FFR and could be used at a more central stage to extract pitch via a mechanism like the ACF.

Pitch predictions derived from FFR spectra using a distortion product

Wile and Balaban (2007) measured the FFR for a 300-Hz F_0 harmonic complex containing harmonics 2–4 and for the same complex with the three partials each shifted in frequency by ±25 or ±50 Hz. A 300-Hz-wide band-pass noise centered at 300 Hz was presented continuously, with a spectrum level 10 dB below the level of the individual primaries, in order to mask that component of distortion products that propagated on the basilar membrane to its characteristic frequency place. The idea was that any remaining periodicity should have its origin in activity at or close to the place of the primaries on the basilar membrane. Wile and Balaban (2007) proposed a "place-gated combination of neural timing information from the envelope and fine structure of a sound relayed via the midbrain and brainstem." They suggested a model that derives the predicted pitch as the weighted average of two peaks in the FFR spectrum. The first one is the peak at the envelope repetition rate in the spectrum of the addition waveform of the FFR. The second one is the peak at the distortion product frequency closest to the envelope rate, i.e., at $2F_1 - F_2$, in the spectrum of the subtraction waveform of the

FFR. They argued that their use of a low-pass masking noise would have eliminated distortion products that propagated along the basilar membrane and that remaining distortion products therefore arose from a place that combined responses from two or more primaries. These portions of the basilar membrane are close to those that respond to the envelope, and the $2F_1-F_2$ distortion product and envelope-related peaks were weighted with their relative amplitudes. The resulting pitch predictions corresponded well with the pitch matches of their subjects, who in turn "performed similarly to those described in previous experiments, with their pitch percepts conforming to de Boer's rule (de Boer 1956; Patterson 1973)."

Pitch predictions for the present stimuli, derived from our FFR measurements in the absence of masking noise and otherwise based on Wile and Balaban's (2007) model, are shown in Figure 5 (turquoise upward-pointing triangles). They agree reasonably well with the psychophysical pitch estimates, even though no masking noise was used in the present experiments. This could be because, even without noise, the propagated component of the distortion products contributed little relative to that originating at the place of the primaries. Another possibility is that Wile and Balaban's (2007) noise not only (partly) masked the propagated distortion component at $2F_1-F_2$ in the subtraction waveform of the FFR, but also reduced the peak height at the envelope period in the addition waveform. The latter could be due to (partial) masking of the propagated component of a quadratic distortion product at F_2-F_1 (if the phase of this distortion product is not inverted with inversion of the stimulus polarity) and/or to masking/suppression of the lowest primary (or primaries) itself. The current data do not make it possible to distinguish between these possibilities.

In summary, Wile and Balaban's (2007) model intends to combine temporal phase locking information related to the envelope with temporal phase locking information related to the temporal fine structure. Application of this model to the present FFR data gives reasonably good pitch predictions for the frequency-shifted tones. The success of this model depends on the presence of spectral peaks in the general vicinity of the pitch percept in the magnitude spectrum of the FFR. In the next experiment, we employed stimuli for which this prerequisite is not expected to hold.

EXPERIMENT 2: COMPLEX TONE WITH HARMONICS PRESENTED DICHOTICALLY

Rationale

Listeners can combine harmonics presented to opposite ears to derive a residue pitch (Houtsma and Goldstein 1972; Bernstein and Oxenham 2003). Thus, pitch perception must involve neural processes occurring at stages of the auditory system where information from the two ears has been combined. More recently, Gockel et al. (2011) demonstrated that a residue pitch can be derived by combining a Huggins pitch (Cramer and Huggins 1958) component (derived solely from binaural interaction) with a spectral component for which no binaural processing is required. Gockel et al. (2011) argued that this suggests the existence of a single central pitch mechanism for the derivation of residue pitch from binaurally created components and from spectral components, operating at the earliest at the level of the dorsal LL or IC, which receive inputs from the medial superior olive, the level at which temporal information from the two ears is first combined. As the suggested generation site of the FFR is at the level of the IC or LL, the FFR could in theory reflect binaural pitches, such as Huggins pitch, and also the residue pitch of complex tones with harmonics presented to opposite ears *if* these pitches are based on temporal pattern of activity. The latter qualification is of course necessary since the FFR reflects sustained phase-locked activity in an assembly of neurons within the rostral brainstem, and if pitch-relevant information had already been extracted into some other format, e.g., a rate place code, this would not be reflected in the FFR. The Huggins pitch is rather weak in a noise background and thus might be difficult to observe in the FFR (see Plack et al. 2010). In contrast, the pitch of dichotic tone complexes—with harmonics presented to opposite ears—can be more robust and the stimulation does not include a simultaneous wideband noise, and thus should be reflected in the FFR *if* the FFR does indeed reflect processing at the level of pitch extraction and if, at this stage, residue pitch information is (still) based on temporal pattern of activity. Experiment 2 tested whether the FFR does reflect the perceived residue pitch of complex tones when harmonics are presented dichotically.

