Noise tolerance in human frequency-following responses to voice pitch

Ximing Li and Fuh-Cherng Jeng
Communication Sciences and Disorders, Ohio University, Grover Center W224, Athens, Ohio 45701
xl285608@ohio.edu, jeng@ohio.edu

Abstract: Speech communication usually occurs in the presence of background noise. This study examined noise tolerance in the brainstem’s processing of voice pitch, as reflected by the scalp-recorded frequency-following response (FFR) from 12 normal-hearing adults. By systematically manipulating signal-to-noise ratio (SNR) across three different stimulus intensities, the results indicated that Frequency Error, Slope Error, and Tracking Accuracy remained relatively stable until SNR was degraded to 0 dB or lower (i.e., a turning point). This turning point not only provided physiological evidence supporting pitch tolerance of noise but also allowed recommendation of a minimal SNR when evaluating pitch processing in difficult-to-test patients.

1. Introduction
In daily life, speech communication often occurs in noisy environments. However, humans are capable of extracting the voice-pitch information in the presence of background noise without much difficulty when signal-to-noise ratio (SNR) is as low as −10 dB.1 This tolerance of background noise in pitch perception shows the adaptation of the human auditory system to acoustic environments. Auditory neurophysiological studies in animal models have shown that some neurons in the auditory cortex can tolerate the presence of background noise with certain response properties unaltered.2 Neurons at the subcortical level also exhibit a certain extent of resistance to background noise, likely due to the neural phase-locking to the periodicity of an incoming sound.3

Recent electrophysiological studies on human participants demonstrate that frequency-following response (FFR) to voice pitch can be used as an objective measure to evaluate pitch processing mechanisms at the subcortical level.4–9 Using Mandarin tones as stimuli, Krishnan et al. (2004)4 showed that the fundamental frequencies (F0s) of time-varying stimuli were accurately tracked via phase-locking and preserved in the FFRs. Several recent studies also examined the relation between behavioral perception of speech-in-noise and the subcortical neural representations in normal-hearing children8 and adults.9 These findings indicated that the subcortical neural processing of speech signals in noise could partly account for speech perception in noise. However, effects of SNR on pitch processing at the subcortical level had not been examined systematically.

With this question in mind, we examined the tolerance in pitch processing by looking into the effects of SNR at various stimulus intensities. Pitch processing was evaluated objectively by recording the FFR to a speech token with a rising pitch in continuously presented background noise (Gaussian white noise). Both the SNR and stimulus intensity were set at multiple levels. By systematically manipulating the stimulus token across several SNR levels at different stimulus intensities, we hypothesized that the FFR to voice pitch would tolerate the degradation of the stimulus SNR to a certain extent and start to decrease when the tolerance limit was exceeded.

Author to whom correspondence should be addressed.
2. Method

Twelve Chinese adults [five males; age: mean ± standard deviation (SD) = 24.00 ± 1.17 yr] were recruited from the Ohio University and local communities. All participants had hearing sensitivity ≤20 dB hearing level (HL) for octave frequencies between 250 and 8000 Hz. The experimental protocols used in this study were approved by Ohio University Institutional Review Board.

Six SNR conditions (no noise, 12, 6, 0, −6, and −12 dB) at three stimulus intensities [70, 55, and 40 dB sound pressure level (SPL)] (i.e., 18 testing conditions in total) were randomly presented to each participant. The lowest SNR level was set at −12 dB, because for normal-hearing adults, their performance of tone recognition in noise dropped greatly when SNR was −10 dB or lower. According to the recommended maximum level of noise exposure, the maximal level of sound presentation should not exceed 85 dB A. Given that the lowest SNR level was −12 dB, the stimulus intensity was limited to 70 dB SPL.

