An Innovative *Voice Analyzer* "*VA*" Smart Phone Program for Quantitative Analysis of Voice Quality

*Tsuyoshi Kojima, *Shintaro Fujimura, *Ryusuke Hori, *Yusuke Okanoue, *Kazuhiko Shoji, and †Masato Inoue, **Nara, and* †*Tokyo, Japan*

Summary: Objective. The '*VA*' Windows program that we developed in 2011 for analyzing voice quality quantitatively uses zerocross picking to find individual basic pitch periods. It has a simple and user-friendly user interface and high accuracy. This program determines the fundamental frequency, jitter, shimmer, PPQ, APQ, and signal-to-noise ratio (Ra). It needs only a general-purpose Windows PC, USB audio interface and a microphone. The aim of this study is to improve the version of the *VA* Windows program in English and to develop a *VA* smart phone program to allow wider use of objective acoustic analysis.

Study Design. Cross-sectional study.

Methods. Sustained vowel /a/ sounds from 40 subjects without evident vocal problems, and 40 subjects with slight hoarseness, were examined. We compared the analyzed data with data from other software (MDVP and Praat). For a comparison between *VA* for Windows and *VA* for a smart phone, sustained vowel /a/ sounds from six subjects without hoarseness were recorded with each system simultaneously.

Results. The normal voice and slightly hoarse voice data analyzed with *VA* showed a high correlation with most parameters from both *MDVP* and *Praat*. There was a strong correlation between the Windows and smart phone versions of *VA* in terms of the fundamental frequency and Ra.

Conclusions. The results showed that the *VA* software was not inferior to the other acoustic analysis software tested. The simple and easy to use smart phone version may facilitate our goal of creating an objective, widely available method to evaluate hoarseness.

Key Words: Voice analyzer—Hoarseness—Acoustic analysis—Smart phone—Tablet PC.

INTRODUCTION

Voice disorders are a common condition encountered by otolaryngologists. To assess voice disorders, an auditory-perceptual evaluation of voice quality is the most popular method.[1,2] Perceptual evaluation is often considered one of the reliable assessments for documenting voice disorders among voice clinicians, since vocal quality is essentially an acoustic perceptual phenomenon, and is thus compatible with perceptual evaluations. Although training for evaluating voice disorders is required to obtain accurate data, it is not necessary to undergo strict technical training. There is no doubt that perceptual ratings of voice are subjective, although there is abundant evidence regarding the reliability of perceptual voice evaluations.[2,3] Ratings sometimes differ between examiners and it is difficult to detect small changes in voice disorders. Thus, it is important that there are multiple clinical methods of voice evaluation, including objective methods related to voice fundamental frequency (f_0) and strength, aerodynamic tests, and acoustic analysis tests. Unfortunately, objective and quantitative analyses of voice disorders are not used widely as "standard" examinations because of their inconvenience and associated complications, even though voice disorders are common. The objective analyses are usually performed only by voice clinicians using expensive instruments.

In this study, we focused on acoustic analysis tests employing objective methods to assess voice disorders. We developed *Voice Analyzer* or "*VA*," a Microsoft Windows program for analyzing voice quality quantitatively in 2011.[4] *VA* has a simple and user-friendly interface and high accuracy and uses zero-cross picking to find individual basic phonatory cycle periods. The voice waveform is analyzed under a running window of two cycles to calculate the acoustic energy of harmonics separately from that of noise components. The program determines the f_0, percent jitter (jitter), percent shimmer (shimmer), pitch period perturbation quotient (PPQ), amplitude perturbation quotient (APQ), and signal-to-noise ratio (SNR; Ra), which is one of the ratios of harmonics to noise (HNR). *VA* needs only a general-purpose Windows computer and a Universal Serial Bus (USB) audio interface. In this study, we improved the *VA* Windows program in English for accurate acoustic analysis. We also developed a smart phone version of *VA* to perform such acoustic analyses more readily and easily.

MATERIALS AND METHODS

Outline of *VA*

We developed the *VA* Windows program with the C# language (Windows version), which was developed by Microsoft (Redmond, WA). The interface of *VA* is user-friendly and no manual is required (Figure 1). Fundamental frequency, jitter, shimmer, PPQ, APQ, and Ra (an SNR measurement method) can be measured with the original f_0 extraction algorithm (details provided Algorithm). The top panel of *VA* shows the real-time voice waveform, and an optionally selected part of the waveform (at about 0.25 second) is able to be analyzed. The voice data are easily recorded as a WAV file (16 bit,

