

# Applications of Static and Dynamic Iterated Rippled Noise to Evaluate Pitch Encoding in the Human Auditory Brainstem

Jayaganesh Swaminathan, *Student Member, IEEE*, Ananthanarayan Krishnan, Jackson T. Gandour\*, and Yisheng Xu

**Abstract**—This paper presents a new application of the dynamic iterated rippled noise (IRN) algorithm by generating dynamic pitch contours representative of those that occur in natural speech in the context of EEG and the frequency following response (FFR). Besides IRN steady state and linear rising stimuli, curvilinear rising stimuli were modeled after pitch contours of natural productions of Mandarin Tone 2. Electrophysiological data on pitch representation at the level of the brainstem, as reflected in FFR, were evaluated for all stimuli, static or dynamic. Autocorrelation peaks were observed corresponding to the fundamental period ( $\tau$ ) as well as spectral bands at the fundamental and its harmonics for both a low and a high iteration step. At the higher iteration step, both spectral and temporal FFR representations were more robust, indicating that both acoustic properties may be utilized for pitch extraction at the level of the brainstem. By applying curvilinear IRN stimuli to elicit FFRs, we can evaluate the effects of temporal degradation on 1) the neural representation of linguistically-relevant pitch features in a target population (e.g., cochlear implant) and 2) the efficacy of signal processing schemes in conventional hearing aids and cochlear implants to recover these features.

**Index Terms**—Auditory brainstem, cochlear implant, experience dependent plasticity, frequency following response, iterated rippled noise, pitch, signal processing.

## I. INTRODUCTION

**P**ITCH is one of the most important information-bearing parameters of species-specific vocal signals [1]. Most periodic complex sounds including speech evoke a sensation of low pitch associated with their fundamental frequency ( $f_0$ ) [2]. There is abundant evidence to suggest that pitch extraction is accomplished by an autocorrelation analysis on the neural spike trains elicited from complex auditory stimuli [3]–[6]. These neural responses are phase-locked to the dominant interval corresponding to the  $f_0$ . Using a wide variety of periodic complex sounds, the most frequent interspike interval in the auditory nerve has been shown to correspond closely to the

perceived low pitch [3], [4]. Thus, we may conclude that neural phase-locked activity related to  $f_0$  plays a dominant role in the encoding of pitch associated with complex sounds.

The scalp-recorded human frequency following response (FFR) reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem [7]–[9]. Its generator site is presumed to be the inferior colliculus. Based on what we know about neuronal responses in the auditory nerve to periodic stimuli [3], [4], any periodicities observed in the FFR are also likely to be present in single units and local ensembles. Thus, the FFR provides a noninvasive window to view neural processing of pitch at the level of the auditory brainstem.

Indeed, pitch information has been shown to be preserved in the phase-locked neural activity generating the FFR not only for steady-state complex tones [10] but also for time-varying pitch contours of Mandarin Chinese lexical tones [11]. A subsequent crosslanguage FFR study of Mandarin speech stimuli further shows that Mandarin tones elicit stronger pitch representation and smoother pitch tracking by Chinese listeners as compared to English listeners [12]. This experience-dependent effect, however, appears to occur only when the speech stimuli reflect prototypical, *curvilinear* dynamic contours representative of Mandarin tones as opposed to *linear* dynamic approximations [13]. These findings together lead us to the question of how fine-grained is this specificity for pitch encoding within the brainstem. To eliminate any potential lexical bias for native listeners, however, we need to be able to generate auditory stimuli that preserve the perception of pitch, but do not have waveform periodicity or highly modulated stimulus envelopes that are characteristic of speech stimuli.

Such auditory stimuli exist in the form of iterated rippled noise (IRN), which preserves the temporal regularity of the stimulus without having to repeat the waveform in a periodic manner. An IRN stimulus is generated using a broadband noise which is delayed and added to itself repeatedly, and therefore does not have a prominent modulated envelope [14], [15]. However, by introducing temporal regularity into the fine structure of the noise, the delay-and-add process does change the envelope structure producing a ‘ripple’ in the long-term power spectrum of the waveform. There are also autocorrelation peaks in IRN envelopes that parallel those found in the entire waveform. The perceived pitch corresponds to the reciprocal of the delay, and the pitch salience increases with the number of iterations of the delay-and-add process. IRN stimuli have been

Manuscript received August 24, 2006. Asterisk indicates corresponding author.

J. Swaminathan, A. Krishnan, and Y. Xu are with the Department of Speech Language and Hearing Sciences, Purdue University, West Lafayette, IN USA 47907-2038 USA (e-mail: rkrish@purdue.edu).

\*J. T. Gandour is with the Department of Speech Language and Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038 USA (e-mail: gandour@purdue.edu).

Digital Object Identifier 10.1109/TBME.2007.896592

used in auditory neurophysiological [16]–[18] and psychophysical [19] studies with animals. With humans, they have been employed in psychophysical studies, showing that the pitch and pitch strength of IRN stimuli can be accounted for using temporal processing models based on autocorrelation analysis [14], [15], [20], [21]. In brain imaging studies (PET, positron emission tomography; fMRI, functional magnetic resonance imaging), IRN stimuli have been exploited to show that the processing of temporal pitch begins as early as the cochlear nucleus and continues up to auditory cortex [22]–[24]. By employing sequences of IRN steady-state stimuli, it is possible to investigate the processing of temporal pitch and melody information in the *musical domain* [24]. However, because of inherent limitations of the IRN algorithm, studies have yet to be conducted on the processing of temporal pitch and melody information in the *language domain*.

