Exponential Modeling of Human Frequency-Following Response to Voice Pitch

Fuh-Cherng Jeng[1], Hsiung-Kwang Chung[2], Chia-Der Lin[2], Brenda Dickman[1], Jiong Hu[1]

[1] Communication Sciences and Disorders, Ohio University, U.S.A.

[2] Department of Otolaryngology-HNS, China Medical University Hospital, Taiwan

Keywords: frequency-following response, voice pitch, exponential model, number of sweeps

Abbreviations: *f0*, fundamental frequency; FFR, frequency-following response; A_{AS}, asymptotic amplitude; A_{noise}, amplitude of noise; τ, time constant; SNR, signal-to-noise ratio; RMS, root-mean-square; ABR, auditory brainstem response

Corresponding Author:
Fuh-Cherng Jeng, M.D., Ph.D.
Grover Center W224
School of Hearing, Speech and Language Sciences
Ohio University
Athens, OH 45701
USA
Phone: (740) 593 4157
Fax: (740) 593 0287
E-mail: jeng@ohio.edu

STRUCTURED ABSTRACT

Objective: Recent studies have shown that the frequency-following response (FFR) to voice pitch can be a useful method to evaluate the signal-processing mechanisms and neural plasticity in the human brainstem. The purpose of this study was to examine the quantitative properties of the FFR trends with an exponential curve-fitting model.

Design: FFR trends obtained with increasing number of sweeps (up to 8000 sweeps) at three stimulus intensities (30, 45 and 60 dB nHL) were fit to an exponential model that consisted of estimates of background noise amplitude, asymptotic response amplitude and a time constant. Five objective indices (*Frequency Error*, *Slope Error*, *Tracking Accuracy, Pitch Strength* and *RMS Ratio*) were used to represent different perspectives of pitch processing in the human brainstem.

Study Sample: Twenty-three native speakers (16 males; age=24.7±2.1 years) of Mandarin Chinese were recruited.

Results: The results demonstrated that the exponential model provided a good fit (r^2=0.89±0.10) to the FFR trends with increasing number of sweeps for the five objective indices.

Conclusions: The exponential model, combined with the five objective indices, can be used for difficult-to-test patients and may prove to be useful as an assessment and diagnostic method in both clinical and basic research efforts.

INTRODUCTION

The scalp-recorded frequency-following response (FFR) to voice pitch has become widely accepted as a useful method for studying the signal-processing mechanisms and the neural plasticity of the human brainstem for normal and pathological populations. Recent studies have shown that neurons in the human brainstem are malleable elements and can be affected by the listener's language experience (Krishnan et al., 2005, 2009, 2010; Swaminathan et al., 2008), long-term musical training (Johnson et al., 2008; Musacchia et al., 2007; Strait et al., 2009; Wong et al., 2007) and short-term auditory training (Russo et al., 2005; Song et al., 2008). As measured through FFR, some children with autism spectrum disorders (Russo et al., 2008) and reading and spelling difficulties (Chandrasekaran et al., 2009) have shown decreased accuracy in tracking changes in voice pitch. Developmental trajectories of the FFR to voice pitch have also been described through studies in normal-hearing children (Johnson et al., 2008) and infants (Jeng et al., 2010). Given the increasing clinical utilities of the FFR to voice pitch, the amount of time needed to complete a recording becomes an important issue to address. The purpose of this study was to quantify the dependency of the FFR to voice pitch on the number of sweeps through the use of an exponential model. We recognize that the major factor affecting our study in determining the dependency of the FFR to voice pitch will be the quality of the recordings as measured by the signal-to-noise ratio (SNR) of the recordings. Because the SNR is affected by the number of sweeps averaged, we have chosen as a matter of ease and convenience to use the number of sweeps as our quantitative metric fully understand that for any individual participant a specific number of sweeps cannot guarantee a given SNR.

FFR trends with increasing number of sweeps

One challenge in recording an FFR to voice pitch is the relatively low SNR of the response waveform taken from an individual. The FFR reflects a small amplitude response, usually on the order of hundreds of nanovolts (Gardi et al., 1979; Krishnan et al., 2004; Jeng et al., 2010; Li & Jeng, 2011), whereas the background noise (physiological and non-physiological) is larger, usually in the range of 10-20 µV. Among many possible ways to improve the SNR of a recording, signal averaging is one of the most commonly used approaches in clinics and research laboratories. Signal averaging takes advantage of the time-locked feature between the onset of the stimulus and that of the computer analysis sweeps. With progressively more sweeps included in the average, the background noise (due to its nature of randomness) will be averaged toward a zero mean. In contrast, neural responses to the external auditory event are time-locked to the stimulus onset and will be summed with increasing number of sweeps. That is, in order to successfully identify an FFR, a certain number of stimulus presentations and recording sweeps will be needed to reduce the background noise to an extent such that the FFR is distinguishable from the background noise. Although it is almost impossible to pre-determine an exact number of sweeps for a given population, a reasonable range can be estimated using a *post hoc* analysis of recordings containing a large number of sweeps. Effects of the number of sweeps can then be assessed by including progressively higher numbers of sweeps into the average (e.g., comparing sub-averages of the first 500, 1000, 1500, 2000 and up to 8000 sweeps).

For conventional auditory evoked potentials such as the auditory brainstem response (ABR) to click stimuli, it has been reported that the SNR changes according to the following formula (Hood, 1998; Hall, 2006; Thornton, 2007).

$$SNR = \frac{response\ \ amplitude}{noise\ \ amplitude}\ x\ (number\ of\ \ sweeps)^{1/2} \qquad (1)$$

That is, SNR changes with the square root of the number of sweeps. SNR will increase quickly in the beginning of a recording and reach an asymptotic value when more sweeps are included. Also, according to this formula, SNR is affected by at least three factors. The first factor is the response amplitude, which is inherent to the integrity and excitability of the auditory system for each individual. Although the amplitude of a response varies between individuals (due to the properties of the individual's volume conduction and the location and orientation of the ABR sources relative to the electrode montage), response amplitude usually does not vary dramatically within an individual and is primarily affected by the intensity of the stimulus token. The second factor is the amplitude of noise, which can be physiological (e.g., ongoing brain activities that are not synchronized to the stimulus) or non-physiological (e.g., ambient noise) in origin. Note that, in deriving the SNR, the units for measuring the response and noise amplitudes are the same. The third factor is number of sweeps. As stated above, noise amplitude in the average declines with an increasing number of sweeps and, consequently, SNR is improved.

