Voice Low Tone to High Tone Ratio - A New Index for Nasal Airway Assessment

Guoshe Lee[1,4], Cheryl C. H. Yang[2] and Terry B. J. Kuo[3]

[1]*Department of Otolaryngology*
Hualien General Hospital Department of Health
Hualien 970
[2]*Department of Physiology*
[3]*Institute of Neuroscience and*
[4]*Institute of Medical Science, Tzu Chi University*
Hualien 970, Taiwan, R.O.C.

Abstract

There are several methodology based on voice analysis to evaluate nasal airway. Here we introduce a new quantitative index based on voice spectrum analysis to evaluate nasal obstruction. Ten subjects of nasal blockage were instructed to produced the sustained consonant-vowel syllable /m∂/ at comfortable levels of speech for at least 5 seconds. After nasal decongestant treatment, the second voice sample was collected. Sound spectrum was obtained by the algorithm of fast Fourier transform and the fundamental frequency (F0) was calculated by the method of autocorrelation. Voice low tone to high tone ratio (VLHR) was defined as the division of low frequency power (LFP) into high frequency power (HFP) of the sound power spectrum and was finally expressed in decibels. The cut-off frequency was the product of F0 and $\sqrt{}(4 \times 5)$. The VLHR after nasal decongestant treatment increased significantly as compared with that before treatment (P < 0.01). VLHR is a new index derived from sound spectral analysis and that may detect the changes in frequency characteristics of voice during treatment for nasal obstruction. The index is quantitative, non-invasive, and potentially useful for basic researches and clinical applications.

Key Words: voice, nasal airway, nasal decongestion, power spectral analysis

Introduction

Hyponasality resulting from acute rhinitis during upper airway infection (URI) or acute attacks of nasal allergy can be promptly identified by experienced experts with their sole ears. The impaired nasal airway can be also detected by sound analysis like nasometric measurement (3). Nevertheless, the voice change in patients with hyponasality can be easily detected by the human ears, and the acoustic mechanisms coupled to human ears are basically derivations of Fourier transform. Knowing this, power spectrum analysis of a voice is potential to provide a sensitive way of evaluating hyponasality and nasal airway obstruction.

Thanks to major progress in computer science in recent years, various medical diagnostic facilities, such as imaging systems, ECG, audiometry, etc., have been relayed to the digital computer systems to obtain sophisticated data analyses. Sometimes automatic diagnoses can be even made. The purpose of this study was to develop a quantitative index for nasal obstruction and even nasality by utilizing personal computer (PC)-based equipment. Here we offered a hypothesis - the spectral content of human voice is related the patency of nasal airway and thus can be altered by nasal decongestion.

Materials and Methods

Patients Selection

Ten patients were enrolled in this study, 5 males and 5 females. The mean age of the subjects was

Corresponding author: Dr. Terry B. J. Kuo, Institute of Neuroscience, Tzu Chi University, 701, Section 3, Chung Yang Road, Hualien 970, Taiwan. Tel: 886-3-8565301 ext. 7513, Fax: 886-3-8580639, E-mail: tbjkuo@ms33.hinet.net.
Received: March 7, 2003; Revised: May 1, 2003; Accepted: May 10, 2003

24.1 years with a range of 14 to 44 years. They complained of nasal obstruction during acute rhinitis at our outpatient department. Subjects with a history of hearing impairment, significant heart disease for which epinephrine nasal packing was contraindicated, or speech disorders were excluded from the study. The nasal cavities were carefully examined, and subjects with obvious anatomical and/or pathological abnormalities such as severely deviated nasal septum, nasal polyposis, or synechia were also excluded.

Sampling of Sound

All subjects were instructed to phonate the consonant-vowel (CV) syllable /m∂:/ for at least 5 s at a comfortable phonation level. Sound samples were collected with a dynamic microphone (JVC model MD480A), which was hooked to an IBM-compatible PC, in a quiet room. Recordings were made in digital format. Sampling rates were adjusted to 22 kHz, and the data width was set to 16 bits. After the first voice recording, bilateral nasal cavities of patients were packed with cotton pads, which were previously soaked in a 1:1000 epinephrine solution, for 10 min to ensure adequate decongestion of the nasal mucosa. The nasal obstruction of all patients improved subjectively after treatment, and patency of the nasal airway was also confirmed under direct inspection. Nasal secretions were cleaned by suctioning, and another voice recording was made under the same recording conditions.