All of the aforementioned IRN studies have been limited to fixed delays that produce a constant pitch or sequences thereof [24]. However, stimuli with *dynamic* spectral and temporal ripples, similar to IRN, have been employed to characterize the spectral and dynamic properties of auditory receptive fields [25]–[27], and to evaluate the auditory system's ability to extract perceptual information from a complex spectral profile [28]. The IRN algorithm was recently generalized to allow multiple time dependent delays [29] over a range of iteration steps, making it possible to detect pitch in *dynamic* iterated ripple noise by humans. However, if we are to replicate human speech perception in a language context, it is imperative that the IRN algorithm be modified to handle pitch contours that are ecologically representative of what occurs in natural speech. From the perspective of auditory neuroethology [30], [31], this adjustment of the IRN algorithm would enable us to investigate neural mechanisms underlying the processing of *pitch contours that are of linguistic relevance* comparable to those underlying the processing of behaviorally relevant sounds in other non-primate and non-human primate animals.

With regard to human perception and communication, periodic sounds produce strong pitch sensations, and the controlled variation of the period of the sound is the basis of melody in music and prosody in speech. Since communication almost always occurs against a noisy background, it is important that neural mechanisms that extract relevant features of the speech signal are relatively less susceptible to degradation. IRN stimuli can provide a noninvasive window on those neural mechanisms that detect and extract the temporal regularity of periodic and quasi-periodic sounds from a noisy background. Importantly, we can evaluate the susceptibility of these neural mechanisms to temporal degradation by systematically varying the temporal regularity of IRN stimuli.

Accordingly, the primary objective of this paper is to present preliminary investigations of three stimulus sets (steady-state, linear rising, curvilinear) using a generalized form of the IRN algorithm that can generate multiple nonlinear time varying temporal correlations. The curvilinear stimuli most closely approximate the time varying characteristics of speech sounds. Besides descriptions of the stimuli, we also present electrophysiological data on pitch representation at the level of the brainstem, as reflected in the FFR, in response to the static and dynamic IRN

stimuli. In so doing, we are able to observe fine-grain timing information in physiological responses extracted from a noisy background that underlie the encoding of pitch at the level of the human brainstem.

## II. METHODS

### A. Subjects

Three young, adult native speakers of English ( $M = 28.0$ ,  $SD = 3.6$ ) participated in the IRN steady state and IRN linear rising experiments. None of them had any significant exposure to Chinese or any other tone language. We chose native English speakers in order to replicate earlier work by Yost [15] and Denham [29]. Three young, adult native speakers of Mandarin Chinese ( $M = 31.7$ ,  $SD = 3.2$ ) participated in the IRN curvilinear rising experiment. Chinese subjects were born and raised in mainland China; classified as late-onset, medium proficiency Mandarin/English bilinguals, not having received formal instruction in English until the age of 11; and resided in the USA for at least 1 year but not more than 4 years. We chose native Chinese speakers because pitch contours are linguistically significant at the syllable level in Mandarin. Our own work has been directed most recently to electrophysiological responses elicited by Mandarin tones in speech contexts at the level of the brainstem [11], [12]. All subjects were musically naive, as determined by a questionnaire about formal training and experience, including self-rating on a musical ability scale. Their hearing sensitivity was better than 15 dB HL for octave frequencies from 500 to 8000 Hz. Subjects were enrolled at Purdue University at the time of testing. They gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

### B. Stimuli

The duration of the IRN steady state, linear rising, and curvilinear rising stimuli was fixed at 250ms with 10ms rise/fall time.

1) *IRN Steady State*: The steady state IRN stimuli were generated using the “add same network” algorithm where the delayed and added noise is added to the original waveform at each iteration or stage of the process [15]. This is represented mathematically in (1):

$$\begin{aligned} y_i(t) &= y_{i-1}(t) + gy_{i-1}(t - \tau); \text{ for } i = 1, 2, \dots, n \\ y_0(t) &= x(t), \text{ the input signal} \end{aligned} \quad (1)$$

where  $\tau$  is the delay (in ms),  $n$  is the number of iteration steps and  $g$  is the gain ( $-1 \leq g \leq 1$ ).

Steady state IRN stimuli were generated to produce a flat  $f_0$  at 250 Hz using a delay of 4 ms at 5 different iteration steps ( $n = 4, 8, 16, 32, 64$ ) (Fig. 1). The gain was set to 1.

2) *IRN Linear Rising*: The IRN algorithm has been generalized to allow time dependent delays [29]. Instead of adding a copy of the original noise back at some fixed delay, a mapping function is created so that each point can be delayed by a different amount (2):

$$\begin{aligned} y_i(t) &= y_{i-1}(t) + gy_{i-1}(t - \tau(t)); \text{ for } i = 1, 2, \dots, n \\ y_0(t) &= x(t), \text{ the input signal} \end{aligned} \quad (2)$$

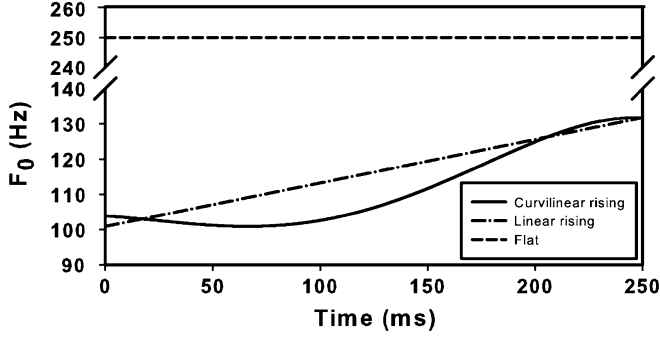


Fig. 1. Time-invariant (flat) and time-variant (linear rising, curvilinear rising)  $F_0$  contours created by means of a generalized IRN algorithm. The curvilinear rising contour represents a close approximation of the citation form of Mandarin Chinese Tone 2 from natural speech. IRN = iterated ripple noise.

