TÜBİTAK

Original Article

Turk J Med Sci
2011; 41 (5): 835-841

E-mail: medsci@tubitak.gov.tr
doi:10.3906/sag-0909-290

Comparison of results in two acoustic analysis programs: Praat and MDVP

Haldun OĞUZ[1], Mehmet Akif KILIÇ[2], Mustafa Asım ŞAFAK[1]

Aim: To compare acoustic analysis results obtained by 2 computer programs, Praat and the Multi-Dimensional Voice Program (MDVP). Different voice analysis programs use similar descriptions to define voice perturbation measures.

Materials and methods: A total of 47 voice samples reflecting a spectrum of normal and pathological voices were randomly selected from a database, and the same voice samples were used to obtain mean fundamental frequency, jitter, shimmer, and noise-to-harmonics ratio results from 2 acoustic analysis programs.

Results: The results obtained for mean fundamental frequency and shimmer were not significantly different between the 2 computer programs. The results for jitter and noise-to-harmonics ratio, however, were significantly different between Praat and MDVP ($P < 0.001$). There was a strong correlation for mean fundamental frequency and jitter values. The correlations for shimmer values and the noise-to-harmonics ratio were moderate.

Conclusion: The numerical values obtained for mean fundamental frequency were comparable between the 2 computer programs. The values obtained for shimmer were not significantly different, but the correlation was moderate. The jitter values and noise-to-harmonics ratio were not comparable between the 2 acoustic analysis programs.

Key words: Voice, acoustic analysis, voice analysis, MDVP, Praat

İki akustik analiz programı sonuçlarının karşılaştırılması: Praat ve MDVP

Amaç: Ses pertürbasyon ölçümleri ifade edilirken, değişik ses analiz programları tarafından benzer tanımlamalar kullanılmaktadır. Bu çalışmanın amacı, Praat ve Multi-Dimensional Voice Program (MDVP) adlı bilgisayar programları ile elde edilen akustik analiz sonuçlarının karşılaştırılmasıdır.

Yöntem ve gereç: Normal ve patolojik sesleri yansıtacak şekilde veri tabanından 47 ses örneği seçilmiştir. Her iki akustik analiz programında tıpatıp aynı sesler kullanılarak ortalama temel frekans, jitter, shimmer ve gürültü harmonik oranı sonuçları elde edilmiştir.

Bulgular: İki bilgisayar programından elde edilen ortalama temel frekans ve shimmer değerleri arasında anlamlı fark yoktu. Jitter ve gürültü-harmonik oranı değerleri ise Praat ve MDVP arasında anlamlı olarak farklı idi ($P < 0,001$). Ortalama temel frekans ve jitter değerlerinde kuvvetli korrelasyon mevcuttu. Shimmer değerleri ve gürültü-harmonik oranı için korrelasyon orta düzeyde idi.

Sonuç: Ortalama temel frekans için elde edilen rakamsal değerler iki bilgisayar program arasında karşılaştırılabilirdir. Shimmer için elde edilen değerler anlamlı olarak farklı olmamakla birlikte, orta derecede korrelasyon göstermektedir. Jitter değerleri ve gürültü-harmonik oranı iki akusitk analiz programı arasında karşılaştırılabilir değildir.

Anahtar sözcükler: Ses, akustik analiz, ses analizi, MDVP, Praat

Received: 18.09.2009 – Accepted: 25.11.2010

[1] Department of Otolaringology, Ankara Education and Research Hospital, Ankara - TURKEY
[2] Department of Otolaringology, Faculty of Medicine, Kahramanmaraş Sütçü İmam University, Kahramanmaraş - TURKEY
Correspondence: *Haldun OĞUZ, Department of Otolaringology, Ankara Education and Research Hospital, Cebeci, Ankara - TURKEY*
E-mail: drhoguz@gmail.com

Introduction

Speech is the most valuable tool a person uses to express his or her thoughts and feelings. Speech and voice disorders, and their impact on quality of life, are attracting more interest in today's communication-based society. This brings the need for an objective definition of normal and abnormal findings obtained during patient examinations in voice clinics. In addition to other objective instrumentation techniques used in the voice laboratory such as videolaryngostroboscopy, aerodynamic assessment, and pH monitorization, objective acoustic analysis has become an indispensable tool for patient evaluation. Objective acoustic analysis gives the clinician the chance to collect documentation for diagnosis and follow-up. Acoustic analysis is a useful tool. However, in order to report valuable results, each component of the analysis equipment must be defined and standardized. The requirements for the recording environment, recording equipment (microphone and recording device), transformation conditions (digitization of samples), and the type of signal obtained have been adequately defined in the literature (1-3).