Measurement of Voice Low Tone to High Tone Ratio (VLHR)

The onset of the voice signal was determined by an abrupt rise in sound energy, which was calculated by the root mean square (RMS) of the sound wave data segmented by 1 ms, from background noise. Often, the energy of the signal produced by the semivowels surged as high as 2-fold over the background noise and could be clearly identified. Data for the first 500 ms after voice onset were omitted, and data of the next 3 seconds were segmented for further analysis.

The fundamental frequency (F0) of phonation was then calculated by the method of autocorrelation. The power spectrum of sound wave was obtained by the algorithm of fast Fourier transform (FFT) without data segment overlapping. The power spectrum was then divided into a low-frequency band and a high-frequency band by the value (fcut) of the fundamental frequency times the geometric average of 4 and 5, that is $\sqrt{(4 \times 5)} = 4.47$ F0. The power of the low-frequency band (LFP) was defined as the summation of the power of each frequency component ranging from 65 Hz to less than 4.47 F0, while the power of the high-frequency band (HFP) was defined as the summation of the power of each frequency component ranging from 4.47 F0 to 8000 Hz. The equations were listed below:

$$LFP = \sum_{i=65}^{fcut} p_i \quad [1]$$

$$HFP = \sum_{i=fcut}^{8000} p_i \quad [2]$$

Where p_i represents the power of the frequency component i Hz. When LHF was divided into HFP, a new index, the voice low tone to high tone ratio, was obtained. The VLHR of each data window in the entire 3-s samples were averaged together to represent the averaged VLHR ($VLHR_{ave}$) of this 3-s sound.

$$VLHR_{ave} = (\sum_{i=0}^{3sec} R_i)/n \quad [3]$$

Where R_i is the VLHR of the i seconds, n is the total calculation windows of FFT in the entire 3-s sound. This ratio was eventually expressed in decibels - VLHR like that was commonly used for description of sound energy.

$$VLHR = \log_{10}(VLHR_{ave}) \times 10 \quad [4]$$

Statistical Analysis

The data before and after nasal treatment were compared with the paired Student's *t*-test. Statistical significance was assumed for $P < 0.01$.

Results

Two snapshots of the sound power spectrogram of one subject before and after nasal decongestant treatment are shown in Fig. 1. A slight decrease in the fundamental frequency was observed after nasal treatment, and thus the cut-off frequency between the high-frequency band and the low-frequency band decreased concurrently. The LFP (the area under the curve and to the left of the cut-off frequency) obviously increased after nasal treatment.

Continuous analysis of the voice before and after nasal decongestant treatment is illustrated in Fig. 2. In this example, the sound intensity, which was expressed in RMS, after nasal treatment were less than those before nasal treatment; however, the VLHR increased significantly and independently of the sound intensity. The correlation coefficient between RMS and VLHR was as low as 0.006 and the *P* value was 0.80 (Fig. 3). The fundamental frequency