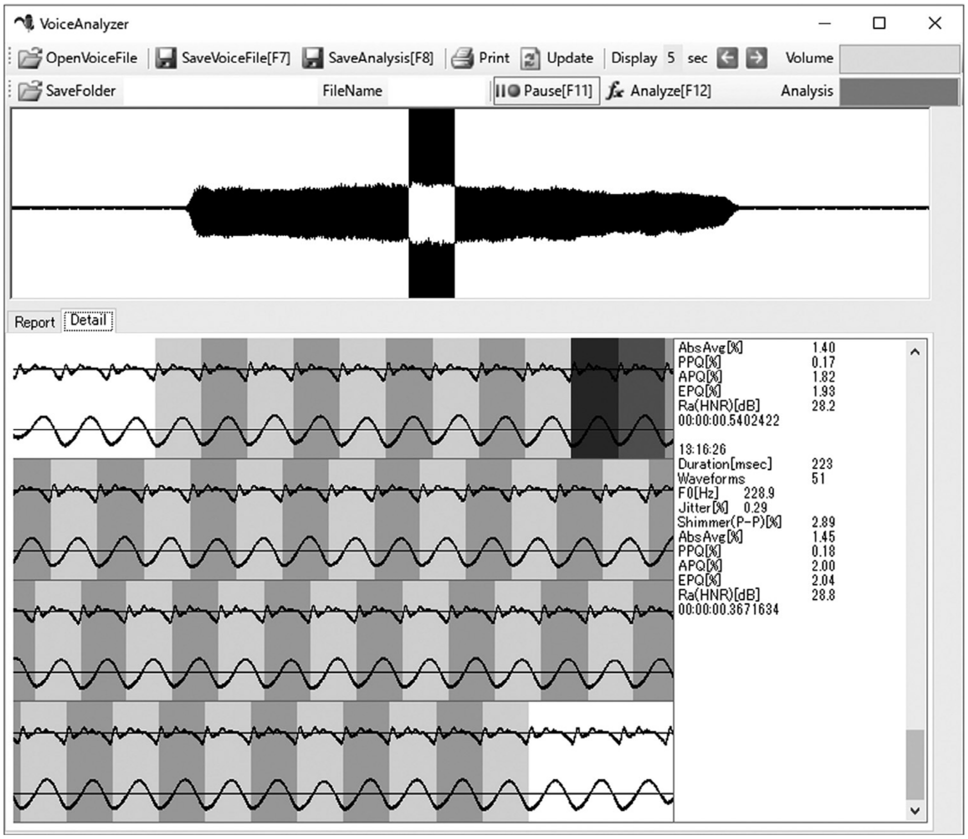
Accepted for publication January 30, 2018.

From the *Department of Otolaryngology, Tenri Hospital, Nara, Japan; and the †Department of Electrical Engineering and Bioscience, School of Advanced Science and Engineering, Waseda University, Tokyo, Japan.

Address correspondence and reprint requests to Tsuyoshi Kojima, Department of Otolaryngology, Tenri Hospital, 200 Mishima-cho, Tenri-shi, Nara 632-8552 Japan. E-mail: t_kojima@ent.kuhp.kyoto-u.ac.jp

Journal of Voice, Vol. 33, No. 5, pp. 642–648
0892-1997

https://doi.org/10.1016/j.jvoice.2018.01.026



```
AbsAve[%]            1.40
PPQ[%]               0.17
APQ[%]               1.82
EPQ[%]               1.93
Ra(HNR)[dB]          28.2
00:00:00.5402422

13:16:26
Duration[msec]       223
Waveforms            51
F0[Hz]      228.9
Jitter[%]   0.29
Shimmer(P-P)[%]      2.89
AbsAve[%]            1.45
PPQ[%]               0.18
APQ[%]               2.00
EPQ[%]               2.04
Ra(HNR)[dB]          28.8
00:00:00.3671634
```

FIGURE 1. User interface of the *Voice Analyzer* (*VA*) Windows program. The real-time voice waveform is shown in the *top panel*. An arbitrary part is selected for analysis. If only the pause button is pushed, it can be selected in one's own time. The *bottom-left panel* shows the details of the selected part. The *bottom-right panel* shows the results of the analysis. Results are shown immediately after selecting a part. The pair of pitch periods shown in *dark gray* was selected as that with the best signal-to-noise ratio (SNR) over 0.25 second.

48 kHz, 20 seconds). A recorded WAV file can be imported and analyzed. Next, we developed a smart phone version of *VA* in JAVA (Android version), which was developed by Sun Microsystems (Menlo Park, CA). This program works on the Android operating system (OS), which is used commonly by smart phones and tablet personal computers (PCs). The algorithm for the voice analysis in the smart phone version of *VA* is the same as that of the Windows version. However, the user interface is actually simpler than that of the Windows version and is suitable for a small device with a touch screen. The smart phone version of *VA* also requires no manual; moreover, both nonmedical and medical personnel can use it (Figure 2).

