Running head: FFR – RELATIVE POWER OF HARMONICS

刪除: TO VOICE PITCH

刪除: FUNDAMENTAL VERSUS

RELATIVE POWER OF DIFFERENT HARMONICS IN HUMAN FREQUENCY-FOLLOWING RESPONSES ASSOCIATED WITH VOICE PITCH IN AMERICAN AND CHINESE ADULTS[1]

刪除: HUMAN

刪除: TO

刪除: : RELATIVE CONTRIBUTIONS OF THE FUNDAMENTAL FREQUENCY AND ITS HARMONICS

FUH-CHERNG JENG, CASSIE E. COSTILOW, DANIELA P. STANGHERLIN

Communication Sciences and Disorders

Ohio University

CHIA-DER LIN

Department of Otolaryngology-HNS,

China Medical University Hospital, Taiwan

[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).

Summary.– **(≤ 150 words)** When the fundamental frequency (*f0*) is removed from a complex stimulus, the pitch of the *f0* is still perceived by the listener. Through the use of the scalp-recorded frequency-following response, this study examined the relative contributions of the *f0* and its harmonics in pitch processing by systematically manipulating the speech stimulus to remove component frequencies. Twelve American and 12 Chinese adults were recruited. Two-way ANOVA (language x experimental-condition) showed a significant difference in *pitch strength* (*p*=0.001, *F*=4.550) and *frequency error* (*p*=0.020, *F*=2.865) for the experimental-condition factor. A *post hoc* Tukey-Kramer analysis demonstrated significantly larger responses to the harmonics-only conditions than those obtained in the *f0*-only and control conditions. No statistically-significant difference was observed between the two groups of participants. These findings indicate that neural responses associated with individual harmonics dominate the pitch processing in the human brainstem, irrespective of whether the listener's native language is nontonal or tonal.

刪除: The phenomenon of the "missing fundamental" has shown that w

刪除: FFRs were recorded in 12

刪除: with 7 experimental and 1 control conditions. The results showed

刪除: FFR

刪除: (*p*=0.017)

刪除: No statistical difference was observed between the two groups of participants in the overall trends of the response across all conditions.

刪除: d

刪除: the

刪除:

刪除: and such harmonic predominance was independent from the listener's linguistic background.

The human brain is capable of discriminating subtle changes in voice pitch from speech signals. Speech signals, like other harmonic complex sounds, consist of a fundamental frequency (*f0*) and component frequencies that are integer multiples of the *f0*, known as harmonics. While the *f0* is known to carry vital information of the sound, the harmonics also play an important role in pitch processing, as shown by the phenomenon of "missing fundamental." This phenomenon has revealed that when the *f0* is removed from a complex stimulus the pitch of the *f0* is still perceived (Moore & Glasberg, 1986; Ballantyne, 1990). Perception of the missing *f0* develops early in life (He & Trainor, 2009). Real-life examples of this phenomenon, such as the telephone, indicate the redundancy of the *f0* energy in pitch perception. Research using behavioral (Plomp, 1967; Ritsma, 1967) and electrophysiological methods (Galbraith, *et al.*, 2004; Krishnan, Xu, Gandour, & Cariani, 2004; Dajani, Purcell, Wong, Kunov, & Picton, 2005; Kraus & Nicol, 2005; Aiken & Picton, 2006; Musacchia, Sams, Skoe, & Kraus, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007; Jeng & Schnabel, 2009; Skoe & Kraus, 2010) have indicated that harmonics provide adequate information for pitch processing to occur in absence of the *f0*, but further exploration into the neural processes behind this phenomenon will give us a more detailed understanding of how the brain processes pitch information.

The ability for normal-hearing adults to process changes in voice pitch has been studied with the use of the frequency-following response (FFR) (Krishnan, *et al.*, 2004; Dajani, *et al.*, 2005; Aiken & Picton, 2006, 2008). The FFR is a scalp-recorded gross auditory electrical potential that mainly reflects the summed activity of whole neuronal populations in the auditory brainstem and midbrain (Moushegian, Rupert, & Stillman, 1973). The FFR follows the temporal patterning of phase-locking neuronal responses in these neural populations. Due to phase locking, neuronal responses in the early auditory pathway faithfully reflect stimulus periodicities up to a

刪除: , which is a necessary ability for the development of language

刪除: Voice pitch carries linguistic and supra-segmental information that is needed to understand the speaker's message.

註解 [N1]: The author should expand upon the description of the missing *f0* phenomenon, include real-life examples of this phenomenon, such as the telephone, and state that the perception of the missing fundamental develops early in life (He and Trainor, J. Neurosci, 2009).

刪除: the

刪除: frequency

刪除: Two of the classic theories of hearing are the place and temporal theories. The classic place theory proposes that the spatial discharge rate patterns in the tonotopically organized neural maps are used to represent the stimulus power spectrum. Spectrally based pattern recognition mechanisms will then analyze the patterns of excitation that are indicative of the *f0* and its harmonics. However, the classic place theory cannot account for the missing fundamental phenomenon since the perceived pitch is not physically present. The classic temporal theory proposes that the inter-spike intervals within single auditory neurons in all frequency regions are used to produce population interval distributions and generates the perception of the f0 from a set of harmonics. That is, all parts of the cochlea respond to all frequencies and the sensory hair cells transmit all frequency parameters of the stimulus; analysis is then performed at higher auditory levels which could account for this missing fundamental phenomenon.

刪除:

刪除: evoked

few kHz. The FFR reflects the synchronized component of summed neural population responses. FFR (1) does not require the participant's behavioral response, (2) is an objective and non-invasive method, (3) is a brainstem response that is not affected by the listener's arousal state (i.e., it makes no differences whether the participant is sleeping or awake), and (4) is sensitive to changes in *f0* contours, therefore making it an ideal method for exploring the relative contributions of the *f0* and its harmonics on pitch processing in the human brainstem.

Previous studies (Dajani, *et al.*, 2005; Aiken & Picton, 2006) have investigated the effects of removing the *f0* component, but thus far more systematic investigations of the contributions of upper harmonics have not been undertaken. The main purpose of this study is to assess the relative contributions of different sets of upper harmonics to the FFR. A secondary goal is to determine whether a listener's linguistic background (nontonal vs. tonal language) might influence the relative contributions of periodicities associated with voice pitch.

In 2004, Krishnan and colleagues demonstrated that FFR spectrograms obtained for normal-hearing adults contain energy bands at several harmonics. In their study, native speakers of Mandarin were presented with a set of four Chinese syllables with typical language-specific pitch contours: flat (yi^{1}), rising (yi^{2}), bidirectional (yi^{3}) and falling (yi^{4}). They determined that FFR energy is most robust at *f0* and the 2nd harmonic (Krishnan, *et al.*, 2004). Dajani and colleagues (2005) removed the *f0* of a natural vowel /a/ through the use of a high-pass filter and successfully recorded synchronous neural activities from normal-hearing adults. Aiken and Picton (2006) presented a natural vowel /Λ/ to normal-hearing adults either with no *f0* or with only the *f0*. They found that the removal of the *f0* does not reduce the response amplitude, whereas the removal of the harmonics produced a much smaller response amplitude than the other experimental conditions. In 2008, Aiken and Picton used FFR and successfully measured

刪除: pitch contour of a complex stimulus, reflecting neural phase-locked activity

刪除: (Moushegian, Rupert, & Stillman, 1973).

刪除: 3

刪除: voice pitch

刪除: phenomenon of the "missing fundamental" and the relative

刪除: exploring this phenomenon have reported the effects of removing the *f0* from the stimulus with no investigation into the role each harmonic plays in an individual's ability to process pitch information. What remain unknown are those frequency properties inherent within the stimulus that contribute to the strength and quality of the FFR.