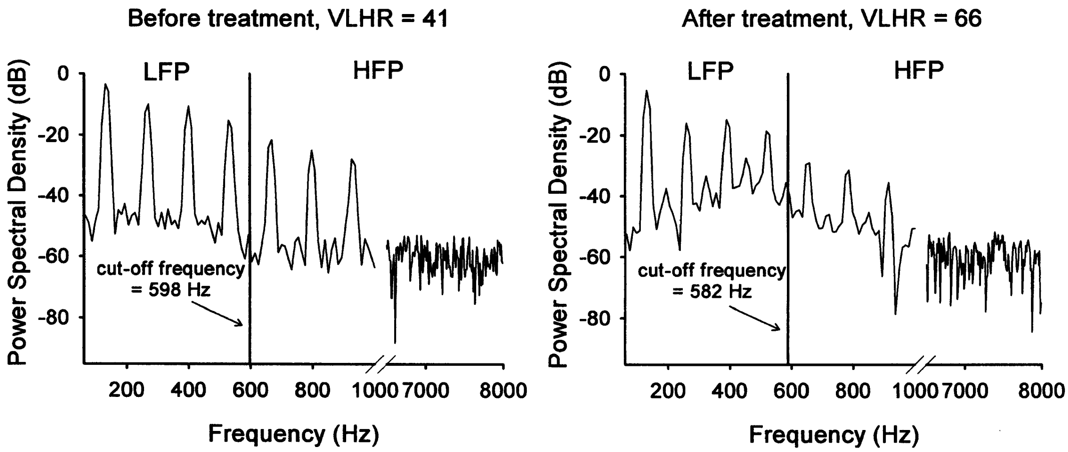

Fig. 1. Two snapshots of the sound sequence in the frequency domain before and after nasal packing in a study subject. The low-frequency power (LFP) and high-frequency power (HFP) was separated by the cut-off frequency, which is the geometric average of the frequencies of the 4th and 5th harmonics. The LFP increased significantly after nasal packing. Therefore, the voice low tone to high tone ratio (VLHR) also increased after nasal packing.

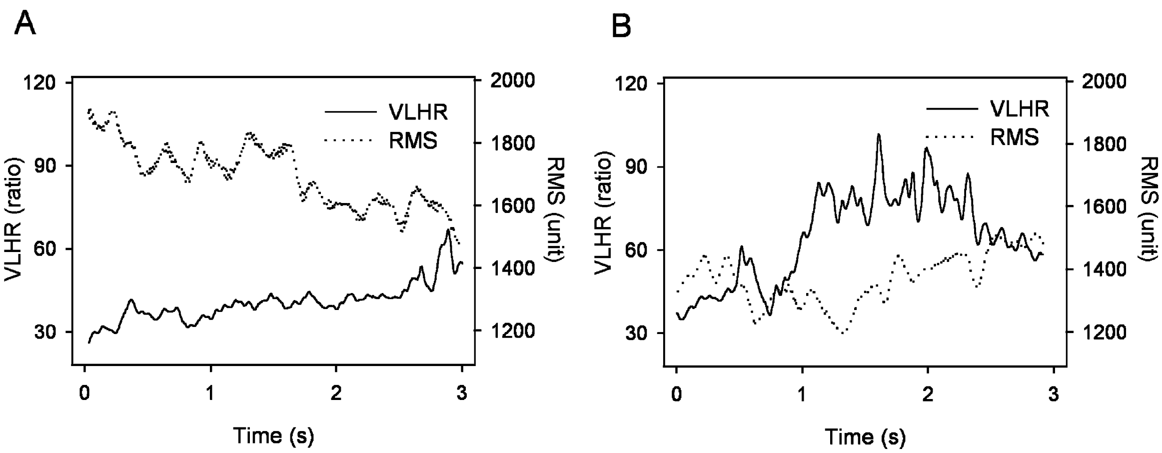

Fig. 2. The sound intensity and voice low tone to high tone ratio (VLHR) before nasal decongestion treatment (A) and after nasal decongestion treatment (B) in a study subject. The sound intensity is quantified by root mean square (RMS).

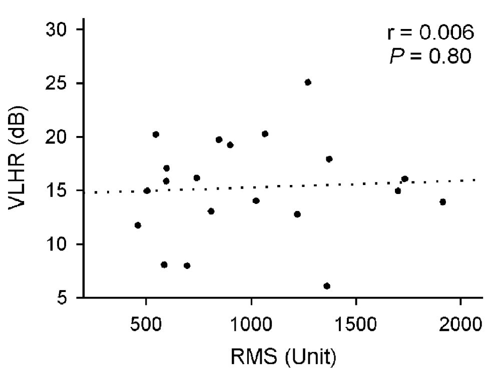

Fig. 3. Correlation analysis of sound intensity and VLHR. The sound intensity is quantified by root mean square (RMS). There was no significant correlation between RMS and VLHR. The correlation coefficient (r) was 0.006 and P value was 0.80.

and sound intensity before and after nasal treatment are listed in Table 1. The cut-off frequency of our study population was 729 ± 210 Hz (mean ± SD). The *P* values of the paired t-test for fundamental frequency and sound intensity before and after nasal decongestion were 0.68 and 0.60, respectively, and no significant difference was found. When the VLHR before and after nasal treatment was examined by paired *t*-test, the *P* value of was 0.0018 (Fig. 4).

The frequency resolution is related to the window sizes used for FFT. VLHR became more stationary when frequency resolutions were less than 100 Hz resolution. The window size of 2048 dots was selected in this study and the frequency resolution was estimated about 10 Hz.

Table 1. Fundamental frequency and sound intensity before and nasal decongestant treatment.

