Original Article

J Chin Med Assoc
2004;67:179-184

Voice Acoustic Analysis of Normal Taiwanese Adults

Chen-Chi Wang
Hui-Tsu Huang

Department of Otorhinolaryngology, Head and Neck Surgery, Taichung Veterans General Hospital, Taichung, Taiwan, R.O.C.

Key Words
acoustic analysis;
fundamental frequency;
harmonics to noise ratio;
maximal phonation time;
perturbation

Background. Because of advances in voice research, voice acoustic analysis including fundamental frequency (F0), sound pressure level (SPL), jitter, shimmer, harmonics-to-noise ratio (H/N ratio), and maximum phonation time (MPT) can now be easily recorded and analyzed with a computer. Because these systems are widely used in clinical practice, this study was designed to establish the normal acoustic analysis parameters in normal Taiwanese adults.

Methods. From Mar. 2002 to Dec. 2002, 45 Taiwanese women and 45 Taiwanese men younger than 50 years old were recruited as subjects for this study. The commercially available Computer Speech Lab and Aerophone II system manufactured by Kay Elemetrics Corp. were used to record the aforementioned acoustic data under comfortable phonation. Each gender was separated equally into 3 age subgroups. Then differences between gender and age subgroups were investigated by statistics software SPSS 10.0. Our results were compared with data from previous reports.

Results. The value of F0 (counting; vowel/a/) was greater for females (203.2 ± 21.7; 213.4 ± 25.4 Hz) than for males (118.3 ± 17.3; 121.3 ± 16.4 Hz). Conversely, the value of MPT was greater for males (28.0 ± 9.4 sec) than for females (22.6 ± 7.6 sec). There were no significant differences in average SPL and jitter between female (77.8 ± 5.5 dB; 0.66 ± 0.27%) and male (77.5 ± 5.5 dB; 0.56 ± 0.23%) subgroups. Except for shimmer and H/N ratio in the male subgroup, there were no differences in parameters within both gender subgroups with respect to age.

Conclusions. We have developed a body of normal data for various parameters of acoustic analysis in different age groups and genders. It seems that the majority of voice characteristics of adults were relatively stable and did not change with aging between 20 and 49. But the shimmer and H/N ratio were variable in different genders and age groups. However, the voice characteristics of adults older than 50 years old were not recorded in this study and therefore require further investigation.

Examination of voice disorders can be carried out through a variety of techniques. Evaluation of the voice by the ears of speech pathologists or laryngoscopy is almost always subjective. Hence, voice acoustic analysis has been introduced clinically for recording objective data before and after surgery of voice disorders to help us better understand the pathophysiology of voice production.

Voice acoustic analyses include several parameters: (1) average fundamental frequency (F0) and sound pressure level (SPL) to reflect habitual pitch and loudness (intensity or amplitude); (2) perturbation (jitter and shimmer) and harmonics-to-noise ratio (H/N ratio) to characterize voice quality; and (3) maximum phonation time (MPT) to evaluate maximum vocal capabilities.[1] The measure of fundamental frequency, reflecting the vibratory rate of the vocal folds, can be measured during production of sustained vowels or during a reading passage. During sustained vibration, the vocal folds will exhibits slight variation of fundamental frequency and amplitude from 1 cycle to the next; these phenomena are called frequency perturbation (jitter) and amplitude perturbation (shimmer).[2] In addition to jitter and shimmer, degree of hoarseness can be evaluated by judging the ex-

Received: May 28, 2003.
Accepted: February 20, 2004.

