

Human frequency-following response: representation of pitch contours in Chinese tones

Ananthanarayan Krishnan ^{a,*}, Yisheng Xu ^a, Jackson T. Gandour ^a, Peter A. Cariani ^b

^a Auditory Electrophysiology Laboratory, Department of Audiology and Speech Sciences, 1353 Heavilon Hall, Purdue University, 500 Oval Drive, West Lafayette, IN 47907-2038, USA

^b Eaton Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114, USA

Received 27 May 2003; accepted 1 December 2003

Abstract

Auditory nerve single-unit population studies have demonstrated that phase-locking plays a dominant role in the neural encoding of both the spectrum and voice pitch of speech sounds. Phase-locked neural activity underlying the scalp-recorded human frequency-following response (FFR) has also been shown to encode certain spectral features of steady-state and time-variant speech sounds as well as pitch of several complex sounds that produce time-invariant pitch percepts. By extension, it was hypothesized that the human FFR may preserve pitch-relevant information for speech sounds that elicit time-variant as well as steady-state pitch percepts. FFRs were elicited in response to the four lexical tones of Mandarin Chinese as well as to a complex auditory stimulus which was spectrally different but equivalent in fundamental frequency (f_0) contour to one of the Chinese tones. Autocorrelation-based pitch extraction measures revealed that the FFR does indeed preserve pitch-relevant information for all stimuli. Phase-locked interpeak intervals closely followed f_0 . Spectrally different stimuli that were equivalent in F_0 similarly showed robust interpeak intervals that followed f_0 . These FFR findings support the viability of early, population-based ‘predominant interval’ representations of pitch in the auditory brainstem that are based on temporal patterns of phase-locked neural activity. © 2004 Elsevier B.V. All rights reserved.

Key words: Time-variant pitch; Voice pitch; Phase-locking; Autocorrelation; Speech prosody; Chinese tones

1. Introduction

Voice pitch is a fundamental auditory perceptual attribute that is important for the perception of speech and music. The evaluation of neural mechanisms underlying pitch perception provides an avenue to understand the neural basis of processing auditory information. Pitch perception and its physiological bases remain topics of controversy up to the present. Most periodic complex sounds (including speech) evoke low pitches associated with their fundamental frequency, sometimes termed periodicity pitch (deBoer, 1976; Evans, 1978; Moore, 1989). Energy may or may not be present at the fundamental frequency. In contrast, place or spectral pitch is associated with individual fre-

quency components (Goldstein, 1973; Terhardt, 1973; Burns and Viemeister, 1976; Moore and Glasberg, 1986).

Several classes of neural information processing models have been proposed to account for the pitch of complex tones. Rate place neural models use spatial discharge rate patterns along tonotopically organized neural maps to represent the stimulus spectrum. Pitch is then extracted by spectrally based pattern recognition mechanisms that detect patterns of excitation produced by harmonically related components (Goldstein, 1973; Terhardt, 1973). Temporal place models utilize local discharge synchrony information between neighboring neurons (Young and Sachs, 1979), or interspike intervals within single neurons (Srulovicz and Goldstein, 1983) to form a frequency-based central spectrum representation. This frequency domain representation is then analyzed by the pattern recognition mechanisms (Whitfield, 1970; Miller and Sachs, 1984).

* Corresponding author. Tel.: +1 (765) 494-3793; Fax: +1 (765) 494-0771.

E-mail address: rkrish@purdue.edu (A. Krishnan).

Purely temporal models use the population interval distribution derived by combining interspike intervals of single auditory neurons over a broad range of characteristic frequencies. It has long been appreciated that discharge periodicities and interspike intervals related to the fundamental are present in the responses of auditory nerve fibers (Young and Sachs, 1979; Rose, 1980; Delgutte, 1980; Voigt et al., 1982; Evans, 1983; Miller and Sachs, 1984; Greenberg, 1986; Palmer et al., 1986). The predominant interval hypothesis holds that the perceived pitch corresponds to the most frequent interspike interval present in the auditory nerve at any given time (Licklider, 1951; Moore, 1980; Meddis and Hewitt, 1991; Cariani and Delgutte, 1996). Using computer simulations of the auditory nerve, Meddis and Hewitt (1991) concretely demonstrated the plausibility of the hypothesis. In their electrophysiological study, Cariani and Delgutte (1996) recorded responses of cat auditory nerve fibers and combined interval distributions from many fibers to form an estimate of population interval distribution in the entire auditory nerve. Many deep correspondences between features of these interval distributions and to patterns of human pitch judgments for a variety of complex sounds were found in the two studies. It thus appears that a central processor capable of analyzing these intervals can provide a unified explanation for many different aspects of pitch perception (Meddis and O'Mard, 1997; Cariani, 1998). Thus, neural phase-locking plays a dominant role in the neural encoding of low pitch associated with complex sounds. Neural phase-locking in the auditory nerve and cochlear nucleus neurons has also been implicated in the temporal encoding of the spectra of steady-state and time-variant speech sounds (Young and Sachs, 1979; Sachs et al., 1983; Miller and Sachs, 1983, 1984; Palmer et al., 1986; Blackburn and Sachs, 1990; Keilson et al., 1997; Rhode, 1998; Recio and Rhode, 2000).

The scalp-recorded human frequency-following response (FFR) reflects sustained phase-locked 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). Because the FFR encompasses responses of multiple neural subpopulations with different best frequencies and response latencies, more stimulus-related temporal information may be available in single units and local neuronal ensembles than in the population response as a whole. Stimulus-related temporal structure observed in the FFR therefore forms the lower limit of the neural timing information potentially available for neuronal information processing at the rostral brainstem level.

We recently demonstrated that the phase-locked activity underlying the FFR does indeed preserve spectral peaks corresponding to the first two formants of both

steady-state speech-like sounds (Krishnan, 1999, 2002) and time-variant speech-like sounds (Krishnan and Parkinson, 2000; Plyler and Ananthanarayan, 2001). The human FFR preserves pitch-relevant information about complex sounds that produce *time-invariant* pitch (Greenberg et al., 1987), leading them to conclude that pitch-relevant neural activity is based on the temporal pattern of neural activity in the brainstem.

In light of these earlier findings, it is postulated that the phase-locked activity underlying the FFR generation is also sufficiently dynamic to encode *time-varying* pitch of speech sounds. The specific aims of this study are to determine whether the phase-locked activity underlying FFR generation (1) is sufficiently dynamic to represent the pitch of stimuli that produce a more complex range of pitch percepts, including changes in trajectory and direction of pitch change; (2) is more robust for rising versus falling pitch trajectories; (3) preserves certain spectral features of the complex stimuli; and (4) supports the predominant interval hypothesis by showing phase-locked interval bands for stimuli that are equivalent in pitch but differ in their spectra.

