Perceptual and Motor Skills, 2010, 111, 3, 765-784.

EARLY MATURATION OF FREQUENCY-FOLLOWING RESPONSES TO VOICE PITCH IN INFANTS WITH NORMAL HEARING[1]

FUH-CHERNG JENG, ELIZABETH A. SCHNABEL, BRENDA M. DICKMAN, JIONG HU, AND XIMING LI

School of Rehabilitation and Communication Sciences
Ohio University

CHIA-DER LIN AND HSIUNG-KWANG CHUNG

Department of Otolaryngology-HNS
China Medical University Hospital, Taiwan

Summary.—Neural plasticity of pitch processing mechanisms at the human brainstem, as reflected by the scalp-recorded frequency-following response (FFR) to voice pitch, has been reported for normal-hearing adults. Characteristics and maturation of such a response during the first year of life have remained unclear. The purpose of this study was to examine the characteristics of FFR to voice pitch in normal-hearing infants and to make a direct comparison with adults using the same stimulus and recording parameters. 9 infants and 9 adults were recruited. A Chinese monosyllable that mimics the English vowel /i/ with a rising pitch was used to elicit the FFR to voice pitch. The results demonstrated that infant FFRs showed slightly larger Pitch Strength but comparable Frequency Error, Slope Error, and Tracking Accuracy to those obtained from adults. Early maturation of FFRs was also observed in the infants starting from 1 to 3 mo. of age.

Voice pitch carries important cues for speech perception in human languages. Objective measures of the brainstem's processing of voice pitch are particularly important because auditory comprehension and language acquisition occur very early in life. For individuals who can provide reliable behavioral feedback, such as normal-hearing adults, the listener's processing of voice pitch can be examined through behavioral testing. For individuals who are unable to provide reliable behavioral feedback, such as infants, examination of a listener's processing of voice pitch poses a great challenge. Traditionally, infants' processing of certain acoustic features of speech stimuli has been evaluated through behavioral measurements, such as sucking behavior and eye gaze patterns. To date, there have been a number of studies suggesting that distinctive patterns of voice pitch can be detected and produced relatively early in life. Bever, Fodor, and Weksel (1965) stated that infants effectively master a variety of voice pitch patterns before they learn any words at all. Kaplan (1970) showed that normal patterns of falling and rising voice pitch can be discriminated

[1]Address correspondence to Fuh-Cherng Jeng, School of Rehabilitation and Communication Sciences, Ohio University, Athens, OH 45701-2979 or e-email (jeng@ohio.edu).

DOI 10.2466/10.22.24.PMS.111.6.765-784 ISSN 0031-5125

by infants at about 8 mo. of age. Crystal (1973) and Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992) also demonstrated that the time period most likely showing emergence of voice pitch perception occurs at approximately 6 mo. of age. Behavioral measurements, however, are limited by an infant's overall maturity and ability to participate actively throughout the duration of testing. Given the possibility that behavioral measurements underestimate infants' maturation of voice-pitch perception, it is hypothesized that infants' perception of voice pitch is mature at a stage earlier than 6 mo. of age.

Electrophysiological measurements have an important advantage over behavioral tests in that electrophysiological techniques can be used to evaluate the voice-pitch processing mechanisms in infants and do not require their active participation. Several neurophysiological responses such as Auditory Brainstem Responses (ABR; Hecox & Galambos, 1974; Mochizuki, Go, Ohkubo, Tatara, & Motomura, 1982; Salamy, 1984; Stapells & Mosseri, 1991), Auditory Evoked Middle Latency Responses (AMLR; Kraus, Reed, Smith, Stein, & Cartee, 1987; Fifer & Sierra-Irizarry, 1988), Auditory Evoked Late Latency Responses (ALLR; Ohlrich & Barnet, 1972; Barnet, Ohlrich, Weiss, & Shanks, 1975; Wunderlich, Cone-Wesson, & Shepherd, 2006), and Auditory Steady-State Responses (ASSR; Aoyagi, Kiren, Furuse, Fuse, Suzuki, Yokota, *et al.*, 1994; Levi, Folsom, & Dobie, 1995; John, Brown, Muir, & Picton, 2004) have been used and reported in infant testing. These responses are typically used to examine the brain's response to certain features of sound. For example, the first three alone (ABR, AMLR, and ALLR) have been used to assess the brain's responses to the onset and offset of a stimulus token, and ASSR has been used to study the brain's responses to the steady-state portion of sound. However, none of the above-mentioned neurophysiological measures reflects the listener's following of the voice pitch of speech sounds.

Recording the frequency-following response (FFR) to voice pitch in normal-hearing adults was not achieved until recently (Krishnan, Xu, Gandour, & Cariani, 2004; Dajani, Purcell, Wong, Kunov, & Picton, 2005; Aiken & Picton, 2006; Wong, Skoe, Russo, Dees, & Kraus, 2007; Song, Skoe, Wong, & Kraus, 2008; Hu & Jeng, 2009). Krishnan and associates (2004) used a set of four different Mandarin tones with different pitch contours and found that the frequency-following response to voice pitch was distinguishable at the fundamental frequency and harmonics of the stimulus. Wong and colleagues (2007) found that the frequency-following response to voice pitch was recordable in musicians and nonmusicians. All these studies indicated that, when the fundamental frequency and harmonics of the stimulus were changed over time, the phase-locked frequency-following response reflected these low frequency pitch contours. Hu and Jeng (2009) further reported the proper sensitivity and specificity of detecting the presence of a frequency-following response to voice pitch in normal-hearing adults.

Although the FFR to voice pitch has been recorded from normal-hearing adults, it remains unclear whether it is possible to measure such a response from infants. If such a response is also recordable in infants, this technique would provide an objective and noninvasive method to study pitch processing mechanisms in infants as well as to examine how the response changes with age. This technique, then, could also be used to characterize the infant FFRs as well as make direct comparisons with adults to gain insight on the point of maturity for voice pitch processing in infants. There has been only one paper published reporting the feasibility of measuring the FFR in infants. Gardi, Salamy, and Mendelson (1979) documented that infants' FFRs evoked by 20-msec. tone bursts exhibited longer latencies at 250 and 500 Hz than those recorded in adults. The data reported by Gardi and colleagues supported the concept that it is feasible to measure FFR in infants from the pitch of a tone burst. Based on literature of the FFR to voice pitch in adults and the FFR to tone bursts in neonates, it was also hypothesized that the FFR technique can be a valid method to evaluate and index the maturation of the brainstem's processing of voice pitch during the first year of life.