Algorithm

Fundamental frequency extraction algorithm

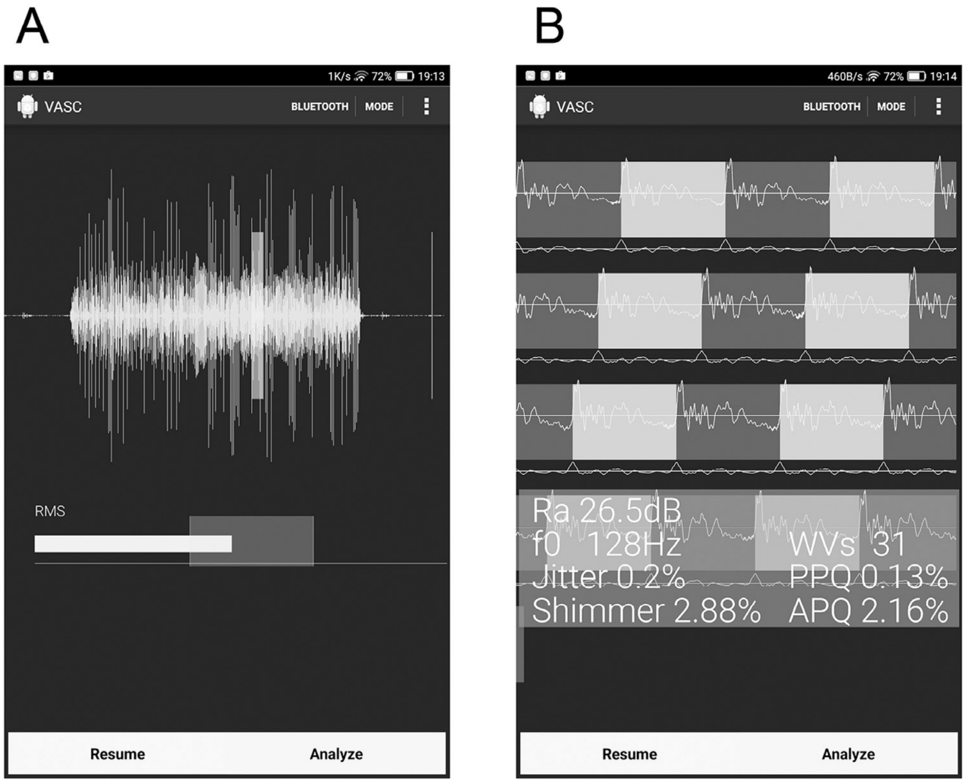
Because accurate pitch-picking is needed for precise determination of fundamental frequency and jitter, it is an important factor in voice analysis, and many algorithms for f_0 extraction have been developed. They can be divided roughly into methods that are period-synchronous or not. Those that are period-synchronous include peak picking and zero-cross picking methods. Peak picking is difficult to estimate the actual peak point from two points that cross the real peak point. *VA* uses zero-cross picking, because the real zero-cross point can be estimated from two points that cross at zero. The problem with this, however, is that there may be several zero-cross points in a given period. Thus, the peaks of the integrated speech waveform are all zero-crossing points of the original speech waveform. It is possible to estimate the position of the zero-cross point for pitch picking by picking the peak with the highest and sharp peak. To increase the accuracy, each zero-cross point was also scored by subtracting the recent dip from the peak in the original waveform, which is based on differential filtering. Erroneous points are discarded based on this score and, finally, the zero-crossing points are determined precisely by linear interpolation of two samples interposing the zero-crossing points extracted along the way (Figure 3).

Fundamental frequency, jitter, shimmer, PPQ, and APQ

The f_0 is computed using the extraction algorithm described above. Jitter (local) and PPQ are related to period-to-period variability in the pitch. Shimmer (local) and APQ are related to period-to-period variability in the peak-to-peak amplitude. Algorithms for each of these metrics were described previously.[5–7]

SNR algorithm

The voice signal contains both harmonics and noise. The HNR has been used previously as a voice analysis metric.[8] In this study, we used the improved iteration of Ra, the original version of which was published in 1980.[9] Using Fourier transform,



Ra 26.5dB
f0 128Hz WVs 31
Jitter 0.2% PPQ 0.13%
Shimmer 2.88% APQ 2.16%

FIGURE 2. User interface of the *Voice Analyzer* (*VA*) smart phone program. (**A**) The real-time voice waveform is shown in the *top panel*. The *middle column* (RMS [root mean square]) displays the sound pressure level to ensure an adequate sound level. The pause button interrupts importation of the signal and changes to become a resume button. The resume button is used for continuing to import the signal. An arbitrary part of the waveform can be selected after pushing the pause button. If the analyze button is pushed, the screen changes to that shown in B. (**B**) Results are shown immediately after the analyze button is pushed. If the screen is touched again, an arbitrary part of the voice waveform can be reselected.

which refers to the frequency domain representation of the original signal, it is possible to decompose the voice signal into harmonic and noise components. If a longer time window is used for the calculation, the error is larger because of the unstable period. We used a running time window of two periods to reduce the error and the SNR for every pair of adjacent periods within optionally selected 0.25 second is considered and the best one is selected as the best SNR.

Hardware and environment

Acoustic signals were recorded using an AKG C747 hyper cardioid condenser microphone (AKG Acoustics, Vienna, Austria), placed 20 cm below the mouth opening at 30°, and digitized using a Rolland UA5 USB audio interface (Rolland Corp., Hamamatsu, Japan) with a notebook Windows PC. To evaluate the *VA* smart phone software, an Android device (MediaPad M3, Huawei Technologies Co. Ltd., Shenzhen, China) using only its built-in equipment was employed for recording of voice data (the Android system). Sustained vowel /a/ sounds were examined in a sound-treated room.

