Variability in Voice Fundamental Frequency of Sustained Vowels in Speakers With Sensorineural Hearing Loss

*,†,‡Guo-She Lee, *†‡Taipei, Taiwan

Summary: In a previous study, the low-frequency modulation extent (LFP) of the vocal fundamental frequency (F_0) showed a significant increase in the presence of binaural noise masking for the healthy individuals. This study was to investigate the F_0 of subjects with sensorineural hearing loss (SNHL) using sustained phonations to explore the changes of F_0 modulations in SNHL. Twenty-three SNHL subjects and 14 age-matched subjects without hearing loss were enrolled in the study. Sustained vocalizations of vowel /a/ for more than 5 seconds were digitally recorded. The F_0 contour of each phonation was acquired using digital signal processing. The modulation extent at different frequencies was obtained using Fourier transformation of F_0 contour. The LFP of F_0 (<3 Hz) was significantly greater for the SNHL subjects ($P < 0.001$, independent samples t test). Although the correlation analysis was limited to the auditory-evoked brainstem response (ABR) thresholds because of their disagreement with the pure-tone thresholds in some subjects with functional hearing disorder, the correlation between LFP and ABR thresholds was significant ($\rho = 0.45$, $P = 0.03$, Spearman's correlation analysis). The LFPs of F_0 were significantly greater for the SNHL subjects and the changes of F_0 modulations could be detected using power spectral analysis of F_0. The method may be used for evaluation of audio-vocal feedback in SNHL.

Key Words: Vocal fundamental frequency–Power spectrum analysis–Sensorineural hearing loss–Audio-vocal reflex.

INTRODUCTION

Human phonation control is closely related to auditory function. In individuals with hearing loss, several abnormalities of speech production have been observed, such as greater vocal intensity[1,2] and a higher vocal fundamental frequency (F_0).[3] In addition, measures of the variability of F_0, such as frequency perturbations and standard deviation (SD) of F_0 (pitch sigma), are greater for the hearing-impaired subjects than the normally hearing controls.[1] In hearing-impaired subjects, the vocal F_0 is higher and after cochlear implantation is significantly lowered approaching normative values.[4–7] These studies provide evidence of partial phonation control by the auditory system.

To achieve precise control of pitch, loudness, and intelligibility when speaking, both open-loop and closed-loop neurological controls are involved. The pitch-shift reflex is a reflex that produces a "compensatory" response in voice F_0 that is opposite in direction to a change in voice pitch feedback (pitch-shift stimulus), thus correcting for the discrepancy between the intended voice F_0 and the feedback pitch.[8] This feedback revealed by the pitch shift stimulus (PSS) is an example of F_0 control using both open-loop and closed-loop controls.[9,10] Both types of the control mechanisms help achieve and maintain F_0 at the desired level. However, even in a steady-as-possible sustained vocalization, the phonation system is not able to keep the F_0 perfectly and constantly at a given level. There exist instabilities in phonation airflow, neuromuscular activity, and articulations that cause fluctuations in the F_0 of vocal fold vibrations. F_0 jitter and pitch sigma are two measurement examples used to describe F_0 variability.

Control of pitch is affected when auditory functions are impaired. According to previous studies that have measured F_0 responses to binaural noise masking, modulations of the F_0 increase significantly in a low-frequency range of less than 3 Hz.[11,12] In this study, using power spectral analysis of vocal F_0, the F_0 modulations of subjects with sensorineural hearing loss (SNHL) were compared with the normally hearing subjects to investigate if there are similar changes of F_0 modulations as the normally hearing subjects in binaural masking. Moreover, the possibility of accessing the phonation changes associated with hearing impairment is tested using this method, as well.

PATIENTS AND METHODS

Subjects and voice sampling

Subjects with hearing loss, who were requesting a certification of their hearing handicap at an outpatient clinic, were enrolled in this study. Exclusion criteria included a medical history of neurological disorders, such as Parkinson's disease or stroke. All denied having any experience using a hearing aid. All together 23 SNHL subjects (17 male, 6 female; age 43–79 years, median 66 years) were recruited. The unaided hearing threshold levels of the subjects were measured with a clinical audiometer (GSI 16; Lucas-Grason-Stadler Inc., Littleton, MA) in a sound-treated booth, at pure-tone threshold levels of 250; 500; 1,000; 2,000; 4,000; and 8,000 Hz. The hearing thresholds at 500; 1,000; and 2,000 Hz were averaged to represent the average hearing threshold for each subject. To detect response of the functional hearing disorder, auditory-evoked brainstem responses (ABRs) using click stimuli (Navigator Pro Loader Version 3.00; Bio-Logic System Corp., Mundelein,

Accepted for publication October 6, 2010.
From the *Department of Otorhinolaryngology, Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan; †Institute of Brain Science, School of Medicine, National Yang-Ming University, Taipei, Taiwan; and the ‡Department of Otorhinolaryngology, Ren-Ai Branch, Taipei City Hospital, Taipei, Taiwan.
Address correspondence and reprint requests to Guo-She Lee, Department of Otorhinolaryngology, Faculty of Medicine, School of Medicine, National Yang-Ming University, No. 155, Section 2, Li-Nong Street, Bei-Tou District, Taipei 112, Taiwan. E-mail: guosheli@ms12.hinet.net, gslee@ym.edu.tw
Journal of Voice, Vol. 26, No. 1, pp. 24-29
0892-1997/$36.00

doi:10.1016/j.jvoice.2010.10.003

IL) were used to evaluate the hearing thresholds in all the SNHL subjects. The study protocols were approved by the institutional review board of National Yang-Ming University (IRB No. 960014).