Although the square-root relationship between the amplitude of noise and the number of sweeps has been well established (Don & Elberling, 1994, 1996), in the presence of an evoked potential, response trends of a specific aspect of the evoked potential will likely deviate from the original square-root relationship, which is solely determined by the decrement of noise. Several mathematical models have been used to capture the response trends of various neural activities in the auditory system (Miller et al., 2006; Nourski et al., 2005). For example, Don and Elberling (1996) reported the usefulness of employing quantitative measures of ABR peak amplitude and residual background noise in the decision to stop averaging. Nourski and colleagues (2005) also successfully used an exponential model to describe the time course of the effects of acoustic noise on electrically evoked auditory compound action potentials in guinea pigs' auditory nerves. For responses elicited by the sustained portion of a stimulus, such as the FFR to voice pitch, it has not yet been determined if the same formula would provide a good fit to the FFR trends.

Quantitative analyses and objective indices of the FFR to voice pitch

There are two general approaches that can be used to quantify the repeating pattern (i.e., periodicity) of a sampled signal. One is to measure the strength of the overall periodicity of a sampled signal in the temporal domain by using an autocorrelation algorithm (Krishnan et al., 2005; Wong et al., 2007; Jeng et al., 2011). Briefly, this method employs an autocorrelation function that multiplies a sampled signal with a time-shifted copy of itself. Strength of the overall periodicity of the sampled signal is then determined by calculating the peak-to-trough amplitude within a certain range of time shifts in the normalized autocorrelation output. The other approach is to examine how accurately the spectral energy of a response follows the fundamental frequency (*f0*) contour of the stimulus by using a narrow-band spectrogram algorithm (Russo et al., 2008; Song et al., 2008; Jeng et al., 2011). Briefly, this algorithm analyzes the spectral components of an incoming signal by using a sliding-window technique. In this study, we used a window size of 50 ms and a step size of 1 ms when plotting the narrow-band spectrogram of a sampled signal. For each time bin (i.e., each windowed segment of the sampled signal), this algorithm searches for the frequency that contains the largest spectral density in a pre-defined frequency range. An *f0* contour of the sampled signal was then constructed by concatenating the fundamental frequencies estimated from each of the time bins. When an FFR is present, spectral components of a recording that are in close proximity to the *f0* contour of the stimulus would have relatively larger and distinguishable spectral energy than the frequency components that are further from the *f0* contour of the stimulus. Thus, small spectral energies in the frequency range around the *f0* contour of the stimulus token could be quantitatively analyzed.

To better quantify the response trends of the FFR to voice pitch from the time and frequency domains, both pitch-extraction algorithms were used. Specifically, five objective indices: *Frequency Error*, *Slope Error*, *Tracking Accuracy, Pitch Strength* and *RMS (root-mean-square) Ratio* were included (Krishnan et al., 2005; Russo et al., 2008; Skoe & Kraus, 2010; Song et al., 2009; Wong et al., 2007) to represent different aspects of pitch processing in the human brainstem. The first index, *Frequency Error*, represents a measure of the accuracy of pitch-encoding during stimulus presentation. *Slope Error* indicates the brainstem's ability to preserve the overall shape of the pitch contour of the stimulus signal. *Tracking Accuracy* reflects the overall faithfulness of pitch tracking between the stimulus and response *f0* contours. *Pitch Strength* denotes the robustness of the phase-locking phenomenon in the human brainstem. *RMS Ratio* represents the dB relationship of the RMS amplitude of a response to that of noise.

MATERIALS AND METHODS

Experimental protocols and procedures used in this study were approved by the China Medical University Hospital (Taichung, Taiwan) Institutional Review Board. All recordings were obtained in an acoustically-treated chamber in the Auditory Electrophysiology Laboratory at China Medical University Hospital.

Participants

Twenty-three adult participants (16 males; mean ± S.D. = 24.7 ± 2.1 years) with hearing sensitivity ≤ 20 dB HL at octave frequencies from 125 to 8000 Hz were recruited. All participants were native speakers of Mandarin Chinese.

Stimulus parameters and calibration

A monosyllabic Mandarin Chinese speech token /yi/, meaning *aunt*, with a rising pitch (117-166 Hz) was utilized to evoke the FFR. This stimulus token had a duration of 250 ms with 10-ms rise and fall times of the stimulus envelope. Stimulus presentation and trigger synchronization was controlled by custom-made software written in LabView 8.0 (National Instruments, Austin, TX). For each recording, the stimulus token was presented up to about 8800 times with a silent interval of 45 ms between the offset of a stimulus token and the onset of the next. All stimulus tokens were routed through a 12-bit digital-to-analog converter (National Instruments, DAQ 6062E) and low-pass filtered through a Wavetek Filter model 442 (cutoff frequency: 20 kHz, slope 24 dB/octave). The stimulus tokens were delivered monaurally via a MAICO MA42 audiometer to an electromagnetically-shielded insert earphone (Bio-logic, TIP300). Three blocks of stimulus tokens at 30, 45 and 60 dB nHL were presented in a random order across participants. Because each block of stimulus presentations was about 44 minutes (295 ms x 8800 sweeps ≈ 44 minutes), stimulus intensities greater than 60 dB nHL (approximately 75.6 dB SPL in a 2 c.c. coupler) were excluded to avoid possible damage to the listener's hearing such as temporary or permanent threshold shifts.

Stimulus intensity of the acoustic token was calibrated using both biological (i.e., dB nHL) and electro-acoustical (i.e., dB SPL) methods. The dB nHL was based on the mean behavioral threshold of a group of eight normal-hearing adults (3 males; mean ± S.D. = 24.9 ± 2.4 years). The modified Hughson-Westlake procedure (Carhart & Jerger, 1959) was used to determine the behavioral threshold for the stimulus token. Blocks of the stimuli were presented to each participant through the electromagnetically-shielded insert earphone with a step size of 5 dB at supra-threshold levels and 2 dB at near threshold levels. Each presentation block consisted of three stimulus tokens. Within each presentation block, adjacent stimulus tokens were separated by silent intervals of 45 ms (i.e., the same silent interval used in the FFR recordings). The dB SPL of the stimulus token was also measured using a Larson & Davis system 824 model sound level meter (dB flat weighting) bridged to a 2 c.c. coupler (GRAS RA0038). By using the same stimulus token and silent interval, the highest stimulus intensity (i.e., 60 dB nHL) used in this study corresponded to 75.6 dB SPL in the 2 c.c. coupler.