Time varying IRN stimuli were generated to produce a linear rising  $F_0$  ramp at increasing iteration steps ( $n = 4, 8, 16, 32, 64$ ) (Fig. 1). The gain was set to 1.  $F_0$  onset and offset were set at 101 Hz and 132 Hz, respectively. The rate of change of  $F_0$  was 0.1233 Hz/ms.  $F_0$  offset and slope were set to equal  $f_0$  offset and average slope of Mandarin Tone 2. These values of  $F_0$  parameters yielded the closest linear approximation to the curvilinear rising  $f_0$  contour (Fig. 1). In a categorical perception experiment [32], it has been demonstrated that this linear rising ramp can be identified as Mandarin Tone 2 with a high level of accuracy.

3) *IRN Curvilinear Rising*: Equation (3) gives the generalized IRN algorithm to generate multiple nonlinear time varying correlations:

$$\begin{aligned} y_i(t) &= y_{i-1}(t) + g y_{i-1}(t - 1/f(t)); \text{ for } i = 1, 2, \dots, n \\ y_0(t) &= x(t), \text{ the input signal} \end{aligned} \quad (3)$$

where  $f(t)$  represents a polynomial equation of any degree modeling a nonlinear  $f_0$  contour.

Time varying IRN stimuli were generated to produce a curvilinear rising  $f_0$  contour at increasing iteration steps (4, 8, 16, 32, 64) (Fig. 1). The gain was set to 1.  $F_0$  onset and offset were set at 103 Hz and 132 Hz, respectively. These stimuli were modeled after natural speech productions of Mandarin Tone 2 [33] using a fourth-order polynomial equation [34]:

$$\begin{aligned} f(t) &= 103.85 - (8.45/d)t - (76.32/d^2)t^2 \\ &\quad + (297.91/d^3)t^3 - (185.34/d^4)t^4 \end{aligned} \quad (4)$$

where  $d$  is the duration of the stimuli.

### C. Data Acquisition

Subjects reclined comfortably in an acoustically and electrically shielded booth. They were instructed to relax and refrain from extraneous body movements to minimize movement artifacts. The order of presentation of stimuli, i.e., iteration steps 4, 8, 16, 32, and 64, was randomized within subjects. All stimuli were controlled by a signal generation and data acquisition system (Tucker-Davis Technologies, System II). The stimulus files were routed through a digital to analog module and presented binaurally to each ear at 60 dB nHL at a repetition rate of 3.33/s through magnetically shielded insert earphones (Biologic, TIP-300).

For all three stimulus conditions (IRN steady state, IRN linear rising, IRN curvilinear rising), FFRs were recorded from each subject from both the right ear and the left ear. These evoked responses were recorded differentially between scalp electrodes placed on the midline of the forehead at the hairline and the linked mastoid. Another electrode placed on the mid-forehead (Fpz) served as the common ground. The inter-electrode impedances were maintained below 3000  $\Omega$ . The EEG inputs were amplified by 200 000 and band-pass filtered from 150 to 3000 Hz (6 dB/octave roll-off, RC response characteristics) for the IRN steady state stimuli and from 60 to 3000 Hz for the IRN linear rising and curvilinear rising stimuli. Each FFR response waveform represents an average of 2000 stimulus presentations over a 300-ms analysis window using a sampling rate of 25 kHz.

### D. Data Analysis

The ability of the FFR to follow the pitch trajectory in all three stimuli was evaluated by extracting the  $F_0$  contour from the grand-average FFRs using a periodicity detection short-term autocorrelation algorithm [35]. This algorithm enabled us to perform a short-term autocorrelation analysis on a number of small segments or frames extracted from the IRN and FFR signals, yielding estimates of pitch periodicity (time lag associated with the autocorrelation maximum) and pitch strength (magnitude of the normalized autocorrelation peak). Average autocorrelation magnitude (pitch strength) was derived from all the time frames in each FFR waveform.

Short-term autocorrelation functions and running autocorrelograms were computed for all three stimuli. The grand averaged FFRs were used to index variation in FFR periodicities over the duration of the response at a low ( $n = 4$ ) and high ( $n = 32$ ) iteration step. 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  $\tau$ . Thus, it represents the running distribution of all-order intervals present in the population response [3], [11].

The stimulus waveform, recorded FFRs, and spectrogram were plotted for all three stimulus conditions and FFRs at a low ( $n = 4$ ) and high ( $n = 32$ ) iteration step.

## III. RESULTS

### A. IRN Steady State

Temporal and spectral properties of the time-invariant (flat) tone stimulus (left panels) and electrophysiological response at the brainstem (right panels) at selected iteration steps are shown in Fig. 2. The spectrogram has darker bands corresponding to 250 Hz ( $1/\tau$ ) and its harmonics for the high iteration step in both the stimulus and FFR response (2nd row). In the short term autocorrelation functions, peaks at the fundamental period  $1/F_0$  (4 ms) and two other multiples of the fundamental (8, 12 ms) are observed in both the stimulus and FFR data for the low iteration step (3rd row). At the high iteration step, peaks corresponding to higher multiples of the fundamental are also present (16, 20, 24, 28 ms) in both the stimulus and FFR response. The magnitude of the autocorrelation peak reaches its maximum at 4 ms and decreases at higher multiples of the fundamental periodicity. The autocorrelograms show clearer bands of temporal