To address aims 1–3, FFRs were elicited from the four (Mandarin) Chinese tones [similar to, e.g., Tone 1, *ma*^{high level} ‘mother’; Tone 2, *ma*^{high rising} ‘hemp’; Tone 3, *ma*^{low falling–rising} ‘horse’; and Tone 4, *ma*^{high falling} ‘scold’ (Howie, 1976)]. This tonal space provides an optimal window for investigating FFRs in response to time-varying f_0 contours associated with monosyllabic speech sounds. To address aim 4, FFRs in response to Chinese Tone 3, which exhibits a *bidirectional* f_0 contour, are compared to FFRs elicited in response to a complex auditory stimulus that exhibits the same f_0 contour but whose spectral composition is different.

2. Methods

2.1. Subjects

Thirteen adult native speakers of Mandarin, ranging in age from 21 to 27 years, participated in the study. Hearing sensitivity in all subjects was better than 15 dB HL for octave frequencies from 500 to 8000 Hz.

2.2. Stimuli

FFRs were elicited using a set of monosyllabic Chinese syllables that were chosen to contrast the four lexical tones (*pinyin Roman phonemic transcription*): *yi*¹ ‘clothing’, *yi*² ‘aunt’, *yi*³ ‘chair’, *yi*⁴ ‘easy’. This particular stimulus set allows us to address issues related to encoding of spectra and voice pitch. First, all four lexical tones have average pitch trajectories and harmonics that lie within the range of easily recordable FFRs.

Second, the tonal inventory includes f_0 trajectories that exhibit complex contour and directional changes in pitch (unidirectional rising = Tone 2; unidirectional falling = Tone 4; bidirectional falling-rising = Tone 3). Third, these four monosyllables permit us to evaluate

the encoding of spectral components and voice pitch associated with vowels and lexical tones, respectively.

Syllables were created using a synthesis-by-rule scheme (Klatt, 1980). Synthesis parameters were based on voice fundamental frequency (f_0) and duration data

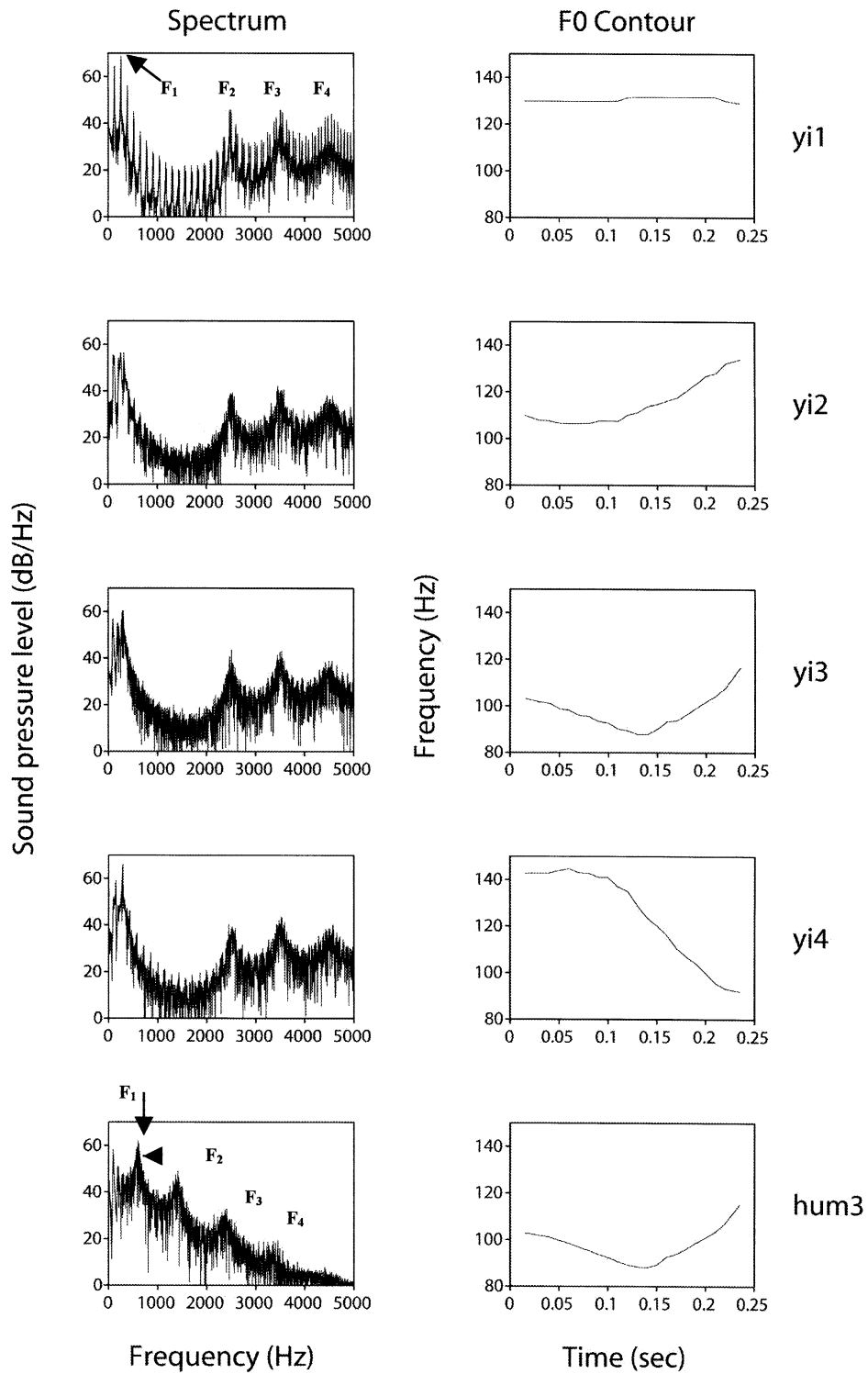


Fig. 1. Acoustic spectra (left panels) and F_0 contours of the stimuli (right panels). The four formants are identified. Note that yi^3 and hum^3 are spectrally different but have identical F_0 contours.

from an adult male speaker (Xu, 1997). Vowel formant frequencies were steady-state and held constant across the four syllables: $F_1 = 300$ Hz; $F_2 = 2500$ Hz; $F_3 = 3500$ Hz; and $F_4 = 4530$ Hz (Howie, 1976). Tonal duration was normalized to 250 ms including the 5 ms linear rise and fall times. These stimuli make it possible to evaluate FFRs elicited by different f_0 contours in a fixed spectral context. The acoustic spectra and f_0 contours of these four stimuli are shown in Fig. 1.