Recording parameters

Three gold-plated recording electrodes were applied to each participant at the high forehead along the midline below the hairline (noninverting), right mastoid (inverting), and left mastoid (ground). All electrode impedances were under 3 kOhm at 10 Hz. Recordings were amplified by an Intelligent Hearing Systems OptiAmp with a gain of 10000. This amplifier also provided an online analog bandpass filter of 10-3000 Hz at 6 dB/octave. Continuous recordings were then digitized at a rate of 20000 samples/s using a 12-bit analog-to-digital converter (National Instruments, DAQ 6062E). Continuous recordings were obtained using custom-written LabView software and stored on a computer for offline analysis.

To enhance the detectability and visibility of the FFR and minimize the contamination of stimulus artifact, 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 noise. 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 better visualize the FFR on a 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). 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 20000 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 data analysis of all recordings. It was also important to note that a control condition (i.e., sound tube plugged and removed from the listener's ear canal) had been used in our previous studies (e.g., Jeng et al., 2010, 2011) and demonstrated that the recordings were physiological in nature. For this current study, the same stimulation and recording techniques were used; however, we were unable to include the control condition due to time constraints.

Data analysis

All data were analyzed using MatLab 2008a (MathWorks, Natick, MA). To better isolate spectral energies at the *f0* contours, continuous recordings were digitally bandpass filtered using a brick-wall, linear-phase finite-impulse-response filter (cutoff frequency 100-1500 Hz, 500th order). Filtered recordings were segmented into sweeps of 295 ms in length. An individual sweep was rejected if it contained voltages greater than ±25 μV. During each recording condition, the rejection rate was less than 10% and a total of 8000 accepted sweeps were included for averaging. Recordings obtained from a distinct number of sweeps, starting from the first sweep, were averaged. The numbers of sweeps used in averaging were 1, 10, 20, 50, 100, 200, 500, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 5000, 6000, 7000 and 8000. Each averaged waveform was subject to the following analytical procedures. First, cross-correlation of the stimulus and an averaged waveform was performed to identify the time shift that produced the maximum cross-correlation value within the 3-10 ms response window (Galbraith et al., 2001; Russo, et al., 2005). Second, a 250-ms segment of the recording was extracted from the averaged waveform starting from the time shift that produced the maximum cross-correlation value. Finally, the same analytical procedures were applied to all other averaged waveforms. Data obtained from each stimulus level were analyzed separately.

Extraction of f0 contours

A narrow-band spectrogram was used to extract the pitch information of a sampled signal. All averaged recordings were first segmented using a 50-ms Hanning window with a step size of 1 ms which resulted in a total of 201 time bins to be analyzed. Each time bin was zero-padded to 1 s and provided a 1-Hz resolution in the spectrogram. For each time bin, the frequency that corresponded to the maximal peak of the spectral density was searched within a pre-defined frequency range and determined as the *f0* estimate for that time bin. This procedure was repeated for all time bins. All *f0* estimates were concatenated to constitute the *f0* contour of an averaged recording. A pre-defined 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. The same technique was applied to the stimulus token and averaged recordings.

Objective indices

Five objective measures (*Frequency Error*, *Slope Error*, *Tracking Accuracy, Pitch Strength* and *RMS Ratio*) were used to quantify the pitch-tracking accuracy and phase-locking magnitude of the responses. These objective indices are described as follows. (1) *Frequency Error* represented the accuracy of pitch-encoding during stimulus presentation. This index was computed as the absolute Euclidian distance between the stimulus and recording *f0* contours for each time bin and averaged across the 201 time bins. (2) *Slope Error* indicated the degree to which the shapes of the pitch contours were preserved in the brainstem, and was derived by subtracting the slope of the regression line of the stimulus *f0* contour from the regression slope of the recording *f0* contour. The estimated slope of the stimulus token used in this study was 275 Hz/s. (3) *Tracking Accuracy* denoted the overall faithfulness of pitch tracking between the stimulus and response *f0* contours and was calculated by finding the linear regression 'r' value on a recording-versus-stimulus f0 contours plot. (4) *Pitch Strength* measured the robustness of phase-locking in the brainstem and was derived from an autocorrelation function that allowed the measurement of overall periodicity of a sampled signal. Specifically, each 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. *f0* 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., periodicity = 1/frequency; e.g., 5 ms = 1/200 Hz). *Pitch Strength* was calculated using the autocorrelation function by finding the peak-to-trough amplitude starting from the maximum positive peak (within the 5-10 ms time shifts) to the following negative trough in the normalized autocorrelation output. Because the *f0* contour of the stimulus token used in this study fell within the frequency range of 100-200 Hz, the time shifts were limited to 5-10 ms when searching for the location of the maximum peak in the autocorrelation output. (5) *RMS Ratio* provided an estimate of the FFR amplitude relative to that of the ongoing neural activity not synchronized to the stimulus. FFR amplitude was calculated by finding the RMS amplitude of the extracted 250-ms segment of an averaged waveform. To obtain an estimate of the background physiological noise, waveforms were extracted from a 10-ms prestimulus interval to determine the amount of brainstem activities not synchronized to the stimulus. *RMS Ratio* was then calculated as the dB ratio of the FFR RMS amplitude relative to that of the noise.