Subjects

Normal voices

In total, 20 females and 20 males, ranging in age from 19 to 49 years, were enrolled in the study (average age, 28 years). They had no subjective or objective symptoms of voice disorders. Voices were recorded at a 48-kHz sampling rate. The data were analyzed with *VA*, *Praat* software (Institute for Phonetic Sciences, University of Amsterdam, The Netherlands), and the *Multidimensional Voice Program* (*MDVP*; KayPENTAX, Montvale, NJ). The fundamental frequency, jitter, shimmer, and HNR (H/N ratio for *Praat* and *MDVP*; Ra for *VA*) were compared between the programs.

Slightly hoarse patients

One auditory-perceptual evaluation method for hoarseness is the Grade, Roughness, Breathiness, Asthenia, Strain (GRBAS) scale of the Japan Society of Logopedics and Phoniatrics, which gives scores ranging from 0 to 3 for the following facets of hoarseness: Grade (G), Roughness (R), Breathiness (B), Asthenia (A), and Strain (S). The G scale represents the overall impression of abnormality in voice. Here, "0" is normal, "1" is a slight degree of hoarseness, "2" is a medium degree, and "3" is a high degree.[1,3,10,11] After two physicians, who were specialist laryngologists, rated the patients for hoarseness, 22 females and 18 males rated as Grade (G) 1 were enrolled and classified as patients with slight hoarseness. They ranged in age from 39 to 79 years (average age, 61 years). Voices were recorded at a 48-kHz sampling rate. The fundamental frequency, jitter, shimmer, and HNR (H/N ratio for *Praat*; Ra for *VA*) were compared.

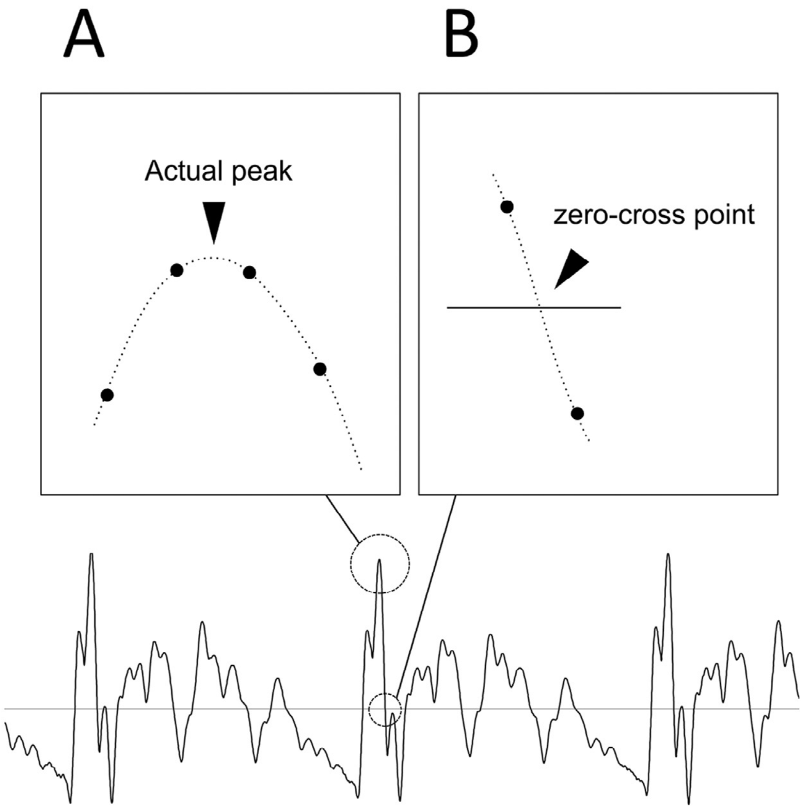


FIGURE 3. Comparison of the peak picking and zero-cross points methods within a given period. (**A**) When using peak picking, it can be difficult to determine the actual peak. (**B**) The zero-cross points method is able to determine the before and after zero-cross points within a given period.

VA *smart phone program*

Three females and three males, ranging in age from 23 to 40 years and who had never smoked, were enrolled (average age, 30 years). They had no subjective or objective symptoms of any voice disorder. Their voices were recorded with the Android and Windows systems simultaneously in a soundproof room. Both waveforms were synchronized and the same precise periods selected and analyzed. The analysis algorithms were the same for both systems, but the recording system differed by system.

Statistical analysis

The *SPSS* software (version 22.0; IBM Corp., Armonk, NY) was used for the data analysis. Spearman's rank correlation coefficient (Spearman's rho) was used to assess the correlation of the fundamental frequency, jitter, shimmer, and HNR between the programs for normal voice data. Spearman's rho was also used to assess the data of patients with slight hoarseness; that is, for comparison of these data between the Windows and Android systems.