A Mandarin syllable /yi/ that mimics the English vowel /i/ with a rising tone was used as the stimulus token. The rising tone was chosen because it was reported to elicit the strongest response in adult participants among the four Mandarin tones. The stimulus token was recorded from a male Chinese speaker at 40 000 samples/s with a duration of 250 ms and a 10 ms rise and fall time for the stimulus envelope. The stimulus F0 and formants followed curvilinear contours (F0: 117–166 Hz, F1: 300–349 Hz, F2: 2058–2043 Hz, and F3: 2899–2857 Hz). A Gaussian white noise wav file, which was used as the continuously presented background noise, was created using MATLAB 2008 (Mathworks, Natick, MA). The noise file was digitized at 40 000 samples/s and low-pass filtered at a cutoff frequency of 5000 Hz with a 24 dB/octave filter slope.

The stimulus token and white noise were presented simultaneously using custom software written in LABVIEW 8.0 (National Instruments, Austin, TX). Both the stimulus token and noise were transmitted to the listener’s right ear through an electro-magnetically shielded insert earphone (Etymotics ER-3A, Elk Grove Village, IL). Although right- versus left-ear stimulation produced similar brainstem responses, right-ear advantages had been reported when recording FFRs to speech sounds. A foam earplug was inserted into the listener’s left ear to block the ambient noise. One trial of 2000 stimulus tokens, with the ongoing background noise, was presented to each participant. A 45-ms interval was set between the offset of each stimulus token and the onset of the next one.

Electroencephalographic activity was recorded in an electrically treated and acoustically attenuated booth. During the recording, each participant rested in a reclining chair with their eyes closed. Three gold-plated electrodes were applied to each participant on the high forehead along the midline below the hair line (non-inverting), the right mastoid (inverting), and the left mastoid (ground). All electrode impedance was below 3 kΩ. Recordings were amplified by Neuroscan SynAmps (gain: 2010), filtered (0.05–3500 Hz, 6 dB/octave), and digitized at 20 000 samples/s.

Procedures for data analysis were reported in our previous publications. Briefly, each recording was band-pass filtered (cutoff frequencies at 100 and 1500 Hz, 500th order), segmented, baseline corrected, artifact rejected (reject criterion: ±25 μV), and averaged. The averaged recording was cross-correlated with the stimulus token to determine the time shift that yielded a maximum cross-correlation value within 3–10 ms. A 250-ms segment was extracted from the recorded waveform starting from the time point corresponding to the maximum cross-correlation value.

The F0 contour of each averaged recording was extracted using a short-term spectrogram algorithm. To generate the spectrogram of the FFR, a sliding Hanning window of 50 ms was applied to the 250-ms waveform at a 1-ms step size, resulting in 201 time bins. In each time bin, the F0 was estimated as the frequency with maximal spectral magnitude within a pre-defined frequency range. Since the F0 of the stimulus token ranged from 117 to 166 Hz, the pre-defined range was set from 107 to 176 Hz, leaving a 10 Hz buffer for possible errors in estimating F0. The F0s from the 201 time bins were concatenated to
construct the F0 contour of a recording. To minimize cross-participant variations in estimating F0s, grand-averaged spectrograms were also computed for the following analyses.

Four objective indices (Frequency Error, Slope Error, Tracking Accuracy, and F0 Amplitude) were adopted to evaluate pitch processing.\textsuperscript{6,14,15} Frequency Error represented the accuracy of pitch-encoding during the course of stimulus presentation. It was computed by finding the absolute Euclidean distance between the F0 contours of the stimuli and recordings and averaging the errors across the 201 time bins. Slope Error indicated the degree to which the shapes of the pitch contours were preserved in the brainstem. It 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. Tracking Accuracy (i.e., $r$ values derived from the regression of the response and stimulus F0 contours) denoted the overall faithfulness of pitch tracking between the two F0 contours. 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. A precise pitch tracking is represented as low values in Frequency Error and Slope Error and high values in Tracking Accuracy and F0 Amplitude.

A repeated-measure two-way analysis of variance (ANOVA) (SNR by intensity) was performed to each of the four indices followed by a post hoc Tukey–Kramer procedure, which was used to isolate the experimental conditions that were significantly different from other conditions.