刪除: In addition, whether the listener's linguistic background would have influence on the relative contributions of the *f0* and its harmonics remain unclear.

刪除: , with energy observed up to the 2nd formant

刪除: This is likely due to the relatively close proximity of *f0* and its first overtone to the 1st formant (peak) of the vocal tract resonances (Dajani, *et al.*, 2005; Aiken & Picton, 2006, 2008).

刪除: to

the response amplitude at the *f0* and 23 higher harmonics. Overall, these results indicate that frequency cues, such as those commonly found in tonal languages (e.g. Mandarin Chinese), can serve as a means to investigate the pitch processing mechanisms in the human brainstem. It was hypothesized that, during systematic removal of the *f0* and its harmonics, the FFR associated with voice pitch would remain relatively stable, regardless of the listener's linguistic background.

METHOD

Participants

Twelve native speakers of American English (9 females, mean age = 21.8 yr., standard deviation [*SD*] = 2.9) and 12 native speakers of Mandarin Chinese (6 females, mean age = 25.3 yr., *SD* = 2.6) were recruited from the student population at Ohio University. Inclusion of both the American and Chinese adults was required in studying effects of the listener's linguistic background on pitch processing. All participants possessed normal hearing as assessed by pure tone audiometry with hearing thresholds of ≤ 20 dB HL for octave frequencies between 125 and 8000 Hz. Experimental protocols and procedures used in this study were approved by the Ohio University Institutional Review Board. All participants provided informed consent prior to participation of this study.

Stimulus Material

A total of seven acoustic stimuli were prepared for this study. A monosyllabic Mandarin Chinese syllable /yi^2/ that mimicked the English vowel /i/ with a rising voice pitch (117 to 166 Hz) was used as the *intact* stimulus. The rising tone was chosen because it was reported to elicit the strongest response than other Mandarin tones (Krishnan et al., 2004). The stimulus token was recorded from a male Chinese speaker at 40000 samples/sec. with a duration of 250 msec. and a 10 msec. rise and fall time for the stimulus envelope. The subsequent stimuli were variations of

刪除: They found substantial FFR responses at harmonic frequencies close to that of the formant peaks. Furthermore, the robustness of the response is also dependent upon the stimulus. For example, responses near formant-related harmonics are most robust in FFR recordings using a rising pitch (yi^2) speech token (Krishnan, *et al.*, 2004).

刪除: .

刪除: It was hypothesized that in support of the temporal theory, as the *f0* and part of its harmonic components were removed from the stimulus, the FFR to voice pitch would remain stable. It was also hypothesized that in support of the place theory, the FFR amplitude would also be identifiable when the *f0* was preserved and all harmonics were removed. .

刪除: **ethod**

刪除: Chinese (6 females, *M* age = 25.3 yr., *SD* = 2.6)

刪除: American (9 females, *M* age = 21.8 yr., *SD* = 2.9) adult participants

刪除: and nearby communities

刪除: 5

the *intact* stimulus, digitally bandpass filtered to remove frequency components of the *f0* and its harmonics. Specifically, the *intact* stimulus was degraded by removing the *f0* and selected harmonics to create a total of seven acoustic stimuli (see Table 1 for details). The high-pass and low-pass filters were brick-wall, linear-phase finite-impulse-response filters. Each experimental condition used one of the seven acoustic stimuli. The cutoff frequency for the *-f0* condition was set at 170 Hz to remove the spectral energies ≤ *f0* and yet to preserve the spectral energies ≥ 2nd harmonic. Cutoff frequencies for the *-h2*, *-h4*, *-h6* and *-h8* conditions were then progressively set at the integer multiples of 170 Hz to remove energies around the harmonics. Due to the fact that neural phase-locking decreases substantially with increasing frequency and the FFR was detectable for frequencies up to about 1000 to 1500 Hz (Aiken & Picton, 2008; Moushegian, *et al.*, 1973), cutoff frequencies of high-pass filtering used in this study were limited to the 8th harmonic at 1360 Hz. The *+f0* condition used a low-pass filter with a cutoff frequency at 170 Hz to remove energies at all harmonics and to preserve the energies at the *f0* and below. To ensure equal loudness across the seven stimuli, all filtered acoustic stimuli were scaled up linearly to have the same root-mean-square (rms) amplitude as that of the *intact* acoustic token. The rms amplitude was calculated from the entire 250-msec. of each stimulus token. Due to time constraints, the *-h3*, *-h5* and *-h7* conditions were excluded from the present study. Fig. 1 shows spectrograms (**A**) and power-spectral density (**B**) of the seven stimuli used in this study. Note energy higher than the 8th harmonic was preserved in all stimuli, except the *+f0* token.

删除: tokens

删除: : intact, *-f0* (high-pass filter with a cutoff frequency at 170 Hz), *-h2* (HPF at 340 Hz), *-h4* (HPF at 680 Hz), *-h6* (HPF at 1020 Hz), *-h8* (HPF at 1360 Hz), and *+f0* (low-pass filter with a cutoff frequency at 170 Hz).

删除: tokens as the

删除: us

删除: tokens

删除: Each acoustic token had a duration of 250 msec. with a rise and fall time of 10 msec.

Stimulus presentations and trigger synchronization were controlled using a custom program written in LabVIEW 9.0 (National Instruments, Austin, TX). Acoustic tokens were delivered via a 16-bit digital-to-analog converter at a rate of 40000 samples/sec. The stimulus tokens were presented monaurally through an electromagnetically-shielded ER-3A insert earphone (Etymotics, ER-3A) to the right ear at 70 dB SPL with an inter-stimulus interval of 45 msec. Although right- versus left-ear stimulation produced similar brainstem responses (Akhoun et al., 2008), right-ear advantages had been reported when recording FFRs to speech sounds (Hornickel et al., 2009). A foam earplug was inserted into the listener's left ear to block ambient noise. Stimulus intensity of the acoustic token was calibrated using a Larson & Davis system 824 model sound level meter (dB flat weighting) bridged to a 2 c.c. coupler (GRAS RA0038).

Recording Parameters

刪除: *FFR*

刪除: *s*

Recording took place in an acoustically and electrically treated sound booth. Three gold-plated electrodes were placed at the midline of the forehead at the hairline (noninverting), right mastoid (inverting), and left mastoid (ground). All electrode impedances were ≤ 3 *k*Ω at 10 Hz. Recordings were amplified (Neuroscan SynAmps^2, 24-bit resolution, least significant bit: 0.15 nV, optically-isolated circuits to ensure patient safety), bandpass filtered (0.05\3500 Hz, 6 dB/octave), and digitized at a rate of 20000 samples/sec. Continuous data were recorded using Neuroscan Acquire 4.4 software (Compumedics, Charlotte, NC) and stored on a computer for offline analysis.

Experimental Protocol

Participants were first evaluated for normal hearing and were then prepared for electrophysiological testing. Data collection for each participant was completed in a single session that was about 1 to 2 hours in duration. All participants were asked to rest in a comfortable recliner. All acoustic stimuli were presented monaurally to the listener's right ear. The order of the seven experimental conditions was randomized within and across participants. A control condition to evaluate the possibility of electrical interference between the stimulation equipment and recording electrodes was conducted at the final portion of the testing session.

刪除: and were permitted to fall asleep during FFR recordings

刪除: The left ear was blocked with an earplug to prevent interference from ambient noise. Each stimulus token was presented at an inaudible level and gradually increased until 70 dB SPL was achieved to not disturb the participant's restfulness.

During the control condition, the sound tube was occluded and moved away from the participant’s ear (and the participant’s ears were then plugged) to prevent audibility of the stimulus sound. Due to time constraints, the control condition was conducted in only six Chinese and seven American participants.