RESULTS

Normal voice

The fundamental frequency determined with *VA* showed a strong correlation with those of *MDVP* and *Praat*. The *VA* shimmer data also showed a strong correlation with those of *MDVP* and *Praat*. There was also a strong correlation between *VA* and *Praat* regarding the HNR. Additionally, the jitter data from *VA* showed a strong correlation with those of *Praat* (Figure 4).

Patients with slight hoarseness

The fundamental frequency, shimmer, and HNR determined with *VA* showed a strong correlation with those of *Praat*. The jitter data also showed a strong correlation with those of *Praat* (Figure 5).

VA smart phone software

An effect of differences in the recording system equipment used was reflected in the form of different voice sound waves between the Android and Windows systems. However, the fundamental frequency and Ra data showed a strong correlation (Figure 6).

DISCUSSION

With *VA*, we aimed to develop an accurate and fast software package, with a simple user interface, for analyzing voice quality quantitatively. The complexity of available software and devices was assumed to be one reason hindering the wide use of objective estimations of voice metrics. The results showed a very strong correlation in terms of fundamental frequency and strong to moderate correlations in other metrics, between the *VA* and *MDVP* software. The results also showed very strong or strong correlations for each metric between the *VA* and *Praat* software. Because these programs are commonly used and generally considered reliable among voice clinicians, it can be concluded that *VA* is also suitable for voice analysis. Indeed, the values of the