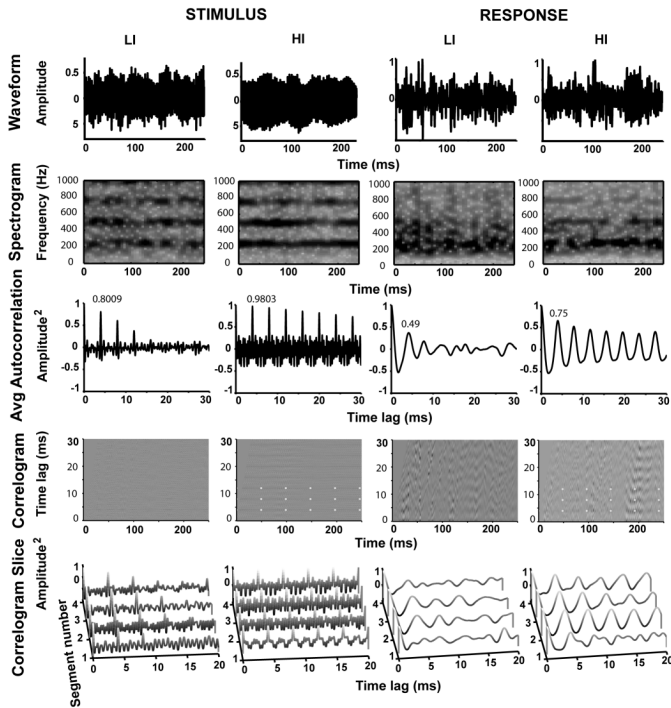


Fig. 2. Waveform, spectrogram, short term autocorrelation function, autocorrelation, and the averaged autocorrelation slice computed from the correlogram at four different time segments (#1: 50–100 ms; #2: 100–150 ms; #3: 150–200 ms; #4: 200–250 ms), in order from top to bottom, of the time-invariant, flat tone stimulus (left panels) and electrophysiological response at the brainstem (right panels) at iteration steps 4 (LI) and 32 (HI). Waveform periodicity (1st row) and resolution of spectral bands at 250 Hz and its higher harmonics are markedly improved at the high iteration step in both stimulus and FFR response (2nd row). The short term autocorrelation function displays a sharper peak at the high iteration step in both stimulus and FFR response (3rd row). Normalized peak autocorrelation values are displayed for both stimulus and response for a 4 ms period. At the high iteration step, there are sharp peaks in the FFR data corresponding to the higher periodicities of the stimuli that are not evident at the low iteration step. The autocorrelograms show clearer bands (black) of temporal regularity in the stimulus and phase-locked activity in the FFR response at the high iteration step (4th row). White dots are superimposed on the bands at every 50 ms to enhance their visualization. The correlogram slice from all four time segments displays a sharper peak at the high iteration step in both stimulus and FFR data (5th row). Note that the centers of the correlation peak align perfectly along the y-axis at 4 ms and its multiple periodicities in the FFR data at the high iteration step. FFR = fundamental frequency – following response; LI = low iteration step; HI = high iteration step.

regularity corresponding to delays of 4, 8 and 12 ms in the stimulus and phase-locked activity in the FFR response at the high iteration step (4th row). Row 5 shows averaged autocorrelations computed from the correlogram at four different time segments. The stimulus autocorrelation peaks align along a straight line at delays of  $\tau$  and  $2\tau$  for the low iteration step and  $\tau$  and higher multiples of  $\tau$  for the high iteration step. Note that there is little alignment of autocorrelation peaks at any particular lag in the FFR for the low iteration step in contrast to close alignment of peaks at lags of  $\tau$  and multiples of  $\tau$  for the high iteration step.

### B. IRN Linear Rising

Temporal and spectral properties of the time-variant linear rising tone stimulus (left panels) and electrophysiological response at the brainstem (right panels) at selected iteration steps are shown in Fig. 3. The FFR response shows greater phase

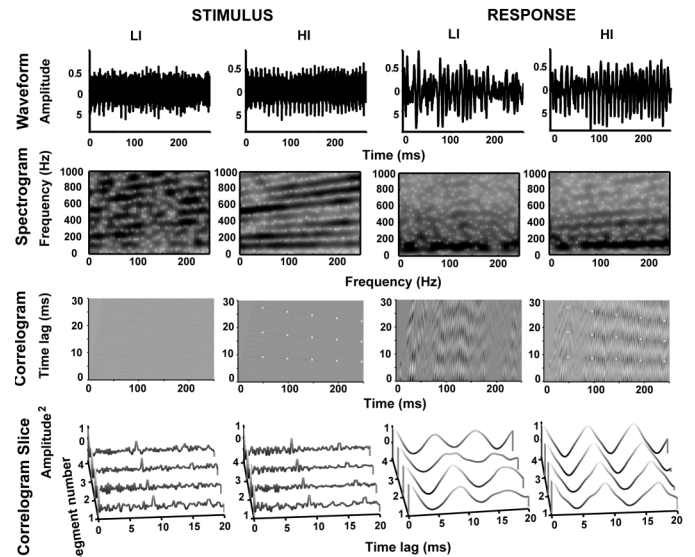


Fig. 3. Waveform, spectrogram, short term autocorrelation function, autocorrelation, and the averaged autocorrelation slice computed from the correlogram at four different time segments (#1: 50–100 ms; #2: 100–150 ms; #3: 150–200 ms; #4: 200–250 ms), in order from top to bottom, of the linear rising tone stimulus (left panels) and electrophysiological response at the brainstem (right panels) at iteration steps 4 (LI) and 32 (HI). Waveform periodicity (1st row) and resolution of linear rising spectral bands (2nd row) are markedly improved at the high iteration step in both stimulus and FFR response. The autocorrelograms show clearer bands (black) of temporal regularity in the stimulus and phase-locked activity in the FFR response at the high iteration step (3rd row). The correlogram slice from all four time segments shows a sharper peak at the high iteration step in both stimulus and FFR data (4th row). Note that the centers of the correlation peak align in a straight line with a slope along the y-axis in both the stimulus and FFR data at the high iteration step. See also caption to Fig. 2.