Data Analysis

Procedures for data analysis were reported in our previous publications (Jeng et al., 2010; Li & Jeng, 2011). Briefly, all data were analyzed using MATLAB 2008b (MathWorks, Natick, MA). To isolate better spectral energies of the response, continuous recordings were digitally bandpass filtered using a brick-wall, linear-phase finite-impulse-response (FIR) filter (85-1500 Hz, 500th order). Filtered recordings were segmented into sweeps of 295 msec. in length. A total of 2200 sweeps were collected for each experimental condition. An individual sweep was rejected if it contained voltages greater than ±25 μV, which was typically caused by patient movement. During each recording condition, the rejection rate was less than 200 sweeps and the remaining sweeps were averaged. To identify the onset of the response, the stimulus tokens were down-sampled to 20000 samples/sec. so that all stimuli and recordings had the same sampling rate. The down-sampled stimulus tokens were used throughout data analyses. Cross-correlation of the stimulus and averaged recording was performed to identify the time shift 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

刪除: control condition (sound tube occluded and moved away from the participant’s ear) was conducted at the end of the testing session to ensure that the recordings were not artifactual.

刪除: A

註解 [N2]: The author should briefly report the values of the best response latencies (time shift that produced maximum cross-correlation)—means, *SD*s, ranges. Also, the author should report the ratio of that value to the 0-time shift cross-correlation.

刪除: ed waveforms

cross-correlation value. For this study, the mean FFR latency for recording obtained from all participants was xx msec. with a SD of msec. and a range from xx to xx msec. A 250-msec. segment of the recorded waveform was extracted from the originally averaged recording starting from the maximum cross-correlation value. The same analytical procedures were applied to all recordings obtained in the experimental and control conditions. Grand-averaged responses were obtained by averaging the raw data (i.e., the 295-msec. recording) for each group of participants. Cross-correlation of the stimulus and grand-averaged recording was used to identify the time shift that yielded the maximum cross-correlation value. A 250-ms segment of the grand-averaged recording was then extracted starting from the maximum cross-correlation value.

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 *f0* estimate for that windowed segment. This procedure was repeated for all windowed segments. All *f0* estimates were concatenated to constitute the *f0* 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

註解 [A3]: Dear co-authors, don't worry about the latency values. I will fill them in later.

刪除: ed waveform

stimulus *f0*) was used to capture possible frequency deviations of the brainstem's responses associated with voice pitch. The same technique was applied to the stimulus token and recorded waveforms. Grand-averaged spectrograms were obtained by averaging spectrograms for each group of participants.

Objective Measures of Pitch Encoding in the Human Brainstem

Six objective measures (*frequency error*, *slope error*, *tracking accuracy*, *pitch strength*, *rms amplitude* and *f0 amplitude*) were used to quantify the pitch-tracking accuracy and phase-locking magnitude of the responses (Krishnan et al. 2005; Russo et al. 2008, Jeng et al. 2010, 2011; Li & Jeng 2011). Each objective measure was meant to represent a specific aspect of pitch processing in the human brainstem. The six objective measures are described as follows.

(1) *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 *f0* contours of the stimulus and recordings and averaging the errors across the 201 windowed segments.

(2) *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 *f0* contour on an *f0*-versus-time plot and then subtracting the slope estimate of the stimulus *f0* contour from that of a recording. Although Mandarin pitch contours are curvilinear, a linear regression was conducted on the entire *f0* contour to represent the degree to which the overall shape of the *f0* contour was preserved in the recording. Slope estimate of the stimulus token used in this study was 272 Hz/sec.

(3) *Tracking accuracy* (i.e., the regression *r* value) denoted the overall faithfulness of pitch tracking between the stimulus and response *f0* contours. To obtain an estimate of *tracking accuracy*, linear regression was conducted on a recording-versus-stimulus *f0* contours plot. Regression *r* value was then denoted as the *tracking accuracy* of the recording.

(4) *Pitch strength* measured the magnitude of neural phase-locking to the *f0* 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. The *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. 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 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.

註解 [N4]: Why is the author not reporting the amplitude of the *f0* as one of the measures? This measure is a nice complement to the rms measure.

(5) *Rms amplitude* represented the response amplitude in the time domain and was calculated from the root-mean-square amplitude of the extracted 250-msec. segment.

(6) *F0 amplitude* represented the amount of energy located at the response *f0* and was calculated by averaging the spectral amplitudes along *F0* contour across time bins.

In order to quantify the FFR associated with voice pitch, responses were analyzed using the methods noted above. Fig. 2 represents an example of the *f0* contour of a response (left panel) and the autocorrelation output (right panel) of a recording obtained using the original acoustic token (i.e., the *intact* condition). 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 msec. 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 24 participants across the seven experimental conditions.

Statistical Analysis

A two-way repeated measures ANOVA was conducted on each objective measure to determine significance across the listeners' linguistic background (American versus Chinese) and the seven experimental conditions. A *post hoc* Tukey-Kramer test was applied to determine which pairs of experimental conditions were significantly different and which were not. A *t* test was used to evaluate the significance between the control and each experimental condition. A *p* value of <0.05 was considered statistically significant.

RESULTS

Spectral and Temporal Representations of FFRs in American and Chinese Participants

The FFR associated with voice pitch was analyzed in both the spectral and time domains. Fig. 3**A** shows grand-averaged spectrograms of the recordings obtained from the 12 Chinese

刪除: Four objective measures (Frequency Error, Slope Error, Tracking Accuracy and Pitch Strength) were used to analyze the data for each experimental condition. Frequency Error represents the accuracy of pitch tracking and was calculated as the average of the absolute Euclidean distance between the stimulus and recording *f0* estimates. *Slope Error* indicates how well the shape of the *f0* contour is preserved. This index was derived by first estimating the slope of the regression line of a stimulus *f0* contour on an *f0*-versus-time plot and then subtracting the slope estimate of the stimulus *f0* contour from that of the recording. Slope estimate of the intact stimulus token used in this study was 272 Hz/sec. *Tracking Accuracy* is a reflection of the overall faithfulness of pitch encoding in the brainstem obtained by plotting the response *f0* estimates as a function of *f0* estimates across time bins, then finding the correlation coefficients of the stimulus and response *f0* estimates. *Pitch Strength* reflects the robustness of neural phase-locking and was derived from the peak-to-trough amplitude of the normalized autocorrelation output for each recording.

刪除: esults

刪除: *the*

刪除: *to Voice Pitch*

刪除: to

刪除: 1

participants. For the *intact* condition, the FFR showed clear energy at the *f0* and its harmonics. When the *f0* was removed from the stimulus (i.e., the *-f0* condition), FFR spectrogram showed similar results to that of the *intact* condition. This finding not only was consistent with the “missing fundamental” phenomenon, but also indicated the importance of the harmonics in pitch processing in the human brainstem. When the harmonics were progressively removed in the *-h2*, *-h4*, *-h6* and *-h8* conditions, spectrograms of the recordings showed clear energy not only at the *f0* but also at the harmonics. This finding indicated that the detection of the *f0* and its harmonics in the human brainstem could be achieved from only the higher harmonics of the stimulus (i.e., ≥ 9th harmonic). In contrast, when all the harmonics were removed and the *f0* energy was preserved in the stimulus (i.e., the *+f0* condition), the FFR spectrogram of the recordings showed clear energy only at the *f0*, but not the harmonics. The spectrogram of the recordings obtained in the control condition showed energy randomly distributed across the frequency range of the *f0* and its harmonics.

刪除: to voice pitch

刪除: up to about the fourth harmonic

刪除: No FFR was observed in the control condition.