Method

Experimental protocols used in this study were approved by the Ohio University Institutional Review Board. All infants' guardians and adult participants signed informed consent prior to data collection. All recordings were obtained in an acoustically-attenuated and electrically-shielded sound booth in the Auditory Electrophysiology Laboratory in the School of Hearing, Speech and Language Sciences at Ohio University.

Participants

Nine infants (2 boys, M age = 5.7 mo., SD = 3.4) and nine adults (2 men, M age = 25.1 yr., SD = 2.8) were recruited. All infants were from native English-speaking households, and all adults were native speakers of English. Parents of all infants reported that their child passed a newborn hearing screening at birth (which was either an ABR or otoacoustic emission) and had no neurological or syndromic disorders. All infants also passed a distortion-product otoacoustic emissions screening (i.e., 6 dB or greater above the noise floor between the frequencies of 2000\8000 Hz) and had a type A tympanogram prior to completion of the following experiments. One of the nine infants (Infant 001) attended five testing sessions at 1, 3, 5, 7, and 10 mo. of age. The other eight infants attended one testing session. All adult participants had hearing sensitivity ≤ 20 dB HL at octave frequencies between 125 and 8000 Hz.

Stimulus Presentation

A Chinese monosyllable which mimics an English vowel /i/ with a rising pitch (117 to 166 Hz) was utilized to evoke the FFR. The speech stimulus was recorded by a male speaker and the token had a duration of 250 msec. with 10 msec. rise and fall times of the stimulus envelope. Stimulus presentation and trigger synchronization were controlled by custom-made software written in LabVIEW 8.0 (National Instruments, Austin, TX). Each acoustic stimulus was followed by a silent interval of 50 msec. before the onset of the next stimulus. This yielded a repetition rate of 3.33/sec., similar to the 3.13/sec. used by Krishnan, *et al*. (2004) and Krishnan, Xu, Gandour, and Cariani (2005). All stimulus tokens were routed through a 16-bit digital-to-analog converter (National Instruments PCI 6221 input-output control card, 40,000 samples/sec.) and low-pass filtered through a Wavetek Filter model 442 (cutoff frequency: 20 kHz, slope 24 dB/octave). The stimulus tokens were presented monaurally through an electromagnetically shielded insert earphone (Etymotics, ER-3A) at 60 dB SPL for adult participants.

To compensate for smaller ear canal volumes commonly observed in infants, the published age-appropriate RECD (real ear to coupler difference) values were used to account for the sound-pressure differences between infants and adults. Specifically, the 1-mo.-old RECD value at 250 Hz is about 5 dB larger (Feigin, Kopun, Stelmachowicz, & Gorga, 1989; Keefe, Bulen, Arehart, & Burns, 1993; Scollie, Seewald, Cornelisse, & Jenstad, 1998; Dillon, 2001) than that measured in adults. To control for these differences in stimulus intensities inside the ear canal, the stimulus token was presented at 55 dB SPL for the nine infants who participated in this study. The RECD value at 250 Hz was used because a value near 125 Hz (closer to the voice pitch of the stimulus token used in this study) was not available.

Recording Parameters

Three gold-plated recording electrodes were applied to the midline of the forehead at the hairline (noninverting), right mastoid (inverting), and left mastoid (ground) when the stimulus was presented to the right ear. For those participants whose left ear was stimulated, the inverting and ground recording electrodes were switched. All electrode impedances were kept under 3,000 Ohm and balanced within 1,500 Ohm at 10 Hz. Recordings were amplified (Neuroscan SymAmps™, 24-bit resolution, gain = 2,010, least significant bit: 0.15 nV), filtered (0.05–3500 Hz, 6 dB/octave), and digitized at a rate of 20,000 samples/sec. A transistor-transistor logic cable was used for trigger connection between the National Instrument PCI 6221 input/output control card and the Neuroscan system box.

All recordings were started at least 3 sec. before the first trigger signal and ended at least 3 sec. after the last trigger. Continuous data were obtained using Neuroscan Acquire Version 4.4 software and stored on a computer for offline analysis.

Experimental Protocol

For all infant participants, testing was conducted during regular nap time. Each infant was either held by a parent or placed in his own carrier seat. Once the infant appeared to be in a restful state, the prerecorded speech stimulus was presented via an insert headphone to the right ear or the left ear, depending on which ear was exposed after the restful state was achieved. The stimulus was initially presented at an inaudible level and slowly increased to testing level. If the infant became restless, the experimenter would wait until the infant was calm before proceeding. When the stimulus reached testing level and the infant still appeared to be in a restful state, recordings were initiated. Guardians of the infants were asked to stay in the sound booth and care for the infant as needed. Adult participants rested in a comfortable recliner with their eyes closed and were encouraged to relax during the testing procedure. Although right- versus left-ear stimulation produced similar brainstem responses (Akhoun, Gallégo, Moulin, Ménard, Veuillet, Berger-Vachon, *et al.*, 2008), right-ear advantages have been reported when recording FFRs to speech sounds (Hornickel, Skoe, Nicol, Zecker, & Kraus, 2009). In this study, acoustic stimuli were delivered monaurally to the right ear for adult participants. For infant participants, ear selection was dependent on the ear that was exposed when the infant fell into a restful state.

A control condition to evaluate the possibility of electrical interference between the stimulation equipment and recording electrodes was conducted at the final portion of each testing session for all infants and adults who participated. During the control condition, the sound tube was occluded and moved away from the participant's ear to prevent audibility of the stimulus sound.

Data Analysis

All data were analyzed using MATLAB 2008a (MathWorks, Natick, MA). To isolate better the spectral energies at the f_0 contour, continuous recordings were digitally bandpass-filtered using a brick-wall, linear-phase finite-impulse-response filter (100\1500 Hz, 500th order). The filtered continuous data were then segmented into sweeps of 300 msec. in length. An individual sweep was rejected if it contained voltages greater than ± 25 μV. A total of 2,400 sweeps (300 msec. × 2,400 sweeps = 12 min.) was collected for each participant. During each recording condition, usually less than 200 sweeps were rejected per trial. The remaining sweeps were averaged.

Cross-correlation of the stimulus and recorded waveforms was performed to identify the time shift (equivalent to the onset of a response) that produced the maximum cross-correlation value between the 3 to 10 msec. response window. Galbraith, Bagasan, and Sulahian (2001) reported that the FFR recorded via a vertical montage had a mean response latency of 4.38 msec. (ranging from 4.08 to 4.79 msec.). Russo, Nicol, Zecker, Hayes, and Kraus (2005) measured the FFR latency by finding the highest cross-correlation values within a response time shift of 6 to 9 msec. To encompass all possible response latencies, a conservative range of 3 to 10 msec. in the response window was used to identify the maximum cross-correlation value. A 250-msec. segment of the recorded waveform was extracted from the originally recorded waveform starting from the maximum cross-correlation value. The same analytical procedures were applied to all recordings obtained in the experimental and control conditions.