In addition, FFRs were elicited from a ‘hum’ version of yi^3 ‘chair’, hereafter referred to as hum^3 . This stimulus was created by extracting the f_0 contour of yi^3 using the Praat speech synthesis and signal analysis software (Praat, Institute of Phonetic Sciences). Duration, f_0 , and intensity characteristics of the speech stimulus were preserved in its hummed counterpart. Formant frequencies were steady-state, albeit different: $F_1 = 610$ Hz; $F_2 = 1425$ Hz; $F_3 = 2380$ Hz; and $F_4 = 3414$ Hz. It was designed to test the predominant interval hypothesis, which asserts that pitch is primarily determined by a purely temporal encoding scheme, thereby achieving pitch equivalence for stimuli that differ in spectra. The acoustic spectrum and f_0 contour of hum^3 is also shown in Fig. 1. All stimuli were controlled by a signal generation and data acquisition system (Tucker-Davis Technologies, System II). The synthesized stimulus files were routed through a digital to analog module, and presented monaurally to the right ear through magnetically shielded insert earphones (Biologic, TIP-300).

2.3. Recording system

Subjects reclined comfortably in an acoustically and electrically shielded booth. Evoked responses were recorded differentially between scalp electrodes placed on the midline of the forehead at the hairline and the 7th cervical vertebra (C7 location). Another electrode placed on the mid-forehead (‘Fpz’) served as the common ground. The inter-electrode impedances were maintained below 3000 Ω . The EEG inputs were amplified by 200 000 and band-pass filtered from 100 to 3000 Hz (6 dB/octave roll-off, RC response characteristics). Each response waveform represents an average of 2000 stimulus presentations over a 260 ms analysis window using a sampling rate of 20 kHz.

2.4. Experimental protocol

FFRs were recorded for each subject for right ear stimulation in all five test conditions – yi^1 , yi^2 , yi^3 , yi^4 , and hum^3 . All stimuli were presented monaurally at 60 dB nHL at a repetition rate of 3.13/s. The order of test conditions was randomized both within and across subjects. All subjects gave informed consent in compli-

ance with a protocol approved by the Institutional Review Board of Purdue University.

2.5. Response evaluation

(1) Narrow-band spectrograms (using a Gaussian window) were obtained for each of the grand averaged FFRs to determine the spectral composition and magnitude of the phase-locked neural activity (see Fig. 5). Grand averaged FFR for each stimulus was derived by averaging the FFRs obtained from each of the 13 subjects. (2) The ability of the FFR to follow the pitch change in the stimuli was evaluated by extracting the pitch contour from the grand averaged FFRs using a periodicity detection short-term autocorrelation algorithm (Boersma, 1993). Essentially, this algorithm performs a short-term autocorrelation analysis on a number of small segments (frames) taken from the signal (stimuli and FFR). This analysis yielded estimates of both pitch period (the time lag associated with the autocorrelation maximum) and pitch strength (the magnitude of the normalized autocorrelation peak expressed as harmonic-to-noise ratio ranging from 0 to 1). Pitch contours were computed for the stimuli (Fig. 1) and from the grand averaged FFRs (Fig. 2). Pitch strength measures obtained from the FFRs recorded from individual subjects for each stimulus (over the duration of the response) were used to construct the average autocorrelation magnitude shown in Fig. 4. We use the term pitch strength to convey the degree of periodicity in the neural activity underlying the FFR and it is used here as a qualitative index of pitch salience. A similar measure of pitch salience has demonstrated good correspondence between the normalized magnitude of the

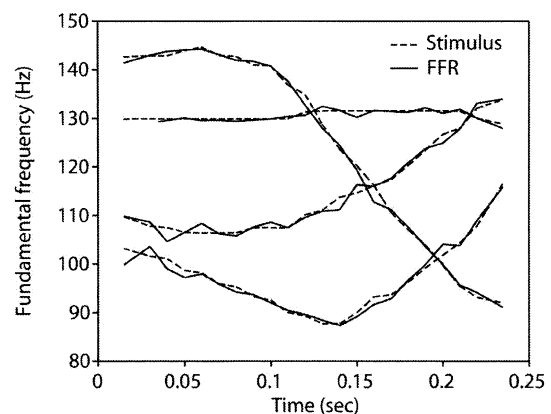


Fig. 2. FFR voice pitch contours (solid lines) superimposed on stimulus F_0 contours (broken lines) for the four Chinese speech sounds (yi^1 : flat; yi^2 : rising; yi^3 : bidirectional; and yi^4 : falling). Pitch was extracted using a short-term autocorrelation algorithm (Boersma, 1993) on multiple frames of the signal utilizing a Hanning window of effective length equal to 0.04 s.

autocorrelation peak and perceived pitch salience for a number of low pitch complex sounds (Cariani and Delgutte, 1996). (3) Short-term autocorrelation functions and running autocorrelograms were computed for the stimuli and the grand averaged FFRs to show the variation of FFR periodicities over the duration of the response (Fig. 3). The autocorrelogram is an expansion of the signal that plots post-stimulus onset time vs. time lag, i.e., $ACG(\tau, t) = X(t) \times X(t - \tau)$ for each time t and time lag τ . Thus the autocorrelogram represents the running distribution of all-order intervals present in the population response. Virtually the same algorithm was used by Cariani and Delgutte (1996) to obtain their autocorrelograms to visualize time-varying changes in all-order interspike interval distributions associated with changing fundamentals. A repeated measures analysis of variance (ANOVA) was conducted to determine whether pitch strength varied as a function of the f_0 contours associated with the four Chinese tones.

3. Results

3.1. Representation of voice pitch

Pitch contours extracted from the FFR (solid lines) to each of the four speech stimuli are superimposed on their corresponding stimulus f_0 contours (broken lines) in Fig. 2. It is clear from this figure that the phase-locked FFR activity carrying pitch-relevant information faithfully follows the pitch changes presented in each stimulus.

The short-term autocorrelation functions and the running autocorrelograms for all five stimuli and their corresponding grand average FFRs are plotted in Fig. 3. The stimulus autocorrelation functions reveal major maxima at the fundamental period $1/f_0$ and its multiples. Minor maxima reflect the formant period $1/F_1$ and its multiples. For example, stimulus yi^1 (top left panel) shows a major peak at about 7.69 ms (which is equal to $1/f_0$ and the pitch heard). A second major peak at about 15.38 ms (right edge of the plot) represents interval spanning two fundamental periods. These interval multiples (e.g., $1/f_0$, $2/f_0$) reflect stimulus periodicities and the autocorrelation-like properties of interval representation. The first minor peak (long arrow) at about 3.33 ms is associated with the formant period $1/F_1$ and the second minor peak at about 13.32 ms corresponds to $4/F_1$. Consistent with its autocorrelation function, the autocorrelogram for this stimulus shows dark bands corresponding to these intervals with the densest band associated with the fundamental period. The stimulus autocorrelation functions for yi^3 and hum^3 are different with hum^3 showing a greater number of f_0 and F_1 intervals. This difference is more clearly indicated in the

autocorrelograms and reflects the spectral difference between these two stimuli.