Exponential curve fitting on the FFR trends of pitch-encoding

Measurements of each of the objective indices (*Frequency Error*, *Slope Error*, *Tracking Accuracy*, *Pitch Strength* and *RMS Ratio*) were analyzed as a function of number of sweeps. For *Frequency Error* and *Slope Error*, which had descending trends with increasing number of sweeps, the following model was used to describe the dependency of the FFR to voice pitch on the number of sweeps included in the averaging procedure.

$$A(n) = A_{noise}\left(e^{-n/\tau}\right) - A_{AS}, \tag{2}$$

where A is an objective measure (i.e., *Frequency Error or Slope Error*) of the FFR to voice pitch; n is the number of sweeps included in the averaging process; A_{noise} represents the amplitude of noise and is derived from the fitted curve of the FFR trend of a specific objective index when the number of sweeps equals 1 (i.e., units of A_{noise} remain the same for each of the five objective indices); A_{AS} represents the asymptotic amplitude of the response and is computed from the fitted curve of the exponential model with the number of sweeps being 8000; e is Euler's number: 2.7182; τ is the "time" constant of the fitted curve that denotes the number of sweeps needed to reach its 63% asymptotic amplitude. Calculation and derivation of the 63% asymptotic amplitude is based on the mathematical principle that the exponential function, with a base of e, is identical to its derivative (Courant & Robbins, 1996; Goldstein et al., 2009). For example, when n equals τ, an ascending exponential function with zero noise will be $A(n) = A_{AS}(1-e^{-n/\tau}) = A_{AS}(1-e^{-1}) = 0.63A_{AS}$.

For *Tracking Accuracy, Pitch Strength* and *RMS Ratio,* which had ascending trends with increasing number of sweeps, an alternative model was used to describe the response trends of these objective indices. Note A_{noise} and A_{AS} were exchanged in place due to the nature of an ascending exponential trend.

$$A(n) = A_{AS}\left(1 - e^{-n/\tau}\right) - A_{noise} \tag{3}$$

RESULTS

Temporal and spectral energies of the FFR to voice pitch were visualized by plotting the averaged time waveforms (**A**) and spectrograms (**B**) of each recording. Figure 1 shows a typical set of the FFR time waveforms and spectrograms at three different stimulus intensities. Each row represents the time waveforms and spectrograms of a recording that were averaged by including a certain number of sweeps. In this example, the FFR was difficult to distinguish from the background noise with less than 1000 sweeps for the three stimulus intensities. However, the FFR became visually identifiable when the number of sweeps was progressively increased up to about 8000 sweeps.

Quantitative analyses of the FFR to voice pitch

In order to quantify the FFR to voice pitch, responses were analyzed using the methods noted above. Figure 2 represents an example of the *f0* contour of a response (left panel) and the autocorrelation output (right panel) of a recording obtained at 60 dB nHL. This response was obtained by including 8000 sweeps in the averaging process. In terms of the accuracy of pitch tracking (left panel), the *f0* contour of the response generally followed the *f0* contour of the stimulus. In terms of the strength of phase-locking, autocorrelation output of the same recording (right panel) demonstrated overall periodicity of the recording. *Pitch Strength* of the response was calculated from the peak-to-trough amplitude starting from the positive peak (within the 5-10 ms time shifts) to the following negative trough in the normalized autocorrelation output. The response *f0* contour and autocorrelation curve seen in this figure are typical of those observed in the 23 participants across the three stimulus intensities.

FFR trends with respect to different objective indices

As the FFR to voice pitch contains enriched information of pitch-encoding mechanisms in the human brainstem, the five objective indices used in this study quantified the FFR from different perspectives. Figure 3 plots the mean *Frequency Error* (**A**), *Slope Error* (**B**), *Tracking Accuracy* (**C)** *Pitch Strength* (**D**) and *RMS Ratio* (**E**) as a function of number of sweeps.

Relatively large values of *Frequency Error* (**A**) were observed when the averaging included only a limited number of sweeps. When the number of sweeps was increased, *Frequency Error* declined dramatically and appeared to reach a steady-state, asymptotic amplitude. At 60 dB nHL, *Frequency Error* was about 18 Hz for ≤10 sweeps, declined with increasing number of sweeps, and then reached a steady-state of about 7 Hz at around 5000-8000 sweeps. *Frequency Error* at 45 and 30 dB nHL showed similar trends and declined from about 16 to 9 Hz and 18 to 11 Hz, respectively. Although three stimulus intensities showed similar values of *Frequency Error* at low numbers of sweeps (e.g., ≤10 sweeps), the *Frequency Error* estimates declined at different rates and reached different asymptotic amplitudes. It was noted that higher stimulus intensities declined faster and reached smaller asymptotic amplitudes than lower stimulus intensities. That is, when a sufficient number of sweeps (e.g., 8000 sweeps) was included in the averaging process, 60 dB nHL produced the least *Frequency Error*, followed by 45 and 30 dB nHL conditions. It should be noted that the asymptotic amplitude of the *Frequency Error*, even at the highest stimulus intensity used in this study, did not reach a mean value of zero.

Slope Error (**B**) showed similar trends to *Frequency Error*. Specifically, when the number of sweeps was increased from 1 to 8000 sweeps, *Slope Error* declined from about 310 to 190, 280 to 100 and 300 to 80 Hz/s at 30, 45 and 60 dB nHL, respectively. High stimulus intensities produced a faster decrement and reached a smaller value of *Slope Error* than low stimulus intensities. It was also noted that *Slope Error* did not reach a mean value of zero at any of the three stimulus intensities.

Tracking Accuracy (**C**) demonstrated an increasing trend when more sweeps were included. At 60 dB nHL, *Tracking Accuracy* was about 0.17 at ≤10 sweeps, which increased substantially with increasing number of sweeps and reached an asymptotic amplitude of about 0.8 at around 6000-8000 sweeps. Data obtained at 45 and 30 dB nHL showed similar trends, but the trends were increased at different rates and reached different asymptotic amplitudes. *Pitch Strength* (**D**) and *RMS Ratio* (**E**) showed similar increasing trends as *Tracking Accuracy*. To better illustrate the dependency of the FFR trends on the SNR estimates, means and standard deviations of the RMS amplitudes of the FFR (i.e., the 250-ms interval of the averaged time waveform), RMS amplitudes of the noise (i.e., the 10-ms prestimulus interval) and the corresponding *RMS Ratios* are summarized in Table 1.