A narrow-band spectrogram was used to extract the pitch information of a sampled signal. All stimulus tokens and recordings were first segmented using a 50-msec. sliding Hanning window with a step size of 1 msec., which resulted in a total of 201 windowed segments to be analyzed. To increase the frequency resolution from 20 Hz to 1 Hz, each 50-msec. segment was zero-padded to 1 sec. before performing fast Fourier transform (Skoe & Kraus, 2010). In the spectrograms and pitch-tracking plots, the time shown on the abscissa was used to represent the midpoint of each 50-msec. windowed segment. For each windowed segment, this algorithm searched for the frequency corresponding to the maximal peak of the spectral amplitudes within a predefined frequency range. The frequency that corresponded to the maximal peak of the spectral amplitudes was determined as the f_0 estimate for that windowed segment. This procedure was repeated for all windowed segments. All f_0 estimates were concatenated to constitute the f_0 contour of a recording. A predefined frequency range (107\176 Hz) was used to fit with the specific pitch contour of the stimulus and allow a buffer of 10 Hz for error measurements. A buffer of 10 Hz (i.e., an extension above and below the frequency range of the stimulus f_0) was used to capture possible frequency deviations of the brainstem's responses to voice pitch. The same technique was applied to the stimulus token and recorded waveforms.

Visualization of FFR to Voice Pitch in Individual Infants

The FFR to voice pitch was visualized by plotting the spectral energies of the recordings as a function of time. Fig. 1 shows the spectrograms of the stimulus (left column) and typical spectrograms recorded from an infant (Infant 003) and an adult (Adult 003; middle column) as well as the spectrograms of the recordings obtained during the control condition (right column). Spectrograms of the stimulus showed clear energy at the

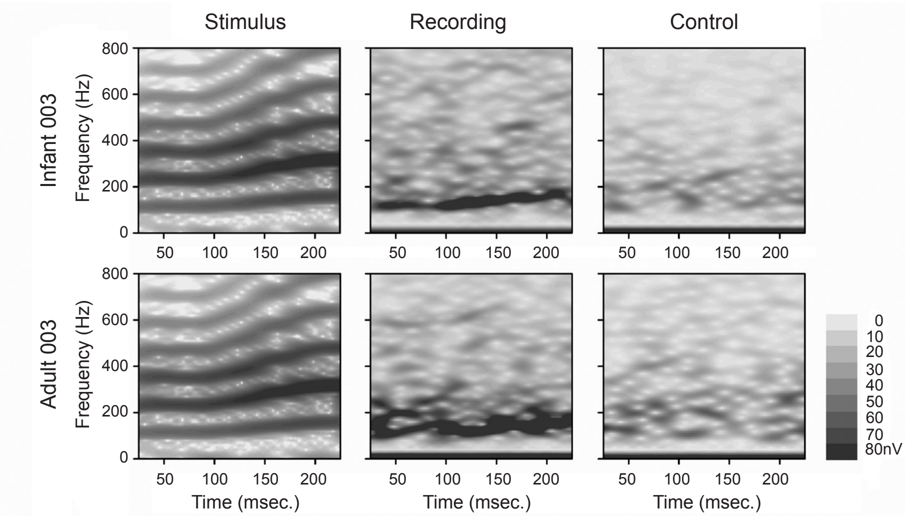

FIG. 1. Spectrograms of the stimulus (left column) and typical recordings of the frequency-following response (middle column) as well as the corresponding recordings obtained during the control condition (right column). The Control condition was conducted with the sound tube occluded and removed from the listener's ear. A gray scale on the right of the spectrograms indicates the spectral amplitude in nV for the recordings and controls obtained from the infant and adult participants. The spectrograms of the stimulus are plotted on a normalized scale ranging from 0 to 1. Spectrograms were obtained using a Hanning window of 50 msec. in length; overlap = 47.5 msec. in length; and a frequency step = 1 Hz.

fundamental frequency and its harmonics. Spectrograms of the recordings taken from the infant and adult participants showed clear FFR energy that followed the fundamental frequency of the stimulus. The FFR energy following the harmonics was not as clear. For recordings taken from individual participants, disruptions of the FFR along the fundamental frequency contour were often observed. Spectrograms of the recordings obtained in the control condition showed no FFR at the fundamental frequency or its harmonics.

Objective Measures

Four objective measures (Frequency Error, Slope Error, Tracking Accuracy, and Pitch Strength) were used to quantify the pitch-Tracking Accuracy and phase-locking magnitude of the responses. Each objective measure was meant to represent a different aspect of pitch processing in the brainstem. The four objective measures are described as follows.

Frequency Error represented the accuracy of pitch-encoding during the course of stimulus presentation. This index was computed by finding the absolute Euclidian distance between the f_0 contours of the stimuli and recordings and averaging the errors across the 201 windowed segments.

Slope Error indicated the degree to which the shapes of the pitch contours were preserved in the brainstem. This index was derived by first estimating the slope of the regression line of a stimulus f_0 contour on an f_0 -versus-time plot and then subtracting the slope estimate of the stimulus f_0 contour from that of a recording. Although Mandarin pitch contours are curvilinear, a linear regression was conducted on the entire f_0 contour to represent the degree to which the overall shape of the f_0 contour was preserved in the recording. Slope estimate of the stimulus token used in this study was 272 Hz/sec.

Tracking Accuracy (i.e., the regression *r* value) denoted the overall faithfulness of pitch tracking between the stimulus and response f_0 contours. To obtain an estimate of Tracking Accuracy, linear regression was first conducted on a recording-versus-stimulus f_0 contours plot. Regression *r* value was then denoted as the Tracking Accuracy of the recording.