Methods

Stimuli were 450-ms three-component harmonic complexes consisting of the second, third, and fourth harmonics of a 244-Hz F_0. In all conditions, all three harmonics were ramped on together (10-ms raised-cosine function) in both ears. Over the next 40 ms, some components were turned off gradually (40-ms raised-cosine function) in one or the other ear so that in the last 400 ms, the harmonics presented were: (1) 2+3+4 to the left ear (condition "mono"); (2) 2+4 to the left and 3 to the right (condition "dichotic"); (3) 2+4 to the left (condition "2+4"); and (4) 3 to the

right (condition "3"). The common onset of all harmonics and gradual fade-out of some of the harmonics was introduced to increase the perceptual fusion of components across ears. Stimuli in conditions "mono" and "dichotic" had the same pitch, and their pitch was not perceived to change over the duration of the stimulus, as established in an informal listening experiment. All harmonics present at the end of the stimulus were ramped off together (10-ms raised-cosine function). The relative starting phases of components were 0°, 120°, and 240° for the bottom, middle, and top components, respectively. Stimuli were presented in quiet, at a level of 70 dB SPL per component, with alternating polarity. They were played with a repetition rate of 1.1 per second, in blocks of 2,400 (valid) trials. One block was run for each condition. Condition dichotic was always tested first, followed by condition mono. Conditions "2+4" and "3" were run in a counterbalanced order across subjects. As effects of attention on the FFR have been reported by some authors (Galbraith and Arroyo 1993; Galbraith and Doan 1995; Galbraith et al. 1998), the dichotic condition was always run first in order to minimize contextual effects; for example, if subjects had listened first to the "2+4" condition, then this may have encouraged them to "hear out" the second and fourth harmonics when they were subsequently presented to the same ear in the dichotic condition. The overall duration of a session, including electrode placement, was about 3 h. The same five subjects as in experiment 1A participated here.

The FFR signal was recorded with a sample rate of 8 kHz over 512 ms (50 ms before stimulus onset as a baseline measure and 462 ms after stimulus onset) for each presentation. Off-line processing was the same as for experiments 1A and 1B, except that further analysis of the FFR was restricted to the time range from 100 to 350 ms after stimulus onset. Thus, all results are for a time range starting 50 ms after the fade-out of some of the components has occurred to allow for the stabilization of brain responses to the part of the stimulus that is specific for that condition. All other methods are identical to those used in previous experiments.

Results and discussion

Figure 6 shows the averaged magnitude spectra of the FFRs for the four conditions; the spectrum of the addition waveform (in blue) and the spectrum of the subtraction waveform (in red) have been shifted very slightly relative to each other on the x-axis to prevent peaks of one being hidden behind the peaks of the other. In condition "mono" (panel A), the magnitude spectrum of the addition waveform (blue) shows the
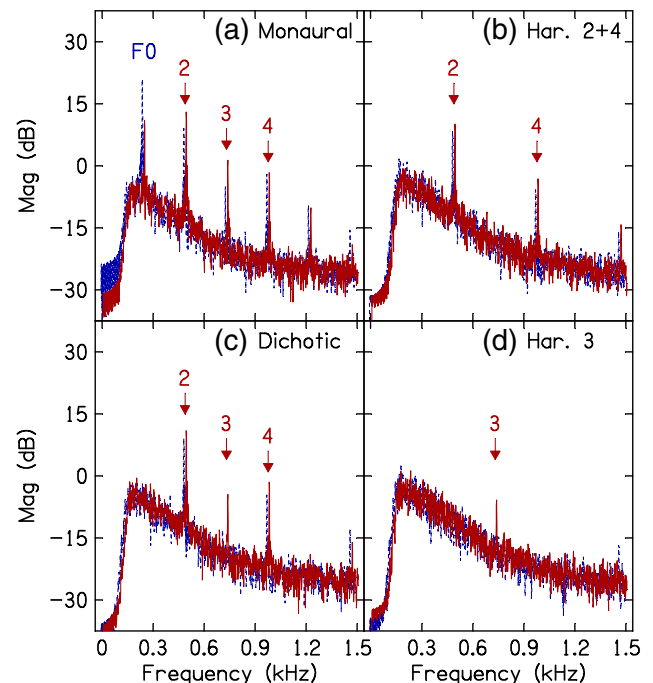