Grand-averaged time waveforms (Fig. 3**B**) of all recordings obtained from the Chinese participants showed similar findings as those observed in the spectrograms. For the *intact* condition, the time waveform of the response showed periodicities similar to that of the stimulus (as reflected from Fig. 3**A** and results of *frequency error*, *slope error* and *tracking accuracy*) and had a response rms amplitude of 63.54 nV. When the *f0* was removed in the *-f0* condition, the time waveform of the response showed periodicities similar to that of the stimulus and had an rms amplitude of 57.99 nV. When harmonics were progressively removed in the *-h2*, *-h4*, *-h6* and *-h8* conditions, the response time waveforms showed similar periodicities to those observed in the *intact* and *-f0* conditions and had rms amplitudes of 60.52, 65.83, 50.03 and 54.40 nV, respectively. For the *+f0* condition, the response time waveform showed similar periodicities

刪除: 1

註解 [N5]: The author states that ‘the response waveforms showed similar periodicities.’ Was this assessment based on a qualitative or quantitative assessment?

刪除: It was noted that the response rms amplitudes decreased when acoustic energy ≥ 6th harmonic was removed from the stimulus tokens.

with an rms amplitude of 36.51 nV that was smaller than those obtained in the *intact*, *-f0*, *-h2*, *-h4*, *-h6* and *-h8* conditions. The time waveform of the recordings obtained in the control condition showed periodicities randomly distributed across the time domain and had an rms amplitude of 26.77 nV. Consistent with results shown in Fig. 3**A** and 3**B**, Fig. 3**C** showed the power-spectral density of each grand-averaged recording. Regardless of the removal of *f0* and harmonics, response energy at *f0* was identifiable through visual inspection of the experimenter for the *intact*, *-f0*, *-h2*, *-h4*, *-h6* and *-h8* conditions.

Grand-averaged spectrograms (Fig. 4**A**), time waveforms (Fig. 4**B**) and power-spectral-density (Fig. 4**C**) of the recordings obtained from twelve American participants demonstrated similar findings with those obtained from the Chinese participants. Similar findings observed in the FFR spectrograms obtained from the two groups of participants indicated the existence of some sort of common pitch-processing mechanisms in the human brainstem among the American and Chinese listeners.

Relative Contributions of the f0 and Its Harmonics

To demonstrate the relative contributions of the stimulus energy at the *f0* and its harmonics on FFR, group mean values of the six objective indices were plotted across the seven experimental and one control conditions (Fig. 5). The overall trend of the *frequency error* (Fig. 5**A**) for recordings obtained from the Chinese and American participants remained relatively stable from the *intact* condition through the removal of up to the 8th harmonic, all falling roughly between 4 and 9 Hz. *Frequency error* began to rise to about 11 Hz at the *+f0* condition, with recordings at the control condition showing the highest *frequency error* at about 15 to 18 Hz. Specifically, for Chinese participants, group mean values of *frequency error* for recordings obtained in the *intact*, *-f0*, *-h2*, *-h4*, *-h6*, *-h8*, *+f0* and control conditions were 9.09, 9.25, 8.04,

刪除: *Influence of Linguistic Experience on the FFR to Voice Pitch*

刪除: 2

註解 [N6]: The description on page 11 of the American participant responses is highly redundant with the preceding paragraph on the Chinese responses and could be shortened considerably.

刪除: For the intact condition, the FFR spectrogram showed clear energy at the *f0* and its harmonics up to about the fourth harmonic. When the *f0* was removed from the stimulus (i.e., the *–f0* condition), FFR spectrogram showed similar characteristics than those observed in the intact condition. When the harmonics were progressively removed from the stimulus tokens, FFR spectrograms of the recordings obtained in the *-h2*, *-h4*, *-h6*, and *-h8* conditions showed clear energy not only at the *f0* but also at the harmonics. When all the harmonics were removed and the *f0* energy was preserved (i.e., the *+f0* condition), FFR spectrogram of the recordings showed clear energy only at the *f0*, but not the harmonics. Spectrogram of the recordings obtained in the control condition showed energy randomly distributed across the frequency bands.

刪除:

Grand-averaged time waveforms (Fig. 2**B**) of the recordings obtained from the American participants demonstrated similar trends from those obtained in the Chinese participants. Specifically, the intact condition had a response rms amplitude of 34.67 nV. When the *f0* was removed in the *–f0* condition, the time waveform of the response had an rms amplitude of 55.10 nV. When harmonics were progressively removed in the *-h2*, *-h4*, *-h6* and *-h8* conditions, the response time waveforms remained relatively stable and had rms amplitudes of 47.56, 43.78, 43.92 and 41.69 nV, respectively. The *+f0* condition had an rms amplitude of 24.93 nV that was smaller than all other experimental conditions. The control condition showed periodicities randomly distributed across the time domain and had an rms amplitude of 19.77 nV. The rms amplitudes of the grand-averaged time waveforms obtained from the Chinese participants in all the experimental conditions was significantly larger (p = 0.02, t = -3.00) than those obtained from the American participants.

5.75, 8.91, 6.67, 11.11 and 15.17 Hz, respectively; whereas the those obtained from the American participants were 9.09, 7.30, 8.83, 8.14, 4.38, 6.75, 12.09 and 17.75 Hz, respectively. Overall trend of the *frequency error* group mean values for recordings obtained from the Chinese participants in the experimental conditions were not significantly different ($p = 0.98$, $t = 0.02$, Cohen's $d = 0.01$) from those obtained from the American participants.

Slope error (Fig. 5**B**) demonstrated an overall trend similar to that observed in the *frequency error*. Specifically, group mean values of *slope error* for recordings obtained from the Chinese participants were -106.27, -99.43, -95.41, -50.97, -126.50, -103.05, -109.42 and -159.67 Hz/sec. for the *intact*, *-f0*, *-h2*, *-h4*, *-h6*, *-h8*, *+f0* and control conditions, respectively; whereas the those obtained from the American participants were -125.05, -90.22, -110.14, -113.30, -49.28, -99.92, -154.10 and -198.34 Hz/sec., respectively. Overall trends of the *slope error* group mean values were not significantly different ($p = 0.56$, $t = -0.59$, Cohen's $d = 0.29$) between Chinese and American participants.

Tracking accuracy (Fig. 5**C**) showed an inverted trend rather than those observed in *frequency error* and *slope error*. Specifically, the overall trend of the *tracking accuracy (r)* for recordings obtained from the Chinese and American participants remained relatively stable from the *intact* condition through the removal of up to the 8th harmonic, all falling roughly between 0.6 and 0.8. *Tracking accuracy* began to drop to about 0.5 at the *+f0* condition, with recordings at the control condition showing the smallest *tracking accuracy* at about 0.2. For the Chinese participants, group mean values of *tracking accuracy* for recordings obtained in the *intact*, *-f0*, *-h2*, *-h4*, *-h6*, *-h8*, *+f0* and control conditions were 0.68, 0.68, 0.70, 0.83, 0.61, 0.77, 0.55 and 0.30, respectively; whereas the those obtained from the American participants were 0.61, 0.73, 0.66, 0.72, 0.89, 0.75, 0.44 and 0.20, respectively. Overall trend of the *frequency error* group mean values for recordings obtained from the Chinese participants was not significantly different ($p = 0.75$, $t = -0.32$, Cohen's $d = 0.16$) from those obtained from the American participants. *Pitch strength* demonstrated a similar trend to that observed in *tracking accuracy*. Overall trends of *pitch strength* group mean values were not significantly different from each other between the American and Chinese participants ($p = 0.95$, $t = -0.07$, Cohen's $d = 0.03$).