The subjects were instructed to make two steady-as-possible sustained 5-seconds vocalizations of vowel /a/ at their comfortable speech level. The voice signals were collected using a dynamic microphone with flat frequency response from 31.5 Hz to 8 kHz (DS-101; Tenmars Electronics, Taipei, Taiwan). Next, the signals were digitally sampled and recorded using an IBM personal computer compatible sound adapter at the rate of 44.1 kHz. The voice sound pressure level was measured with a sound level meter (DS-101; Tenmars Electronics, Taipei, Taiwan) that was maintained at the distance of 15 cm in the front of mouth by the examiner's hand. The vocal intensity was digitally sampled at 100 Hz.

Another 14 subjects (7 male, 7 female; age 48 $\sim$ 72 years, median 58 years) without hearing loss were enrolled as the control group. The pure-tone averages at 500; 1,000; and 2,000 Hz of the control subjects were all less than 25 dB hearing level (HL). The vowel produced, recording devices, and hardware settings were the same as those used for the SNHL group. In a similar manner to the subjects with hearing loss, two /a/ phonations were collected.

Calculation of F_0 and frequency perturbation

To avoid the phonation irregularity at the beginning of a vocalization, the 0.5-seconds signals after the voice onset were bypassed for analysis. The 5-seconds signals after the bypass point were then retrieved for the calculations of F_0. To get every beat of F_0 of a phonation, the first 20-milliseconds voice signals after the bypass point were retrieved as the first window. The fundamental period of the window was then determined by the interval at which the autocorrelation function was maximal. To retrieve the F_0 more correctly, the analytic window should better include at least two glottal cycles. After the first fundamental period was acquired, the analytic window was then shifted by this fundamental period. The following 20-milliseconds signals were submitted to the same calculation process to get the fundamental period of the next window. Consecutively, every fundamental period of the phonation was obtained one by one. All fundamental periods were then transformed to F_0s using their reciprocals. Besides, during the signal processing for retrieving the F_0, each analytic window should include at least two glottal cycles for correctly obtaining the F_0.

Variability of the F_0 is classically expressed by pitch sigma, and the pitch sigma had been found to increase under various vocal pathologies[13–15] and hearing loss.[16] In this study, the variation of the F_0 ($VF_0\%$) was calculated by dividing the pitch sigma by the mean F_0 of each 5-seconds vocalization and was expressed as a percentage. This value represents the F_0 variability relative to the baseline F_0. Jitter is a short-term (cycle-to-cycle) perturbation of F_0, and the relative average perturbation (RAP) is a mean measurement of jitter, which can be used to evaluate a voice clinically.[17,18] The RAP used here was the perturbations over an average of three sequential periods, which is the same as in previous studies.[11,12]

Resampling of F_0 and conversion of cent

An evenly sampled time sequence is necessary for power spectral analysis. However, the beat periods of a phonation are different from window to window. Therefore, the F_0s derived from the reciprocals of fundamental periods are not evenly spaced in time. A correction is required to get an evenly sampled time sequence contour of F_0 using the linear interpolations of F_0 at a constant period, for example, 20 milliseconds in this work. This process is called resampling of F_0 using linear interpolation, and the resampling rate using a 20-millisecond period is 50 Hz.

The lowest tracing of Figure 1 shows the voice signal of a 100-millisecond window in a phonation. In the figure, the F_0s (open circle) were marked at each fundamental period of the vocal fold vibrations. The F_0 contour showed that the F_0s are not evenly time spaced. Therefore, the contour has to be reconstructed by resampling for power spectral analysis. While resampling of F_0 at the rate of 50 Hz, the F_0s have to be presented at the interval of 20 milliseconds. In fact, the F_0 values do not always exist at that interval because the fundamental periods vary from time to time. Interpolations by the preceding and following F_0s were used to obtain a time evenly spaced contour of F_0 (filled circle).

In addition, because the F_0 of adult females is often twice that of adult males, the F_0 values were converted to cents to allow comparisons across the subjects. A similar conversion has been used when investigating the audio-vocal reflex of the normally hearing subjects using the pitch-shift technique.[10,19] Thus, before power spectral analysis, the resampled F_0s were first converted to a sequence of cents using the following equation:

$$\text{cent} = 1,200 \times \log_2\left(\frac{f}{\overline{F}_0}\right)$$

where $\overline{F}_0$ is the mean of all F_0s of a 5-seconds voice sample and f is the frequency to be converted to cent.

Power spectral analysis of the F_0 sequence

The power spectrum of the F_0 was derived using fast Fourier transformation of the F_0 contour (in cent). The division of

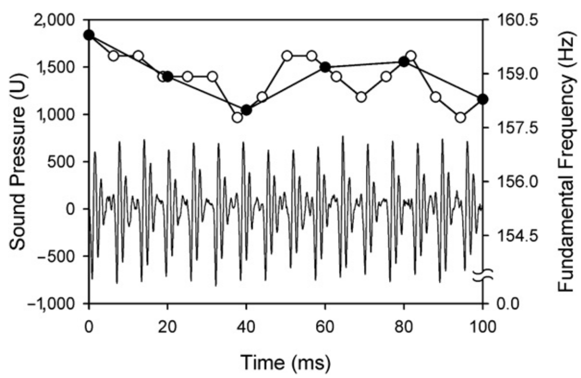

FIGURE 1. The voice signals of a 100-milliseconds window in a phonation (*lowest tracing*), the contour of the individual fundamental frequencies (F_0s) (*open circle*), and the contour of the resampled F_0s (*filled circle*) by interpolations at the interval of 20 ms (50 Hz).