Case No.	Sex	Fundamental Frequency (Hz)		Sound Intensity (RMS unit)	
		Before Treatment	After Treatment	Before Treatment	After Treatment
1	F	151	176	504	597
2	F	234	236	846	1270
3	F	187	181	740	545
4	F	240	224	899	1065
5	F	205	199	459	594
6	M	141	142	808	1022
7	M	118	124	1913	1700
8	M	104	113	1360	1219
9	M	134	130	1732	1372
10	M	110	113	694	584

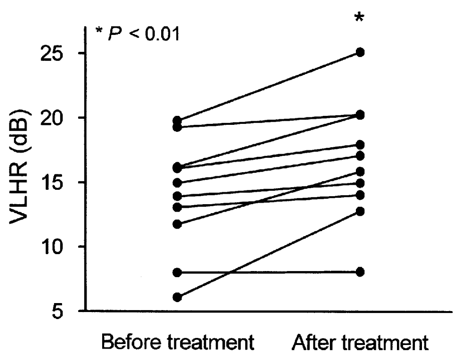

Fig. 4. Effect of nasal decongestion treatment on VLHR of all 10 subjects. *$P < 0.01$ by paired *t*-test.

Discussion

In this communication, we described a new index to detect nasal obstruction. Our methodology is based on power spectral analysis of voice signals. Our results revealed that spectral contents of human voice were closely related to nasal airway. The change of voice spectrum after nasal decongestion can be detected and quantified by VLHR as we described above. With the application of such technique, it will be possible that the evaluation of nasal obstruction and related problems, for example follow-up of nasal obstruction during URI, can be assisted automatically by a personal computer equipped with a microphone.

There have been various methodologies to quantify nasal obstruction. The methodology of sound pressure measurement seems to be more popular and widely used for nasal speech evaluation. The use of nasometry in identifying and evaluating hypernasality has been documented since its invention in 1970 (5). The parameter "nasalance" which was first measured by the oral nasal acoustic ratio system (4) is now widely used for assessment of nasality, velopharyngeal function (2), palatal fistula (7) and management of nasalized speech disorders. The correlation between nasalance and hyponasality was established in the late 1990s. An inverse correlation between nasalance score and nasal airway resistance was proven by the reduction in nasalance scores after decreases in nasal airway resistance induced by topical nasal decongestants (9). However, all these present methodologies call for special equipment and maneuvers thus are not very easily assessable.

Frequency-domain characteristics of voice in nasalized speech, such as a reduction in the intensity of the first formant, the presence of extra-resonances, and increased bandwidth of formants (1), have been reported since the 1950s. However, different vowels have different acoustic "nasalization" characteristics because of the complex interaction of the resonance characteristics of the nasal cavity with variations in vocal tract shaping. Besides, individual variations in vocal tract anatomy lead to differences in acoustic correlates of nasality from one speaker to another. Inconsistencies among speakers and among phones have made power spectral analysis of voice incapable of defining the unique features and quantifying the "expression of nasality". Nevertheless, changes in the intensity of each harmonic of the power spectrum contribute to the perception of nasality, although there are usually concurrent changes in fundamental frequencies. The summation of each harmonic change would be more constant and discriminative than would the change in a single frequency or even a single formant. Based on this, we divided the harmonics into a high-frequency band and a low-frequency band according to the fundamental frequency. The power of each frequency band was individually summed, and 2 values, i.e., the HFP and LFP, were derived. Our experimental data revealed that the LFP to HFP ratio was sensitive to changes in patency of the nasal airway. Similar methods have been applied to analysis of other biological signals such as heart rate variability (6), in which the ratio of lower frequency power to high frequency power on power spectrum of heart rate variability has been used to indicate sympathetic activity (8, 10).

The nasal airway becomes more patent after nasal treatment with decongestant, meanwhile the VLHR rises. Our data imply that the nasal airway acted somewhat like a "low-pass filter" which allows more low-frequency sound to pass through it or acted somewhat like a "high-cut filter" which blocks high-frequency sound to pass through it. At least that is true for the sustained CV syllable /m∂/ in our experiment. Detailed underlying mechanisms warrant further exploration. The VLHR was ultimately expressed in the format of dB to be statistically useful since the sound intensity is also expressed in the format of dB clinically. This index can potentially and easily provide another quantification method for the patency of the nasal airway, although there are several traditional measurements such as rhinomanometry or acoustic rhinometry available in routine clinical applications. The only measurement required here is to record a 3-s voice of a CV syllable.