Overall trend of the group mean values of *rms amplitude* for recordings obtained from the Chinese participants were significantly different ($p < 0.01$, $t = -3.39$, Cohen's $d = 1.69$) than those obtained from the American participants. Overall trend of the group mean values of *f0 amplitude* for recordings obtained from the Chinese participants were also significantly different ($p = 0.02$, $t = -2.55$, Cohen's $d = 1.27$) than those obtained from the American participants.

Results of a Two-Way Repeated Measures ANOVA

Two-way ANOVA (i.e., the language and *f0*-harmonic factors) showed a significant difference in *pitch strength* and *frequency error* for the *f0*-harmonic factor. Specifically, for *pitch strength*, ANOVA showed a significant difference for the *f0*-harmonic factor ($p = 0.001$, $F = 4.550$), but not the language factor ($p = 0.987$, $F = 0.001$) nor the interaction between the two factors ($p = 0.208$, $F = 1.482$). For the *f0*-harmonic factor that reached a significance, the conservative *post hoc* Tukey-Kramer procedure further demonstrated that the *pitch strength* measures obtained from the *intact*, *-h4*, *-h6* and *-h8* conditions were significantly larger than the *+f0* condition ($p < 0.001$). For *frequency error*, ANOVA showed a significance difference for the *f0*-harmonic factor ($p = 0.020$, $F = 2.865$), but not the language factor ($p = 0.732$, $F = 0.127$) nor the interaction between the two factors ($p = 0.366$, $F = 1.122$). A *post hoc* Tukey-Kramer procedure further demonstrated that the *frequency error* measures obtained from the *-h4* and *-h6* conditions were significantly smaller than the *+f0* condition ($p = 0.003$). *Slope error*, *tracking*

刪除: ere

刪除: demonstrated a similar trend to that observed in .

刪除: G

註解 [N7]: Did the author find an interaction between group and stimulus condition for the Pitch Strength measurement? Why did the author collapse *-f0*, *-h2* and –h4 into a single statistical analysis, rather than doing a *post hoc* on each measure? This doesn't appear to be statistically valid.

註解 [N8]: It is also important to state which measure are sensitive to the harmonics in the stimulus. For example, the autocorrelation measure is sensitive to the presence of harmonics in the signal.

刪除: 0.61, 0.62, 0.61, 0.66, 0.53, 0.55, 0.29 and 0.15 for the intact, -, *-h2*, *-h4*, *-h6*, *-h8*, + and control conditions, respectively; whereas the those obtained from the American participants were 0.45, 0.49, 0.44, 0.57, 0.67, 0.58, 0.33 and 0.15, respectively. Although the overall trend of the group mean values of the for recordings obtained from the American participants in the seven experimental conditions was not significantly different ($p = 0.46$, $t = 0.77$) than those obtained from the Chinese participants. Group mean values of the Pitch Strength obtained from the Chinese participants in the intact, *-f0*, *-h2* and –h4 conditions were significantly larger ($p = 0.01$, $t = 4.32$) than those obtained in the American participants.

刪除: *f0 amplitude*

accuracy, rms amplitude and *f0 amplitude* did not show statistical significance in the language or *f0*-harmonic factor (see Table 2, which displays the ANOVA results of the six objective measures).

DISCUSSION

In this study, the relative contributions of the *f0* and its harmonics on the FFR associated with voice pitch were examined through a systematic removal of the *f0* and up to the 8th harmonic. Clear FFRs were observed in the -*f0* condition. This finding is consistent with the "missing fundamental" phenomenon and data reported in the FFR literature (Dajani, *et al.*, 2005; Aiken & Picton, 2006, 2008). Dajani and colleagues (2005) removed the fundamental frequency energy of a natural vowel /a/ through the use of a high-pass filter with a cutoff frequency at 300 Hz and successfully recorded synchronous neural activities from seven normal-hearing adults. Aiken and Picton (2006) presented a natural vowel /Λ/ to normal-hearing adults either with no *f0* (equivalent to the -*f0* condition used in this study) or with only the *f0* (equivalent to the +*f0* condition used in this study). They found that the removal of the *f0* (i.e., the -*f0* condition) does not reduce the response amplitude, whereas the removal of all harmonics (i.e., the +*f0* condition) produced a significantly smaller response amplitude than those obtained in the *intact* and -*f0* conditions. The current study makes one further step and examined the response amplitudes through a systematic removal up to the 8th harmonic of the acoustic stimulus. In this study, the FFR remains relatively stable in the *intact*, -*f0*, -*h2*, -*h4*, -*h6* and -*h8* conditions. This finding indicates that small variations of speech periodicity were accessible to the human brainstem, likely through the process of extracting the *f0* information by resolving the inter-harmonic intervals of the acoustic signal. The results of this study provide additional evidence for this view

刪除: Results of a Two-Way Repeated Measures ANOVA
To illustrate better the dependence of FFR on the stimulus f0 and its harmonics, a complete list of significant conditions (for data obtained from the Chinese and American participants combined) is displayed in Table 1. For Frequency Error, all experimental conditions produced significantly smaller mean values than that of the control condition. The +*f0* condition had a significantly larger mean value than those obtained in the –*f0*, -*h2*, -*h4*, -*h6* and -*h8* condition. Although the +*f0* condition had a larger mean value than that of the intact condition, it did not reach a statistical significance. *Slope Error* demonstrated similar significance findings with those observed in Frequency Error, except that the mean value of the +*f0* condition was only significantly larger than that of the –*f0* condition. For Tracking Accuracy, all experimental conditions had significantly larger mean values than that of the control condition and the +*f0* condition had a significantly smaller mean value than those obtained in all the other experimental conditions. For Pitch Strength, all experimental conditions had larger mean values than that of the control condition, whereas the +*f0* condition had a significantly smaller mean value than those obtained in all other experimental conditions.

刪除: we examined

刪除: to

刪除: While this study has provided physiological evidence for both temporal and place theories of pitch processing, it is the temporal cues that hold more significance.

刪除: to voice pitch

刪除: as

註解 [N9]: A discussion of resolved vs unresolved harmonics and pitch strength (and their FFR correlates) would be appropriate.

of pitch encoding in the human brainstem due to the lack of energy in the stimulus spectrograms that was still well represented in the FFR spectrograms. From these results, it was concluded that the brainstem does indeed rely on temporal aspects of the stimulus waveform, such as small variations of the stimulus periodicity, to extract the pitch information of an incoming signal.

One important finding is the significantly lower response observed for the +*f0* condition compared to the harmonic conditions. There are several reasons why a low frequency pure tone might be less salient than a harmonic complex tone with the same fundamental frequency. In general, pure tones in middle and high frequencies (e.g., 1000 Hz and higher) are as salient as their complex tone counterparts. However, as pure tone frequencies get down below 200 Hz, pitch salience was not as strong as those observed in the middle and high frequencies. The reasons for this phenomenon have to do with the relative efficiencies of driving the widest range of cochlea and the most auditory nerve fibers—although a very low frequency pure tone at 70 dB SPL will drive two-thirds of the cochlear nerve population at above spontaneous rate, the complex tone with many harmonics will drive more fibers into saturation, and hence create more pitch-related interspike intervals. This was observed in the Cariani and Delgutte (1996) study in cats. For example, a 160-Hz pure tone created autocorrelation peaks not quite as high as an amplitude-modulated tone with a carrier frequency of 640 Hz and modulation frequency of 160 Hz. This was raised as a minor discrepancy with the psychophysics that the 160 Hz tone would be at least as salient as the AM fundamental pitch. Another set of factors relevant to the FFR are cochlear delays. For carrier frequencies below 1 kHz, cochlear delays smear temporal firing patterns. Such smearing could reduce FFR magnitudes at low-frequency pure tones, because the FFR is the synchronized component of a population response. Heinz and colleagues (2001) presented a nonlinear computational model to simulate auditory-nerve responses. Dau et al.