Exponential modeling of the FFR trends

To better quantify the decreasing and increasing trends of pitch processing in the human brainstem as a function of number of sweeps, data obtained from the 23 participants were fit to either a descending or ascending exponential model as noted above. Figure 4 shows the exponential curves that best fit the pitch-encoding trends of *Frequency Error*, *Slope Error*, *Tracking Accuracy*, *Pitch Strength* and *RMS Ratio* at three different stimulus intensities. Equations and goodness of fit (r^2) are displayed in each panel. The exponential model used in this study provided a good fit to the pitch-encoding trends in the human brainstem, with a mean r^2 value of 0.89 and standard deviation of 0.10, across the five objective indices and three stimulus intensities. For clarity, dotted and dashed lines were used in each panel to indicate the estimates of the noise amplitude (A_{noise}) and response asymptotic amplitude (A_{AS}) of the fitted curve, respectively. The A_{noise} and A_{AS} values were calculated from the fitted curves of the FFR trends when the numbers of sweeps were 1 and 8000, respectively.

Frequency Error (Figure 4**A**) of the fitted curves had asymptotic amplitudes (i.e., A_{AS}) of 10.78, 7.23 and 6.09 Hz at 30, 45 and 60 dB nHL, respectively. The fact that *Frequency Error* reached a smaller asymptotic value at higher stimulus intensities indicated the dependency of the *Frequency Error* measurement on stimulus intensities. Estimates of the noise amplitudes (i.e., A_{noise}) of the fitted curves for *Frequency Error* were 16.02, 14.31 and 16.37 Hz for 30, 45 and 60 dB nHL, respectively. The unfavorable noise amplitudes of *Frequency Error* reflected the poor signal-to-noise ratios when only a limited number of sweeps was included in the averaging procedures. In addition to the A_{AS} and A_{noise} estimates, another important parameter in our exponential model was the τ value which indicated the number of sweeps needed to reach its 63% asymptotic amplitude of the response. The τ values of the fitted curves for *Frequency Error* were 3940, 3110 and 1578 sweeps at 30, 45 and 60 dB nHL, respectively. It is important to note that the τ values decreased with increasing stimulus intensity. This finding indicated that higher stimulus intensities (e.g., 60 dB nHL) produced a faster improvement in pitch-tracking acuity (i.e., less *Frequency Error*) in the human brainstem than lower stimulus intensities (e.g., 30 dB nHL).

Slope Error (Figure 4**B**) showed similar trends to *Frequency Error*. The asymptotic amplitudes were 182.66, 101.91 and 82.67 Hz/s at 30, 45 and 60 dB nHL, respectively. The asymptotic amplitude of *Slope Error* decreased with increasing stimulus intensity. This finding indicated a better preservation of the shape of the stimulus *f0* contour at higher stimulus intensities (e.g., 60 dB nHL). Noise amplitudes of the fitted curves for *Slope Error* were 263.45, 243.15 and 271.53 Hz/s at 30, 45 and 60 dB nHL, respectively. The unfavorable *Slope Error*s reflected the poor SNRs when only a limited number of sweeps were included. The τ values of the fitted curves for *Slope Error* were 3310 sweeps at 30 dB nHL, decreased to 2030 sweeps at 45 dB nHL and 1932 sweeps at 60 dB nHL. Note that higher stimulus intensities produced a faster improvement (i.e., less *Slope Error*) in preserving the shape of the stimulus *f0* contour in the FFR than lower stimulus intensities.

Tracking Accuracy (Figure 4**C**) of the fitted curves demonstrated an ascending trend with increasing number of sweeps. *Tracking Accuracy* of the fitted curves had asymptotic amplitudes of 0.41, 0.72 and 0.75 at 30, 45 and 60 dB nHL, respectively. Noise amplitudes of the fitted curve for *Tracking Accuracy* were 0.24, 0.30 and 0.15 at 30, 45 and 60 dB nHL, respectively. The τ values of the fitted curves were 5527 sweeps at 30 dB nHL, decreased to 2633 sweeps at 45 and 1299 sweeps at 60 dB nHL. Higher stimulus intensities produced better pitch-tracking accuracy (i.e., larger asymptotic amplitudes) at a faster rate (i.e., smaller τ values) than lower stimulus intensities.

Pitch Strength (Figure 4**D**) showed similar trends as those observed in *Tracking Accuracy*. The asymptotic amplitudes of the fitted curve for *Pitch Strength* were 0.54, 0.54 and 0.63 at 30, 45 and 60 dB nHL, respectively. Noise amplitudes of the fitted curves were 0.20, 0.27 and 0.16 at 30, 45 and 60 dB nHL, respectively. The τ values of the fitted curves for *Pitch Strength* were 9634 sweeps at 30 dB nHL, which decreased to 3737 sweeps at 45 dB nHL and 1767 sweeps at 60 dB nHL. Note that higher stimulus intensities produced a larger enhancement (i.e., larger asymptotic amplitudes) of neural phase-locking in the human brainstem at a faster rate (i.e., smaller τ values) than lower stimulus intensities. *RMS Ratio* (Figure 4**E**) showed similar trends as those observed in *Tracking Accuracy* and *Pitch Strength*.

Dependence of the FFR trends on stimulus intensity and objective index

To better illustrate the dependence of the FFR on stimulus intensity and the choice of objective indices, the asymptotic amplitude (A_{AS}), noise amplitude (A_{noise}) and τ values of the fitted curves are summarized in Table 2. As reflected from the τ values of the fitted exponential model, response trends of pitch-encoding in the human brainstem were dependent on both the stimulus intensity and the choice of objective indices. Specifically, higher stimulus intensities (e.g., 60 dB nHL) demonstrated a faster improvement in SNR (i.e., smaller τ values) with increasing number of sweeps than lower stimulus intensities (e.g., 30 dB nHL). The five objective indices were all feasible and effective in quantifying the FFR trends. Note Tracking Accuracy had the smallest τ value (1229 sweeps at 60 dB nHL) across the five objective indices and the three stimulus intensities.

DISCUSSION

As the FFR has shown its potential in basic research and clinical applications, parameters that can be used to determine the number of sweeps (i.e., amount of time) needed to obtain an FFR become important factors to examine. This study collected up to 8000 accepted sweeps and examined the response trends of the FFR to voice pitch in normal-hearing adults. Results demonstrated that the exponential model used in this study provides a good fit (r^2 value: mean ± S.D. = 0.89 ± 0.10, median = 0.93, range = 0.69-0.98) to the FFR trends. This finding supports the use of an exponential model to mathematically analyze the FFR trends.