With the increased interest in voice analysis, a number of acoustic analysis computer programs have been made available to clinicians and scientists. Most of the acoustic analysis programs use similar descriptions to objectively define fundamental frequency (FF), jitter, shimmer, and the noise-to-harmonics ratio (NHR). The Computerized Speech Lab software, Multi-Dimensional Voice Program (MDVP) (Kay Elemetrics Corporation, Lincoln Park, New Jersey, USA), is the most commonly used and cited acoustic analysis program (4). It reports the findings of analyzed voice samples with definitions for mean, standard deviation, and thresholds of normal for each parameter, which helps the clinician to immediately assess the findings for a particular patient. Praat, designed by Paul Boersma and David Weenink of the Phonetic Sciences Department of the University of Amsterdam, is free software that is used and supported by many clinicians and scientists all over the world (5-8). It has been demonstrated that Praat is very successful at discriminating pathological voices from normal ones in comparative clinical studies (9-11). However, the program does not yet have established values for the thresholds between normal and abnormal voices. The aim of this study was to identify whether the results obtained from the same voice samples are comparable and/or correlative between the MDVP and Praat acoustic analysis computer programs.

Materials and methods

A total of 47 subjects were randomly selected from a voice database to reflect the spectrum of normal and pathologic voices that are usually seen in a voice clinic. Before the voice recordings were collected, each subject in the database underwent a complete otolaryngological examination. Subjects were also evaluated with videolaryngostroboscopy using a 90° rigid scope (Karl Storz laryngostrobe, Tuttlingen, Germany) in order to define possible laryngeal findings leading to voice changes. Diagnoses of the subjects were as follows: 14 (29.8%) with unilateral vocal cord paralysis, 3 (6.4%) with vocal cord cysts, 1 (2.1%) with myasthenia gravis, 1 (2.1%) with a vocal cord nodule, 1 (2.1%) with a pyriform sinus tumor, 2 (4.3%) with essential tremor, 2 (4.3%) with acute laryngitis, 1 (2.1%) with Parkinson's disease, 1 (2.1%) with a polyp, 3 (6.4%) with type I postoperative thyroplasty, and 18 (38.3%) normal subjects.

The voice samples were recorded by the same examiner under identical conditions in a sound-treated room with an ambient noise below 50 dB. The task was demonstrated by the examiner before recording. For each subject, 5 samples of sustained vowel /a/ at a comfortable pitch, constant amplitude, and flat tone were obtained. A Shure C606N cardioid microphone (Shure Inc., Niles, IL, USA) was placed on a stand 8 cm from the subject at an angle of 45° to the subject's mouth to decrease aerodynamic noise from the mouth. Praat software, version 4.2.17, was used for recording voice data for a minimum duration of 5 s on a personal computer with a sampling rate of 22,050 Hz.

Voice samples were saved in .wav format in the database. Because MDVP (Model 5105, Version 2.7.0) does not evaluate voice recordings below a sampling rate of 25 kHz, the perceptually defined best voice sample for each patient was selected and upsampled to 50 kHz by Praat software using the synthesize, convert, resample, and 50,000 Hz commands in the Praat objects window. The resampled voice file was then recorded in .wav format. The voice samples, each of a minimum duration of 5 s, were opened in MDVP. In order to exclude irregularities associated with the onset and offset of phonation, the most stable 3 s of the midvowel segment was chosen by the clinician and recorded in .nsp format. The same trimmed 3-s voice sample was then used in both Praat and MDVP for obtaining the objective acoustic evaluation results. This procedure was repeated for each voice sample. For the comparison and correlation studies, mean FF, jitter local (Jlocal), jitter absolute (Jabs), jitter relative average perturbation (Jrap), jitter period perturbation quotient (Jppq), shimmer dB (SdB), shimmer local (Slocal), shimmer amplitude perturbation quotient (Sapq), and NHR ratio values were obtained for each sample.

Some of the acoustic parameters obtained from MDVP and Praat are named differently in the voice reports of these 2 computer programs. The definitions and abbreviations used in this paper and their equivalents in both computer programs are shown in Table 1.

The statistical analyses for comparison and correlation studies were done with StatCrunch 4.0 (Integrated Analytics LLC) statistics software. An independent samples t-test was used to compare the results of the 2 computer programs for statistical significance.