(2003) combined the Goldstein and Kiang’s convolution concept and Heinz’s nonlinear model and evaluated the importance of cochlear processing on the FFR by comparing predicted response patterns to experimental data. Although they led to the conclusion that the FFR to low-frequency tones were mainly derived from the mid- and high-frequency units at the basal end of the cochlea (Dau, 2003), these models are not applicable in extracting *f0* contours nor detecting the presence of a response.

This study demonstrated that the human brainstem’s ability to process pitch information did not decrease as frequency components are removed from the stimulus, but instead shows a relatively stable response when the *f0* and up to the 8th harmonic were removed. One possible explanation was that the *f0* might be preserved in the amplitude envelope of the stimulus as the lower harmonic frequencies are selectively removed. Neurons in the human brainstem could retrieve the information and produce robust phase-locking to the amplitude envelope. Another possible explanation for the results could be the creation of the stimuli. As frequency components were removed from the stimuli, the total energy of the stimulus token decreased. As a result, the tokens were scaled up to the equal rms amplitude to ensure that all tokens contained the same amount of energy. This process resulted in the re-distribution of more energy among the remaining harmonics and likely produced a stronger response. However, if the rms amplitude when removing the harmonics had not been equalized, acoustic tokens with a fewer number of remaining harmonics would contain a smaller overall amplitude than those with more harmonics remaining. In such a case, results would likely be confounded by the overall stimulus intensity of the acoustic tokens. This is a tradeoff between the choice of equalizing rms amplitude across acoustic tokens rather than maintaining the original amplitude of the *f0* and each of its harmonics. Compensating rms amplitude is beneficial because it allows researchers to compare results

刪除: Although the temporal theory seems to emerge as the leader, the place theory is also of importance. The response at the + condition indicates that there may be a population of neurons that are frequency specific, though the response is not as robust as when temporal cues are provided. The significantly lower response observed for the + condition compared to the harmonic conditions provides evidence for this view. Cariani and Delgutte (1996) examined the temporal discharge patterns of the auditory nerve in anesthetized cats and compared those discharge patterns to human judgments of pitch perception. They found that complex stimuli produced a stronger salience of pitch than pure tones and a stronger perception of pitch to complex tones was also correlated by a more concentrated distribution of the inter-spike interval neural activities in the auditory nerve. Thus, they suggested that a neural response elicited by a complex stimulus (e.g., containing harmonics) would be stronger than that elicited by a pure tone stimulus (e.g., containing only the *f0*).

刪除: One interesting finding is

註解 [N10]: The author should discuss how the temporal envelope of the stimulus changes as the frequencies are selectively removed, and how the brainstem‘s robust phase-locking to the amplitude envelope might contribute to the present results.

刪除: we had not equalized

刪除: In our opinion, c

刪除: us

across testing conditions. This approach, however, has some drawbacks. Had the rms compensation not taken place, there would have been decreased response amplitudes at higher harmonics.

Another important finding is that the overall trends of the response obtained from the two groups of participants were significantly different in *rms amplitude* and *f0 amplitude*, but not in *frequency error*, *slope error*, *tracking accuracy* and *pitch strength*. The FFR literature has shown that pitch processing at the level of the human brainstem can be influenced by the listener's linguistic background (Krishnan, Xu, Gandour, & Cariani, 2005; Swaminathan, Krishnan, & Gandour, 2008; Krishnan, Gandour, & Bidelman, 2010), musical experience (Musacchia, *et al.*, 2007; Wong, *et al.*, 2007), and short-term training (Song, Skoe, Wong, & Kraus, 2008). Consistent with the FFR literature, results of this study showed that Chinese adults demonstrated stronger *rms amplitude* and *f0 amplitude* than American adults at the *intact* condition. This study also demonstrated that the Chinese adults had stronger *rms amplitude* and *f0 amplitude* than American adults when up to the 4th harmonic was removed from the stimulus. The *rms amplitude* and *f0 amplitude* obtained from the Chinese adults, however, exhibited no statistical difference than those obtained in American adults when up to the 6th or the 8th harmonics were removed from the stimulus. When the overall response trends of *rms amplitude* and *f0 amplitude* were considered across all seven experimental and one control conditions, Chinese participants exhibited were significantly larger response amplitudes than those obtained in American adults. Surprising enough, when the overall response trends of *pitch strength*, *frequency error*, *slope error* and *tracking accuracy* were considered across all seven experimental and one control conditions, Chinese and American adults were not significantly different from each other. This finding was consistent with Swaminathan et al. (2008) who had shown that there are conditions

刪除: ; h

刪除: interesting

刪除: Chinese participants

刪除: not

刪除: from those obtained from American participants

刪除: across the all experimental conditions

刪除: This finding is surprising because it is different from what would be expected from the FFR literature.

刪除: Studies have

註解 [N11]: Krishnan's group has shown in several experiments that there are conditions on which the tonal language speakers do not show enhanced pitch tracking. This should be factored into the interpretation.

on which the tonal language speakers do not show enhanced pitch tracking. Taken together, these results indicated that although pitch processing at the level of the human brainstem can be influence by the listener's linguistic background, some mechanisms that underlie the relative contribution of the *f0* and its harmonics on the FFR did not appear to alter with respect to the listener's linguistic background. That is, although Chinese adults might exhibit stronger responses, their "tolerance" to the removal of harmonics from the acoustic stimuli (as reflected by the scalp-recorded FFR) was similar to American adults.

Several models of pitch perception had been proposed in the psychophysical literature. One theory in pattern matching was proposed by Goldstein (1973). Goldstein assumed that an internal Gaussian noise existed in each input frequency channel. After the input frequency components were transmitted and added with noise, the frequency was estimated by using a central processor and assuming that these frequencies were successive harmonics. The best matching harmonic series was calculated as the perceived pitch. Although the pattern matching successfully explained certain aspects of pitch perception, it failed to explain the phenomena caused by unresolved pitch. Another mainstream of pitch perception theory was the autocorrelation, which had been widely used in computerized models. For example, Moore's (2003) schematic model included five procedural steps: creation of a bank of band-pass filters, neural transduction, analysis of inter-spike intervals, a common interval recognizer and a decision machine. For channels of resolved harmonics, the timing intervals were recognized by multiples of the channel periodicity. For channels of unresolved harmonics, the timing intervals were analyzed to identify the common interval across channels. The common interval across channels was treated as the perceived pitch. This model explains several phenomena, including the dominance pitch, the existence region of residue, repetition pitch, and the difference between

刪除: Similar findings were observed in the other three indices of the FFR to voice pitch, such as the Frequency Error, Slope Error and Tracking Accuracy. T

刪除: o

刪除: ay

註解 [N12]: Dau has proposed a model for the generation of the FFR that is relevant to the work here (Dau, *JASA*, 2003).

註解 [N13]: The author should expand upon the discussion of how these results related to pitch perception. If the author would include behavioral data on these subjects it would add tremendously to the manuscript. However, if these data are not available, then a broader literature review would be helpful to discuss the three-way relationship between stimulus features, pitch perception and FFR quality.

註解 [N14]: How does the perception of pitch change as more and more of the stimulus spectrum is removed? This has certainly been addressed in the psychoacoustics literature.

刪除: i

the phase effect on resolved and unresolved pitch. Integration of electrophysiological and psychophysical indices associated with voice pitch had also been reported. Krishnan et al (2010a) examined neural representations of pitch salience in the human brainstem by systematically increasing the pitch salience in the stimulus tokens. They found that behavioral frequency-difference-limens were negatively correlated with the FFR pitch strength, indicating that the acoustic periodicity of the stimulus directly influence the pitch processing in the brainstem. Although this current study has shown the relative contributions of the *f0* and its harmonics on the FFRs, it is anticipated that future studies combining both behavioral and psychophysical measures will help shed light on neural representations of pitch processing in the human brainstem.