Correspondence to: Chen-Chi Wang, MD, Department of Otorhinolaryngology, Head and Neck Surgery, Taichung Veterans General Hospital, 160, Sec. 3, Taichung-Kang Road, Taichung 407, Taiwan.
Tel: +886-4-2359-2525 ext. 5401; Fax: +886-4-2359-6868; E-mail:ccwang@vghtc.gov.tw

tent to which noise replaces the harmonic structure in the spectrogram of a sustained vowel. As the degree of judged hoarseness increases, the noise component appears to a greater degree and replaces the harmonic structure in the spectrogram. The relationship between these 2 components has been quantified as H/N ratio.[3] Maximum phonation time is the maximum time a person can sustain a tone on one continuous expiratory breath. It supposedly is a measure of phonatory control and respiratory "support".[2]

All of the aforementioned parameters can now be easily recorded and analyzed with a computer, a technique that is widely used in clinical practice. Many parameters of normal adults in western countries are well-documented in the literature.[2] However, there are only few papers about the acoustic analysis of voice for normal adults in Taiwan. In 1985, Chen[4] used phonolaryngograph SH-01 and vowel /u/ to analyze fundamental frequency, intensity and air flow rate of Chinese normal adults. However, other parameters such as maximal phonation time, jitter, shimmer and harmonics-to-noise ratio were not discussed. In 2002, Chu *et al*.[5] used Dr. Speech software and vowel /ee/ to record the acoustic data of 20 young male and 20 young female aged between 20 and 26 years old. But in clinical practice, the ages of many patients fall out of this range; whether aging will influence analysis results should be further investigated. Hence we designed a study to gather many acoustic data from different age groups and genders. The results from the aforementioned articles were also summarized and compared to our data. Our goal was to set up tables of normal acoustic data for quick reference in clinical practice and to better understand the voice characteristics of Taiwanese adults.

METHODS

From Mar. 2002 to Dec. 2002, 45 Taiwanese women and 45 Taiwanese men, none of whom had professional training in singing, were recruited as subjects for this study. All test subjects were between 20 and 49 years old. The 45 female subjects were separated equally into 3 age groups: group I for 20 to 29 y/o, group II for 30 to 39 y/o, and group III for 40 to 49 y/o. The 45 male subjects were separated as well into group IV for 20 to 29 y/o, group V for 30 to 39 y/o, and group VI for 40 to 49 y/o. Each group contained 15 subjects. They all met the criteria of being non-smokers and having no upper respiratory tract infection for 3 weeks prior to the test. The second author, an experienced speech therapist, listened to the voice of subjects to make sure that no 1 had perceptual voice abnormalities. In addition, the first author checked their vocal folds with a flexible laryngoscope to confirm that no 1 had organic lesions of the vocal folds.

The second author performed the acoustic examination in a soundproof room with the test subjects in a sitting position. We used the commercially available Computer Speech Lab system (CSL) and Aerophone II (AP II) (Kay Elemetrics Corp., Lincoln Park, NJ) to record the data described below. The subject was instructed to count from 1 to 10 in Mandarin to first obtain the F0 (Hz) of counting. Then the data for F0 (Hz), jitter (%), shimmer (dB) and H/N ratio (dB) were recorded during stable and comfortable phonation of the vowel /a/ for more than 3 seconds. After proper instruction and several test trials, the subjects were instructed to phonate a stable vowel /a/ continuously for as long as possible, using habitual vocal pitch and loudness. Data including MPT (sec) and average SPL (dB) were recorded.

All data were analyzed with the statistics software SPSS 10.0 for Windows. We used the Kruskal-Wallis test to assess the differences between the 3 age groups of each gender. If there was no difference between 3 groups of each gender for a specific parameter, we combined the data from the 3 groups and used Student's *t*-test to compare the overall mean value between genders. If there was difference between the 3 groups for a specific parameter, then we used Mann-Whitney *U* test to analyze gender difference for each age range.

RESULTS

A total of 90 test subjects received voice acoustic analysis in this study; the mean value and standard deviation of each parameter and subject age are summarized in Table 1 for female adults and Table 2 for male adults. Table 3 shows the aggregate data of some parameters

that were not influenced by aging.

For the F0 of counting and vowel /a/, there was no significant difference between the 3 age groups for both genders (Tables 1 and 2). The results show that the F0 was rather stable between ages 20 and 49. Hence we know that the F0 plateaus and does not change before age 50 for adults. The overall mean F0 of the female group was significantly higher than that of the male group (Table 3).

For the average SPL and jitter of sustained vowel /a/, there was no significant difference between the 3 age groups for both genders (Tables 1 and 2). The results show that the average SPL and jitter were also rather stable between ages 20 and 49 for adults. Aging did not change the data of average SPL and jitter before 50 years old. The overall mean SPL and jitter of the female group were not significantly different from those of the male group (Table 3).