Pitch Strength measured the magnitude of neural phase-locking to the f_0 contour of the stimulus waveform. This index was derived from an autocorrelation function that allowed the measurement of overall periodicity of a sampled signal. Specifically, each recording (i.e., the entire 250 msec. of a recording) was multiplied by a copy of itself with increasing time shifts. For each time shift, an autocorrelation value was calculated and expressed between −1 and 1. f_0 was calculated using the output of the autocorrelation function by finding the time shift that yielded the maximum autocorrelation value and taking the inverse of that time shift (i.e., frequency = 1/periodicity; e.g., 200 Hz = 1/5 msec.). Because the f_0 contour of the stimulus token used in this study fell within the frequency range of 100\200 Hz, the time shifts were limited to 5 to 10 msec. when searching for the location of the maximum peak in the autocorrelation output. Pitch Strength was calculated using the autocorrelation function by finding the peak-to-trough amplitude starting from the maximum positive peak (within the 5- to 10-msec. time shifts) to the following negative trough in the normalized autocorrelation output.

Statistical Analysis

The null hypothesis, that the FFR to voice pitch is immature during infancy, was tested using two-sample *t* tests for Frequency Error, Slope Error, Tracking Accuracy, and Pitch Strength for responses obtained from both infants and adults. The two-sample *t* test was conducted under the assumption of unequal population variance, after applying a Bonferroni correction factor (alpha level = 0.05/4 = 0.0125). To determine which of the Infant 001 sessions would be used for infant versus adult statistical comparisons, the mean age of the other eight infants (5.3 mo. old) was taken into consideration. Thus, Session 3 (5 mo.) was used in combination with other participants' data for all statistical comparisons throughout this article.

Results

Spectrograms of the recordings obtained in the nine infants (Infant 001 to Infant 009) are displayed in Fig. 2. Spectrograms are arranged according to the age of each infant. Numbers in parentheses indicate each infant's age in months. Infant 001 had five visits when she was 1, 3, 5, 7, and 10 mo. old, respectively. Data recorded from this infant are displayed in a separate line on the bottom of this figure. Clear FFRs in response to the voice pitch of the stimulus were observed in all recordings, evidenced by distinguishable spectral energies which followed the f_0 contour of the stimulus. The FFR was also observed on the spectrograms up to the contours of the second, third, or fourth harmonic in some recordings (e.g., Infant 001 at 3 and 5 mo. old, Infant 002, Infant 006, and Infant 008). As expected, the FFR recorded from individual listeners was not as robust as the grand-mean averages. Specifically, recordings obtained from individual listeners showed a relatively lower signal-to-noise ratio in the spectrogram, which resulted in occasional disruptions of the FFR in response to the f_0 contour of the stimulus (e.g., Infant 004 and Infant 001 at 1 mo. old).

Despite the intermittent disruption of the FFR in individual recordings, consistent results were obtained from the nine infants; that is, the FFR to voice pitch was identifiable in all nine infants starting from as young as 1 mo. old. Although Infant 001 showed a relatively weak response at 1 mo. of age, in a longitudinal follow-up of this infant the FFR to voice pitch became stronger as the infant advanced in age. This finding indicated a possible maturation trend of how the brain responds to voice pitch at the brainstem in early stages of life.

Early Maturation of Pitch Encoding in Human Brainstem

Fig. 3 presents Frequency Error (a), Slope Error (b), Tracking Accuracy (c), and Pitch Strength (d) as a function of age. Data obtained from infants (left column) were plotted in months of age and data obtained from adults (middle column) were plotted in years of age. A dashed line in each panel indicates the mean of the control recordings (for Infant 001, only Session 3, at 5 mo. old, was included in calculating the mean values). These dashed lines were treated as the noise floor of the FFR measurements in infants and adults. For clarity, data obtained from the five visits of Infant 001 were plotted using a different symbol and connected with solid lines. To illustrate better the differences between the FFR to voice pitch in infants and adults, data from the left and middle columns were replotted in the right column. Additionally, the descriptive statistics and *t* test results of the Frequency Error, Slope Error, Tracking Accuracy, and Pitch Strength are summarized in Table 1 for clarity.

A Student *t* test showed no significant difference in Frequency Error between recordings obtained from the infant and adult participants

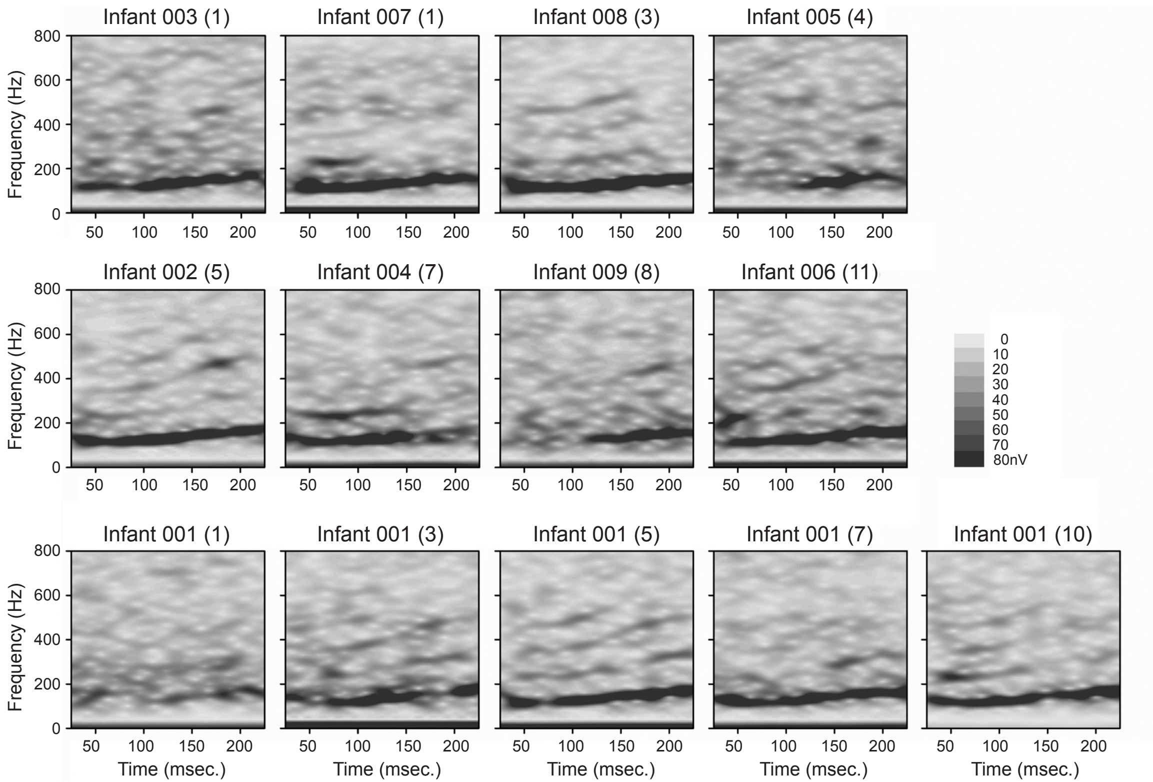

Fig. 2. Spectrograms of the nine infants' frequency-following responses to voice pitch. Longitudinal follow-ups of an infant (Infant 001) are displayed in a separate line on the bottom of this graph. Age in months is indicated by number in parentheses. Spectrograms were obtained using a Hanning window of 50 msec. in length; overlap = 47.5 msec. in length; and a frequency step = 1 Hz.