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.

註解 [N15]: This article is very relevant: Krishnan, *et al.*, (2010) Hearing neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hearing Research* 268(1-2):60-6.

刪除: The results of this study serve to help fill in the gaps of our knowledge of how the brain processes pitch information of complex sounds.

刪除: It also provides additional physiological evidence for the importance of temporal cues in pitch processing. Current hearing aid and cochlear implant technology utilizes only frequency specific place cues in speech processing strategies. By continuing to gain a better understanding of how the brain processes pitch information, we hope to improve upon the current technology of hearing aids and cochlear implants to include both temporal and place cues in their pitch processing strategies.

REFERENCES

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

Aiken, S. J., & Picton, T. W. (2008) Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, 245, 35-47.

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 Neurophysiolology*, 119, 922-933.

Ballantyne, D. (1990) *Handbook of audiological techniques*. Rushden: Butterworth-Heinemann.

Cariani, P. A., & Delgutte, B. (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76, 1698-1716.

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.

Dau, T. (2003) The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *Journal of Acoustical Society of America*, 113(2), 936-50.

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

Goldstein, J. L. (1973) An optimum processor theory for the central formation of the pitch of complex tones. *Journal of Acoustical Society of America*, 54, 1496-1516.

He, C., & Trainor, L. J. (2009) Finding the pitch of the missing fundamental in infants. *Journal of Neuroscience*, 29, 7718-8822.

Heinz, M. G., Zhang, X., Brucc, I. C., & Carncy, L. H. (2001) Auditory-nerve model for predicting performance limits of normal and impaired listeners. *Journal of Acoustical Society of America (Acoustics Research letters Online)*, 2, 91-6. (DOI 10.1121/1.1387155)

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.

Jeng, F-C., & Schnabel, E. A. (2009) Frequency-following response to voice pitch in infants. In *Abstracts of American Auditory Society Annual Meeting*, Scottsdale, AZ, March 5-7. P. 21.

Kraus, N., & Nicol, T. (2005) Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences*, 28, 176-181.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, 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. *Cognitive Brain Research*, 25, 161-168.

Krishnan, A., Bidelman, G. M., & Gandour, J. T. (2010a) Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hearing Research*, 268, 60-66.

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

Moore, B. C. J., & Glasberg, B. R. (1986) The role of frequency selectivity in the perception of loudness, pitch and time. *In:* B. C. J. Moore (ed.) *Frequency Selectivity in Hearing*. London: Academic Press, pp. 251-308.

Moore, B. C. J., & Moore, G. A. (2003) Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hearing Research*, 182, 153-163.

Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973) 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.

Plomp, R. (1967) Pitch of complex tones. *Journal of Acoustical Society of America*, 41, 1526-1533.

Ritsma, R. (1967) Frequency dominant in the perception of the pitch of complex sounds. *Journal of Acoustical Society of America*, 42, 191-198.

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.

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.

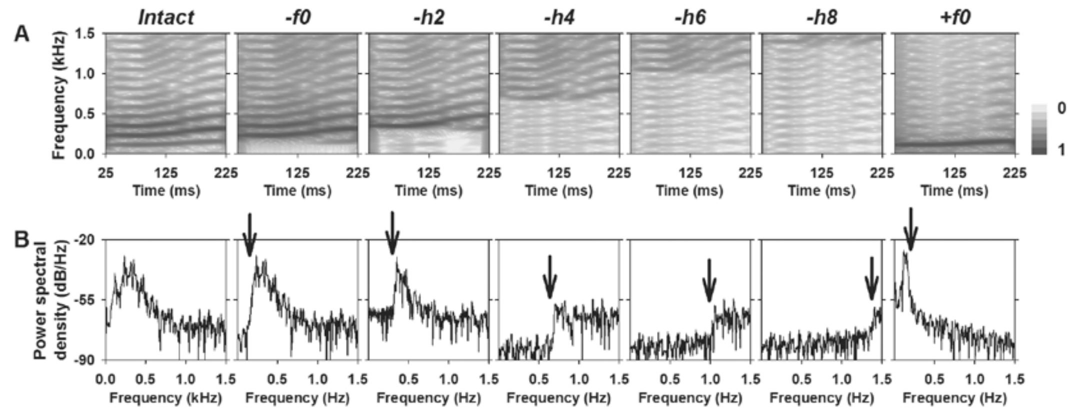


FIG. 1 Spectrograms (**A**) and power-spectral density (**B**) for the seven testing conditions: original token (*intact*), no fundamental frequency (-*f*0), removal of the 2nd harmonic (-*h2*), removal of the 4th harmonic (-*h4*), removal of the 6th harmonic (-*h6*), removal of the 8th harmonic (-*h8*), and fundamental frequency only (+*f0*). Arrows point to the cutoff frequencies of the acoustic stimuli.

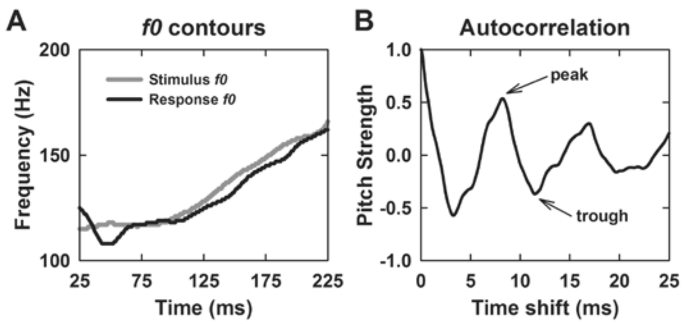


FIG. 2 A typical example of the *f0* contour (**A**) and the autocorrelation output (**B**) of the FFR associated with voice pitch. Arrows in the right panel point to the positive peak and its following trough of the normalized autocorrelation output.