3. Results

Effects of SNR at different stimulus intensities were visualized from grand-averaged spectrograms (Fig. 1). The effects of SNR, as shown horizontally across columns, indicated that the response energy at F0 decayed as SNR was degraded from no noise to $-12$ dB. Across the three stimulus intensities, as shown vertically across rows, the response energy at F0 decreased as the stimulus intensity decreased.

To quantify the effects of SNR on FFR, four objective indices (Frequency Error, Slope Error, Tracking Accuracy, and F0 Amplitude) derived from the grand-averaged spectrograms are shown in Fig. 2. Frequency Error [Fig. 2(A)] remained relatively stable around 5 Hz when SNR decreased from no noise to 0 dB and started to increase when SNR was less than 0 dB. Specifically, Frequency Error increased from 5.67 to 16.02 Hz when stimulus intensity was 40 dB SPL. Frequency Error increased from 5.49 to 18.70 Hz and 3.99 to 9.14 Hz, respectively, when the stimulus intensity was 55 and 70 dB SPL. This finding indicated a possible “turning point” region in the effects of SNR on FFR.
For Frequency Error, the turning points were all located around 0 dB SNR at three intensities. An exponential model was used to fit the data and calculate the 3 dB higher (Frequency Error and Slope Error) or lower (Tracking Accuracy and F0 Amplitude) from the steady value. The SNR that corresponded to the 3-dB location was defined as the turning point. For Frequency Error, the turning points were −2.52, −6.07, and −1.92 dB SNR for 40, 55, and 70 dB SPL, respectively. After the turning points, Frequency Error started to grow, and the growth appeared slower at high intensities than at low intensities. The growth rates after the turning points (i.e., slope estimates across from 0 to −12 dB) were 0.86, 1.10, and 0.43 Hz/dB for 40, 55, and 70 dB SPL, respectively.

Slope Error [Fig. 2(B)] showed similar trends to those of Frequency Error. The turning points (i.e., 3 dB above the steady values) were −1.71, −3.47, and −1.75 dB SNR for 40, 55, and 70 dB SPL, respectively. The growth rates (after the turning points) for 40, 55, and 70 dB SPL were −11.67, −14.59, and −3.06 Hz/dB, respectively. Tracking Accuracy [Fig. 2(C)], in which a high value was indicative of precise phase-locking, showed consistent trends compared to those of Frequency Error and Slope Error. The turning points occurred at −1.99, −4.89, and −1.95 dB SNR for 40, 55, and 70 dB SPL, respectively. The decline rates after the turning points were 0.052, 0.063, and 0.018 per dB for 40, 55, and 70 dB SPL, respectively. Although F0 Amplitude [Fig. 2(D)] did not show a turning point, F0 Amplitude declined from 42.40 to 20.91 nV, 64.57 to 19.66 nV, and 144.43 to 26.85 nV (when SNR was decreased from no noise to −12 dB) for 40, 55, and 70 dB SPL, respectively.

The results from repeated-measure two-way ANOVA are summarized in Table 1. Both SNR and stimulus intensity showed significant effects for all four indices (Frequency Error, Slope Error, Tracking Accuracy, and F0 Amplitude). There were no significant interactions between SNR and stimulus intensity for any of the four indices. Results from the post hoc Tukey–Kramer analysis for pairwise comparisons within the SNR conditions are summarized in Table 2. Note the numbers listed in Table 2 differed slightly from those plotted in Fig. 2. This was due to the fact that the group means in Table 2 were calculated based on data from each individual, whereas the numbers plotted in Fig. 2 were computed from grand-averaged spectrograms. Group means of the four indices appeared to increase (Frequency Error and Slope Error) or decrease (Tracking Accuracy and F0 Amplitude) with decreasing SNR.

Table 1. Repeated-measure two-way ANOVA (SNR by intensity) results for Frequency Error, Slope Error, Tracking Accuracy, and F0 Amplitude.