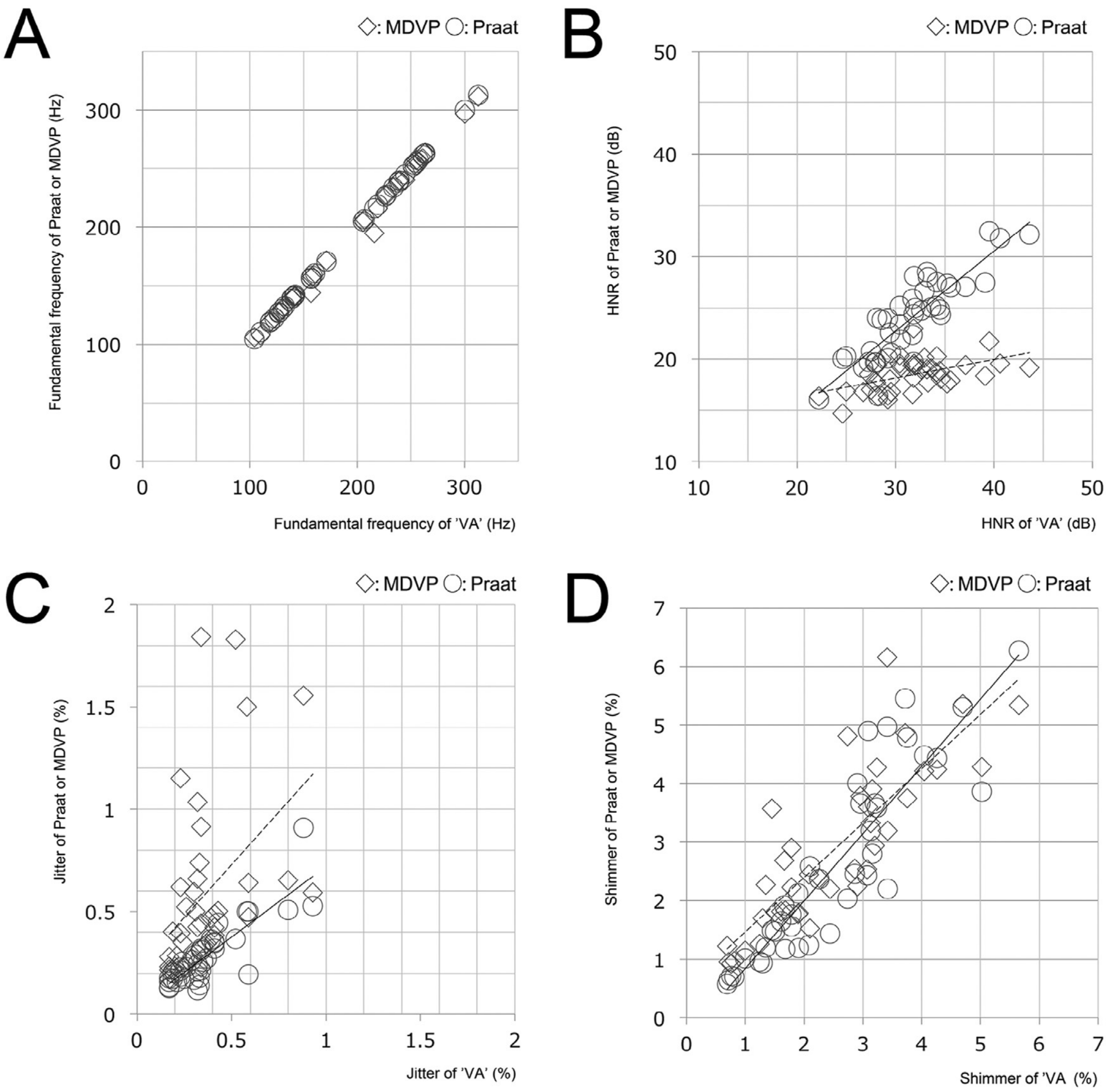


FIGURE 4. Normal voice data. The horizontal axis gives the *Voice Analyzer* (*VA*) data. The vertical axis gives the *Multidimensional Voice Program* (*MDVP*) or *Praat* data. The *squares* denote the *MDVP* data. The *circles* denote the *Praat* data. (**A**) Scatter plot of fundamental frequency. *VA* and *MDVP*: Spearman's rho = 0.998. *VA* and *Praat*: Spearman's rho = 1.000. (**B**) Scatter plot of HNR. *VA* and *MDVP*: Spearman's rho = 0.476. *VA* and *Praat*: Spearman's rho = 0.857. (**C**) Scatter plot of jitter. *VA* and *MDVP*: Spearman's rho = 0.556. *VA* and *Praat*: Spearman's rho = 0.794. (**D**) Scatter plot of shimmer. *VA* and *MDVP*: Spearman's rho = 0.859. *VA* and *Praat*: Spearman's rho = 0.919.

measured metrics were similar between *VA* and *Praat*. Moreover, it was reported previously that *Praat* had better values for the metrics of jitter and the H/N ratio than *MDVP*.[12] The values of such metrics are related to the accuracy of pitch picking. Although the fundamental frequency values showed strong correlations among all tested programs in our study, small differences may produce differences in jitter and the HNR. It is difficult to compare data on the HNR directly, which should have a high value, because the algorithms differ and normal values vary among programs. However, the algorithm for jitter, which should have a small value for a normal voice, is the same. *VA* and *Praat* showed similar values for jitter, both of which were smaller than the value of *MDVP*. Additionally, because the fundamental frequency of *VA* and *Praat* was almost the same, we suggest that both programs have good pitch-picking algorithms and the accuracy of their acoustic analyses is high.

In acoustic analyses, comparison between noise and harmonics is one of the major means of objectively measuring the degree of hoarseness, and many methods have been developed based on this.[8,9,13] The HNR was first introduced in 1982,[8] and constitutes another major method of measuring hoarseness. Although the HNR has been preferred because of its simplicity and ease of computation, the larger perturbation or trend affects the results. Ra, introduced in 1980, is more complex and time-consuming to calculate than the HNR. However, the evolution of computer performance mitigated this weakness while the strength of Ra, that is, its sensitivity, remained. The original Ra measurement used three pitch periods to estimate signal and noise energy. Here, we used instead a running pair of pitch periods to estimate energy, because the difference between the two approaches was not large. Calculation using a running pair of pitch periods can provide the SNR continuously and in real-