While there is a close correspondence between the stimulus and FFR autocorrelation functions for yi^1 , yi^2 (albeit the minor peaks are very small) and yi^4 , the FFR autocorrelation functions for yi^3 and hum^3 show a single broad maximum. Consistent with the stimulus autocorrelation functions, the FFR autocorrelograms for yi^1 , yi^4 and yi^2 (to a lesser extent) show three interval bands that reflect phase-locked activity at $1/f_0$, $1/F_1$ (indicated by the arrowhead), and $4/F_1$, with the densest band for all three stimuli associated with the fundamental period. Similarly, the autocorrelograms for yi^3 and hum^3 exhibit a single interval band that is closely related to the *time-varying* fundamental period $1/f_0$. The interval band corresponding to $1/F_1$, while present, is not easily discernible in the FFR autocorrelograms for yi^3 and the time-varying portion of yi^4 . In general there appears to be a suppression of non- f_0 intervals. The relatively more diffused interval (broader band) at $1/f_0$ for the FFR to yi^3 may simply represent inadequate resolution of three intervals – the fundamental period, one interval above, and one interval below the fundamental period. Also, it is noteworthy that the FFR autocorrelograms for yi^3 and hum^3 were virtually identical given that these stimuli showed differences in spectra, autocorrelation functions, and autocorrelograms.

The average FFR pitch strength for each of the four speech stimuli is plotted in Fig. 4. A repeated measures ANOVA showed a significant effect of stimulus type on pitch strength [$F(4,44) = 7.30$, $P = 0.0009$]. A test for the linear contrast, $(yi^2 + yi^3) - (yi^1 + yi^4)$, indicated that the pitch strengths for yi^2 and yi^3 , both of which contain rising f_0 segments, were significantly greater than those for yi^1 and yi^4 [$F(1,41) = 20.97$, $P < 0.0001$]. Also, there was no significant difference in pitch strength between yi^3 and hum^3 .

3.2. Representation of spectral components

Stimulus spectrograms and grand averaged FFR spectrograms are plotted in Fig. 5. The spectrograms of all five stimuli (left panels) reveal energy bands at several multiples of f_0 with the expected stronger energy bands at F_1 -related harmonics (h_2 , h_3 for stimulus $yi^1 - yi^4$ and h_6 , h_7 for stimulus hum^3). Like the stimulus spectrograms, the FFR spectrograms (right panels) for each stimulus show energy bands at several harmonics and appreciably stronger bands at F_1 -related harmonics. However, unlike the stimulus spectrograms, the response spectrograms also show strong phase-locked neural activity at f_0 and harmonics proximal to $2F_1$. Finally, the FFR tracking of the formant-related harmonics is stronger and continuous over the stimulus

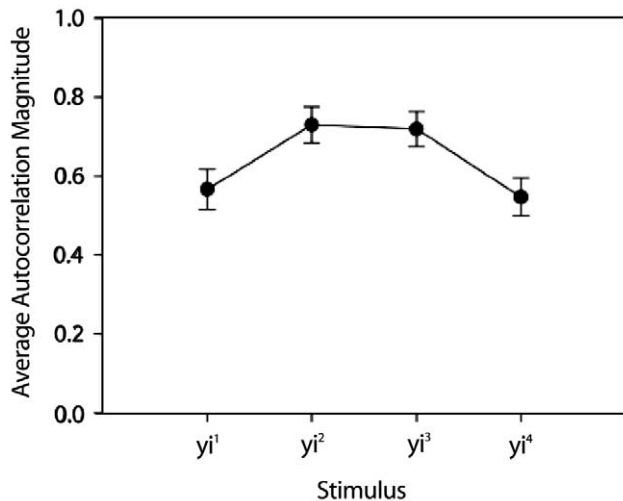


Fig. 4. Mean normalized magnitude of the FFR autocorrelation peak for the four Chinese speech sounds. The error bars indicate ± 1 S.D.

duration for yi² compared to the relatively weaker tracking, particularly for the falling segment in the harmonic, observed for yi⁴.

4. Discussion

4.1. Representation of voice pitch

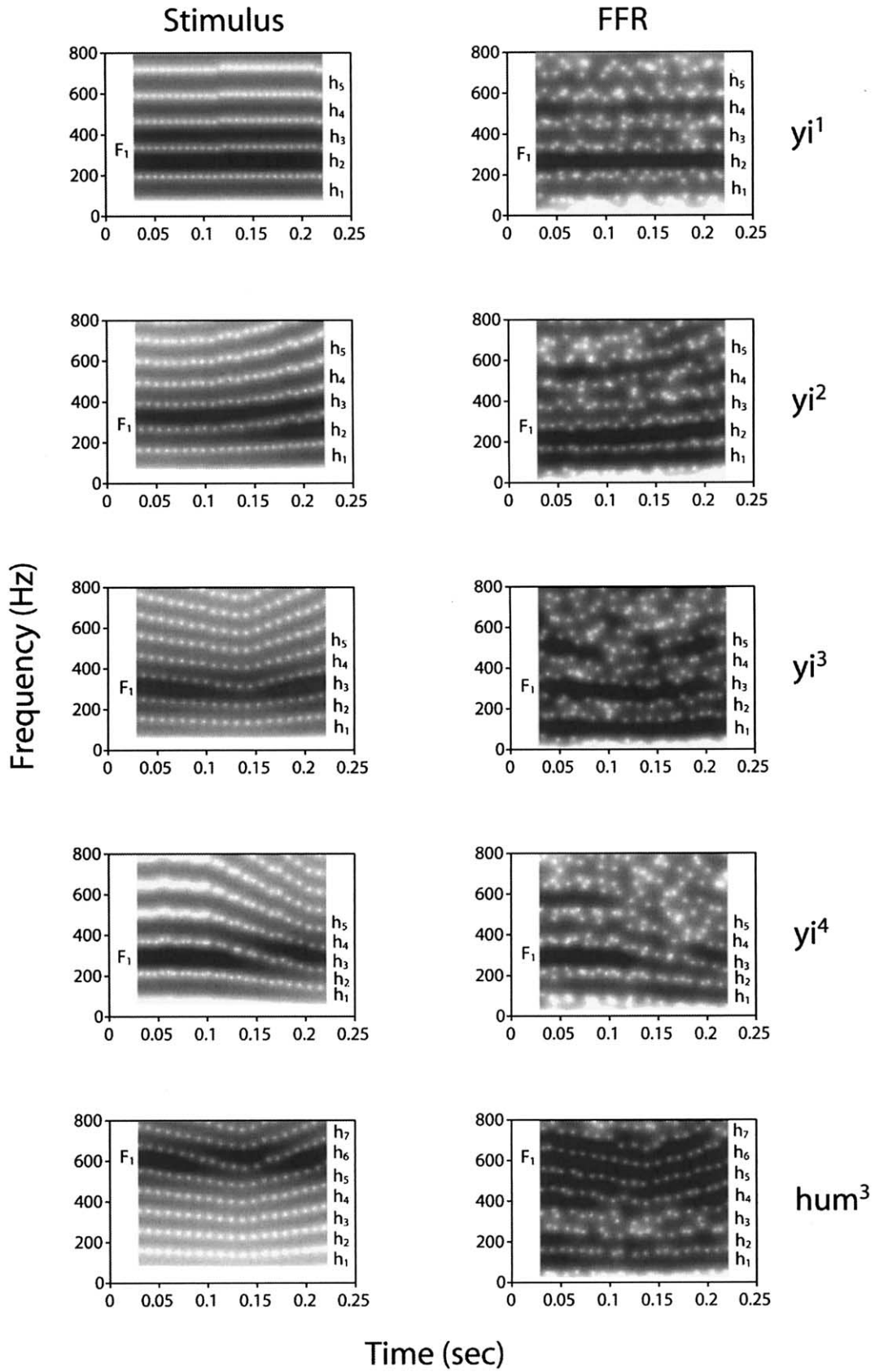
The results of this study clearly demonstrated that for several stimuli with *time-varying* f_0 contours, the prominent interval band in the phase-locked FFR neural activity followed closely the fundamental period ($1/f_0$). These findings suggest that a robust neural temporal representation for pitch is preserved in the phase-locked neural activity of an ensemble of neural elements in the rostral brainstem. These findings are consistent with Greenberg et al. (1987), who reported that the FFR encoded pitch-relevant information to complex stimuli with time-invariant fundamental frequency. They concluded that this pitch-relevant activity reflected in the FFR is based on the temporal discharge patterns of neurons in the rostral brainstem pathways instead of neural activity synchronized to the envelope of the signal waveform (Hall, 1979). Specifically, the FFR amplitude for a three-tone complex was not appreciably altered by changes in the envelope modulation depth.