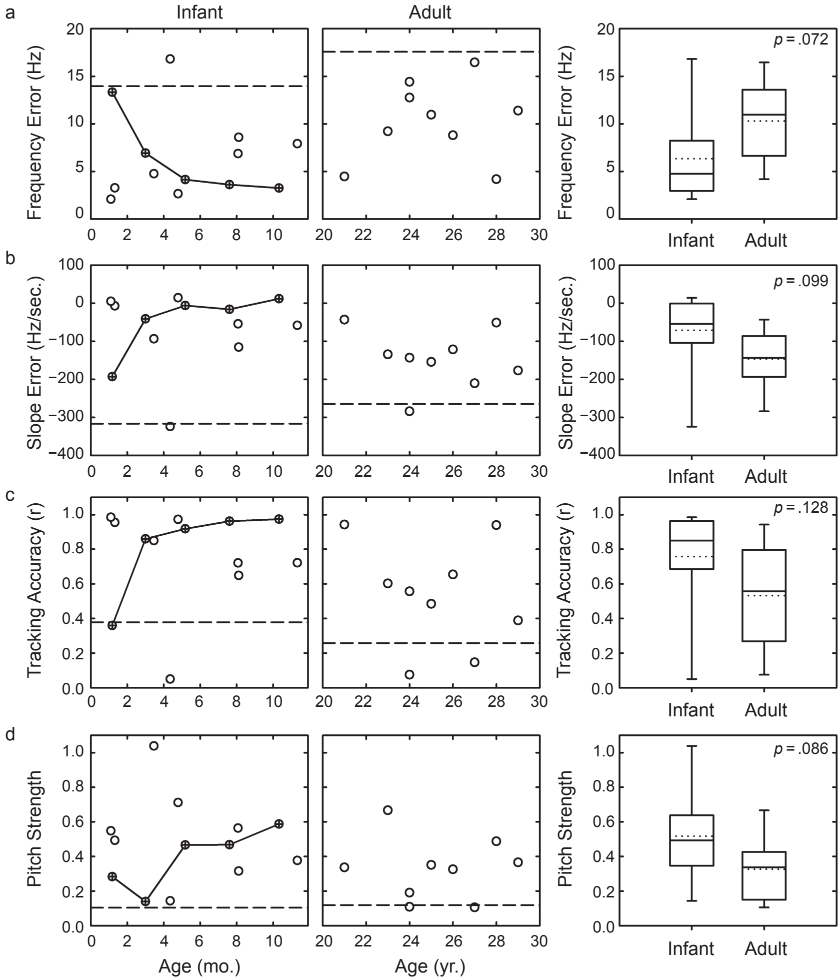

FIG. 3. Scatter plots of Frequency Error (a), Slope Error (b), Tracking Accuracy (c), and Pitch Strength (d) recorded from nine infants (left column) and nine adults (middle column). Longitudinal data obtained from an infant (Infant 001) are presented by dots with crosshairs and connected with solid lines. Dashed lines in the left and middle columns indicate the mean of the control recordings, i.e., the noise floor of the FFR measurements in infants and adults, respectively. Note different abscissa units of age are used for infants and adults. To illustrate the differences better between the two groups of participants, data presented in the left and middle columns were replotted using box plots in the right column. The upper and lower boundaries of each box indicate the 25th and 75th percentiles, respectively. The solid line within the box marks the median and the dotted line represents the mean. Whiskers above and below the box indicate the 5th and 95th percentiles, respectively. For Infant 001, only Session 3 (5 mo. old) was included in the box plots and statistical analysis. *p* values, for group comparisons between the infants' and adults' data, are indicated in the top right corner of each panel in the right column.

TABLE 1

Descriptive Statistics and *t* Test Results For the FFR to Voice Pitch Obtained From Normal-hearing Infants and Adults

Objective Measure	*M*	*SD*	Range	*t*	*p*	ES	95%*CI*
Frequency Error (Hz)							
Infants	6.33	4.29	2.08–16.81	−1.93	0.07	0.87	−8.30, 0.40
Adults	10.29	3.91	4.19–16.45				
Slope Error (Hz/sec.)							
Infants	−71.21	99.03	−324.17–14.16	1.76	0.10	0.72	−16.34, 167.32
Adults	−146.71	70.28	−283.97– −43.54				
Tracking Accuracy (r)							
Infants	0.76	0.28	0.05–0.98	1.60	0.13	0.77	−0.07, 0.52
Adults	0.53	0.29	0.08–0.94				
Pitch Strength							
Infants	0.52	0.24	0.14–1.04	1.84	0.09	0.75	−0.03, 0.41
Adults	0.32	0.17	0.10–0.67				

Note.—ES: Cohen's *d* effect size; *CI*: confidence interval.

($t = -1.87$, $p = 0.96$, $d = 0.87$). This finding indicated an early maturation of FFR to voice pitch, likely at the age of about 1 mo. old. A longitudinal follow-up of Infant 001, who showed a weak response before 1 mo. old, demonstrated a decrement trend for Frequency Error as she advanced in age. This improvement (i.e., less Frequency Error) suggested a possible maturation for infants who did not show robust responses when they were young (e.g., 1 mo.), in "catching-up" in the early stages of life (e.g., 3 mo.). Slope Error, Tracking Accuracy, and Pitch Strength showed trends similar to Frequency Error.

To compare better the spectral and temporal representations of the FFR to voice pitch in infants and adults, Fig. 4 plots the grand-averaged spectrograms (a) and time waveforms (b) recorded from the two groups of participants. In the spectrograms, both the infant and adult participants showed clear response energies at the f_0. Spectral energies at the 2nd, 3rd, and 4th harmonics, although disrupted, were also visually identifiable in the two groups of participants. Response amplitudes in nV along the f_0, 2nd, 3rd, and 4th harmonics of the stimulus for recordings obtained from infant participants showed comparable results to those obtained from adults. In the time waveforms, FFRs were not clear in the prestimulus period, but became evident after the onset of the stimulus. For Infant 001, only Session 3 data, at 5 mo. of age, were included.