FIG. 6. Magnitude spectra of FFRs for all conditions of experiment 2, averaged across subjects, for FFR traces with the two polarities added (*blue dashed line*) or subtracted (*red solid line*). The spectrum of the addition waveform and the spectrum of the subtraction waveform have been shifted very slightly relative to each other on the x-axis to prevent peaks of one being hidden behind the peaks of the other. **A** Harmonic complex tone with F_0 of 244 Hz, containing harmonics 2+3+4, presented monaurally to the left ear. **B** As **A**, but containing harmonics 2+4. **C** As **A**, but with harmonics 2+4 presented to the left ear and harmonic 3 presented to the right ear. **D** As **A**, but for harmonic 3 only, presented to the right ear.

largest peak at 244 Hz (the envelope rate) and decreasing amplitude peaks at all its integer multiples. The magnitude spectrum of the subtraction waveform (red) shows clear peaks at frequencies corresponding to the three primaries (488, 732, and 976 Hz) and to the cubic distortion products ($2F_1 - F_2$ at 244 Hz and $2F_3 - F_2$ at 1,220 Hz). For condition "2+4" (panel B), a similar pattern is observed, but here, the F_0 of the stimulus is 488 Hz, and the first and second harmonics of that F_0 are actually present. For this condition, in the spectrum of the addition waveform, there are peaks at 488 Hz (the envelope rate) and at its integer multiples. The spectrum of the subtraction waveform (red) has peaks at the frequencies of the two primaries (488 and 976 Hz) and the cubic distortion product at $2F_2 - F_1$ (1,464 Hz). For condition "3" (panel D), the spectrum of the subtraction waveform shows a peak at 732 Hz (corresponding to the frequency of that harmonic), and the spectrum of the addition waveform shows a small peak at 1,464 Hz. This results from the addition of two opposite polarity waveforms each of which has undergone a nonlinear transformation. For these

three conditions, the spectral pattern of the FFR is very much as expected.

For the condition of primary interest, condition "dichotic" (panel C), the spectra correspond to the sum of the spectra observed for conditions "2+4" and "3"; the spectrum of the subtraction waveform of the FFR shows peaks corresponding to all three primaries and a peak at 1,464 Hz, the cubic distortion product peak visible for condition "2+4," while the spectrum of the addition waveform shows peaks at 488 Hz and its integer multiples. There are two points to be stressed here. Firstly, all peaks observed for either of the two monaural control conditions appear in the dichotic condition, and they do so with roughly the same amplitude. This indicates that input from both ears is represented in the FFR for the dichotic condition. Secondly, for the dichotic condition, peaks are only observed at frequencies where there also was a peak for one of the two monaural control conditions. This indicates that neural activity underlying the FFR is summed almost linearly across the two ears, perhaps because two different sets of neurons are activated by each of the two ears at the site of FFR generation. Importantly, for the dichotic condition, there is no spectral component in the FFR at a frequency corresponding to the perceived pitch of the complex, either in the addition waveform or in the subtraction waveform. Thus, as for the frequency-shifted tones, pitch is not represented in the FFR with a spectral peak. The absence of any spectral peak in the vicinity of 244 Hz also means that Wile and Balaban's (2007) model is unable to account for the pitch of a dichotically presented harmonic complex tone.