locked activity at the high iteration step (1st row). The spectrogram shows darker rising band in the FFR data corresponding to the high iteration step (2nd row). The average  $f_0$  value of the linear rising tone over the entire duration of the stimulus is 116 Hz, which corresponds to a delay of 8.5 ms. In the autocorrelogram of the FFR response at the high iteration step (3rd row), a time-varying band of phase locked activity (dark band descending from a time lag of about 10 to 6 ms) is observed to closely follow the decreasing fundamental period, which corresponds to increasing  $f_0$  in the linear rising tone. Note that the correlogram slice of the FFR response has better alignment and higher magnitude peaks at the high iteration step (4th row).

### C. IRN Curvilinear Rising

Temporal and spectral properties of the time-variant curvilinear rising tone stimulus (left panels) and electrophysiological response at the brainstem (right panels) at selected iteration steps are shown in Fig. 4. The FFR response shows greater phase locked activity at the high iteration step (1st row). The spectrogram shows darker bands in the FFR data at the high iteration step (2nd row). The onset  $F_0$  value of the curvilinear rising tone is 103 Hz, which corresponds to a delay of approximately 10 ms. In the autocorrelogram of the FFR response at the high iteration step (3rd row), a time-varying band of phase locked activity (dark band descending from a time lag of about 10 to 7 ms) closely follows the duration of the fundamental period,

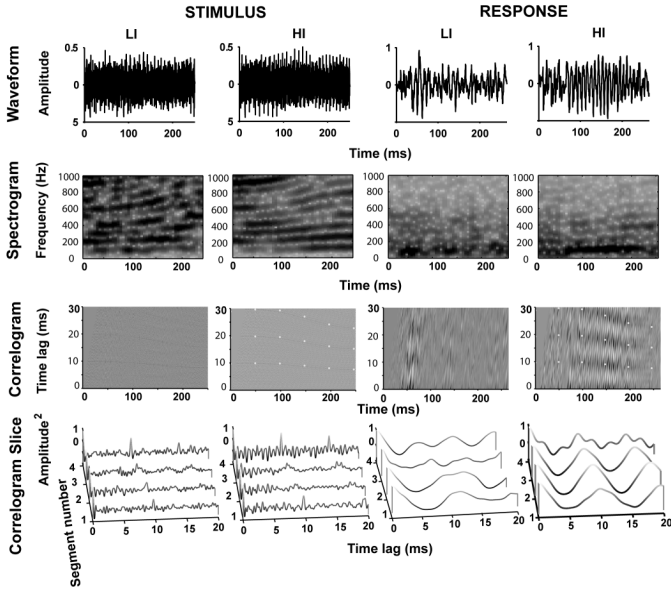


Fig. 4. Waveform, spectrogram, short term autocorrelation function, autocorrelogram, and the averaged autocorrelation slice computed from the correlogram at four different time segments (#1: 50–100 ms; #2: 100–150 ms; #3: 150–200 ms; #4: 200–250 ms), in order from top to bottom, of the curvilinear rising tone stimulus (left panels) and electrophysiological response at the brainstem (right panels) at iteration steps 4 (LI) and 32 (HI). Waveform periodicity (1st row) and resolution of curvilinear spectral bands are markedly improved at the high iteration step in both stimulus and FFR response (2nd row). The autocorrelograms show clearer bands (black) of temporal regularity in the stimulus and phase-locked activity in the FFR response at the high iteration step (3rd row). Correlogram slices show sharper peaks at the high iteration step in both stimulus (segments 1 and 4) and FFR response (segments 1, 2 and 3) (4th row). See also caption to Fig. 2.

which corresponds to falling/rising  $f_0$  movement in the curvilinear rising tone. Correlogram slices show sharper peaks at the high iteration step in both stimulus (segments 1 and 4) and FFR response (segments 1, 2 and 3) (4th row). Note that the peaks are not sharper in segments 2 and 3 of the stimulus, i.e., those segments in which we observe the larger changes in slope.

#### IV. DISCUSSION

This novel generalization of the IRN algorithm makes it possible to generate time-variant, dynamic *curvilinear* pitch contours that are representative of those that occur in natural speech. Herein we applied  $f_0$  polynomial modeling of Mandarin Tone 2 to demonstrate that our modification of the IRN algorithm can generate nonlinear pitch contours that are representative of those that occur in languages of the world. Listed below are the key mathematical steps that can be extended to create  $n$  iterations of dynamic IRN stimuli in general.

Let  $x[n] = [x_1, x_2, x_3 \dots x_n]$  denote the vector of wideband noise.

Let  $f[n] = [f_1, f_2, f_3 \dots f_n]$  denote the vector of  $f_0$  pitch contour [e.g., Mandarin Tone 2, (4)].

Let  $\tau[n] = [\tau_1, \tau_2, \tau_3 \dots \tau_n]$  denote the delays at different time instances computed as

$$[1/f_1, 1/f_2, 1/f_3 \dots 1/f_n].$$

The delayed vector of wideband noise is represented as  $x^t[n] = [x_1 - \tau_1, x_2 - \tau_2, x_3 - \tau_3 \dots x_n - \tau_n]$ .

