Biomedical Signal Processing and Control 4 (2009) 202–205

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A clinical comparison between two acoustic analysis softwares: MDVP and Praat

Ofer Amir [a,*], Michael Wolf [b], Noam Amir [a]

[a] Department of Communication Disorders, Tel-Aviv University, Tel-Aviv, Israel
[b] Department of Otolaryngology, Head and Neck Surgery Department, Sheba Medical Center, Tel-Hashomer, Israel

ARTICLE INFO

Article history:
Received 27 May 2008
Received in revised form 30 October 2008
Accepted 6 November 2008
Available online 24 December 2008

Keywords:
Acoustic analysis
MDVP
Praat
Voice disorders

ABSTRACT

The Multi-Dimensional Voice Program (MDVP) and Praat are computer programs commonly used for acoustic analysis of voice in clinical and research settings. Both softwares extract a set of acoustic parameters, many of which are defined similarly. The purpose of this study was to compare results obtained by both programs, and examine whether they can clinically distinguish among pathological groups differently. Fifty-eight women participated in the study. Of these women, 28 were diagnosed with functional dysphonia and 30 were diagnosed with benign mass-lesions (10 nodules, ten polyps and 10 cysts). Six productions of the vowels /a/ and /i/, were analyzed using MDVP and Praat. Results show similar mean fundamental-frequency (mF0) values for both programs ($P > 0.05$). However, values of jitter, shimmer, noise-to-harmonic ratio (NHR) and degree of unvoiced (DUV) segment were significantly lower using Praat, in comparison with MDVP. Jitter values obtained using MDVP, for the vowel /i/, revealed a significant group difference between the nodule and cyst groups ($P < 0.05$). This group contrast was not observed using Praat. Results demonstrate that although high correlations are found between values obtained by both programs, individual numerical values vary greatly. Therefore, combining results from both programs are not advisable. In addition, there are indications that linear transformation for the results from one program to the other might lead to erroneous conclusions, and should be carried out with caution.

1. Introduction

Acoustic analysis of voice is considered valuable for quantifying measures of voice quality in various experimental as well as clinical settings. The validity of this tool has been challenged by many studies, since it is yet unclear which set of acoustic measures best represents voice quality [1]. Moreover, the relationship between vibratory properties of the vocal folds and specific acoustic measures has not been substantiated yet [2]. While previous studies have included various sets of acoustic measures, the majority of these studies have examined, among other parameters, fundamental frequency (F0), measures of frequency-perturbation (e.g., jitter), measures of amplitude-perturbation (e.g., shimmer) and various noise-indices. For example, a recent study which also compared different commercial softwares used a very similar set of parameters [3].

Validity and reliability of acoustic analysis performed with different tools was previously shown to be affected by many factors. These include, for example, microphone type, noise levels, data acquisition system, sampling rate and software used for analysis [4,5]. Ostensibly, the values of the commonly used frequency- and amplitude-perturbation measures should not be dependent on the software used to obtain them. Jitter and shimmer, for example, are defined by relatively simple and standardized formulas [6]. The differences observed between the numerical values obtained for these measures using the different softwares apparently stem mainly from the raw F0 data on which these calculations are based. Despite the basic nature of this parameter, there is no standardized algorithm for calculation of F0, which has been adopted and implemented by all programs.

While different methods for calculating F0 may yield relatively small differences in mean F0, they can influence the perturbation measures to a far greater extent. This introduces a difficulty for the clinical voice specialist, because the different programs which are available for conducting voice analysis could report different values, when analyzing identical voice samples. Moreover, it is not clear whether normative data which are presented by specific software (e.g., the data used for the radial graph in Multi-Dimensional Voice Program (MDVP)) are comparable with values obtained in other programs. This possible discrepancy between the results obtained by different programs was previously noticed and addressed by various researchers [3,4,7,8].

It is important to note that previous research has indeed established some guidelines for proper recording procedures, in order to ensure that the perturbation measures are reliable and

* Corresponding author at: Department of Communication Disorders, Tel-Aviv University, Sheba Medical Center, 52621 Tel-Aviv, Israel. Fax: +972 3 5352868.
E-mail address: oferamir@post.tau.ac.il (O. Amir).

1746-8094/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bspc.2008.11.002

accurate. The issue of sampling rate, for example, was investigated thoroughly recently by Deliyski et al. [4]. Another study by the same authors [5], examines the influence of various other factors on the values of the perturbation measures, but does not attempt to establish whether one of the different softwares performed better than the other in any sense. Moreover, their data was collected from healthy subjects only. Variability between the results obtained from different programs is likely to increase even more when examining pathological subjects, magnifying algorithmic differences between them. Therefore, it remains that any practicing clinician, confronted with the mass of data in the literature, would benefit from a straightforward comparison between softwares in order to make an informed decision as to which software to adopt.