We chose this CV syllable /m∂/ for tests because this CV is a nasal vowel and can be pronounced easily. We have tried other CVs, and obtained similar results.

The nasal airway of normal subjects was not identical to that of patients after nasal decongestion in this study. We followed the decongestion method of the previous study (3) in order to demonstrate the applicability of this index.

The VLHR was independent of sound intensity since the correlation coefficient (r) was low and *P* value was not significant. The sound intensity varies in different sound recording conditions such as the sensitivity of microphone, the distance between sound source and recorder, the subjects' phonation status .. etc. However, we simply divided the LFP to HFP in order to eliminate the effects of sound intensity on the assumption that the change of sound intensity occurs proportionally in both LFP and HFP. This change would be cancelled by the calculation of "division" into each other. This might explain the robust nature of VLHR to intensity variance, and even "acceptable" noisy conditions.

Several frequency values were used to divide the power spectrum into a high-frequency band and a low-frequency band. The *P* values for 2.45 F0 ($\sqrt{(2\times3)}$ F0) and 3.46 F0 ($\sqrt{(3\times4)}$ F0) were 0.42 and 0.017 respectively. We finally selected 4.47 F0 ($\sqrt{(4\times5)}$ F0) to be the cut-off frequency of choice because the *P* value was the most significant of the three and the results were more consistent and discriminative. The sound samples of the first 0.5 s were omitted because the ratio rose abruptly at voice onset and became more stable 0.5 s later. The power of each frequency component was important for summation. Thus a narrow-band spectral analysis (bandwidth of 10 Hz) was used to obtain fine frequency resolution and stationary values of VLHR.

The VLHR derived from sound spectral analysis sensitively reflects the change in frequency characteristics of voice induced by nasal decongestant in nasal obstructed patients. It can be regarded as a quantification method of nasal patency or obstruction and might be closely related to human perception of nasality since it directly reflects sound spectrum changes. Measuring the VLHR is non-invasive and requires only simple hardware (a personal computer equipped with a microphone). The monitoring of nasal airway will be easy and non-invasive, and more importantly even through the resources of internet in the near future.

Acknowledgments

This study was supported by grants form the National Science Council, ROC (NSC 89-2314-B-320-016).

References

1. Curtis, J.F. The acoustics of nasalized speech. *Cleft Palate J*. 7: 480-496, 1970.
2. Dalston, R.M., Warren, D.W. and Dalston, E.T. The identification of nasal obstruction through clinical judgments of hyponasality and nasometric assessment of speech acoustics. *Am. J. Orthod. Dentofacial Orthop*. 100: 59-65, 1991.
3. Dalston, R.M., Warren, D.W. and Dalston, E.T. A preliminary investigation concerning the use of nasometry in identifying patients with hyponasality and/or nasal airway impairment. *J. Speech Hear. Res*. 34: 11-18, 1991.
4. Fletcher, S.G. Theory and instrumentation for quantitative measurement of nasality. *Cleft Palate J*. 7: 601-609, 1970.
5. Fletcher, S.G. and Bishop, M.E. Measurement of nasality with TONAR. *Cleft Palate J*. 7: 610-621, 1970.
6. Kuo, T.B.J., Lin, T., Yang, C.C.H., Li, C.L., Chen, C.F. and Chou, P. Effect of aging on gender differences in neural control of heart rate. *Am. J. Physiol.* 277: H2233-H2239, 1999.
7. Pinborough-Zimmerman, J., Canady, C., Yamashiro, D.K. and Morales, L., Jr. Articulation and nasality changes resulting from sustained palatal fistula obturation. *Cleft Palate Craniofac. J*. 35: 81-87, 1998.
8. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation* 93: 1043-1065, 1996.
9. Williams, R.G., Eccles, R. and Hutchings, H. The relationship between nasalance and nasal resistance to airflow. *Acta Otolaryngol*. 110: 443-449, 1990.
10. Yang, C.C.H., Kuo, T.B.J. and Chan, S.H.H. Auto- and cross-spectral analysis of cardiovascular fluctuations during pentobarbital anesthesia in the rat. *Am. J. Physiol*. 270: H575-H582, 1996.