Our findings are also consistent with reports of tem-

poral discharges of auditory nerve fibers in response to complex harmonic stimuli, which used interspike interval distributions (Evans, 1983; Greenberg, 1986; Voigt et al., 1982) or period histograms (Delgutte, 1980; Miller and Sachs, 1984; Palmer et al., 1986; Young and Sachs, 1979). These data indicate that interspike intervals and/or discharge periodicities related to the fundamental are present in the responses of many single auditory nerve fibers. Cariani and Delgutte (1996) were the first to combine interval distributions from many fibers to form an estimate of population interval distribution and its correspondence to human pitch judgments. They demonstrated that for a variety of complex stimuli with fundamental frequencies ranging from 80 to 500 Hz, the most frequent interspike interval in the auditory nerve corresponds closely to the perceived low pitch. Thus, neural phase-locking plays a dominant role in the neural encoding of low pitch associated with complex sounds. The relationship between the FFR and temporal discharge patterns of single units is neither overly simple nor complex. Both population response synchronies and interspike interval patterns are direct products of stimulus-locked neuronal activity, and are therefore intimately related. Based on what is known about neuronal responses in the cochlear nuclei to periodic stimuli, it is very likely that any periodicities observed in the FFR would also be present in single units and local ensembles. In the present study, the consistent observation of close correspondence between the best pitch estimates (following closely the f_0) and the prominent band of FFR neural interval in the autocorrelograms of all stimuli strongly supports the predominant interval hypothesis for voice pitch at the brainstem level. Moreover, robust encoding of the Chinese speech stimuli that change not only in pitch trajectory but also in direction suggests that this temporal pitch encoding mechanism is dynamic.

In this study, the FFR autocorrelograms, depicting the prominent phase-locked interval band at $1/f_0$, were essentially similar for stimuli that were spectrally different but produced equivalent pitch percepts (yi³ vs. hum³). Auditory nerve single-unit population responses elicited by a variety of spectrally different sounds similarly produce the same pitch (Cariani and Delgutte, 1996). Specifically, the predominant interval for all stimuli corresponded to their common pitch ($1/f_0$). These findings taken together strongly support a temporal code for voice pitch based on temporal distribu-

Fig. 3. Short-term autocorrelation functions and running autocorrelograms for all stimuli (left pair) and for grand averaged FFR (right pair). The autocorrelation functions exhibit major maxima (short arrows) at the fundamental period ($1/f_0$) and shorter delay minor maxima (long arrows) at the formant-related harmonic ($1/F_1$). Other peaks (more apparent for the stimulus) in the autocorrelation function represent subharmonics and harmonics of f_0 and F_1 . The dark interval bands in the autocorrelograms represent distribution of stimulus and FFR periodicities. The densest band (more apparent for the FFR) corresponds to the fundamental period ($1/f_0$). The arrow heads point to interval bands corresponding to $1/F_1$. Note the general correspondence between the autocorrelation peaks and the location of the prominent interval bands.



tion of phase-locked neural activity in a population of neural elements.

Another interesting finding on voice pitch is that the FFR pitch strengths for yi^2 and yi^3 are greater than for yi^1 and yi^4 . It has previously been demonstrated that the FFR amplitude for falling tonal sweeps is smaller than that for corresponding rising tonal sweeps (Krishnan and Parkinson, 2000). Similar selectivity to rising tonal sweeps has been observed for cochlear microphonics (Shore and Cullen, 1984), eighth nerve compound action potentials (Shore and Nuttall, 1985), and responses of the ventral cochlear nucleus units (Shore et al., 1987). Specifically, relative phase measurements of microphonic potentials suggest that displacements of the cochlear partition towards either scala occur closely together in time for rising tones, but are dispersed in time for falling tones (Shore and Cullen, 1984). Consequently, rising tones produce relatively greater synchrony in the eighth nerve activity than falling tones. In fact, the eighth nerve compound action potential (CAP) evoked by a rising tone is larger in amplitude (reflecting more synchronous activity) than the CAP to falling tones (Shore and Nuttall, 1985). This response asymmetry is further indicated by differential temporal response patterns to rising and falling tones for single units in the ventral cochlear nucleus (Shore et al., 1987). It is plausible that differences in FFR pitch strength between Chinese tones with rising and falling f_0 contours may reflect a differential temporal response pattern for rising and falling tones among the neural elements generating the FFR. Consistent with this notion is our observation of relatively more robust FFR representation of harmonics in the dominant region for pitch for the stimuli with rising contour compared to the stimulus with falling contour. This observation appears to suggest that the greater pitch salience for the rising stimulus may be due to its more effective activation of frequencies in the dominant pitch region (Flanagan and Guttman, 1960a,b; Ritsma, 1967; Plomp, 1967).