For the shimmer and H/N ratio of vowel /a/, there was no significant difference between the 3 female groups (Table 1). The overall mean values of the female subject are shown in Table 3. However, there was significant difference between the 3 male groups. So the data from groups IV to VI were not combined to determine the overall aver-

Table 1. Acoustic analysis data of normal female adults under comfortable phonation

Group		Age (years)	Counting F0 (Hz)	/a/ F0 (Hz)	SPL (dB)	Jitter (%)	Shimmer (dB)	H/N ratio (dB)	MPT (seconds)
I	Mean	24.9	215.1	221.9	76.3	0.69	1.65	12.2	19.5
(n = 15)	SD	2.5	22.3	19.5	4.4	0.23	1.58	3.9	5.6
II	Mean	34.5	200.0	215.1	79.6	0.73	1.84	11.3	24.0
(n = 15)	SD	3.1	20.5	23.1	6.4	0.30	2.96	3.3	9.1
III	Mean	44.4	194.4	203.1	77.6	0.55	1.99	13.1	24.5
(n = 15)	SD	2.6	17.9	30.4	5.4	0.25	3.16	4.2	7.1
K-W test[a] *p* value			0.062	0.194	0.168	0.082	0.753	0.207	0.099

[a] Kruskal-Wallis test.

Table 2. Acoustic analysis data of normal male adults under comfortable phonation

Group		Age (years)	Counting F0 (Hz)	/a/ F0 (Hz)	SPL (dB)	Jitter (%)	Shimmer (dB)	H/N ratio (dB)	MPT (seconds)
IV	Mean	23.6	121.7	125.6	76.2	0.61	0.26	8.7	30.9
(n = 15)	SD	2.2	16.6	17.5	3.9	0.25	0.09	5.3	12.8
V	Mean	33.5	110.9	122.2	79.1	0.51	0.20	7.6	25.5
(n = 15)	SD	3.0	10.5	14.0	6.7	0.23	0.06	4.4	7.0
VI	Mean	44.7	122.3	116.0	77.3	0.56	0.19	14.1	27.6
(n = 15)	SD	2.2	21.7	17.2	5.6	0.22	0.11	2.2	6.9
K-W test *p* value			0.226	0.234	0.429	0.491	0.038[b]	0.000[b]	0.587

[a]Kruskal-Wallis test; [b]$p < 0.05$ indicated significant difference.

Table 3. Comparison of acoustic analysis data between female and male adults under comfortable phonation

Group		Age (years)	Counting F0 (Hz)	/a/ F0 (Hz)	SPL (dB)	Jitter (%)	Shimmer (dB)	H/N ratio (dB)	MPT (seconds)
F	Mean	34.6	203.2	213.4	77.8	.66	1.83	12.2	22.6
(n = 45)	SD	8.5	21.7	25.4	5.5	.27	2.60	3.8	7.6
M	Mean	33.9	118.3	121.3	77.5	.56	[c]	[c]	28.0
(n = 45)	SD	9.0	17.3	16.4	5.5	.23	[c]	[c]	9.4
t-test[a] *p* value			0.000[b]	0.000[a]	0.799	0.069	[d]	[d]	0.004[b]

[a] student's t-test; [b]$p < 0.05$ indicated significant difference; [c] Aggregated data were not recommended to be used due to variation in different age groups. [d] Comparison between genders was not done.

age for male subjects. However, if we investigate gender difference in each age range with Mann-Whitney *U* test, we can find there were always significant differences between genders for shimmer. The *p* value was 0.000 for groups I and IV; 0.001 for groups II and V; and 0.002 for groups III and VI. But for H/N ratio, the condition was variable. Only the groups of age ranged from 30 to 39 years old (groups II and V) had significant difference ($p = 0.016$). Conversely, the *p* value was 0.098 for groups I and IV and 0.653 for groups III and VI.

For the MPT of sustained vowel /a/, there was no significant difference between the 3 age groups for both gender (Tables 1 and 2). The mean MPT of the female group was significantly lower than that of the male group (Table 3).