Figure 7 shows the ACFs of the FFRs averaged across subjects for the four conditions. For condition "mono" (panel A), the ACF of the addition waveform (blue) shows a very large peak (r=0.82) at/close to the delay corresponding to the envelope rate and at all integer multiples of this delay (indicated by the black arrows in this and all other panels). No other peaks are observed, indicating a clear periodicity of 244 Hz in the FFR waveform when processed to enhance information related to the envelope and to suppress information related to the temporal fine structure. The ACF of the subtraction waveform (red) shows major peaks at the same delays as the addition waveform and additional regular smaller peaks in between. These much smaller peaks are at delays equal to odd integer multiples of the delay corresponding to a period of 488 Hz and are small but positive, except the one around 488 Hz which is below zero. They might be the result of the relatively large 488-Hz component in the magnitude spectrum of the subtraction waveform of the FFR. Overall, the subtraction waveform, like the addition waveform, shows
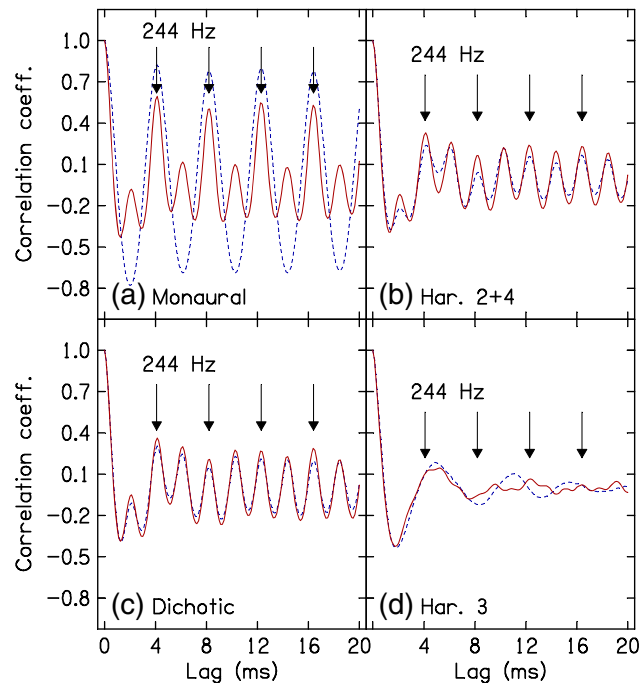

FIG. 7. ACFs of FFRs for all conditions of experiment 2, averaged across subjects, for FFR traces with the two polarities added (*blue dashed line*) or subtracted (*red solid line*). **A–D** as **A–D** in Figure 6.

the major periodicity at 244 Hz, corresponding to the pitch of the complex. Figure 7B shows the ACFs for condition "2+4." Peak heights are generally reduced relative to those observed for condition "mono," indicating less clear periodicity in the FFR, either because the precision of phase locking at the level of the FFR is reduced or because smaller numbers of neurons are involved in producing the FFR response. ACFs of the subtraction and addition waveforms look more similar than they do for condition "mono." As for condition "mono," there are peaks at the delay corresponding to 244 Hz and its integer multiples, but now there are equally large peaks at delays equal to odd integer multiples of the delay corresponding to 488 Hz (except the one around 488 Hz which is again below zero). Thus, peaks at all (odd and even) integer multiples of the delay corresponding to 488 Hz are present with about equal height, indicating the major periodicity at 488 Hz, corresponding to the pitch of this tone complex. For condition "3" (Fig. 7D), no clear periodicity is seen. The small peaks that are present do not follow a regular pattern, and there is no peak corresponding to 732 Hz, the frequency of the pure tone. This is a consequence of the low-pass behavior of the FFR; the spectral peak in the subtraction waveform visible at 732 Hz (see Fig. 6D) is very small relative to the noise floor at lower frequencies. The low-pass behavior of the FFR may be due to a number of factors, including a decrease in phase-locking precision with increasing frequency

and a decrease in the number of neurons phase locking with increasing frequency. For the condition of primary interest, condition "dichotic" (Fig. 7C), the ACFs overall look very similar to those for condition "2+4," except that the peak heights are slightly larger, probably because a somewhat larger ensemble of neurons is contributing to the phase-locked response.

To quantify the comparison of the ACF patterns obtained for condition "dichotic" with those obtained for condition "mono" on the one hand, and those obtained for condition "2+4" on the other hand, we derived a simple measure of the strength of the 244-Hz periodicity within the overall pattern of the ACF. This measure was the average height (average of the r values) of all peaks at the delay corresponding to 244 Hz and integer multiples of this delay, minus the average height of peaks at other delays. In the special case where there was no peak at any other delay (in the ACF of the addition waveform for condition "mono"), minus 1 was subtracted. This was done in order to derive a consistent measure; the value obtained when there is no peak at any other delay should be larger than the value obtained when there are peaks with negative r values at other delays. Note, however, that this is not crucial for the general pattern of the results. The values of this measure are plotted in Figure 8, for the individual subjects and averaged across subjects. For all subjects, the strength of the 244-Hz periodicity for the dichotic condition is weak (the value is small) and is more similar to that for conditions "2+4" and "3" than to that for the monaural condition. For the latter, there was a strong 244-Hz periodicity. This is true for both the addition (Fig. 8A) and the subtraction waveform (Fig. 8B). The results of a sign test show that the strength of the 244-Hz periodicity is significantly larger ($p<0.05$) for condition "mono" than for condition "dichotic" and is not significantly different between conditions "dichotic" and "3" for both the addition and the subtraction waveforms.