Finally, the autocorrelograms, particularly for the time-varying segments of yi^4 (falling contour) and, to a lesser extent, yi^2 (rising contour), showed suppression of components surrounding f_0 . While the basis of this suppression is unclear, at least it does not appear to be simply an artifact introduced by the analysis algorithm (given that the autocorrelograms for all stimuli clearly show multiple bands corresponding to intervals other than f_0). Consistent with the spectral data (Fig. 5) it

is possible that the absence of phase-locked bands (or the presence of lighter weaker bands) at intervals other than f_0 (particularly for the changing contour in yi^4) in the autocorrelogram may simply reflect degraded phase-locking at these non- f_0 frequencies resulting from disruption of neural synchrony. The systematic characteristic frequency-dependent latency shifts due to cochlear delays in the low frequency region could produce this disruption. That is, the $1/f_0$ interval dominates because other intervals are more widely distributed and therefore are smeared out. However, it is not clear why a falling contour would be more susceptible to such a disruption. While the underlying mechanisms are not clear these findings collectively suggest that the auditory system is more sensitive to upward sweeping tones.

Human psychophysical experiments do, in fact, support this view. The just-noticeable difference (JND) for change in frequency of sweep tone stimuli is smaller for rising tones than for falling tones (Shore et al., 1987), and detection thresholds for rising tones are lower than those for falling tones (Collins and Cullen, 1978; Cullen and Collins, 1979; Nabelek, 1978). Discrimination of falling sweeps requires longer durations and/or higher sweep rates than rising sweeps (Schouten, 1985). Turning to naturally spoken Dutch stimuli, JNDs indicate that changes in the size of pitch rises are easier to discriminate than changes in the size of otherwise identical pitch falls ('t Hart, 1974). Using synthetic speech stimuli, Klatt (1973) demonstrated that the JNDs for detection of changes in slope of linear F_0 ramps show greatest sensitivity when one ramp is rising and the other is falling. Based on multidimensional scaling analyses of synthetic speech stimuli, the underlying perceptual dimension related to direction of pitch change in the stimulus space separates primarily rising vs. non-rising f_0 movements (Gandour and Harshman, 1978; Gandour, 1983).

4.2. Representation of spectral components

For all stimuli in this study, the FFR spectrograms show clearly discernible peaks at the fundamental frequency and its harmonics. These data suggest that the sustained neural activity in a population of auditory brainstem neurons is phase-locked to the individual harmonics of the complex Mandarin Chinese stimuli. It is well established that the FFR is a sustained response reflecting phase-locking in a population of neural elements within the rostral brainstem (Worden and

←
Fig. 5. Narrow-band spectrograms of stimuli (left panels) and grand averaged FFRs (right panels). Spectrograms were obtained utilizing a Gaussian window of 29 ms length; frequency step = 20 Hz; and a dynamic range = 50 dB. First formant (F_1) frequency and f_0 -related harmonics (h_1 – h_7) are identified.

Marsh, 1968; Marsh et al., 1974; Smith et al., 1975; Glaser et al., 1976). Phase-locking, as reflected in the FFR, has been demonstrated for pure tones (Moushegian et al., 1973; Gardi et al., 1979; Ananthanarayan and Durrant, 1992), complex steady-state stimuli (Hall, 1979; Greenberg et al., 1987; Krishnan, 1999, 2002), and time-variant stimuli (Krishnan and Parkinson, 2000; Plyler and Ananthanarayan, 2001). The human FFR spectrum, in response to several steady-state synthetic vowels, indeed shows clear and distinct peaks corresponding to the two formant frequencies (Krishnan, 2002).

Auditory nerve single-unit population studies have clearly established that neural phase-locking is a primary basis for encoding synthetic speech-like sounds (Reale and Geisler, 1980; Young and Sachs, 1979; Voigt et al., 1982; Sachs et al., 1983; Miller and Sachs, 1983, 1984). Phase-locked activity to the formant harmonics dominates the temporal response patterns. Moreover, a good estimate of the stimulus spectrum can be derived from these response patterns. Separate populations of auditory nerve neurons are involved in encoding the first and the second formant of the speech-like sounds. Place-specific FFRs have been reported in response to a moderate intensity 500 Hz tone burst (Ananthanarayan and Durrant, 1992) and complex tones (Greenberg, 1980; Greenberg et al., 1987); the presence of distinct peaks not only at F_1 and F_2 but also at a frequency corresponding to the $2f_1 - f_2$ distortion product in the FFR to two-tone vowels (Krishnan, 1999, 2002). A similar conclusion has been drawn about tonotopic specificity of the FFR to complex tones based on latency measures for different carrier frequencies (Greenberg, 1980; Greenberg et al., 1987). The ability of the FFR to systematically follow the frequency change presented in an upward and a downward swept tone burst (Krishnan and Parkinson, 2000) argues strongly for separate populations of neurons contributing to the multiple peaks present in the FFR to complex stimuli. That is, each harmonic forming the complex stimuli used in this study would produce place-specific activity in the cochlea, which then would activate an ensemble of neurons in the brainstem with different characteristic frequencies. Bledsoe and Moushegian (1980) interpreted the presence of multiple peaks in their 500 Hz FFR to suggest that multiple brainstem sites contributed to its generation. Thus, the spectral peaks in the FFR at several different harmonics for all stimuli may very well reflect phase-locked activity from distinct populations of neurons. However, it is also possible that the broadly tuned single units in the lower brainstem with similar characteristic frequencies (Rhode, 1994) could contribute to the encoding of frequency change presented in the sweeping stimuli used in our study.

Another observation in the results of this study was that the FFR spectrum for the five different stimuli was dominated by the response peaks at F_1 harmonics with the higher harmonics showing appreciably smaller response magnitudes. Similar response behavior has been observed for both the auditory nerve fiber response to tonal approximations of vowels (Reale and Geisler, 1980) and synthetic speech sounds (Young and Sachs, 1979); and for the human scalp-recorded FFR responses to two-tone vowel approximations (Krishnan, 1999) and steady-state synthetic speech sounds (Krishnan, 2002). Young and Sachs (1979), using synthetic steady-state vowels, observed that the auditory nerve single-unit temporal response to the first formant was the largest, particularly for the higher intensity levels. These authors explain that at low intensities, the temporal response to each stimulus component is maximal among units with characteristic frequencies near the frequency of the component – i.e., the response is place-specific. But, as sound level is increased, the response to the formants, particularly the first formant, not only increases near their place but also spreads primarily toward a place characterized by units with higher characteristic frequency (synchrony spread). Thus, it is possible that the additional contribution to the first formant by these higher characteristic frequency neurons could account for the dominance of the response at the first formant harmonics, not only in the single-unit population data but also in the FFR data presented in this paper. However, it is also plausible that the smaller amplitude for the higher harmonics in our FFR data may, at least in part, reflect the decreasing phase-locking ability with increasing frequency observed for both FFR (Gardi et al., 1979) and single-unit responses (Palmer et al., 1986).