One iteration of dynamic IRN is created as  $y[n] = x[n] + x^t[n]$ .

In contrast, almost all previous applications of IRN stimuli in psychoacoustic [15], [19]–[21], physiological [16]–[18] and cortical imaging [22], [24] experiments have used fixed delays that produce a constant pitch. Such steady-state stimuli are of limited value when investigating pitch contours that are linguistically relevant. Denham’s modification of the IRN algorithm [29] enable us to generate time variant delays that yield falling and rising  $f_0$  contours. Linear dynamic pitch trajectories have been demonstrated to evoke different perceptual responses depending on the language background of the listener [36]. But electrophysiological responses at the level of the brainstem reveal that language experience influences the processing of dynamic curvilinear [12], but not linear [13], pitch contours. Thus, only if we evaluate the neural encoding of both static and dynamic (linear, curvilinear) pitch cues at different subcortical and cortical levels along the auditory pathway are we able to develop a comprehensive profile of pitch processing from the cochlea to the cortex.

Our electrophysiological data, as reflected in the FFR, show that both static and dynamic IRN stimuli can elicit fine-grained measures of pitch representation at the level of the brainstem. In so doing, we are able to observe fine-grain timing information in physiological responses extracted from a noisy background that underlie the encoding of pitch at the level of the human brainstem. Our preliminary results from all three stimulus sets show autocorrelation peaks corresponding to  $\tau$  for both a low and high iteration. FFR responses at the higher iteration are more robust and temporally more accurate not only at  $\tau$  but also at multiples of  $\tau$ . In agreement with earlier perceptual [[15], [19]–[21]] and physiological [16]–[18] data, our FFR responses correspond to a delay equal to  $\tau$ . Thus, it appears that the ensemble phase locked neural activity reflected in the FFR is able to preserve not only information about pitch and pitch tracking accuracy, but also information about the improvement of pitch salience as temporal regularity is increased from lower to higher iteration steps.

In addition to temporal properties in the FFR responses, we observe robust spectral bands in the lower harmonics at higher iteration steps. These spectral bands correspond to the presence of spectral ripples in the IRN stimuli, which leads us to infer that spectral properties of the response must also play an important role in the extraction of pitch relevant information at the level of the rostral brainstem. Indeed, the FFR spectrograms in our data appear similar to the computed midbrain output to a spectrotemporally modulated noise [25].

In this experiment, however, we are unable to tease apart the temporal and spectral contributions to the pitch representation reflected in our FFR data because spectral components were not restricted to the unresolved spectral region. The absolute bandwidth of auditory filters increases with center frequency. When an input stimulus comprising of a tone complex undergoes a cochlear frequency analysis, lower harmonics are separated out in the cochlea and are said to be “resolved”, whereas the higher harmonics (above about the 8th harmonic) are not separated out by the cochlea and are said to be “unresolved”. Consequently, demonstration of a clear pitch percept for IRN stimuli with only unresolved spectral components would counter spectral argu-

ments for pitch processing and favor a temporal basis for pitch encoding. In the case of high-pass filtered IRN stimuli, it is unlikely that spectral ripples are resolved in the spectral region above the eighth harmonic of the reciprocal of the delay, yet IRN stimuli restricted to this spectral region produce a strong pitch percept suggesting the operation of a temporal pitch encoding scheme [[14], and references therein].

The scalp recorded FFR can provide a noninvasive neural index of representation of segmental and suprasegmental *features* of speech sounds at the brainstem level. The use of dynamic IRN stimuli described here to elicit FFRs allows us to systematically vary the temporal regularity of these acoustic features thereby enabling us to evaluate the effects of acoustic feature degradation on (i) the neural representation of these features in a target population; (ii) retraining protocols to improve neural representation; and (iii) signal processing schemes utilized in conventional hearing aids and/or cochlear implants to recover degraded representation of these features. Such dynamic stimuli enable us to evaluate the sensitivity of FFR responses to speech-like pitch contours in a parametrically controllable way without lexical-semantic confounds. They should also prove to be a useful tool for assessing the efficacy of different signal processing strategies for cochlear implants. Indeed, FFRs have already been used to assess the integrity of neural representation of certain acoustic features of consonant-vowel syllables and to evaluate the effectiveness of a retraining program for learning-disabled children [37]. With respect to cochlear implants, current speech processing algorithms extract only envelop information of narrow band signals and omit fine spectral information thereby limiting pitch encoding. This limitation is particularly detrimental for native speakers of a tonal language like Mandarin which exploits variations in pitch to signal differences in meaning at the level of the syllable [38]. Dynamic IRN stimuli representing native-like pitch contours can be used to evaluate the role of temporal and spectral cues in speech perception by cochlear implant users [39].

Although it is possible to perceive pitch changes in time-varying IRN stimuli, the ability for human subjects to discriminate dynamic changes from a non-changing stimulus is limited to rates slower than 5–10 Hz [40]. In both linear and curvilinear IRN stimuli, the time-varying correlations vary in accordance with the polynomial equations (Fig. 1). However, at the level of the brainstem, as reflected in our FFR data, pitch relevant information about the direction of pitch changes are clearly preserved (Figs. 3 and 4). Similarly, pitch sweeps are reported to be audible if the dynamic IRN is generated from sorted delays as opposed to random delays [29] In view of the ‘sluggishness’ of the auditory system, further research is warranted to determine precisely what the constraints are for using such stimuli in the study of auditory perception.