Two main factors may contribute to this clear pattern of results. The first is the extreme low-pass characteristic of the FFR: the ACF of a physical complex tone containing harmonics 2–4 at equal levels shows regularly spaced high peaks at the delay corresponding to the F_0 and all integer multiples of it, with only small peaks in between. However, the ACF of a complex tone containing the same three harmonics but with relative levels roughly corresponding to the ones seen in the magnitude spectrum of the subtraction waveform of the FFR (in condition monaural, or condition dichotic, or conditions "2+4" and "3") shows peaks of nearly equal height at the delay corresponding to the F_0 (and its integer multiples) and the delay corresponding to twice the F_0
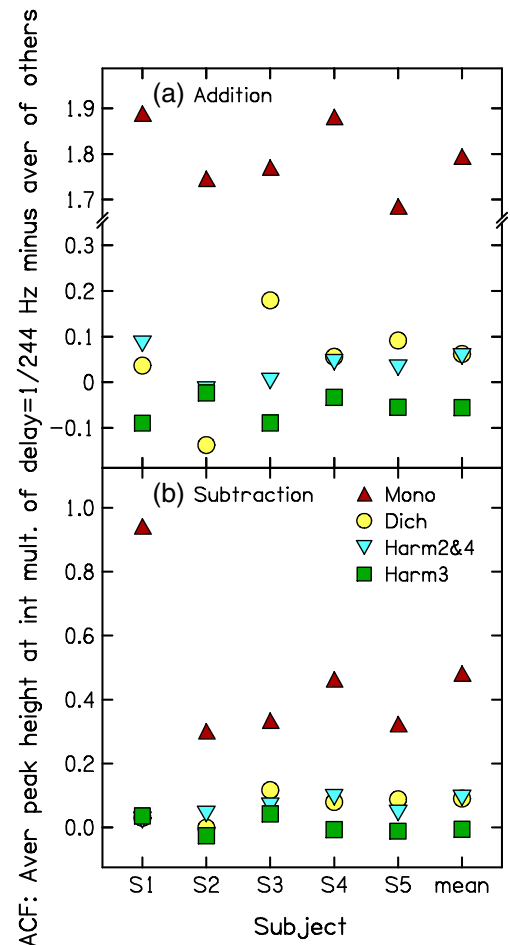

FIG. 8. Measure of the strength of a 244-Hz periodicity within the overall ACF patterns for the FFRs observed in the four conditions of experiment 2, for individual listeners and averaged across listeners. **A** FFR traces with the two polarities were added. **B** FFR traces with the two polarities were subtracted.

(and its integer multiples). This, in combination with the high noise floor of the FFR at low frequencies, means that the difference between the ACF patterns of the FFR of condition "2+4" on the one hand and the pooled ACF of conditions "2+4" and "3" on the other hand is expected to be very small. The second factor contributing to the overall pattern of results is the occurrence of distortion products in the FFR in the monaural condition. Specifically, the presence of the relatively high-level cubic distortion product at 244 Hz, when added to harmonics 2–4 with relative levels as indicated in the magnitude spectrum of the subtraction waveform, results in an ACF distinctly different from the other ACFs, with regularly spaced high peaks at the delay corresponding to 244 Hz and all integer multiples of this delay, with only small peaks in between (as in Fig. 7A). Thus, it seems that for the FFR—as recorded from the scalp—to be a useful input for a residue pitch extraction scheme that is based on temporal information rather than

place information, the presence of lower frequency distortion products is crucial when the fundamental is absent from the stimulus.

Let us consider briefly what it would mean (or not mean) if, in the dichotic condition, we had observed the major peak of the ACF of the subtraction waveform at 1/244 Hz and its integer multiples. If the FFR faithfully represented the stimulus, then the ACFs of the subtraction waveform should, in condition "mono," show peaks at multiples of $1/F_0$. This, of itself, would not necessarily imply any pitch extraction by the auditory system; because the ACF is a mathematical operation that extracts periodicity, a peak at F_0 is also present in the ACF of the raw stimulus. Similarly, if two separate populations of neurons responded faithfully to the input to each ear in the dichotic condition, and if the FFR represented the sum of these responses, then a periodicity at $1/F_0$ would necessarily appear for this condition too. Hence, one could not tell from the ACF whether or not any additional processing, arising from the combination of tones across ears to derive the "missing fundamental" pitch, had occurred.