DISCUSSION

The aforementioned results demonstrate that for adults between 20 and 49 years old, the F0, SPL, jitter, and MPT are relatively stable and does not change with aging in each gender. Thus, except for shimmer and H/N ratio of males, the aggregate data (Table 3) of these parameters for each gender can represent the normal data for adults younger than 50 years old.

In clinical practice, there is a large discrepancy between subjective complaints about voice problems and vocal fold findings after examination. The need for an objective, quantifiable measure of voice quality is clearly important to establish correlation between treatments and outcomes. The goal of this study was to empirically measure vocal characteristics with respect to age and gender to establish baseline normal levels in healthy Taiwanese adults.

Fundamental frequency is an acoustic measure that directly reflects the vibrating rate of the vocal folds and is expressed in Hz. It can be measured during production of sustained vowels or while reading a passage. For simplification of the procedure, we use counting instead of reading a passage. This study revealed that the F0 of female adults was higher than that of male adults for both counting and producing vowel /a/. Because vocal pitch is regulated mainly by 3 factors: tension, mass and length of the vocal folds; in the example case of a violin, a thinner string, a shorter string, or a combination of both produces higher pitch. Therefore, the vocal pitch of women is usually higher than that of men due to the vocal folds in men being longer and thicker than those in women.[6] Comparing with other reports (Table 4), the F0 of vowel /u/ in the study of Chen[4] and the F0 of vowel /ee/ in thestudy of Chu *et al.*[5] were all higher than the F0 of vowel /a/ in our study. This could be explained by the phenomenon of "intrinsic pitch of vowels"- the tendency of high vowels (e.g., /i/ and /u/) to be produced with higher F0 than low vowels (e.g., /a/).[7] Furthermore, if we compare our data with those from Western countries (Table 4), we find there is no large discrepancy. With respect to age, Chen observed a trend of higher F0 in males from 20 to 79 years old, but a decreasing trend in F0 for females from 20 to 79. However, the difference between age groups was not significant for subjects before 50 years old in both genders. In our study, there was no significant difference between age groups either. This demonstrates aging has no influence on F0 in subjects up to 50 years old.

Various instruments employed to measure the SPL

Table 4. Comparison of fundamental frequency data between other studies and the present study

gender	age	Present study[a] counting F0 (Hz) (n = 15)	Present study[a] /a/ F0 (Hz) (n = 15)	Chen[4b] /u/ F0 (Hz) (n = 10)	Chu *et al*[5c] /ee/ F0 (Hz) (n = 20)	Studies from Western countries[2]
F	20-29	215.1 ± 22.3	221.9 ± 19.5	226 ± 25.0	238.7 ± 28.5	224
	30-39	200.0 ± 20.5	215.1 ± 23.1	216 ± 18.4	No data	196
	40-49	194.4 ± 17.9	203.1 ± 30.4	208 ± 16.9	No data	189
M	20-29	121.7 ± 16.6	125.6 ± 17.5	125 ± 18.3	141.7 ± 20.8	120
	30-39	110.9 ± 10.5	122.2 ± 14.0	126 ± 13.5	No data	112
	40-49	122.3 ± 21.7	116.0 ± 17.2	126 ± 16.5	No data	107

[a]Computer Speech Lab (CSL) manufactured by Kay Elemetrics, Inc.; [b]Phonolaryngograph SH-01; [c]Dr. Speech software. The age of subjects ranged from 20-26 y/o.

can analyze vocal intensity during sustained phonation. The SPL of a sound in decibels (dB) is 20 times the logarithm of the ratio of the pressure of the sound to the reference pressure. The reference pressure is explicitly stated and is most commonly 2×10^{-4} microbars (0.0002 dyne/cm^2).[8] Accurate SPL measurement can provide a highly reliable means of tracking vocal function.[1] The average SPL results in this and Chen's studies are displayed in Table 5. Although Chen's data were consistently higher than ours in each group, the discrepancy was small. The minor difference might be due to different analysis systems and procedures, because intensity could be influenced by the distance of the microphone from the lips of the speaker. However, in our study, the data was gained by using the Aerophone II system. Every subject used the same mask during test; hence the distance between microphone and the lips of the speaker should be constant. According to our study, there was no difference between gender and age groups, and the mean SPL was about 77 to 78 dB.