The FFR data also showed a dominance of the harmonics closest to the first formant frequency and its multiples (compared to the non-formant harmonics). Similar results were observed for FFR to steady-state synthetic vowels (Krishnan, 2002). This observation may be suggestive of the formant capture phenomenon described for the single-unit population data (Delgutte and Kiang, 1984; Ghitza, 1988; Greenberg, 1988, 1994, 1996; Young and Sachs, 1979). That is, the responses of fibers with best frequencies near a formant frequency are captured by the largest harmonic near the formant. Both the synchrony spread and the capture phenomenon reflect non-linear signal processing in the cochlea and are believed to contribute to the robust nature of the temporal encoding process.

5. Implications

Our knowledge about processing of speech sounds in

the mammalian nervous system is largely derived from animal single-unit population studies at the level of the auditory nerve and cochlear nucleus. These studies have demonstrated that the temporal place code is indeed preserved at these auditory loci. However, it is not known if the temporal place scheme is preserved at more rostral levels in the brainstem where neural phase-locking is limited to frequencies below about 2000 Hz. For any scheme of neural representation to be effective in information transfer it has to be preserved in some manner all along the auditory neuraxis up to the decoding site. To the extent that the FFR reflects sustained phase-locked neural activity restricted to the rostral brainstem, the encoding of both spectral components and voice pitch-relevant information of Chinese monosyllables in the FFR suggests that neural encoding based on phase-locking is indeed preserved at higher levels (presumably, at the level of the lateral lemniscus and/or inferior colliculus) in the human brainstem.

Thus, the scalp-recorded FFR provides a non-invasive window to view neural processing of voice pitch in human speech sounds at the level of the auditory brainstem. Directions for future research include cross-linguistic studies which compare the pitch strength of Chinese tones when presented to native Chinese speakers, speakers of non-tone languages (e.g., English), and speakers of other tone languages (e.g., Thai). If based strictly on acoustic properties of the stimulus, FFRs in response to time-varying f_0 contours at the level of the brainstem are expected to be homogeneous across listeners regardless of language experience. If, on the other hand, FFRs vary to some extent depending on long-term language learning, they may be somewhat heterogeneous depending on how f_0 cues are used to signal pitch contrasts in the listener's native language. Another interesting question is whether language-specific ear asymmetries for cortical voice pitch representation, as seen in dichotic listening studies of lexical tones (Van Lancker and Fromkin, 1973; Wang et al., 2001), are already evident at the brainstem level.

Finally, autocorrelation-based analysis methods appear to be not only more robust than simple FFT methods for extracting pitch-relevant information, but also are readily applicable to FFR data.

Acknowledgements

This research was supported in part by a research grant from the National Institutes of Health (R01 DC04584-04; J.T.G.). We are grateful to the anonymous reviewers whose insightful suggestions have appreciably improved the manuscript.

References

- Ananthanarayan, A.K., Durrant, J.D., 1992. The frequency-following response and the onset response: Evaluation of frequency specificity using a forward-masking paradigm. *Ear Hear.* 13, 228–233.
- Blackburn, C.C., Sachs, M.B., 1990. The representation of steady-state vowel sound /e/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *J. Neurophysiol.* 63, 1191–1212.
- Bledsoe, S.C., Moushegian, G., 1980. The 500 Hz frequency-following potential in kangaroo rat: An evaluation with noise masking. *Electroencephalogr. Clin. Neurophysiol.* 48, 654–663.
- Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc. Inst. Phon. Sci.* 17, 97–110.
- Burns, E.M., Viemeister, N.F., 1976. Nonspectral pitch. *J. Acoust. Soc. Am.* 60, 863–869.
- Cariani, P., 1998. Neural computations in the time domain. Poster, ARO Midwinter Meeting.
- Cariani, P.A., Delgutte, B., 1996. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716.
- Collins, M.J., Cullen, J.K., 1978. Temporal integration of tone glides. *J. Acoust. Soc. Am.* 63, 469–473.
- Cullen, J.K., Collins, M.J., 1979. Rate effects in the detection of short duration tonal glides. *J. Acoust. Soc. Am.* 65 (Suppl. 1), S38.
- 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.
- Delgutte, B., 1980. Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *J. Acoust. Soc. Am.* 68, 843–857.
- Delgutte, B., Kiang, N.Y.S., 1984. Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *J. Acoust. Soc. Am.* 75, 879–886.
- 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.
- Flanagan, J.L., Guttman, N., 1960a. On the pitch of periodic pulses. *J. Acoust. Soc. Am.* 32, 1308–1319.
- Flanagan, J.L., Guttman, N., 1960b. Pitch of periodic pulses without fundamental component. *J. Acoust. Soc. Am.* 32, 1319–1328.
- Gandour, J., 1983. Tone perception in Far Eastern languages. *J. Phon.* 11, 149–175.
- Gandour, J., Harshman, R., 1978. Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language Speech* 21, 1–33.
- Gardi, J., Salamy, A., Mendelson, T., 1979. Scalp recorded frequency following responses in neonates. *Audiology* 18, 494–506.
- Ghitza, O., 1988. Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment. *J. Phon.* 16, 109–124.
- Glaser, E.M., Suter, C.M., Dasheiff, R., Goldberg, A., 1976. The human frequency following response: its behavior during continuous stimulation. *Electroencephalogr. Clin. Neurophysiol.* 40, 25–32.
- Goldstein, J.L., 1973. An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 54, 1496–1516.
- Greenberg, S., 1980. Neural Temporal Coding of Pitch and Vowel Quality. UCLA Working Papers in Phonetics, Volume 52 [Doctoral Dissertation] Los Angeles, CA.