In summary, the results of experiment 2 showed that the FFR of a harmonic complex tone with dichotic presentation of its harmonics reflects the input from both ears, but the neural activity underlying the FFR seems to be added independently across ears. There is no spectral component in the FFR corresponding to the perceived pitch, and there is no intermodulation between components presented to opposite ears. As a consequence, neither Wile and Balaban's (2007) model nor the ACF of the FFR is able to account for the residue pitch of the dichotically presented complex.

GENERAL DISCUSSION

Summary of results

1. The FFR of frequency-shifted complex tones did not show a spectral peak at the frequency corresponding to behavioral pitch matches, in contrast to the finding of Greenberg et al. (1987).
2. The FFR of frequency-shifted complex tones showed spectral peaks related to the envelope repetition rate in the addition waveform and spectral peaks related to the primary components and cubic distortion products in the subtraction waveform.
3. The ACF of the subtraction waveform of the FFR of frequency-shifted complexes indicates periodicity corresponding to the shifted pitch previously observed in psychophysical experiments.
4. The summary ACF of neural phase locking derived from an auditory nerve model applied to frequency-shifted complex tones also indi-

cates periodicity corresponding to the shifted pitch, indicating that low-pass-filtered peripheral temporal information was preserved at the level of the FFR.
5. The FFR of a three-component harmonic complex with harmonics 2 and 4 presented to one ear and harmonic 3 presented to the opposite ear does not reflect its pitch. While the magnitude spectrum of the FFR indicates the presence of information from both ears, there is no spectral peak at the F_0, and the ACF of the FFR, which is commonly used as an indicator for neural encoding of pitch-related information, does not reflect the residue pitch of the dichotically presented harmonic complex.

Comparison with animal experiments

Recently, Shackleton et al. (2009) recorded the responses of multi-unit clusters in the central nucleus of the IC of guinea pigs (GP) to complex tones consisting of a large number of harmonics. One of the main aims of their study was to search for evidence of binaural integration of dichotically presented components. Therefore, they not only included conditions where all harmonics were presented diotically or monaurally but also conditions where the even-numbered harmonics were presented contralaterally to the ear receiving the odd-numbered harmonics, which were presented either to the right or to the left ear. They reasoned that if the harmonics are resolved by the peripheral auditory system, then, if binaural integration of pitch occurs at or before the IC, they should observe neural responses phase-locked to the F_0 of the complex in the dichotic conditions. Instead, responses in these conditions were predominantly phase-locked to $2F_0$ and were consistent with the neural clusters being mainly driven by the input to the contralateral ear. However, Shackleton et al. (2009) were unable to draw any firm conclusions because a combination of factors limited the number of responses to resolved harmonics that they could obtain. AN fibers of GPs are quite broadly tuned, so that, in order to record from an IC cluster that was driven by resolved harmonics, that cluster of neurons would have to have a low CF and the complex would need a high F_0. Shackleton et al. (2009) recorded only from a limited number of low-CF clusters, and unfortunately, phase locking was weak to the highest F_0 (400 Hz) that they studied. As they pointed out, when harmonics are unresolved, a dichotic complex will produce a response in the AN fibers of each ear that reflects the envelope of the stimulus in that ear, and this will correspond to $2F_0$. Under such circumstances, one would not expect the IC to phase lock to

F_0. They concluded that the responses that they measured primarily reflected an envelope response to the contralateral stimulus.

Experiment 2 addressed a similar question to that posed by Shackleton et al. (2009), but used exclusively low-numbered harmonics that are known to be resolved by the human peripheral auditory system. The results indicated that neural activity underlying the FFR is summed (almost) linearly across the two ears. This could be because two different sets of neurons are activated by each of the two ears at the site of FFR generation, and it is consistent with neurons in the upper brainstem responding predominantly to the contralateral ear, as reported by Shackleton et al. (2009). The results are not consistent with the output of a pitch extraction mechanism whereby neurons phase lock at a rate corresponding to the perceived pitch of dichotically presented harmonics.