Exponential modeling of the FFR trends of pitch-encoding in the human brainstem

It has been reported that amplitude of noise decreases with the square root of the number of sweeps (Hood 1998; Hall, 2006; Thornton, 2007). However, when a response is present in addition to noise, response trends with increasing number of sweeps may be further influenced by the physiological properties of the response. That is, in the presence of a specific neural potential, trends of the response will likely deviate from the original square-root relationship that is solely determined by the decrement of noise. Several exponential models have also been reported to track the time course of various neural activities in the auditory system (Miller et al., 2006; Nourski et al., 2005). For example, Nourski and colleagues (2005) successfully used an exponential model to describe the time course of the effects of acoustic noise on electrically evoked auditory compound action potentials in the guinea pig's auditory nerve. Our study expands the use of the exponential model to track the changes of the FFR to voice pitch with increasing number of sweeps in normal-hearing adults. Additionally, the model used in this study provides a good fit of the FFR trends across the five objective indices and three stimulus intensities. This finding supports the use of an exponential model to delineate the response trend of pitch-encoding in the human brainstem. One important advantage of utilizing an exponential model to describe the FFR trends is that it allows an objective method to mathematically analyze and compare the amplitude of such a response across different testing conditions and populations of interest. It is also important to note that, although each of the objective indices represents a different aspect of pitch processing in the human brainstem, they may not be totally independent from each other. For example, averaged recordings with lower values of *Frequency Error* will likely have lower *Slope Error* and higher *Tracking Accuracy*.

This modeling approach also improves the consistency and proficiency in selecting a pre-determined threshold criterion (or a combination of several criteria) for recording FFRs. It should be noted that a recording can be terminated for different reasons. In developing a statistical model for ABR, Don and Elberling (1996) proposed that a recording could be terminated based on several conditions: (1) when averaging has reached a point where it is possible to visually identify a "normal" response from the background noise, (2) when the targeted neural response with a given amplitude has been established, (3) when the afforded time has been exhausted before sufficient averaging occurs, (4) when a given residual averaged background noise level has been reached, and (5) when a given criterion of a quantitative detector has been achieved. Although a statistical model has not been used in the present study, successful results of the exponential curve fitting to the FFR trends permits the use of a normal quantitative analysis. The absence of an FFR would be apparent in conditions where the criterion for residual averaged background noise is met or when afforded test time is exhausted. It is also important to note that the FFR trends of the five objective indices may change if a different set of experimental parameters are used. However, one would not need to conduct their own modeling to evaluate FFR trends but could achieve the same results if FFR and noise amplitudes are comparable to the numbers listed in Table 2.

One interesting finding derived from the exponential model is that the asymptotic amplitude of the response (i.e., A_{AS}) does not reach a zero mean at any of the three stimulus intensities. For example, the A_{AS} value of *Frequency Error* is largest at 30 dB nHL and smallest at 60 dB nHL. The A_{AS} values of *Slope Error* show a similar trend to *Frequency Error*. (i.e., do not reach a zero mean at the three stimulus intensities). This finding indicates that at higher stimulus intensities (e.g., 60 dB nHL), neural responses are more synchronized to the stimulus frequency when compared to lower stimulus intensities (e.g., 30 dB nHL). This phenomenon is also observed in the A_{AS} values of *Tracking Accuracy* and *Pitch Strength*. Although the non-zero A_{AS} values of *Frequency Error*, *Slope Error*, *Tracking Accuracy* and *Pitch Strength* at 60 dB nHL can be explained by the residual errors of the less synchronized neural responses in the brainstem, it is possible that higher stimulus intensities (e.g., 80 dB nHL) will produce A_{AS} values that are very close or equal to zero. If so, the non-zero A_{AS} values observed in this study simply represent the FFR dependence on stimulus intensity. It is also possible that, even at stimulus intensities higher than 60 dB nHL, the A_{AS} values still do not reach a zero mean. In this case, the non-zero A_{AS} values would further represent the upper limits of the pitch-tracking acuity and phasing-locking phenomenon in the human brainstem.

Effects of number of sweeps and stimulus intensity

When recording an auditory evoked potential, it is necessary to identify the physiological response of small-amplitude from the relatively large-amplitude background noise. While the amplitude of a response can be enhanced by a careful design of experiments, amplitude of background noise can be reduced through a variety of techniques. The most efficient way of reducing noise is likely to eliminate the noise from its source, which can be physiological (e.g., muscle artifact) or non-physiological (e.g., environmental noise and stimulation artifact) in nature. For example, muscle artifact can be reduced by making sure that the participant is relaxed and receives appropriate head support during experiments. Environmental noise can be minimized by conducting experiments in an acoustically-isolated and electrically-treated chamber. Stimulus artifact can be minimized through careful selection of the stimulation and recording parameters as well as the use of an electromagnetically-shielded earphone. After the various sources of background noise have been eliminated, SNR of a recording can be further improved by including more sweeps in the averaging process (Hood, 1998; Hall, 2006; Thornton, 2007). In order to determine the appropriate range of number of sweeps needed to reduce the noise to an acceptable level, this study used an exponential model to examine the response trends of the FFR to voice pitch. It is found that, given the proviso of equal background noise levels, the number of sweeps needed to stop a recording is dependent on the stimulus intensity and the choice of objective indices. For example, if FFR recordings are performed at 60 dB nHL and *Tracking Accuracy* is used to signal the presence of a response, the exponential model provides the specific A_{AS}, A_{noise} and τ values that can be used to compute the number of sweeps needed to complete a recording. For example, if 75% of the response asymptotic value is desired, approximately ($1.39\ \tau = 1.39 \times 1229 = 1708$) sweeps will be needed. If 90% or more is satisfactory, approximately ($2.30\ \tau = 2.30 \times 1229 = 2827$) sweeps will be required. Similarly, if anywhere between 75-90% of the asymptotic value is set as the threshold criterion, roughly 1700 to 2800 sweeps will be needed to obtain an FFR to voice pitch. This result is consistent with the number of sweeps that are commonly used in FFR literature (Galbraith et al., 1994, 2000, 2001; Jeng et al., 2010, 2011; Krishnan, 2007; Krishnan et al., 2004, 2005, 2010; Li & Jeng, 2011; Skoe & Kraus, 2010; Song et al., 2009; Wong et al., 2007).