Discussion

The feasibility of measuring the FFR to voice pitch in infants is indicated in the characteristics of the infant FFR and direct comparisons between the infants' and adults' FFRs using the same stimulus and recording parameters. Comparable results between these two groups allow the

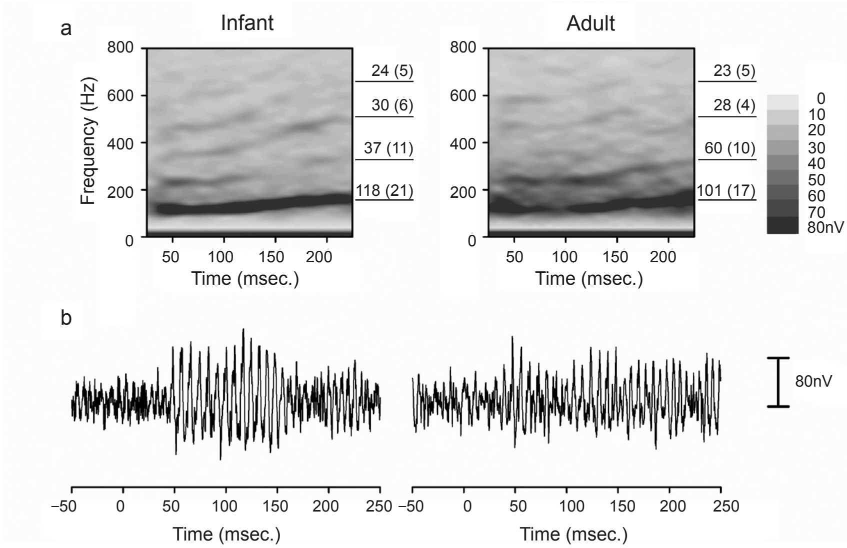

Fig. 4. Grand-averaged spectrograms (a) and time waveforms (b) for recordings obtained from normal-hearing infants and adults. A gray scale on the right of the spectrograms indicates the spectral amplitude in nV. Underlined numeric values on the right of each spectrogram represent the means and standard deviations (in parentheses) of the spectral amplitudes in nV at the fundamental frequency, 2nd, 3rd, and 4th harmonics of the grand-averaged spectrograms. Spectral amplitudes of the harmonics were determined by finding the spectral peaks closest to those of the stimulus. Spectrograms were obtained using a Hanning window of 50 msec. in length, overlap = 47.5 msec. in length, and a frequency step = 1 Hz. The response time waveforms include a 50-msec. recording prior to the stimulus onset. The ordinate scale unit of the response waveform is indicated by a vertical bar on the right.

inference that maturation of FFR to voice pitch is early. It is worthwhile to note that in a longitudinal follow-up of an infant (Infant 001) who did not show robust FFR at 1 mo. old, an improvement in pitch encoding at age 3 mo. was evident. Such findings agree with the hypothesis that pitch processing and neural encoding mechanisms mature in the human brainstem at a relatively early stage of life.

Visualization of the FFR in Individual Infants

One goal in studying the electrophysiological responses of the brain has been to develop a technique for use in evaluating the pitch-processing mechanisms of individual listeners who cannot provide reliable behavioral feedback. Recordings obtained from each individual, however, have less favorable signal-to-noise ratios than those derived from the grand-averaged waveforms. One challenge in detecting the presence of an FFR is the relatively low signal-to-noise ratio of the recordings taken from each individual. The FFR reflects a small-amplitude response, usually on the order of several hundreds of nV (Gardi, *et al*., 1979), whereas the environmental and physiological background noises are substantially greater than the FFR amplitude. To enhance the detectability and visibility of the FFR, a few procedural steps and precautions were exercised in this study. First, all waveforms were recorded in an acoustically attenuated and electrically shielded sound booth to reduce environmental noises. Second, the insert earphone and the stimulation cable were electromagnetically shielded to minimize electromagnetic leakage from the stimulation equipment to the recording cables. Finally, to visualize better the FFR on the spectrogram, a high-order bandpass filter was used to "extract" the spectral energy within the frequency region of interest (e.g., 100 to 1500 Hz). In this study, a 500-pole finite-impulse-response filter with a passband of 100 to 1500 Hz was used to reduce the spectral energy at the low (i.e., <100 Hz) and high (i.e., >1500 Hz) frequency regions, thereby to contrast better (i.e., enhance) the spectral energy at the frequency region of interest. One drawback of applying a high-order filter is that it introduces a noticeable filter delay in the output. To accommodate the 250 data-point filter delay (i.e., 12.5-msec. delay with a recording sampling rate of 20,000 samples/sec.) created by the 500-pole digital filter on the continuous data, all recordings were started at least 3 sec. before the first stimulus token was delivered to the listener's ear. The filter delay was corrected in the latency analysis of all recordings. The fact that discernible FFR was observed in each of the nine infants supports with the potential of measuring such a response in individual infants. This finding confirmed that the FFR in response to voice pitch is measurable in individual infants and is also consistent with Gardi, *et al*.'s results (1979), who recorded the FFR in response to tone bursts in neonates.

Early Maturation of FFR to Voice Pitch in Infants

Accurate encoding of voice pitch and its change over time is critical for listeners to perceive different lexical meanings and prosodic cues embedded in a speech signal. When infants are just born, they are capable of detecting differences between sounds, including the sounds they have never heard (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Also, infants exhibit a similar pattern of sound perception regardless of the language environment in which they are born (Eimas, *et al*., 1971; Kuhl, *et al*., 1992; Carral, Huotilainen, Ruusuvirta, Fellman, Naatanen, & Escera, 2005). These evidences indicate that the perception of human speech is strongly influenced by "innate factors." On the other hand, the "specific language environment" that infants are exposed to has a profound effect on the perception of speech sounds. Data show that exposure to a specific language environment during the early stages of life results in a reduction in the ability to perceive differences between speech sounds of other languages.

Kuhl and colleagues (1992) analyzed 6-mo.-old American and Swedish infants' perception of both native- and foreign-language vowel sounds and reported that the ability to hear differences among many of the sounds that are not used in the infant's language is lost by 6 mo. of age. In addition, as infants approach 12 mo. of age, after gaining experience listening to speech, they begin to comprehend the meaning of some words well before they begin to produce them (Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994). There are also a number of studies suggesting that distinctive intonation patterns can be detected and produced relatively early in life. Bever and associates (1965) stated that infants effectively master the intonation pattern of a language before they learn any words at all. Kaplan (1970) showed that normal patterns of falling and rising voice pitch can be discriminated by infants at about 8 mo. of age. Crystal (1973) concluded that 6 mo. of age is the most likely period for the emergence of voice pitch perception. Kuhl and colleagues (1992) also demonstrated that by 6 mo. of age, well before the acquisition of language production, exposure to a specific language environment alters infants' perception of phonemic lexicons. Thus, it is important to examine the neural mechanisms underpinning the early maturation of the FFR to voice pitch in infants with normal hearing.