Theoretical and clinical implications

The results of the present study suggest that the FFR preserves a low-pass-filtered version of monaural temporal information that is conveyed from the AN to the upper brainstem rather than reflecting a representation of pitch *per se*; the extraction of periodicity via the ACF is not inherent in the FFR, but is an additional process that may, or may not, take place at a later stage. However, for dichotic complexes as used here, the scalp-recorded FFR does not carry temporal information necessary to extract residue pitch, at least not via a mechanism like the ACF or via a mechanism like the one suggested by Wile and Balaban (2007). This does not, of course, mean that the FFR is of no value, not least because there is no other method of recording sustained phase-locked activity from the human auditory brainstem. For monaurally or diotically presented complex tones, the scalp-recorded FFR carries temporal information sufficient for the extraction of residue pitch. However, it is not necessarily the case that pitch is derived from this temporal pattern of activity, and it is possible that, at the level of the IC, pitch-relevant information may have, for example, already been extracted into some other format (e.g., a place–rate code; see Shackleton et al. 2009) that would not be reflected in the FFR.

If, as we suggest, the FFR reflects information that might be used to extract pitch, rather than the result of pitch extraction, this raises interesting questions for the interpretation of both training studies and for studies of clinical populations. In the former case, the question arises as to what is being modified by experience and/or training. If the FFR does not reflect pitch extraction, then neither can its modification by training or experience. Instead, it

may occur at an earlier stage, either in terms of the response to the stimulus envelope or to the individual frequency components. For studies that compare FFR to particular sounds between "experts" (e.g., speakers of tonal languages, musicians) and non-specialists, there is the additional possibility that the differences may be genetic (leading some people to study music, or for languages to evolve to meet the auditory abilities of its speakers). Studies of impaired populations (e.g., Russo et al. 2008) and of modification of the FFR by training may reflect differences at peripheral sites.

Open Access

REFERENCES

BENDOR D, WANG X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161–1165

BERNSTEIN JG, OXENHAM AJ (2003) Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? J Acoust Soc Am 113:3323–3334

CARCAGNO S, PLACK CJ (2011) Subcortical plasticity following perceptual learning in a pitch discrimination task. J Assoc Res Otolaryngol 12:89–100

CARIANI PA, DELGUTTE B (1996a) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol 76:1698–1716

CARIANI PA, 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

CRAMER EM, HUGGINS WH (1958) Creation of pitch through binaural interaction. J Acoust Soc Am 30:413–417

DARWIN CJ, CARLYON RP (1995) Auditory grouping. In: Moore BCJ (ed) Hearing. Academic, San Diego, pp 387–424

DE BOER E (1956) On the "residue" in hearing. PhD thesis, University of Amsterdam

GALBRAITH GC (1994) Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. Electroencephalogr Clin Neurophysiol 92:321–330

GALBRAITH GC, ARROYO C (1993) Selective attention and brainstem frequency-following responses. Biol Psychol 37:3–22

GALBRAITH GC, DOAN BQ (1995) Brain-stem frequency-following and behavioral-responses during selective attention to pure-tone and missing fundamental stimuli. Int J Psychophysiol 19:203–214

GALBRAITH GC, BHUTA SM, CHOATE AK, KITAHARA JM, MULLEN TA (1998) Brain stem frequency-following response to dichotic vowels during attention. Neuroreport 9:1889–1893

GARCIA D, HALL DA, PLACK CJ (2010) The effect of stimulus context on pitch representations in the human auditory cortex. Neuro-Image 51:808–816

GLASBERG BR, MOORE BCJ (2002) A model of loudness applicable to time-varying sounds. J Audio Eng Soc 50:331–342

GLASER EM, SUTER CM, 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

GOBLICK TJ, PFEIFFER RR (1969) Time-domain measurements of cochlear nonlinearities using combination click stimuli. J Acoust Soc Am 46:924–938

GOCKEL HE, CARLYON RP, PLACK CJ (2011) Combination of spectral and binaurally created harmonics in a common central pitch processor. J Assoc Res Otolaryngol 12:253–260

GREENBERG S, SMITH JC, MARSH JT, BROWN WS (1978) Human frequency following response to synthetic vowels. J Acoust Soc Am 63:S76

GREENBERG S, MARSH JT, BROWN WS, SMITH JC (1987) Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. Hear Res 25:91–114

GREENWOOD DD (1990) A cochlear frequency-position function for several species—29 years later. J Acoust Soc Am 87:2592–2605

HALL DA, PLACK CJ (2007) The human 'pitch center' responds differently to iterated noise and Huggins pitch. Neuroreport 18:323–327

HALL DA, PLACK CJ (2009) Pitch processing sites in the human auditory brain. Cereb Cortex 19:576–585

HOUTSMA AJM, GOLDSTEIN JL (1972) The central origin of the pitch of complex tones: evidence from musical interval recognition. J Acoust Soc Am 51:520–529

KRISHNAN A (2002) Human frequency-following responses: representation of steady-state synthetic vowels. Hear Res 166:192–201