Results

There were 26 female subjects (55.3%) and 21 (44.7%) male subjects.

According to the classification system of the National Center for Voice and Speech, all of the evaluated voice samples were type 1 signals (1). Type 1 signals are nearly periodic voice samples, and performing acoustic analysis for perturbation parameters on such samples is reliable (1). It is not recommended to perform acoustic analysis if a voice sample is not a type 1 signal.

Acoustic analysis results obtained by the 2 analysis programs and a comparison of them are shown in Table 2 and Figures 1-4. There was no statistically significant difference between the absolute values of the 2 computer programs for mean FF (2-tailed P = 0.996) (Figure 1, Table 2). The variance interval ± SD was also nearly identical for both programs.

Table 1. Definitions and abbreviations used in this paper and their equivalents in the voice reports of 2 acoustic analysis programs.

Definitions and abbreviations	MDVP report: voice report	Praat voice report
Mean fundamental frequency (mean FF)	Mean fundamental frequency	Mean pitch
Jitter local (Jlocal)	Jitter percent	Jitter (local)
Jitter absolute (Jabs)	Absolute jitter	Jitter (local, absolute)
Jitter relative average perturbation (Jrap)	Relative average perturbation	Jitter (rap)
Jitter period perturbation quotient (Jppq)	Pitch perturbation quotient	Jitter (ppq5)
Shimmer dB (SdB)	Shimmer in dB	Shimmer (local, dB)
Shimmer local (Slocal)	Shimmer percent	Shimmer (local)
Shimmer amplitude perturbation quotient (Sapq)	Amplitude perturbation quotient	Shimmer (apq11)
Noise-to-harmonics ratio (NHR)	Noise-to-harmonic ratio	Mean noise-to-harmonics ratio

Table 2. Acoustic analysis results obtained by MDVP and Praat acoustic analysis programs; * indicates a statistical significance of P < 0.001.

Parameter	Software	Mean ± standard deviation
Mean FF	MDVP Praat	229.743 ± 78.803 229.828 ± 78.878
Jabs *	MDVP Praat	79.883 ± 82.780 28.148 ± 33.339
Jlocal *	MDVP Praat	1.618 ± 1.503 0.550 ± 0.581
Jrap *	MDVP Praat	0.963 ± 0.876 0.303 ± 0.310
Jppq *	MDVP Praat	0.971 ± 0.942 0.342 ± 0.401
SdB	MDVP Praat	0.481 ± 0.291 0.505 ± 0.345
Slocal	MDVP Praat	5.423 ± 3.160 5.590 ± 3.686
Sapq	MDVP Praat	3.996 ± 2.410 4.055 ± 2.784
NHR *	MDVP Praat	0.144 ± 0.050 0.028 ± 0.045

There were no statistically significant differences between the absolute results of the 2 computer programs for Slocal, SdB, and Sapq (Figure 2, Table 2). For Slocal, SdB, and Sapq, respectively, 2-tailed statistical significance values (P) of 0.813, 0.717, and 0.914 were obtained. Mean values and variance interval were slightly higher with Praat.

There were statistically significant differences for Jlocal, Jabs, Jrap, and Jppq between the 2 computer programs (P < 0.001 each). The mean values and variance were significantly lower with Praat for each of the jitter parameters (Figure 3, Table 2).

There was a statistically significant difference for NHR (P < 0.001) between Praat and MDVP. The mean value was significantly higher with MDVP (Figure 4, Table 2).

There was perfect positive correlation between MDVP and Praat for mean FF values (r = 0.999) (Figure 5). There was a strong positive correlation for the jitter values. The correlation was 0.921 for Jabs, 0.899 for Jlocal, 0.889 for Jrap, and 0.897 for Jppq (Figure 5). The NHR (r = 0.804) and shimmer values (r = 0.734 for SdB, r = 0.685 for Slocal, and r = 0.770 for Sapq) were positively correlated, but to a lesser extent (Figure 5).

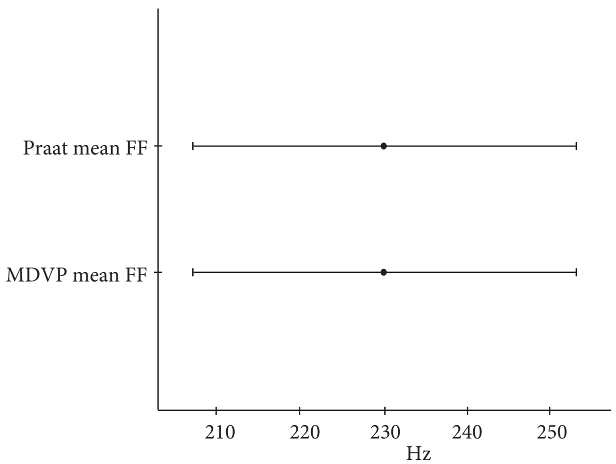

Figure 1. Graphic representation shows mean ± 2 standard deviations of values of mean fundamental frequency (FF) for both software programs.