Because of the dramatic advances in voice research, acoustic perturbation analyses of sustained vowels, including jitter, shimmer and H/N ratio, have been widely used in determining phonatory stability characteristics. Increased jitter or shimmer values have been associated with phonatory instability due to aging, [9]amyotrophic lateral sclerosis,[10] and various laryngeal pathologies.[11] Hence, voice perturbation seems to be a probable indicator for a physiological disorder. However, we should note that several factors, including frequency, intensity, and vowel selection, would affect various phonatory stability measures and that the highest perturbation values almost always occur in low frequency-low intensity situations.[12] In this study, we used voice within comfortable frequency and intensity ranges to establish our norm in order to minimize bias. Based on our study, jitter was quite stable and did not change with aging in subjects from 20 to 49 years old, and there was no significant difference between genders. In addition, shimmer and H/N ratio showed differences between age groups in males, but not in females. At the same time, there were also significant gender differences in some age ranges for shimmer and H/N ratio. But we cannot explain the reason due to the limits of our study design, and it warrants further investigation. However, this suggests that we should be extremely careful if we want to use shimmer or harmonic-to-noise ratio as a clinical parameter for comparing voice quality, because they are variable in different genders and age ranges.

According to Bielamowicz *et al.*,[13] commercially available acoustical analysis programs agreed well, but not perfectly, in their measures of F0. However, measures of perturbation in the various analysis packages use different algorithms, provide results in different units, and often yield values for voices that violate the assumption of quasi-periodicity. As a result, poor rank order correlations between programs using similar measures of perturbation were noted. This might explain the large discrepancy in data between Chu's[5] study and our study (Table 6). Therefore, if our data are used as a baseline for normal adults, we recommend that comparisons be made only with data gained from using the same procedures and analytical systems.

Table 5. Comparison of SPL data between Chen's study and the present study

Gender	Age	Present study[a] /a/ SPL (dB) (n = 15)	Chen [4b] /u/ SPL (dB) (n = 10)
F	20-29	76.3 ± 4.4	80 ± 3.3
	30-39	79.6 ± 6.4	81 ± 3.8
	40-49	77.6 ± 5.4	79 ± 2.5
M	20-29	76.2 ± 6.7	82 ± 3.5
	30-39	79.1 ± 5.6	83 ± 3.4
	40-49	77.3 ± 5.5	84 ± 4.3

[a]Computer Speech Lab (CSL) manufactured by Kay Elemetrics, Inc.; [b]Phonolaryngograph SH-01.

Table 6. Comparison of acoustic perturbation data between Chu's study and the present study

Gender	Present study[a] /a/ (Group I) (n = 15) (age 20-29)			Chu *et al*[5] [b] /ee/ (n = 20) (age 20-26)		
	Jitter (%)	Shimmer (dB)	H/N ratio (dB)	Jitter(%)	Shimmer (%)	H/N ratio (dB)
F	0.69 ± 0.23	1.65 ± 1.58	12.2 ± 3.9	0.22 ± 0.11	1.19 ± 0.52	26.3 ± 3.7
M	0.61 ± 0.25	0.26 ± 0.09	8.7 ± 5.3	0.21 ± 0.08	1.34 ± 0.64	25.8 ± 3.4

[a]Computer Speech Lab (CSL) manufactured by Kay Elemetrics, Inc; [b]Dr. Speech software.

MPT is the maximum duration a person can sustain a tone. Arnold[14] employed the measurement of phonation time routinely during phoniatric examinations and observed that the MPT was frequently reduced to a few seconds in paralytic dysphonia. Hence, this method can be utilized intraoperatively before and after the vocal fold medialization procedure.[8] Treole and Trudean[15] quoted Hirano's research on MPT in adults and noted that MPT values were greater for males (25-35 sec) than for females (15-25 sec). Our study agrees with his result, with the mean value for males (28.0 ± 9.4 sec) being greater than that for females (22.6 ± 7.6 sec). The pulmonary vital capacity of males is larger than that of females, which might explain this phenomenon. Interestingly, the data from Chu's[5] report showed no difference between male (19.2 ± 3.6 sec) and female (19.7 ± 3.7 sec) adults.