- Greenberg, S., 1986. Possible role of low and medium spontaneous rate cochlear nerve fibers in the encoding of waveform periodicity. In: Moore, B., Patterson, R. (Eds.), *Auditory Frequency Selectivity*. Plenum, New York, pp. 241–251.
- Greenberg, S., 1988. The ear as a speech analyzer. *J. Phon.* 16, 139–150.
- Greenberg, S., 1994. Speech processing: Auditory models. In: Asher, R.E. (Ed.), *The Encyclopedia of Language and Linguistics*. Pergamon, Oxford, pp. 4206–4227.
- Greenberg, S., 1996. Auditory processing of speech. In: Lass, N. (Ed.), *Principles of Experimental Phonetics*. Mosby, St. Louis, MO, pp. 362–407.
- 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 tone. *Hear. Res.* 25, 91–114.
- Hall, J.W., 1979. Auditory brainstem frequency following responses to waveform envelope periodicity. *Science* 205, 1297–1299.
- Howie, J.M., 1976. *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge University Press, New York.
- Keilson, S.E., Richards, V.M., Wyman, B.T., Young, E., 1997. The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: Evidence for a periodicity-tagged spectral representation. *J. Acoust. Soc. Am.* 102, 1056–1071.
- Klatt, D.H., 1973. Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *J. Acoust. Soc. Am.* 53, 8–16.
- Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* 67, 971–995.
- Krishnan, A., 1999. Human frequency-following responses to two-tone approximations of steady-state vowels. *J. 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., Parkinson, J., 2000. Human frequency-following responses: representation of tonal sweeps. *J. Audiol. Neurootol.* 5, 312–321.
- Licklider, J.C.R., 1951. A duplex theory of pitch perception. *Experientia* 7, 128–134.
- Marsh, J.T., Brown, W.S., Smith, J.C., 1974. Differential brainstem pathways for the conduction of auditory frequency following responses. *Electroencephalogr. Clin. Neurophysiol.* 38, 415–422.
- Meddis, R., Hewitt, M.J., 1991. Virtual pitch and phase-sensitivity studies using a computer model of auditory periphery: I pitch identification. *J. Acoust. Soc. Am.* 89, 2866–2882.
- Meddis, R., O'Mard, L., 1997. A unitary model of pitch perception. *J. Acoust. Soc. Am.* 102, 1811–1820.
- Miller, M.I., Sachs, M.B., 1983. Representation of stop consonants in the discharge patterns of auditory-nerve fibers. *J. Acoust. Soc. Am.* 74, 502–517.
- Miller, M.I., Sachs, M.B., 1984. Representation of voice pitch in discharge patterns of auditory-nerve fibers. *Hear. Res.* 14, 257–279.
- Moore, B.C.J., 1980. Neural interspike intervals and pitch. *Audiology* 19, 363–365.
- Moore, B.C.J., 1989. *Introduction to the Psychology of Hearing*, 3rd edn. Academic Press, London.
- Moore, B.C.J., Glasberg, B.R., 1986. The role of frequency selectivity in the perception of loudness, pitch and time. In: Moore, B.C.J. (Ed.), *Frequency Selectivity in Hearing*. Academic Press, London, pp. 251–308.
- Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp recorded early responses in man to frequencies in the speech range. *Electroencephalogr. Clin. Neurophysiol.* 35, 665–667.
- Nabelek, I., 1978. Temporal summation of constant and gliding tones at masked auditory threshold. *J. Acoust. Soc. Am.* 64, 751–763.
- Palmer, A.R., Winter, I.M., Darwin, C.J., 1986. The representation of steady-state vowel sounds in the temporal discharge patterns of guinea pig cochlear nerve and primary-like cochlear nucleus neurons. *J. Acoust. Soc. Am.* 79, 100–113.
- Plomp, R., 1967. Pitch of complex tones. *J. Acoust. Soc. Am.* 41, 1526–1533.
- Plyler, P.N., Ananthanarayan, A.K., 2001. Human frequency following Responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. *J. Am. Acad. Audiol.* 12, 423–533.
- Reale, R.A., Geisler, C.D., 1980. Auditory-nerve fiber encoding of two-tone approximations to steady-state vowels. *J. Acoust. Soc. Am.* 67, 891–902.
- Recio, A., Rhode, W.S., 2000. Representation of vowel stimuli in the ventral cochlear nucleus of the chinchilla. *Hear. Res.* 146, 167–184.
- Rhode, W.S., 1994. Temporal coding of 200% amplitude modulated signal in the ventral cochlear nucleus of cat. *Hear. Res.* 77, 43–68.
- Rhode, W.S., 1998. Neural encoding of single-formant stimuli in the ventral cochlear nucleus of the chinchilla. *Hear. Res.* 117, 39–56.
- Ritsma, R.J., 1967. Frequencies dominant in the perception of the pitch of complex tones. *J. Acoust. Soc. Am.* 42, 191–198.
- Rose, J.E., 1980. Neural correlates of some psychoacoustical experiences. In: McFadden, D. (Ed.), *Neural Mechanisms of Behavior*. Springer, New York, pp. 1–33.
- Sachs, M.B., Voight, H.F., Young, E.D., 1983. Auditory nerve representation of vowels in background noise. *J. Neurophysiol.* 50, 27–45.
- Schouten, M., 1985. Identification and discrimination of sweep tones. *Percept. Psychophys.* 37, 369–376.
- Shore, S.E., Cullen, J.K., 1984. Cochlear microphonic responses of the peripheral auditory system to frequency-varying signals. *Am. J. Otolaryngol.* 5, 34–42.
- Shore, S.E., Nuttall, A.L., 1985. High synchrony compound action potentials evoked by rising frequency swept tonebursts. *J. Acoust. Soc. Am.* 78, 1286–1295.
- Shore, S.E., Clopton, B.M., Yolande, N.A., 1987. Unit responses in ventral cochlear nucleus reflect cochlear coding of rapid frequency sweeps. *J. Acoust. Soc. Am.* 82, 471–478.
- Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency following responses: evidence for the locus of brainstem sources. *Electroencephalogr. Clin. Neurophysiol.* 39, 465–472.
- Srulovicz, P., Goldstein, J.L., 1983. A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *J. Acoust. Soc. Am.* 73, 1266–1276.
- Terhardt, E., 1973. Pitch, consonance and harmony. *J. Acoust. Soc. Am.* 55, 1061–1069.
- 't Hart, J., 1974. Discriminability of the size of pitch movements in speech. *Inst. Percept. Res. Annu. Prog. Rep.* 9, 56–63.
- Van Lancker, D., Fromkin, V., 1973. Hemispheric specialization for pitch and tone: evidence from Thai. *J. Phon.* 1, 101–109.
- Voigt, H.F., Sachs, M.B., Young, E.D., 1982. Representation of whispered vowels in the discharge patterns of auditory-nerve fibers. *Hear. Res.* 8, 49–58.
- Wang, Y., Jongman, A., Sereno, J., 2001. Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain Language* 78, 332–348.
- Whitfield, I.C., 1970. Neural integration and pitch perception. In: Anderson, P., Jansen, P. (Eds.), *Excitatory Synaptic Mechanisms*. Universitetsforlaget, Oslo, pp. 277–285.
- Worden, F.G., Marsh, J.T., 1968. Frequency following (microphonic-like) neural responses evoked by sound. *Electroencephalogr. Clin. Neurophysiol.* 25, 45–52.
- Xu, Y., 1997. Contextual tonal variations in Mandarin. *J. Phon.* 25, 61–83.
- Young, E.D., Sachs, M.B., 1979. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J. Acoust. Soc. Am.* 66, 1381–1403.