KRISHNAN A (2006) Frequency-following response. In: Burkard RF, Don M, Eggermont JJ (eds) Auditory evoked potentials: basic principles and clinical application. Lippincott, Williams, and Wilkins, Philadelphia

KRISHNAN A, PLACK CJ (2011) Neural encoding in the human brainstem relevant to the pitch of complex tones. Hear Res 275:110–119

KRISHNAN A, XU Y, GANDOUR JT, CARIANI PA (2004) Human frequency-following response: representation of pitch contours in Chinese tones. Hear Res 189:1–12

KRISHNAN A, XU Y, GANDOUR J, CARIANI P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res 25:161–168

LICKLIDER JCR (1956) Auditory frequency analysis. In: Cherry C (ed) Information theory. Academic, New York, pp 253–268

LOPEZ-POVEDA EA, MEDDIS R (2001) A human nonlinear cochlear filterbank. J Acoust Soc Am 110:3107–3118

MARSH JT, BROWN WS, SMITH JC (1975) Far-field recorded frequency-following responses: correlates of low pitch auditory perception in humans. Electroencephalogr Clin Neurophysiol 38:113–119

MEDDIS R, HEWITT M (1991a) Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. I: Pitch identification. J Acoust Soc Am 89:2866–2882

MEDDIS R, HEWITT M (1991b) Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. II: Phase sensitivity. J Acoust Soc Am 89:2883–2894

MEDDIS R, O'MARD L (1997) A unitary model of pitch perception. J Acoust Soc Am 102:1811–1820

MOORE BCJ (2003) An introduction to the psychology of hearing, 5th edn. Academic, San Diego

MOORE BCJ, GOCKEL HE (2011) Resolvability of components in complex tones and implications for theories of pitch perception. Hear Res 276:88–97

MOORE GA, MOORE BCJ (2003) Perception of the low pitch of frequency-shifted complexes. J Acoust Soc Am 113:977–985

MUSACCHIA G, SAMS M, SKOE E, KRAUS N (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci U S A 104:15894–15898

PATTERSON RD (1973) The effects of relative phase and the number of components on residue pitch. J Acoust Soc Am 53:1565–1572

PATTERSON RD, WIGHTMAN FL (1976) Residue pitch as a function of component spacing. J Acoust Soc Am 59:1450–1459

PATTERSON RD, UPPENKAMP S, JOHNSRUDE IS, GRIFFITHS TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776

PENAGOS H, MELCHER JR, OXENHAM AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci 24:6810–6815

PLACK CJ, FITZPATRICK S, CARLYON RP, GOCKEL HE (2010) A temporal code for Huggins pitch? In: Lopez-Poveda EA, Palmer AR, Meddis R (eds) Advances in auditory physiology, psychophysics and models. Springer, New York, pp 191–199

RUSSO NM, SKOE E, TROMMER B, NICOL T, ZECKER S, BRADLOW A, KRAUS N (2008) Deficient brainstem encoding of pitch in children with autism spectrum disorders. Clin Neurophysiol 119:1720–1731

SCHOUTEN JF, RITSMA RJ, CARDOZO BL (1962) Pitch of the residue. J Acoust Soc Am 34:1418–1424

SHACKLETON TM, LIU LF, PALMER AR (2009) Responses to diotic, dichotic, and alternating phase harmonic stimuli in the inferior colliculus of guinea pigs. J Assoc Res Otolaryngol 10:76–90

SKOE E, KRAUS N (2010) Auditory brain stem response to complex sounds: a tutorial. Ear Hear 31:302–324

SMITH JC, MARSH JT, BROWN WS (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr Clin Neurophysiol 39:465–472

SMITH JC, MARSH JT, GREENBERG S, BROWN WS (1978) Human auditory frequency-following responses to a missing fundamental. Science 201:639–641

SONG JH, SKOE E, WONG PC, KRAUS N (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. J Cogn Neurosci 20:1892–1902

SUMNER CJ, LOPEZ-POVEDA EA, O'MARD LP, MEDDIS R (2002) A revised model of the inner-hair cell and auditory–nerve complex. J Acoust Soc Am 111:2178–2188

WIEGREBE L (2001) Searching for the time constant of neural pitch extraction. J Acoust Soc Am 109:1082–1092

WILE D, BALABAN E (2007) An auditory neural correlate suggests a mechanism underlying holistic pitch perception. PloS one 2:e369

WONG PC, SKOE E, RUSSO NM, DEES T, KRAUS N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci 10:420–422