Development and evaluation of an objective method are particularly important when clinicians and researchers are trying to apply such a technique to populations who cannot provide reliable feedback such as infants, children, and difficult-to-test patients. For example, it has been reported that the FFR to speech sounds provided an objective method for understanding pitch processing mechanisms for typically developing children and children with deficient auditory function (Russo, Nicol, Musacchia, & Kraus, 2004). In addition to its possible effect on the listener's auditory comprehension and language development, the brainstem's processing of voice pitch is also related with the listener's reading and social skills. Some children with autism spectrum disorders have shown less pitch-tracking accuracy than typically developing children (Russo, Skoe, Trommer, Nicol, Zecker, Bradlow, *et al.*, 2008). It is also reported that with short-term training on specific linguistic pitch contours, listeners not only improve their behavioral response correctness but also express enhanced pitch-tracking accuracy reflected through scalp-recorded FFRs (Song, *et al.*, 2008). Johnson, Nicol, Zecker, and Kraus (2008) also reported that preschool age children (e.g., 3 to 4 years old) exhibited less synchronized frequency-following responses to a speech token when compared with those obtained from 5- to 12-year-old school-age children. Hornickel and colleagues (2009) recently reported that the FFR in response to speech sounds was related with reading and speech-in-noise perception for 8- to 13-year-old school-age children. The human brainstem's accurate response to the changes of voice pitch, as reflected by scalp-recorded FFRs, has been reported in adults with normal hearing who spoke tonal (Gandour, Wong, & Hutchins, 1998; Krishnan, *et al.*, 2005; Swaminathan, Krishnan, & Gandour, 2008) and nontonal languages (Galbraith, Amaya, de Rivera, Donan, Duong, Hsu, *et al.*, 2004). The brainstem's accurate encoding of voice pitch and delivery of such information to the neocortex is also important for listeners to process and appreciate music. Recent studies (Musacchia, Sams, Skoe, & Kraus, 2007; Wong, *et al.*, 2007) have shown that musical training enhanced the acuity of pitch tracking in the human brainstem as reflected by scalp-recorded FFRs to voice pitch in musicians versus nonmusicians. These findings support the notion that FFR to voice pitch can be a viable, objective, and noninvasive neurophysiological index of the brainstem's processing of voice pitch. Most importantly, these findings also have potential clinical applications for diagnostic and remediation strategies for normal and pathological populations.

Despite the fact that the four objective measures (Frequency Error, Slope Error, Tracking Accuracy, and Pitch Strength) used in this study were indicative of different aspects of pitch processing in the human brainstem, similar maturational trends were observed across the four measurements. Specifically, pitch-tracking acuity and phase-locking magnitude in infants were comparable to those in adults. One may then infer an early maturation of pitch encoding in the human brainstem which likely occurs at 1 mo. of age or earlier. A longitudinal follow-up of an infant (Infant 001) who did not show robust FFRs at 1 mo. old, but clearly had caught up at 3 mo. of age, was another exciting finding. This infant, who did not show clear FFRs at 1 mo. of age, demonstrated a clear "catch-up." This improvement corroborates the idea that pitch-processing mechanisms mature during early stages of life, likely about 1 to 3 mo. of age. Although this study provides preliminary results, a larger scale of normative data with a longitudinal design will be needed to further understand the mechanisms underpinning the early maturation of the FFR to voice pitch.

With an attempt to avoid waking the infant participants from their restfulness during testing while recording distinguishable responses, in this study a moderate stimulus level of 55 dB SPL was used. Gardi, *et al.* (1979) measured FFR in response to tone bursts in neonates and showed a larger response amplitude in the 250-Hz tone burst FFR for stimulus levels at 55 dB nHL or higher. The present moderate stimulus level seems to have been sufficient to elicit clear FFRs in all infant participants. It has been reported that sleep or wakefulness has little or no effect on the brainstem responses such as ABR (Jewett & Williston, 1971) and 80-Hz ASSR (Cohen, Richards, & Clark, 1991; Picton, John, Purcell, & Plourde, 2003).

Given the brainstem origin of the FFR to voice pitch (Moushegian, Rupert, & Stillman, 1973), it is anticipated that differences in the wakefulness state between infants and adults have minimal effects on the data presented in this study.

The four objective measures used in this study show similar, but slightly different response characteristics and distribution patterns. For example, Infant 005, who had a small FFR shown in Fig. 2, also had Frequency Error, Slope Error, and Tracking Accuracy that fell beyond the respective control line of the noise measurement (Fig. 3). The Pitch Strength value obtained from this infant, however, does not fall below the noise floor. Similarly, two adults had Tracking Accuracy and Pitch Strength that fell outside of the control lines. Their Frequency Error estimates, however, were within a "response" range and only one of the two adults' Slope Error fell below the noise floor. This finding indicates the potential usefulness of the four objective indices in evaluating the pitch processing mechanisms in the human brainstem. It is possible that a combination of two or more indices may yield a better separation of the distribution patterns between infants and adults. For example, one particular index (e.g., Pitch Strength ≥ 0.10 for infants) could be used as an initial screening criterion, followed by an additional one or two indexes (e.g., Frequency Error ≤ 13.94 Hz and Tracking Accuracy ≥ 0.38 for infants) as the final decision-making criteria to decide whether a certain participant's data shall be included in the group comparisons.

Clinical Implications

In this study, data are presented to document characteristics of the FFR to voice pitch from infants and adults with normal hearing. The current techniques can be used in difficult-to-test patients and may be useful as an assessment and diagnostic method in both clinical and basic research activities. Specifically, this technique can help assess the pitch-processing mechanisms at the brainstem, and help diagnose abnormal signal processing of voice pitch in patients who cannot provide reliable behavioral feedback. Data about the FFR to voice pitch in infants with normal hearing are provided in this study. These data can serve as a starting point to help patients with communication disorders, such as patients with central auditory processing disorders and hearing loss. Moreover, children with a certain period of hearing deprivation can be assessed by this response to evaluate the effect of hearing deprivation on pitch encoding at the brainstem. This response may also be used to monitor the change and progression of hearing (re)habilitation after appropriate amplification has been applied.