As a potential window on experience-dependent neural plasticity at the level of the brainstem, a large-scale crosslanguage (Chinese, English, Thai) study is already underway in our laboratory to determine if pitch representations in Chinese listeners are more robust in the presence of changing temporal regularity in IRN stimuli than in Thai (tone language) or English (non-tone language) listeners. Using IRN *curvilinear*  $f_0$  contours representative of the four Mandarin tones, we hypothesize that FFR

responses to the Mandarin tones presented at varying degrees of temporal regularity will emerge at earlier iteration steps and display larger magnitudes in Chinese listeners as compared to non-native listeners (Thai, English). Although crosslanguage differences in FFR responses may emerge from linguistic experience, the effects of such experience are not specific to speech perception. Our expectations are that FFR responses to IRN stimuli that contain linguistically-relevant speech parameters will show that neural mechanisms are targeting particular features of speech rather than speech per se.

#### ACKNOWLEDGMENT

The authors are very grateful for the assistance of W. A. Yost and two anonymous reviewers for their constructive comments on earlier versions of the manuscript.

#### REFERENCES

- [1] C. J. Plack, A. J. Oxenham, and R. R. Fay, “Pitch: Neural coding and perception,” in *Springer Handbook of Auditory Research*. New York: Springer, 2005, vol. 24, p. xvi, 364 p.
- [2] B. C. Moore, *Introduction to the Psychology of Hearing*, 3rd ed. London, U.K.: Academic Press, 1989.
- [3] P. A. Cariani and B. Delgutte, “Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,” *J. Neurophysiol.*, vol. 76, pp. 1698–716, Sep. 1996.
- [4] P. A. Cariani and B. Delgutte, “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.*, vol. 76, pp. 1717–34, Sep. 1996.
- [5] J. C. R. Licklider, “A duplex theory of pitch perception,” *Experientia*, vol. 7, pp. 128–134, 1951.
- [6] R. Meddis and L. O’Mard, “A unitary model of pitch perception,” *J. Acoust. Soc. Am.*, vol. 102, pp. 1811–20, Sep. 1997.
- [7] E. M. Glaser, C. M. Suter, R. Dasheiff, and A. Goldberg, “The human frequency-following response: Its behavior during continuous tone and tone burst stimulation,” *Electroencephal. Clinical Neurophysiol.*, vol. 40, pp. 25–32, Jan. 1976.
- [8] J. T. Marsh, W. S. Brown, and J. C. Smith, “Differential brainstem pathways for the conduction of auditory frequency-following responses,” *Electroencephal. Clinical Neurophysiol.*, vol. 36, pp. 415–24, Apr. 1974.
- [9] J. C. Smith, J. T. Marsh, and W. S. Brown, “Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources,” *Electroencephal. Clinical Neurophysiol.*, vol. 39, pp. 465–72, Nov. 1975.
- [10] S. Greenberg, J. T. Marsh, W. S. Brown, and J. C. Smith, “Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones,” *Hearing Res.*, vol. 25, pp. 91–114, 1987.
- [11] A. Krishnan, Y. Xu, J. T. Gandour, and P. A. Cariani, “Human frequency-following response: Representation of pitch contours in Chinese tones,” *Hearing Res.*, vol. 189, pp. 1–12, Mar. 2004.
- [12] A. Krishnan, Y. Xu, J. Gandour, and P. Cariani, “Encoding of pitch in the human brainstem is sensitive to language experience,” *Brain Res. Cogn. Brain Res.*, vol. 25, pp. 161–8, Sep. 2005.
- [13] Y. Xu, A. Krishnan, and J. T. Gandour, “Specificity of experience-dependent pitch representation in the brainstem,” *Neuroreport*, vol. 17, pp. 1601–5, Oct. 23, 2006.
- [14] R. D. Patterson, S. Handel, W. A. Yost, and A. J. Datta, “The relative strength of the tone and noise components in iterated ripple noise,” *J. Acoust. Soc. Am.*, vol. 100, pp. 3286–3294, 1996.
- [15] W. A. Yost, “Pitch of iterated rippled noise,” *J. Acoust. Soc. Am.*, vol. 100, pp. 511–8, Jul. 1996.
- [16] V. Neuert, J. L. Verhey, and I. M. Winter, “Temporal representation of the delay of iterated rippled noise in the dorsal cochlear nucleus,” *J. Neurophysiol.*, vol. 93, pp. 2766–76, May 2005.
- [17] W. P. Shofner, “Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla,” *J. Acoust. Soc. Am.*, vol. 90, pp. 2450–66, Nov. 1991.
- [18] W. P. Shofner, “Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms,” *J. Neurophysiol.*, vol. 81, pp. 2662–74, June 1999.