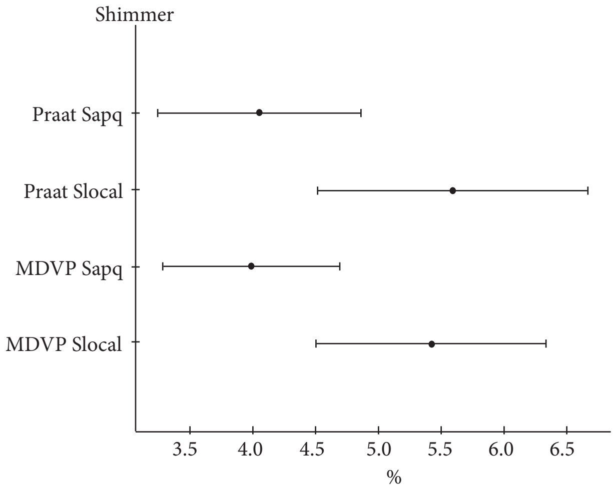

Figure 2. Graphic representation shows mean ± 2 standard deviations of shimmer amplitude perturbation quotient (Sapq) and shimmer local (Slocal) for both acoustic analysis programs.

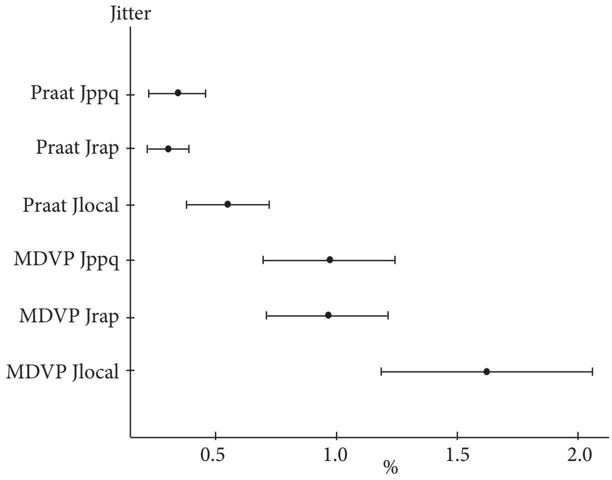

Figure 3. Graphic representation shows mean ± 2 standard deviations of jitter relative average perturbation (Jrap), jitter local (Jlocal), and jitter period perturbation quotient (Jppq) for both computer programs.

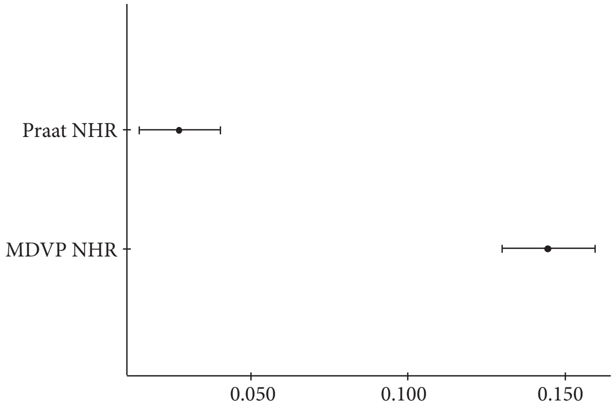

Figure 4. Graphic representation shows mean ± 2 standard deviations of noise-to-harmonics ratio (NHR) for Praat and MDVP.

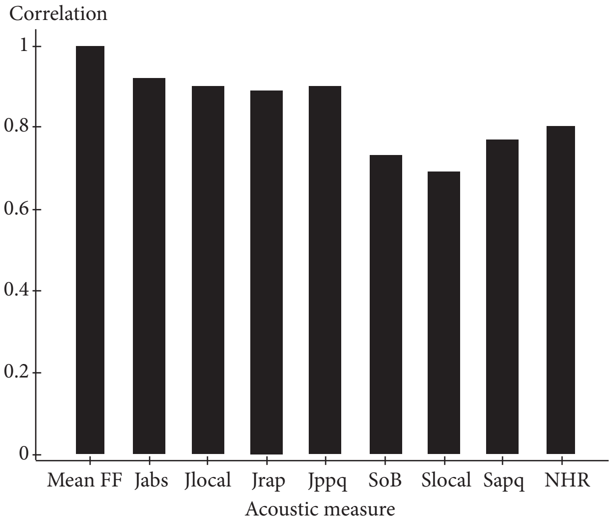

Figure 5. Correlation of the voice parameters between 2 acoustic analysis programs.