As the aforementioned results and discussions demonstrate, we have developed reliable normal data for various parameters of acoustic analysis in different age groups and genders in Taiwanese adults. Importantly, this body of data can be used as a convenient reference to evaluate the voices of patients. In summary, the value of F0 (counting; vowel/a/) was greater for females (203.2 ± 21.7; 213.4 ± 25.4 Hz) than for males (118.3 ± 17.3; 121.3 ± 16.4 Hz). Conversely, the value of MPT was greater for males (28.0 ± 9.4 sec) than for females (22.6 ± 7.6 sec). There were no significant differences in average SPL and jitter between females (77.8 ± 5.5 dB; 0.66 ± 0.27%) and males (77.5 ± 5.5 dB; 0.56 ± 0.23%). Except for shimmer and H/N ratio of males, there were no differences in parameters between age groups of both genders. Hence, a majority of voice characteristics of adults seem to be relatively stable and do not change with aging between ages 20 to 49. Additionally, we should be extremely careful if we want to use shimmer or harmonic-to-noise ratio as a clinical parameter for comparing voice quality. However, the voice characteristics of adults older than 50 were not measured in this study and therefore require further investigation.

ACKNOWLEDGEGMENTS

We would like to thank all of the volunteers who participated in this research. We also appreciate the financial sponsorship from Taichung Veterans General Hospital (grant No. TCVGH 917002A).

REFERENCES

1. Zeitels SM, Montgomery WW, Hillman RE. Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Curr Opin Otolaryngol Head Neck Surg* 1997;5:172-5.
2. Colton RH, Casper JK. Some normative data on the voice. In: Colton RH, Casper JK. *Understanding Voice Problems*. Maryland: Williams & Wilkins, 1990:352-61.
3. Yumoto E. The quantitative evaluation of hoarseness. *Arch-Otolaryngol* 1983:109:48-52.
4. Chen SH. The use of phonolaryngograph SH-01 and voice analysis in Chinese normal adults. *J Chin Med Assoc* 1985; 35:41-54.
5. Chu YH, Hsiung MW, Lin CS, Lee MH, Wang HW, Su WY. Voice analysis in normal young men and women. *J Taiwan Otolaryngol Head Neck Surg* 2002;37:159-62.
6. Nobuhiko I. Physiology of speech production. In: Nobuhiko. *Phonosurgery: Theory and Practice.* Tokyo: Springer-Verlag, 1989:5-20.
7. Shimon S. The intrinsic pitch of vowels: theoretical, physiological, and clinical considerations. *J Voice* 1989:3:44-51.
8. Nobuhiko I. Assessment of vocal function. In: Nobuhiko. *Phonosurgery J Voice: theory and practice.* Tokyo: Springer-Verlag, 1989:35-58.
9. Linville, SE. Acoustic-perceptual studies of aging voice in women. 1987;1:44-8.
10. Kent J, Kent R, Rosenbek J, Weismer G, Martin R, Sufit R, Brooks B. Quantitative description of dysarthria in women with amyotrophic lateral sclerosis. *J Speech Hear Res* 1992; 35:723-33.
11. Wolfe V, Cornell R, Palmer C. Acoustic correlates of pathologic voice types. *J Speech Hear Res* 1991;34:509-16.
12. Gelfer MP. Fundamental frequency, intensity, and vowel selection: effects on measures of phonatory stability. *J Speech Hear Res* 1995;38:1189-98.
13. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke G. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res* 1996;39:126-34.
14. Arnold GE. Vocal rehabilitation of paralytic dysphonia: II. Acoustic analysis of vocal function. *AMA Arch Otolaryng* 1955;62:593-601.
15. Treole K, Trudeau MD. Changes in sustained production tasks among women with bilateral vocal nodules before and after voice therapy. *J Voice* 1997;11:462-9.