- [19] W. P. Shofner, "Perception of the periodicity strength of complex sounds by the chinchilla," *Hearing Res.*, vol. 173, pp. 69–81, Nov. 2002.
- [20] W. A. Yost, "Pitch strength of iterated ripple noise," *J. Acoust. Soc. Am.*, vol. 100, pp. 3329–3335, 1996.
- [21] W. A. Yost, "Pitch strength of iterated rippled noise when the pitch is ambiguous," *J. Acoust. Soc. Am.*, vol. 101, pp. 1644–8, Mar. 1997.
- [22] T. D. Griffiths, C. Buchel, R. S. Frackowiak, and R. D. Patterson, "Analysis of temporal structure in sound by the human brain," *Nature Neurosci.*, vol. 1, pp. 422–7, Sep. 1998.
- [23] T. D. Griffiths, S. Uppenkamp, I. Johnsrude, O. Josephs, and R. D. Patterson, "Encoding of the temporal regularity of sound in the human brainstem," *Nature Neurosci.*, vol. 4, pp. 633–7, June 2001.
- [24] R. D. Patterson, S. Uppenkamp, I. S. Johnsrude, and T. D. Griffiths, "The processing of temporal pitch and melody information in auditory cortex," *Neuron*, vol. 36, pp. 767–76, Nov. 14, 2002.
- [25] T. Chi, P. Ru, and S. A. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," *J. Acoust. Soc. Am.*, vol. 118, pp. 887–906, Aug. 2005.
- [26] N. Kowalski, D. A. Depireux, and S. A. Shamma, "Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra," *J. Neurophysiol.*, vol. 76, pp. 3524–34, Nov. 1996.
- [27] N. Kowalski, D. A. Depireux, and S. A. Shamma, "Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra," *J. Neurophysiol.*, vol. 76, pp. 3503–23, Nov. 1996.
- [28] W. A. Yost and M. J. Moore, "Temporal changes in a complex spectral profile," *J. Acoust. Soc. Am.*, vol. 81, pp. 1896–905, June 1987.
- [29] S. Denham, "Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model," *Biosystems*, vol. 79, pp. 199–206, Jan.–Mar. 2005.
- [30] N. Suga, M. S. Gazzaniga and E. Bizzi, Eds., "Processing of auditory information carried by complex species-specific sounds," in *The Cognitive Neurosciences*. Cambridge, MA: MIT Press, 1994, pp. 295–318.
- [31] N. Suga, X. Ma, E. Gao, M. Sakai, and S. A. Chowdhury, "Descending system and plasticity for auditory signal processing: Neuroethological data for speech scientists," *Speech Commun.*, vol. 41, pp. 189–200, 2003, 2003/8.
- [32] Y. Xu, J. Gandour, and A. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *J. Acoust. Soc. Am.*, vol. 120, pp. 1063–1074, 2006.
- [33] Y. Xu, "Contextual tonal variations in Mandarin," *J. Phonetics*, vol. 25, pp. 61–83, 1997.
- [34] Y. Xu, J. Gandour, T. Talavage, D. Wong, M. Dziedzic, Y. Tong, X. Li, and M. Lowe, "Activation of the left planum temporale in pitch processing is shaped by language experience," *Human Brain Mapping*, vol. 27, pp. 173–83, Feb. 2006.
- [35] P. Boersma and V. van Heuven, "Speak and unSpeak with PRAAT," *Glott. Int.*, vol. 5, pp. 341–347, 2001.
- [36] J. Gandour, "Tone perception in Far Eastern languages," *J. Phonetics*, vol. 11, pp. 149–175, 1983.
- [37] N. M. Russo, T. G. Nicol, S. G. Zecker, E. A. Hayes, and N. Kraus, "Auditory training improves neural timing in the human brainstem," *Behavioural Brain Res.*, vol. 156, pp. 95–103, Jan. 6, 2005.
- [38] N. Lan, K. B. Nie, S. K. Gao, and F. G. Zeng, "A novel speech-processing strategy incorporating tonal information for cochlear implants," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 752–60, May 2004.
- [39] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, pp. 303–4, Oct. 13, 1995.
- [40] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. Boston, MA: Academic, 2003.



**Jayaganesh Swaminathan** (S'04) was born in Chennai, India, in 1981. He received the B.E. degree in electrical engineering from Madurai Kamaraj University, Tamilnadu, India, in 2002, and the M.S. degree in electrical and computer engineering from the University of Rhode Island, Kingston, RI, in 2004. From 2002 to 2004, he worked in the Experimental Signal Processing Laboratory at University of Rhode Island, as a Research Assistant, implementing speech signal processing algorithms for automatic speech recognition. Presently, he is pursuing the Ph.D. degree in the Department of Speech Language Hearing Sciences at Purdue University, West Lafayette, IN.

His research interests are in the areas of auditor neuroscience, speech signal processing, cochlear implants and neural computation.



**Ananthanarayan Krishnan** received the B.Sc. degree in Speech And Hearing Science from the University of Mysore, India, the M.A. degree in audiology from the University of Memphis, and the Ph.D. degree in hearing science from the University of Texas at Dallas, in 1973, 1980, and 1983, respectively.

He is currently an Associate Professor in hearing science at Purdue University, West Lafayette, IN.

Dr. Krishnan is an active member of the Association for Research in Otolaryngology and the American Speech Language Hearing Association. He is an

assistant editor for the *Journal of the American Academy of Audiology*.



**Jackson T. Gandour** was born in Sistersville, WV, in 1941. He received the B.A. degree in political science from Wheeling Jesuit University, Wheeling, WV, the M.A. degree in linguistics from the University of Pittsburgh, Pittsburgh, PA, and the Ph.D. degree in linguistics from the University of California, Los Angeles, in 1963, 1968, and 1976, respectively.

He was a Postdoctoral Fellow at Bell Laboratories, Murray Hill, New Jersey, from 1976 to 1977. He joined the faculty of the Department of Speech Language Hearing Sciences at Purdue University, West Lafayette, IN, in 1977, and is currently a Full Professor. He served in the Peace Corps in Thailand from 1964 to 1966. He was a visiting Fulbright Lecturer at Niigata University, Japan, in 1969, and a Senior Fulbright Research Scholar at Mahidol University, Bangkok, Thailand, in 1988–1989. He was Interim Head of the Department of Speech Language Hearing Sciences from 1993 to 1995, and Head from 1995 to 1998. He has published 147 journal articles and 17 book sections. He currently serves on the editorial board of *Brain and Language* and *Aphasiology*. His research interests focus primarily on the effects of language experience in pitch processing at cortical and subcortical levels of the human brain.

Prof. Gandour is a member of the Cognitive Neuroscience Society and the Organization for Human Brain Mapping.

**Yisheng Xu**, photograph and biography not available at the time of publication.