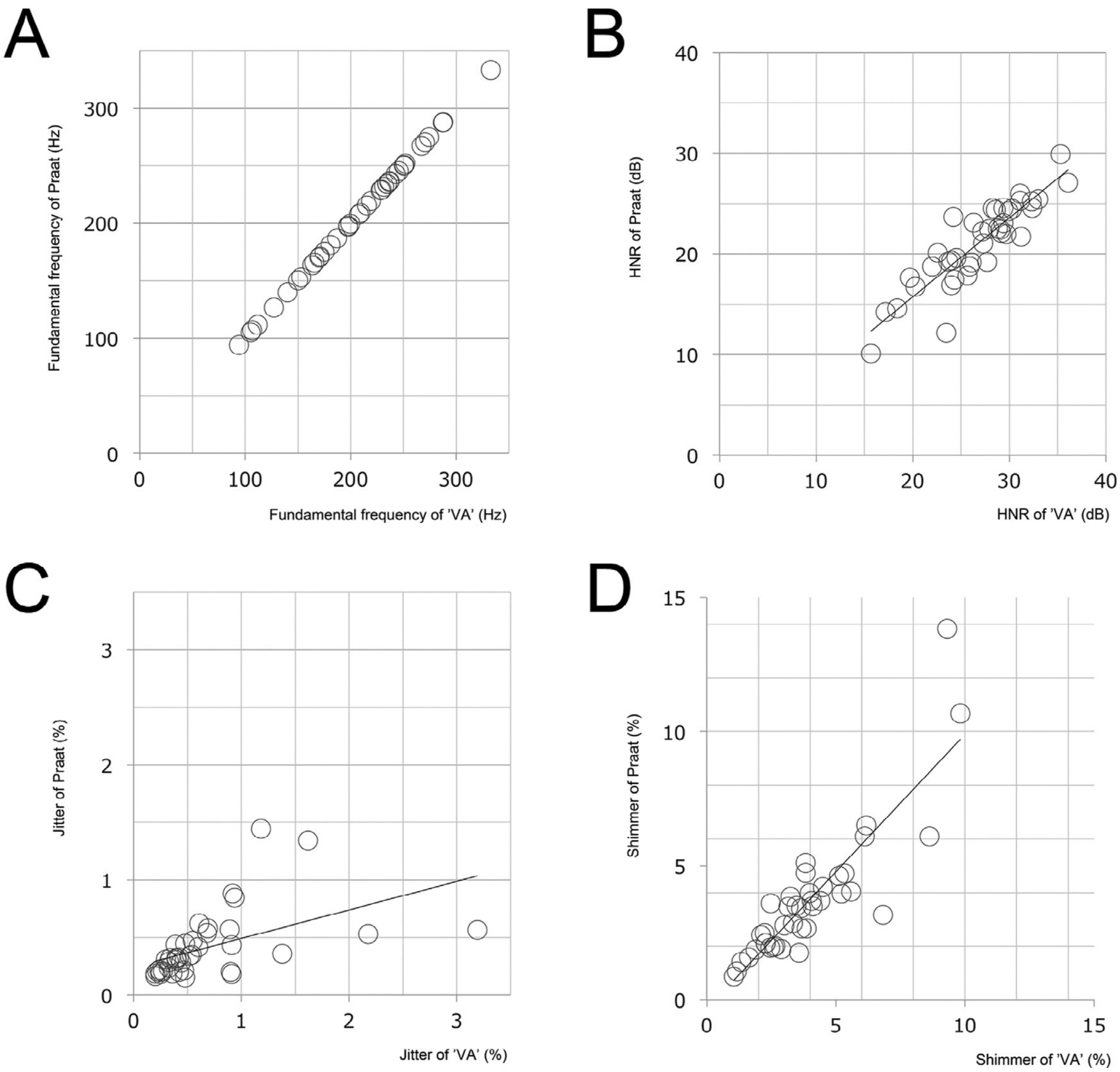


FIGURE 5. Patients with slight hoarseness. The horizontal axis gives the *Voice Analyzer* (*VA*) data. The vertical axis gives the *Praat* data. (**A**) Scatter plot of fundamental frequency. *VA* and *Praat*: Spearman's rho = 1.000. (**B**) Scatter plot of HNR. *VA* and *Praat*: Spearman's rho = 0.870. (**C**) Scatter plot of jitter. *VA* and *Praat*: Spearman's rho = 0.693. (**D**) Scatter plot of shimmer. *VA* and *Praat*: Spearman's rho = 0.856.

time. The provision of superior assessment in real-time is one reason why we prefer this measure for analyzing the HNR. Another reason is that a more precise HNR is considered useful in evaluating slight hoarseness. Most patients are concerned about the type of individual-specific hoarseness due to nodules, inflammation, or atrophy, for example. Because there are differences among individuals in terms of the "normal" voice, it is sometimes difficult to detect a slight degree of hoarseness. However, objective assessment of the voice before and after treatment is useful for understanding whether a treatment is appropriate for a given individual's voice disorder. In this study, the *VA* and *Praat* programs showed very strong or strong correlations in terms of the slight hoarseness data, as well as for the normal voice data. Evaluation of the slight hoarseness data showed "worse" values than those of the normal voice data, in terms of jitter, shimmer, and the HNR, although there was overlap between groups. Because the HNR, especially, is considered a good method to evaluate hoarseness, it is possible that the sensitivity to Ra of the *VA* software will be helpful for evaluating voice disorders characterized by slight hoarseness. An auditory-perceptual evaluation of voice quality is limited for evaluating subtle changes, although it is possible to evaluate moderate to severe hoarseness with this method. However, with acoustic analyses, it can be difficult to determine appropriate values for severe hoarseness, because it is challenging to identify the optimal period. Thus, acoustic analyses are typically performed together with auditory-perceptual evaluations for understanding voice disorders.

Another aim of this study was the development of a program for tablet PCs and smart phones. We developed an Android version of *VA*, programmed in JAVA, which works with the Android OS. It runs on tablets using the Android OS and the Windows version works on tablet PCs running Windows. Because these platforms use the same calculation algorithm, the same voice data produced identical results. Although the central processing unit power differs between platforms, the calculation speeds are equally acceptable. However, there are some difficulties in achieving the same values as in the original Windows