Dependence of the FFR to voice pitch on stimulus intensity is also observed (i.e., FFR amplitude increases with stimulus intensity). An interesting finding observed in this study is that higher stimulus intensities (e.g., 60 dB nHL) produce a faster SNR improvement than lower stimulus intensities (e.g., 30 dB nHL). This finding can be explained, at least partially, by the fact that high stimulus intensities produce better neural-firing efficiency and less temporal jitter in single neuron recordings in the auditory nerve (Miller et al., 2006; Imennov & Rubinstein, 2009) and brainstem nuclei (Keller & Takahashi, 2000; Voytenko & Galazyuk, 2008). Briefly, firing efficiency can be computed as a ratio of the number of neuronal spikes elicited and the number of times the stimulus is presented. Jitter is often considered as the temporal uncertainty of spike timing and can be calculated as the standard deviation of the spiking times. At high stimulus intensities, neurons in the brainstem will likely produce a larger number of spikes (i.e., greater firing efficiency) that are closely synchronized with the onset of the stimulus (i.e., less temporal jitter). As the scalp-recorded FFR to voice pitch requires synchronized neural responses, higher stimulus intensities will likely produce a "cleaner" response. Such clean responses from individual neurons in response to high intensities will likely build up and reveal the presence of a response more quickly than those obtained at low stimulus intensities. In the current study, all five objective indices showed a faster SNR improvement (i.e., smaller τ values) at high stimulus intensities than low stimulus intensities. This finding is consistent with the effect of stimulus intensity reported in FFR literature (Gardi et al., 1979; Krishnan & Parkinson, 2000). For example, Gardi and colleagues (1979) recorded FFRs to 10-ms tone bursts in normal-hearing adults and neonates at 25-65 dB nHL and found that the largest FFR amplitude was produced at 65 dB nHL for both the neonates and adults. Krishnan and Parkinson (2000) recorded FFRs to 80-ms frequency sweeps (400-600 Hz) at 65-95 dB nHL in normal-hearing adults and found that the largest response amplitude was produced at 95 dB nHL. Although higher stimulus intensity produces larger response amplitude, we limited our stimulus to $\leq$ 60 dB nHL (due the relatively long duration of stimulus presentation used in this study; e.g., 295 ms x 8000 sweeps ≈ 39 minutes) in order to avoid any possible damage to the listener's hearing.

Clinical Implications

Although Mandarin tones are used to elicit FFRs in this study, the exponential model could realistically be applied to any complex sound with a variable pitch contour; thus, it has utility beyond a Mandarin speaking population and can be useful to any clinician interested in obtaining an objective measurement of pitch processing in the human brainstem. It is important to note that different populations (e.g., musicians, native speakers of tonal versus non-tonal languages, normal-hearing children and infants, children with specific hearing or language disorders) and stimuli with different pitch contours (e.g., a falling pitch rather than a rising pitch) may have different response properties in pitch processing and therefore exhibit FFR trends with different A_{AS}, A_{noise} and τ values. Equations derived from these trends of a specific population can be useful in developing objective methods and experimental protocols to determine the presence of an FFR and to complete a recording by applying a pre-determined stopping criterion or a combination of them. It is anticipated that future studies focusing on examining the exponential trends of the FFR in a variety of populations will shed light on signal-processing mechanisms and neural plasticity of the human brainstem.

ACKNOWLEDGMENTS

This study was supported in part by (1) Advancing Academic-Research Career (AARC) Award from the American Speech-Language-Hearing Association, U.S.A., (2) Research Incentive Grant (DMR-99-048) from the Department of Medical Research at China Medical University Hospital, Taiwan, and (3) the Clinical Trial and Research Center of Excellence Funds (DOH100-TD-B-111-004) from Taiwanese Department of Health. Preliminary results of this study were presented at the American Auditory Society Annual Meeting, March 4-6, 2010. The authors thank Cheng-Han Chiu for his assistance in data collection.

REFERENCES

Carhart, R. & Jerger, J. F. 1959. Preferred method for clinical determination of pure-tone thresholds. J *Speech Hear Disord*, 24, 330-345.

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T. & Kraus, N. 2009. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*, 64, 311-319.

Courant, R. & Robbins, H. (eds.) 1996. *What is Mathematics? An Elementary Approach to Ideas and Methods* (2nd edition, revised by Stewart, I.). New York: Oxford University Press.

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 Trans Biomed Eng*, 52, 1614-8.

Don, M. & Elberling, C. 1994. Evaluating residual background noise in human auditory brainstem responses (ABRs). *J Acoust Soc Am*, 96, 2746-2757.

Don, M. & Elberling, C. 1996. Use of quantitative measures of auditory brain-stem response peak amplitude and residual background noise in the decision to stop averaging. *J Acoust Soc Am*, 99, 491-499.

Galbraith, G.C. 1994. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalogr Clin Neurophysiol*, 92, 321-330.

Galbraith, G.C., Amaya, E.M., de Rivera, J.M., Donan, N.M., Duong, M.T., et al. 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. *Percept Mot Skills*, 92, 99-106.

Galbraith, G.C., Threadgill, M.R., Hemsley, J., Salour, K., Songdej, N., et al. 2000. Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci Lett*, 292, 123-127.

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

Goldstein, L.J., Schneider, D.I., Lay, D.C. & Asmar, N.H. (eds.) 2009. *Calculus and Its Applications* (12th edition). New Jersey: Prentice-Hall.

Hall, J.W. 2006. *New Handbook of Auditory Evoked Responses*. Boston, MA: Allyn and Bacon.

Hood, L.J. 1998. *Clinical Applications of the Auditory Brainstem Response*. San Diego, CA: Singular Publication Group.

Imennov, N.S. & Rubinstein, J.T. 2009. Stochastic population model for electrical stimulation of the auditory nerve. *IEEE Trans Biomed Eng*, *56*(10), 2493-2501.