<table>
<thead>
<tr>
<th></th>
<th>SNR (dB)</th>
<th>Intensity (dB SPL)</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-statistic</td>
<td>P</td>
<td>F-statistic</td>
</tr>
<tr>
<td>Frequency Error (Hz)</td>
<td>23.98</td>
<td>&lt;0.001b</td>
<td>4.49</td>
</tr>
<tr>
<td>Slope Error (Hz/s)</td>
<td>10.18</td>
<td>&lt;0.001b</td>
<td>5.63</td>
</tr>
<tr>
<td>Tracking Accuracy (r)</td>
<td>19.24</td>
<td>&lt;0.001b</td>
<td>6.16</td>
</tr>
<tr>
<td>F0 Amplitude (nV)</td>
<td>4.99</td>
<td>&lt;0.001b</td>
<td>5.60</td>
</tr>
</tbody>
</table>

a Signal-to-noise ratio (SNR).
b Significance level: <0.05.
decrease (Tracking Accuracy and F0 Amplitude) systematically as SNR decreased from no noise to −12 dB. For Frequency Error, Slope Error, and Tracking Accuracy, it was noted that the increment or decrement was not statistically significant when SNR was greater than 0 dB; however, it became statistically significant when SNR was 0 dB or lower. This finding was consistent with the existence of turning points as noted earlier.

4. Discussion

This study provides physiological evidence indicating that the subcortical, pre-attentive processing of voice pitch in the human brainstem is relatively tolerant of background noise. Specifically, effects of the SNR became significant only when the SNR was degraded to roughly 0 dB or lower. This finding is particularly important when evaluating auditory integrity in infants, children, and difficult-to-test populations. For example, the SNR turning point around 0 dB indicates that the intensity of stimulus token is recommended to be at least equal to that of the background noise if proper audibility of the stimulus token is to be ensured.

This finding is consistent with the fact that Russo et al. (2004)5 and Song et al. (2010)9 were able to record steady-state neural responses from the human brainstem by using a speech syllable, /da/, at 80 dB SPL with an SNR of 5 and 10 dB, respectively. The sustained responses are typically evoked by the steady-state portion in a stimulus, such as vowel portions. Using a vowel–consonant–vowel syllable to evoke neural responses from guinea pigs, Cunningham et al. (2002)16 demonstrated that the sustained responses evoked by the vowel portions were immune to noise interruption. For human listeners, Song et al. (2010)9 also demonstrated that the amplitude of steady-state portions of the response remained relatively stable in noise conditions.

Different types of noise may have differential influence on the FFR to voice pitch. For example, Song et al. (2010)9 used multi-talker babbles as background noise to provide different levels of masking, with the six-talker babble spectrally denser than the two-talker babble. They observed that although SNR was unchanged, responses to the formant transition of the stimulus token degraded more in the six-talker babble than in the two-talker babble condition. However, the steady-state portion of the response did not change significantly in the multi-talker babble conditions. It is anticipated that data obtained from systematic evaluation of the effects of different types of noise (e.g., multi-talker babbles versus white noise) can shed light on pitch processing mechanisms in the human brainstem.
The phenomenon of noise tolerance has also been observed through behavioral psychoacoustic experiments. For example, Patel et al. (2010) reported that speech intelligibility in noise was greater for Mandarin speakers when a natural pitch contour was used compared to a flat F0. Recent FFR literature has also reported that the human brainstem is better at tracking pitch contours that are specific to the listener’s native language. Krishnan et al. (2009) recently reported that the human brainstem was more sensitive only to a naturally rising pitch contour that was specific to the listener’s native language and was less sensitive to unnatural pitch contours that were manipulated and deviated from the natural pitch contours. The human brainstem also had a differential preference to specific pitch contours over others. Krishnan et al. (2009) recorded FFRs in response to tone 2 with naturally rising, linear-ramped, and inverted-curvilinear pitch contours. They found that the Tracking Accuracy in a group of Chinese participants was larger than the English group in response to the naturally rising pitch contour but not with the linear-ramped or inverted-curvilinear pitch contours. These recent findings support a top-down mechanism that has been found evident in electrophysiological studies for human participants.

References and Links