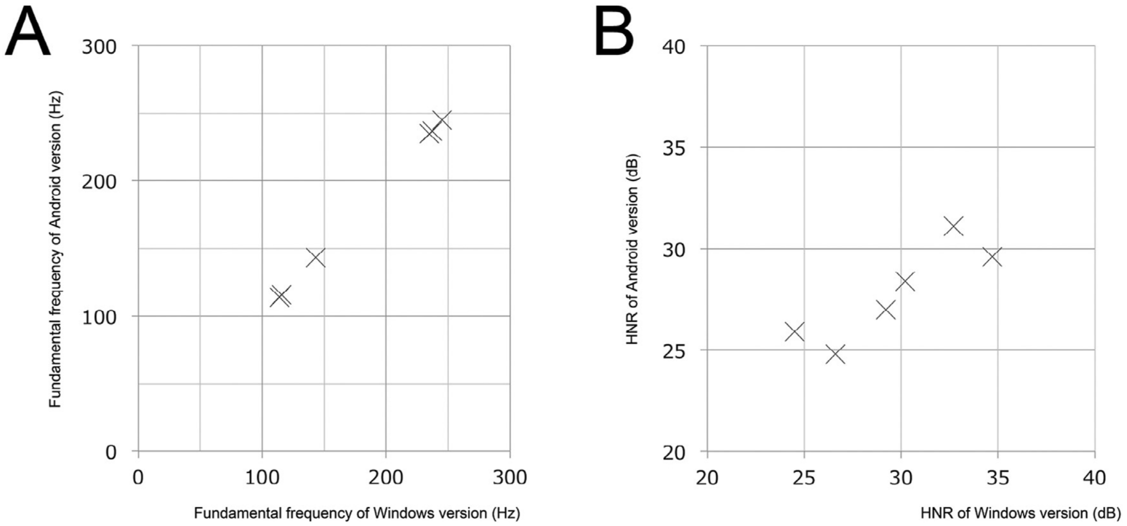


FIGURE 6. *Voice Analyzer* "*VA*" smart phone program. The horizontal axis gives the data from the Windows version of *VA*. The vertical axis gives the data from the Android version of *VA*. (**A**) Scatter plot of fundamental frequency. Windows and Android versions: Spearman's rho = 1.000. (**B**) Scatter plot of Ra. Windows and Android versions: Spearman's rho = 0.886.

version with the Android OS, due to the recording system equipment, including the microphone and audio user interface; most small devices have a simplified audio interface and microphone. A Windows PC, or tablet PC running Windows, typically has an easy-to-use general-purpose USB audio interface, as well as a microphone, which may be of professional quality. The system may be too complex for use in a clinic or at home, although this is not a serious issue when used in a sound-treated room. The Android version of *VA* showed a very strong correlation with the Windows version in terms of the fundamental frequency and Ra metrics. Because the HNR is the most important factor in evaluating hoarseness, both the Android and Windows versions of *VA* are sufficient to evaluate voice disorders. The simplicity of the Android version may enable acoustic analyses in various situations, such as by a speech pathologist evaluating voice changes during rehabilitation. Furthermore, patients may be able to analyze their own voices objectively and make good use of voice training performed at home. We typically use the Android version of *VA* as an aid for evaluating voices during thyroplasty, which is performed under local anesthesia. Beyond the surgeon's subjective opinion, an objective evaluation performed each time the vocal cord position is adjusted can be helpful in achieving relatively high rates of success for thyroplasty.

An absolute comparison between systems is sometimes difficult; however, performing at least a relative evaluation is useful. Of course, the recording environment is one factor in any such evaluation, because the noise level affects the results. Also, differences in equipment may affect the results, because different devices are equipped with different recording systems. A further goal is to improve the *VA* program by making it less sensitive to such differences, through adding a noise-cancelling system and a system for performing adjustments between devices.

CONCLUSIONS

We improved the *VA* Windows program in English and developed an equivalent *VA* smart phone program for acoustic analysis of voice disorders. *VA* was not inferior to any of the other acoustic analysis software tested. This simple and easy-to-use system may help us achieve our goal of making objective evaluation of hoarseness a common practice not only for specialists, but also for general practitioners. The *VA* software for smart phones will allow easy access to a means of voice analysis in clinical and voice therapy settings.

REFERENCES

1. Yamaguchi H, Shrivastav R, Andrews ML, et al. A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatr Logop*. 2003;55:147–157.
2. Oates J. Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatr Logop*. 2009;61:49–56.
3. Nemr K, Simões-Zenari M, Cordeiro GF, et al. GRBAS and Cape-V scales: high reliability and consensus when applied at different times. *J Voice*. 2012;26:812.e17–812.e22.
4. Mizuta M, Shoji K, Kojima T, et al. New VA software program quantitatively analyzes voice quality. *Pract Otorhinolaryngol*. 2011;104:297–302.
5. Bielamowicz S, Kreiman J, Gerratt BR, et al. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res*. 1996; 39:126–134.
6. Koike Y. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Stud Phonol*. 1973;VII:17–23.
7. Titze IR, Horii Y, Scherer RC. Some technical considerations in voice perturbation measurements. *J Speech Hear Res*. 1987;30:252–260.
8. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Speech Hear Res*. 1984;27:1544–1550.
9. Kojima H, Gould WJ, Lambiase A, et al. Computer analysis of hoarseness. *Acta Otolaryngol*. 1980;89:547–554.
10. Omori K. Diagnosis of voice disorders. *Japan Med Assoc J*. 2011;54: 248–253.
11. Minoru H. *Psycho-acoustic evaluation of voice. Clinical Examination of Voice (Disorders of Human Communication)*. Springer-Verlag; 1981: 81–84.
12. Oğuz H, Kiliç MA, Şafak MA. Comparison of results in two acoustic analysis programs: Praat and MDVP. *Turk J Med Sci*. 2011;41:835–841.
13. Kasuya H, Ogawa S, Mashima K, et al. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J Acoust Soc Am*. 1986;80:1329–1334.