Jeng, F.-C., Hu, J., Dickman, B., Lin, C.-Y., Lin, C.-D., Wang, C.-Y., Chung, H.-K. & Li, X. 2011. Evaluation of two algorithms for detecting human frequency-following responses to voice pitch. *Int J Audiol*, 50(1), 14-26 (DOI:10.3109/14992027.2010.515620).

Jeng, F.-C., Schnabel, E.A., Dickman, B.M., Hu, J., Li, X., Lin, C.-D. & Chung, H.-K. 2010. Early maturation of frequency-following responses to voice pitch in infants with normal hearing. *Percept Mot Skills*, 111, 765-784.

Johnson, K.L., Nicol, T., Zecker, S.G. & Kraus, N. 2008. Developmental plasticity in the human auditory brainstem. *J Neurosci*, 28(15), 4000-4007.

Keller, C.H. & Takahashi, T.T. 2000. Representation of temporal features of complex sounds by the discharge patterns of neurons in the owl's inferior colliculus. *J Neurophysiol*, 84(5), 2638-2650.

Krishnan, A. 2007. Frequency-following response. In R. F. Burkard, J. J. Eggermont, M. Don (Eds.). *Auditory Evoked Potentials: Basic Principles and Clinical Application* (pp, 313-335). Philadelphia, PA: Lippincott Williams & Wilkins.

Krishnan, A., Xu, Y., Gandour, J.T. & Cariani, P. 2004. Human frequency-following response: Representation of pitch contours in Chinese tones. *Hear Res*, 189, 1-12.

Krishnan, A., Xu, Y., Gandour, J.T. & Cariani, P. 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn Brain Res*, 25, 161-8.

Krishnan, A., Gandour, J.T. & Bidelman, G.M. 2010. The effects of tone language experience on pitch processing in the brainstem. *J Neurolinguist*, 23, 81-95.

Krishnan, A. & Parkinson, J. 2000. Human frequency-following response: representation of tonal sweeps. *Audiol Neuro-Otol*, 2000, 5, 312-321.

Krishnan, A., Swaminathan, J. & Gandour, J.T. 2009. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci*, 21, 1092-105.

Li, X. & Jeng, F.-C. 2011. Noise tolerance in human frequency-following responses to voice pitch. *J Acoust Soc Am Express Letters*, 129(1), EL21-26 (DOI:10.1121/1.3528775).

Miller, C.A., Abbas, P.J., Robinson, B.K., Nourski, K.V., Zhang, F., et al. 2006. Electrical excitation of the acoustically sensitive auditory nerve: Single-fiber responses to electric pulse trains. *J Assoc Res Otolaryngol*, 7(3), 195-210.

Musacchia, G., Sams, M., Skoe, E. & Kraus, N. 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc natl Acad Sci*, 104(40), 15894-8.

Nourski, K.V., Abbas, P.J., Miller, C.A., Robinson, B.K. & Jeng, F.-C. 2005. Effects of acoustic noise on the auditory nerve compound action potentials evoked by electric pulse trains. *Hear Res*, 202, 141-153.

Rosenhamer, H., Lindstrom, B., & Lundborg, T. 1978. On the use of click evoked electric brainstem responses in audiologic diagnosis: I. The variability of the normal response. *Scandinavian Audiology*, 7, 193-205.

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

Russo, N.M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., et al. 2008. Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clin Neurophysiol*, 119(8), 1720-1731.

Skoe, E. & Kraus, N. 2010. Auditory brain stem response to complex sounds: A tutorial. *Ear Hear*, 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. *J Cogn Neurosci*, 20(10), 1892-1902.

Strait, D.L., Kraus, N., Skoe, E. & Ashley, R. 2009. Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci*, 29, 661-668.

Swaminathan, J., Krishnan, A. & Gandour, J.T. 2008. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport,* 19, 1163-7.

Thornton, A. R. D. 2007. Instrumentation and Recording Parameters. In R. F. Burkard, J. J. Eggermont, M. Don (Eds.). *Auditory Evoked Potentials: Basic Principles and Clinical Application* (pp, 42-72). Philadelphia, PA: Lippincott Williams & Wilkins.

Voytenko, S. V. & Galazyuk, A. V. 2008. Timing of sound-evoked potentials and spike responses in the inferior colliculus of awake bats. *Neuroscience*, 155(3), 923-936.

Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T. & Kraus, N. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci*, 10(4), 420-422.

FIGURE LEGENDS

Figure 1 A typical example of the time waveforms (**A**) and spectrograms (**B**) of the frequency-following response to voice pitch. The three columns on the left represent the time waveforms obtained at three different stimulus intensities (30, 45 and 60 dB nHL), whereas the three columns on the right display the corresponding spectrograms at each intensity. Each row represents the time waveforms and spectrograms after averaging, including the number of sweeps incorporated into the average as indicated on the right. Numbers in parentheses indicate the amount of time (minutes: seconds) that is needed to obtain recordings with the associated numbers of sweeps. The vertical bar in the bottom left corner indicates the ordinate scale and unit of the response time waveforms; the gray scale in the bottom right corner indicates the scale and unit of the spectral amplitudes of the spectrograms.

Figure 2 A typical example of the *f0* contour (left panel) and the autocorrelation output (right panel) of an FFR to voice pitch that represents an average of 8000 recording sweeps at 60 dB nHL. Arrows in the right panel point to the positive peak and its following trough of the normalized autocorrelation output.

Figure 3 FFR trends revealed by plotting the mean values of *Frequency Error* (**A**), *Slope Error* (**B**), *Tracking Accuracy* (**C**), *Pitch Strength* (**D**) and *RMS Ratio* (**E**) as a function of number of sweeps. Stimulus intensities are plotted using different symbols. Vertical error bars indicate one standard error.

Figure 4 Exponential curve-fitting to the FFR trends with respect to five objective indices: *Frequency Error* (**A**), *Slope Error* (**B**), *Tracking Accuracy* (**C**), *Pitch Strength* (**D**) and *RMS Ratio* (**E**). The three columns represent data obtained at three different stimulus intensities. Data of each index were fit to an exponential model with descending or ascending trends (solid curves). The fitted equation, along with the coefficient of determination (r^2), is shown in each panel. Dotted and dashed lines in each panel indicate the estimated amplitude of background noise (A_{noise}) and asymptotic amplitude of the response (A_{AS}), respectively.