Discussion

Definition of parameters

Jitter is one of the main measures of microinstability in vocal cord vibrations (12). It refers to a cycle-to-cycle, short-term perturbation in the fundamental frequency of the voice (1). Jabs is the average absolute difference between consecutive periods and is defined in microseconds (13). Jlocal is the average absolute difference between consecutive periods, divided by the average period (13). It is the relative evaluation of the very short-term variability of the pitch within the analyzed voice sample (14). Jrap is the average absolute difference between a period and the average of it and its 2 neighbor periods (smoothing factor of 3 periods), divided by the average period (13,14). Jppq is the average absolute difference between a period and the average of it and its 4 closest neighbor periods (smoothing factor of 5 periods), divided by the average period (13,14). Jlocal, Jrap, and Jppq are defined in percentages.

Shimmer is a cycle-to-cycle, short-term perturbation in amplitude of voice (1). SdB is the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20 (13). It is defined in dB. Slocal is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude (13). It is the relative evaluation of very short-term variability of peak-to-peak amplitude within the analyzed voice sample (14). Sapq is the 11-point amplitude perturbation quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its 10 closest neighbors (smoothing factor of 11 periods), divided by the average amplitude (13,14). Slocal and Sapq are defined in percentages.

Increased jitter and shimmer reflect both diminished laryngeal control and degenerative changes in laryngeal tissue (12). In addition to short-term period and amplitude variations, inconsistent or absent vocal cord closure leads to air leakage through the glottis, which is acoustically characterized as noise (15). The NHR is the average ratio of the inharmonic spectral energy to the harmonic spectral energy (14). This is a general evaluation of noise in the analyzed signal and is not specific to any cyclic parameter (14). It includes contributions from both perturbations of amplitude and frequency. The measure correlates best with the overall perception of noisiness or roughness in the signal (1).

The algorithms of each of these parameters for both computer programs are already present in the literature (13,14,16-18).

Comments on results

Today's computer and software technology provides for the ability to transfer voice files between computers and laboratories. Currently, voice samples are usually recorded in digital format. Different file format types such as .wav, .nsp, .aiff, .aifc, and .nist can be used to record and save data. Most of these file forms are supported by different computer programs, which means that a sound file recorded by one software package may be easily transferred to and analyzed by other software. However, this transportability is not yet valid for the obtained acoustic analysis results.

Defining normal values for a voice is very difficult. Due to well-known anatomic and physiological differences, child, female, and male voices differ significantly. It was also shown that aging and different hormones have important effects on voice quality (19,20). In addition to these personal variables, factors relating to environmental conditions and data acquisition devices have been covered in the literature (1-3). Sustained vowels are usually used for obtaining perturbation parameters in order to decrease linguistic and dialectical variations and increase subject consistency. However, even studies with different vowels may result in significant differences (21).

In our study, by using the same voice samples, all of the above-mentioned factors that can lead to variations in results were excluded. This method gave us the chance to objectively compare Praat and MDVP acoustic analysis programs for the first time in the literature. It was observed that mean FF, Slocal, SdB, and Sapq results are numerically comparable between the Praat and MDVP computer programs. However, because our study group contained a spectrum of pathologic and normal voices, it is questionable whether these results are comparable for normal voices or a specific group of patients. Further investigation of this hypothesis must include studies on normal voices only, as well as separate studies regarding specific voice diseases such as vocal cord paralysis or a particular neurological disease. It was also observed that Jlocal, Jabs, Jrap, Jppq, and NHR values obtained from the 2 computer programs could not be numerically compared. Although both programs use similar definitions for these parameters, the reason for this variation in results may be caused by the different voicing strategies and algorithms used by the 2 computer programs (13,14).

The observed strong to moderate correlations for even numerically incomparable parameters lead to an interpretation that both computer programs may have similar decision strategies for normal and pathologic voices. Further studies are needed to establish stronger conclusions on this view.