REFERENCES

Aiken, S. J., & Picton, T. W. (2006) Envelope following responses to natural vowels. *Audiology & Neurotology,* 11, 213-232.

Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., Collet, L., & Thai-Van, H. (2008) The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology*, 119, 922-933.

Aoyagi, M., Kiren, T., Furuse, H., Fuse, T., Suzuki, Y., Yokota, M., & Koike, Y. (1994) Effects of aging on amplitude-modulation following response. *Acta Oto-laryngologica*, 511 (Suppl.), 15-22.

Barnet, A. B., Ohlrich, E. S., Weiss, I. P., & Shanks, B. (1975) Auditory evoked potentials during sleep in normal children from ten days to three years of age. *Electroencephalography and Clinical Neurophysiology*, 39, 29-41.

Bever, T. G., Fodor, J. A., & Weksel, W. (1965) On the acquisition of syntax: a critique of "contextual generalization." *Psychological Review*, 72, 467-482.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Naatanen, R., & Escera, C. (2005) A kind of auditory "primitive intelligence" already present at birth. *European Journal of Neuroscience*, 21, 3201-3204.

Cohen, L. T., Richards, F. W., & Clark, G. M. (1991) A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. *Journal of the Acoustical Society of America*, 90, 2467-2479.

Crystal, D. (1973) Linguistic mythology and the first year of life: an edited version of the sixth Jansson Memorial Lecture. *British Journal of Disorders of Communication*, 8, 29-36.

Dajani, H. R., Purcell, D., Wong, W., Kunov, H., & Picton, T. W. (2005) Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, 52, 1614-1618.

Dillon, H. (2001) *Hearing aids*. New York: Thieme.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971) Speech perception in infants. *Science*, 171, 303-306.

Feigin, J. A., Kopun, J. G., Stelmachowicz, P. G., & Gorga, M. P. (1989) Probe-tube microphone measures of ear-canal sound pressure levels in infants and children. *Amplification and Aural Rehabilitation*, 10, 254-258.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994) Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59 (Serial No. 242).

Fifer, R. C., & Sierra-Irizarry, B. (1988) Clinical application of the auditory middle latency response. *American Journal of Otology*, 9, 47-56.

Galbraith, G. C., Amaya, E. M., de Rivera, J. M., Donan, N. M., Duong, M. T., Hsu, J. N., Tran, K., & Tsang, L. P. (2004) Brain stem evoked response to forward and reversed speech in humans. *NeuroReport*, 15, 2057-2060.

Galbraith, G. C., Bagasan, B., & Sulahian, J. (2001) Brainstem frequency-following response recorded from one vertical and three horizontal electrode derivations. *Perceptual and Motor Skills*, 92, 99-106.

Gandour, J., Wong, D., & Hutchins, G. (1998) Pitch processing in the human brain is influenced by language experience. *NeuroReport*, 9, 2115-2119.

Gardi, J., Salamy, A., & Mendelson, T. (1979) Scalp-recorded frequency-following responses in neonates. *Audiology: Journal of Auditory Communication*, 18, 494-506.

Hecox, K., & Galambos, R. (1974) Brain stem auditory evoked responses in human infants and adults. *Archives of Otolaryngology*, 99, 30-33.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009) Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 13022-13027.

Hu, J., & Jeng, F-C. (2009) An automatic algorithm for detecting human frequency-following response to voice pitch. In *Abstracts of American Auditory Society Annual Meeting*, Scottsdale, AZ, March 5–7. P. 20.

Jewett, D., & Williston, J. (1971) Auditory-evoked far fields averaged from the scalp of humans. *Brain*, 94, 681-696.

John, M. S., Brown, D. K., Muir, P. J., & Picton, T. W. (2004) Recording auditory steady-state responses in young infants. *Ear and Hearing*, 25, 539-553.

Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008) Developmental plasticity in the human auditory brainstem. *Journal of Neuroscience*, 28, 4000-4007.

Kaplan, E. (1970) Intonation and language acquisition. *Papers and Reports on Child Language Development*, 1, 1-21.

Keefe, D. H., Bulen, J. C., Arehart, K. H., & Burns, E. M. (1993) Ear-canal impedance and reflection coefficient in human infants and adults. *Journal of the Acoustical Society of America*, 94, 2617-2638.

Kraus, N., Reed, D., Smith, D. I., Stein, L., & Cartee, C. (1987) High-pass filter setting affects the detectability of MLRs in humans. *Electroencephalography and Clinical Neurophysiology*, 62, 234-236.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25, 161-168.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, 189, 1-12.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.

Levi, E. C., Folsom, R. C., & Dobie, R. A. (1995) Coherence analysis of envelope-following responses (EFRs) and frequency-following responses (FFRs) in infants and adults. *Hearing Research*, 89, 21-27.

Mochizuki, Y., Go, T., Ohkubo, H., Tatara, T., & Motomura, T. (1982) Developmental changes of brainstem auditory evoked potentials (BAEPs) in normal human subjects from infants to young adults. *Brain Development*, 4, 127-136.

Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973) Laboratory note: scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology*, 35, 665-667.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 15894-15898.

Ohlrich, E. S., & Barnet, A. B. (1972) Auditory evoked responses during the first year of life. *Electroencephalography and Clinical Neurophysiology*, 32, 161-169.

Picton, T. W., John, M. S., Purcell, D. W., & Plourde, G. (2003) Human auditory steady-state responses: the effects of recording technique and state of arousal. *Anesthesia and Analgesia*, 97, 1396-1402.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004) Brainstem responses to speech syllables. *Clinical Neurophysiology*, 115, 2021-2030.

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005) Auditory training improves neural timing in the human brainstem. *Behavioral Brain Research*, 156, 95-103.

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008) Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, 119, 1720-1731.

Salamy, A. (1984) Maturation of the auditory brainstem response from birth through early childhood. *Journal of Clinical Neurophysiology*, 1, 293-329.

Scollie, S. D., Seewald, R. C., Cornelisse, L. E., & Jenstad, L. M. (1998) Validity and repeatability of level-independent HL to SPL transforms. *Ear and Hearing*, 19, 407-413.

Skoe, E., & Kraus, N. (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear and Hearing*, 31, 302-324.

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, 20, 1892-1902.

Stapells, D. R., & Mosseri, M. (1991) Maturation of the contralaterally recorded auditory brain stem response. *Ear and Hearing*, 12, 167-173.

Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008) Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport*, 19, 1163-1167.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10, 420-422.

Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. (2006) Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, 212, 185-202.

Accepted November 5, 2010.