In the present study, we perform a preliminary examination, from a clinical perspective, of the comparison between the results of the acoustic analyses performed by two programs: MDVP (Kay Elemetrics) and Praat (Boersma & Weenink). These programs are commonly used for acoustic analysis in clinical as well as research settings, and while MDVP is a commercial package, Praat is distributed for free use. Both softwares provide a calculation of a set of parallel acoustic measures. As noted earlier, previous studies have already shown that the two programs could present different values for the same acoustic measures [11]. Therefore, the present study was designed to address two questions, which were defined from a clinical perspective: (1) do the two programs provide similar or different values for a basic set of commonly used acoustic measures and (2) do the results obtained by any of the programs distinguish better between specific pathological groups. Our hypothesis was that the two programs would yield different values for the same acoustic measures, but it was unknown whether, despite these numerical differences, similar group contrasts would be obtained with the two programs.

2. Methods

2.1. Participants

Fifty-eight women who were examined in the Voice Clinic at the "Sheba" Medical Center, Tel-Hashomer, were included in the study after obtaining the approval from our institutional review board and written consent of all participants. All patients were women over the age of 18, and all had undergone a laryngeal stroboscopy and a voice evaluation. Of these women, 28 were diagnosed with functional dysphonia (i.e., patients were dysphonic, with no observed organic finding). Thirty women were diagnosed with vocal fold benign mass-lesions. Of these women, 10 were diagnosed with vocal nodules, 10 with polyps and the remaining 10 were diagnosed with cysts. Table 1 presents group means and overall means of age and physical characteristics of the participants in the four study groups.

Although physical characteristics are not considered to directly affect voice quality, the different groups were tested for differences in physical features, to assess the possibility of a bias effect. To this end, three separate analyses of variance were performed. No significant differences were found among the four study group for age, height or weight: $F(3, 54) = 0.42$, $P = 0.742$; $F(3, 52) = 0.07$, $P = 0.976$; $F(3, 52) = 0.31$, $P = 0.818$; respectively.

2.2. Recordings

Each woman was recorded, individually, while seated in a quiet room. Recordings were performed using a Sennheiser PC160 headset microphone, connected directly to a computer, with a sampling rate of 48 kHz. Each subject was recorded producing the vowels /a/ and /i/ repeatedly six times.

2.3. Acoustic analyses

All recordings were analyzed twice: using MDVP and using Praat. The MDVP analyses were performed manually. The Praat analyses were performed automatically, controlled by a Matlab program written Ad hoc for this study. During analyses, F0 identification range was set between 110–500 Hz, to minimize octave errors, and to avoid erroneous tracking of F0. Although the two programs provide extensive sets of acoustic parameters, only five parallel measures that are calculated by both programs were included in the present analysis. These measures included mean fundamental frequency (mF0), jitter, shimmer, noise-to-harmonic ratio (NHR) and percentage of unvoiced segments (referred to as degree of unvoiceness (DUV) in the MDVP program, and as DEG in the Praat program).

Both programs calculate F0 using algorithms based on the autocorrelation method [9,10]. Nevertheless, there are differences between the two implementations, which evidently cause noticeable differences between the results obtained by the two programs. The details of the implementations are well documented, though to the best of our knowledge, there is no comparison of their absolute accuracy. Because absolute accuracy is not well defined when dealing with pathological voices, we decided to use, for the purpose of this study, a clinical criterion of differentiating among specific pathological groups.

Fig. 1 illustrates an example of the differences between the two programs in tracking F0. In this figure, the calculated F0 points are presented over a short segment of approximately 0.8 s, which was extracted from a single file that was included in this study. Mean F0 values calculated for this voice segment were 181.07 Hz using MDVP and 181.16 Hz using Praat. Nonetheless, the figure clearly illustrates that the MDVP presents a larger spread of values in comparison with the values obtained with Praat. This is further corroborated in the following section.

2.4. Statistical analyses

Prior to the statistical analysis, preliminary data reduction was performed by averaging the results of the acoustic analyses of the repeated recordings of each vowel produced by each participant. Separate analyses of variance were performed for each vowel. In these analyses, group (nodule, polyp, cyst and functional) was treated as a main factor, and programs (MDVP and Praat) was treated as a repeated factor. In addition, Pearson correlation coefficients [12] were calculated to compare between the results obtained by the two programs. In all analyses, significance levels

Table 1
Means and standard deviations (in parentheses) of age, height and weight for the four study groups.