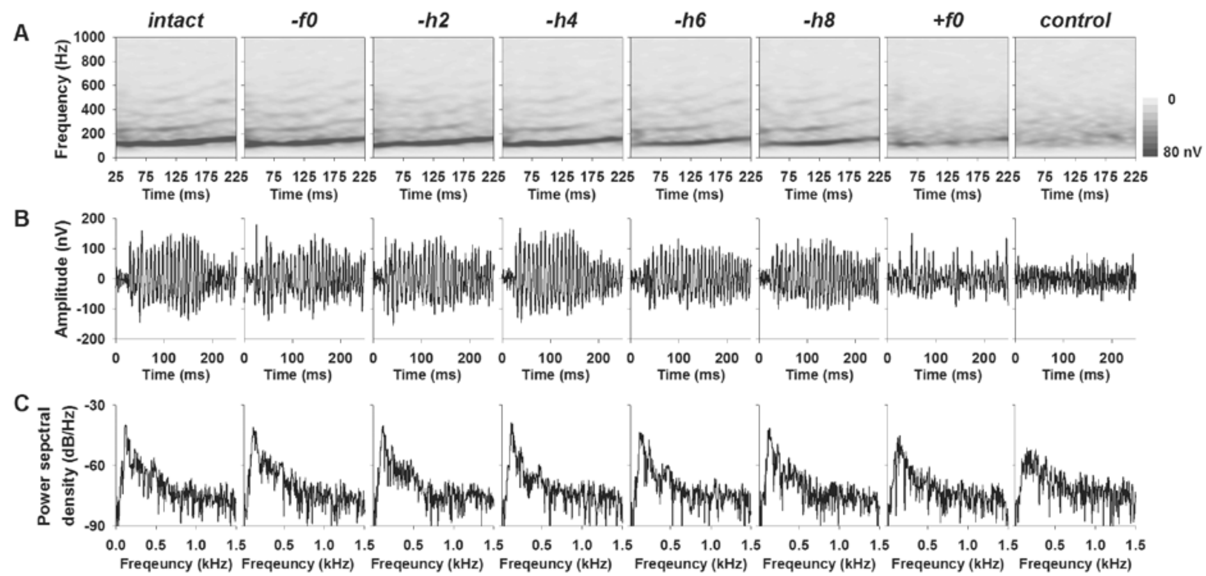
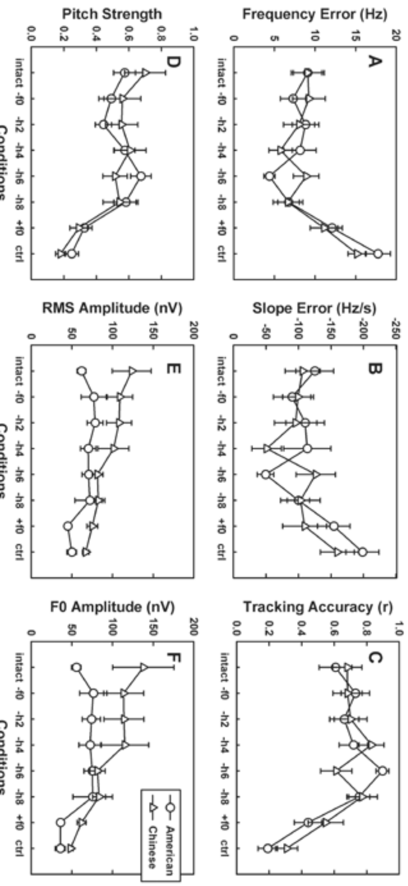


FIG. 3 Grand-averaged spectrograms and time waveforms of the FFR to voice pitch recorded from 12 Chinese participants for all of the testing conditions, as well as a control condition. The testing conditions included: *intact*, no fundamental frequency (*-f0f0*), removal of the 2nd harmonic (*-h2*), removal of the 4th harmonic (*-h4*), removal of the 6th harmonic (*-h6*), removal of the 8th



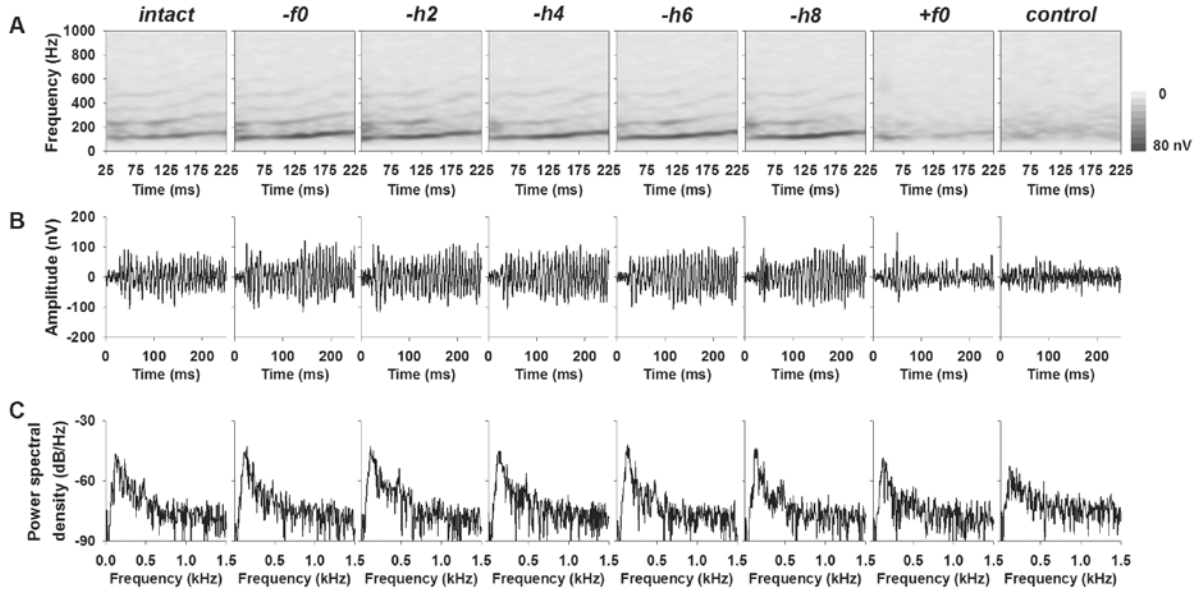


FIG. 4 Grand-averaged spectrograms and time waveforms of the FFR to voice pitch recorded from 12 American participants for all of the testing conditions, as well as a control condition. The testing conditions included: *intact*, no fundamental frequency (-*f0*), removal of the 2nd harmonic (-*h2*), removal of the 4th harmonic (-*h4*), removal of the 6th harmonic (-*h6*), removal of the 8th harmonic (-*h8*), and fundamental frequency only (+*f0*).

FIG. 5 Objective Measures: Group data obtained from the American (open circles) and Chinese (open triangles) participants are graphed into *frequency error*, *slope error*, *tracking accuracy*, *pitch strength*, *rms amplitude* and *f0 amplitude*. Conditions are as follows: *intact*, no fundamental frequency (-*f0*), removal of the 2nd harmonic (-*h2*), removal of the 4th harmonic (-*h4*), removal of the 6th harmonic (-*h6*), removal of the 8th harmonic (-*h8*), fundamental frequency only (+*f0*), and control (*ctrl*).

Table 1 Acoustic stimuli used to examine relative contributions of the *f0* and its harmonics in pitch processing.

Conditions	**Cutoff frequencies**	**Harmonics removed**	**Harmonics preserved**
intact	none	none (i.e., the original acoustic token)	*f0* and all harmonics are preserved
-f0	HPF at 170 Hz	*-f0*	> *f0*
-h2	HPF at 340 Hz	*-f0, -h2*	> *h2*
-h4	HPF at 680 Hz	*-f0, -h2, -h3, -h4*	> *h4*
-h6	HPF at 1020 Hz	*-f0, -h2, -h3, -h4, -h5, -h6*	> *h6*
-h8	HPF at 1360 Hz	*-f0, -h2, -h3, -h4, -h5, -h6, -h7, -h8*	> *h8*
+f0	LPF at 170 Hz	all harmonics are removed, except *f0*	only *f0* is preserved

HPF = high-pass filter; LPF = low-pass filter; *f0* = fundamental frequency; *h* = harmonic. Stimuli were degraded by systematically removing (-) the fundamental frequency (*f0*) and selected harmonics (*h*).

Table 2 Two-way ANOVA results for the frequency-following responses recorded from 12 American and 12 Chinese adults.

Objective Measures	Language		Condition		Interaction	
	F	***p***	***F***	***p***	***F***	***p***
***frequency error* (Hz)**	**0.127**	**0.732**	**2.865**	**0.020***	**1.122**	**0.366**
***slope error* (Hz/sec.)**	**0.029**	**0.869**	**0.917**	**0.492**	**1.338**	**0.262**
***tracking accuracy* (r)**	**0.028**	**0.871**	**2.124**	**0.071**	**1.603**	**0.170**
pitch strength	**0.001**	**0.987**	**4.550**	**0.001†**	**1.482**	**0.208**
***rms amplitude* (nV)**	**0.826**	**0.394**	**1.427**	**0.227**	**1.071**	**0.395**
***f0 amplitude* (nV)**	**0.512**	**0.497**	**1.334**	**0.264**	**1.448**	**0.220**

F: *F* statistics; * $p < 0.05$; † $p < 0.01$