Many different factors, such as the patient turnover at a clinic, the adequacy of the personnel, and the limits of monetary resources, may affect the selection of data acquisition devices and acoustic analysis programs used in a voice clinic. Under these variable circumstances, with respect to the general and important obligations to decrease environmental, equipmental, intersubject, and intrasubject effects on analysis quality, each laboratory may evaluate their findings according to their own normal dataset and report their findings in a similar way. In previous studies, both of the computer programs used in this study were shown to effectively discriminate between pathologic and normal voices (9-11,19). We hope that further investigation and improvements in voice analysis programs will allow voice clinicians to share and compare the data obtained from different software packages.

Conclusion

It was shown that values obtained for mean FF, SdB, Slocal, and Sapq may be numerically compared between Praat and MDVP. The values obtained for Jlocal, Jabs, Jrap, Jppq, and NHR are not numerically comparable. There is a strong to moderate correlation between the results of the 2 computer programs. Further studies are needed to show this comparability and correlation for specific vocal pathologies and normal subjects.

Acknowledgements

The authors would like to thank to Robert Hillman, Ph.D., CCC-SLP, and Steven M. Zeitels, M.D., F.A.C.S., associate professors (co-director/research director and director, respectively) of the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA, for their kind support in this study.

References

1. Titze IR. Workshop on acoustic voice analysis: summary statement. Iowa City (IA): National Center for Voice and Speech; 1995. http://www.ncvs.org/ncvs/info/rescol/sumstat/sumstat.pdf; accessed 25 June 2008.
2. Deliyski DD, Shaw HS, Ewans MK. Adverse effects of environmental noise on acoustic voice quality measurements. J Voice 2005; 19: 15-28.
3. Deliyski DD, Ewans MK, Shaw HS. Influence of data acquisition environment on accuracy of acoustic voice quality measurements. J Voice 2005; 19: 176-86.
4. Smits I, Ceuppens P, De Bodt MS. A comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab. J Voice 2005; 19: 187-96.
5. Praat: doing phonetics by computer. http://www.praat.org; accessed 25 June 2008.
6. Praat Tutorial Version 4.3, by Pascal van Lieshout. http://ots.utoronto.ca/users/vanlieshout/; accessed 8 August 2005.
7. Praat beginner's manual, by Sidney Wood. http://www.ling.lu.se/persons/Sidney/praate/frames.html; accessed 25 June 2008.
8. Praat Users Group. http://uk.groups.yahoo.com/group/praat-users/; accessed 25 June 2008.
9. Oguz H, Tarhan E, Korkmaz M, Yilmaz U, Safak MA, Demirci M et al. Acoustic analysis findings in objective laryngopharyngeal reflux patients. J Voice 2007; 21: 203-10.
10. Oguz H, Demirci M, Safak MA, Arslan N, Islam A, Kargin S. Effects of unilateral vocal cord paralysis on objective voice parameters. Eur Arch Otorhinolaryngol 2007; 264: 257-61.
11. Oguz H, Tunc T, Safak MA, Inan L, Kargin S, Demirci M. Objective voice changes in non-dysphonic Parkinson's disease patients. J Otolaryngol 2006; 35: 349-54.
12. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity. J Commun Disord 1997; 30: 403-15.
13. Praat manual. Version 4.2.17. Paul Boersma and David Weenink, Phonetic Sciences Department, University of Amsterdam, the Netherlands.
14. MDVP manual. Version 2.7.0, Kay Elemetrics Corporation, Lincoln Park, New Jersey, USA.
15. Reijonen P, Soderlund SL, Rihkanen H. Results of fascial augmentation in unilateral vocal fold paralysis. Ann Otol Rhinol Laryngol 2002; 111: 523-29.
16. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. J Speech Hear Res. 1996; 39: 126-34.
17. Boersma P. Stemmen meten met Praat. Stem-, Spraak- en Taalpathologie 2004; 12: 237-51.
18. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. IFA Proceedings 1993; 17: 97-110.
19. Akcam T, Bolu E, Merati AL, Durmus C, Gerek M, Ozkaptan Y. Voice changes after androgen therapy for hypogonadotrophic hypogonadism. Laryngoscope 2004; 114: 1587-1591.
20. Kandogan T, Seifert E. Influence of aging and sex on voice parameters in patients with unilateral vocal cord paralysis. Laryngoscope 2005; 115: 655-660.
21. Kilic MA, Ogut F, Dursun G, Okur E, Yildirim I, Midilli R. The effects of vowels on voice perturbation measures. J Voice 2004; 18: 318-324.