	Group				
	Nodule	Polyp	Cyst	Functional	Overall
Age (years)	36.30 (9.56)	42.50 (6.79)	38.20 (13.98)	38.07 (15.22)	38.55 (12.89)
Height (cm)	164.40 (5.78)	163.50 (7.13)	163.90 (11.76)	163.29 (4.81)	163.63 (6.79)
Weight (kg)	66.10 (16.59)	62.88 (7.68)	61.30 (13.78)	65.29 (13.10)	64.38 (13.10)

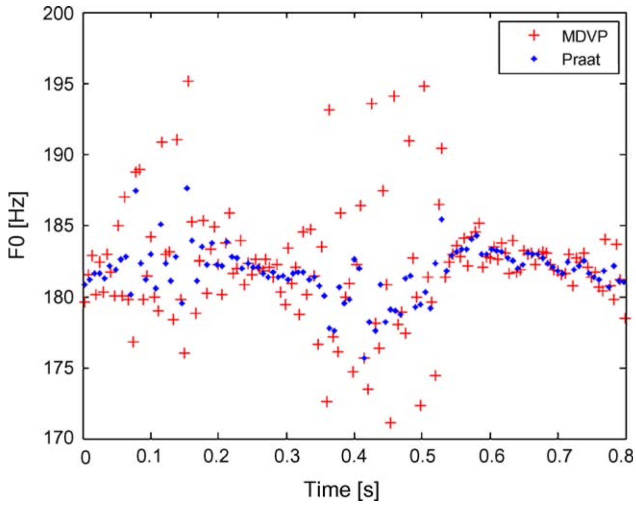

Fig. 1. F0 values calculated using Praat and MDVP over a short segment of the vowel /a/ produced by one of the women who participated in the study.

were set at $P = 0.05$. All statistical analyses were conducted using SPSS 15.0.1 (SPSS Inc., Chicago, IL).

3. Results

Table 2 presents the results of the acoustic analyses performed using the two programs for the four study groups. Results show that similar numerical values were obtained for mF0 using the two programs. However, the values obtained for the jitter, shimmer, NHR and DUV measures were, in general, higher in MDVP than those obtained using Praat.

Statistical analyses revealed significant differences between the results obtained from two programs for jitter [$F(1, 53) = 68.84$, $P < 0.001$; $F(1, 53) = 49.29$, $P < 0.001$; for /a/ and /i/ respectively], shimmer [$F(1, 53) = 3.61$, $P = 0.063$; $F(1, 53) = 5.11$, $P = 0.028$; for /a/ and /i/ respectively], NHR [$F(1, 53) = 336.16$, $P < 0.001$; $F(1, 53) = 408.48$, $P < 0.001$; for /a/ and /i/ respectively] and for DUV [$F(1, 53) = 26.70$, $P < 0.001$; $F(1, 53) = 32.88$, $P < 0.001$; for /a/ and /i/ respectively]. No significant differences were found between the two programs for the mF0 measure [$F(1, 53) = 0.467$, $P = .497$; $F(1, 53) = 0.039$, $P = 0.845$; for /a/ and /i/ respectively).

No significant main effect was found for group, for any of the acoustic measures tested. A significant program × group interaction was found only for the jitter measure in the vowel /i/ ($F(1, 53) = 3.88$, $P = 0.014$). Post hoc analysis, using Tukey's HSD (honest significant difference) [12], revealed a significant group difference between the nodule group (mean = 1.67, S.D. = 1.40) and the cyst group (mean = 3.16, S.D. = 1.77), when analysis was performed using the MDVP program (adjusted $P < 0.05$). This group contrast was not observed when analysis was performed using the Praat program.

Finally, high correlation coefficient values were observed between the results obtained in the two programs. Correlations for mF0 ranged between $0.963 < r < 0.970$. Correlations for the perturbation measures ranged between $0.719 < r < 0.932$. However, correlations for the DUV measure were moderate ($0.481 < r < 0.672$). This parameter could be influenced to a great degree by the "voicing threshold" parameter in Praat, which governs the voiced/unvoiced decision. In the present research it was left at its default value, though it might be of interest to compare results for several different values. It should be noted, though, that although high correlation coefficients were obtained for most parameters, further inspection of the data revealed additional information. Fig. 2, for example, presents the correlation between the jitter values for the vowel /a/, obtained using MDVP and Praat. It is evident that a high correlation coefficient value was obtained when computing the correlation over the entire range of values ($r = 0.82$). However, when the sample was limited to stimuli with relatively lower jitter values (0–3%), the correlation decreased to 0.39, although this range covered the majority of values. In contrast, when voice samples with higher jitter values were examined, the correlation coefficient was high ($r = 0.87$), although the sample size was smaller. Similar findings were observed for all other parameters and vowels.

4. Discussion

The results of our study support previous findings, suggesting that different programs present different values of acoustic measures [4,5]. This is attributed to algorithmic differences between the programs (see Boersma & Wininck, Praat manual). On the one hand, our data show that in most cases, similar group differences (or lack of differences) were obtained using both programs, and that strong correlations were found between the two programs. Furthermore, mean F0 values are also similar for the two programs. These findings could be interpreted to support common use of both programs. On the other hand, values of the perturbation and noise measures were notably different between the two programs, and in a specific condition (jitter for the vowel /i/) MDVP appeared to differentiate among pathological groups better than Praat. The latter finding suggests that combining results from the two programs, for clinical purposes, is not recommended, despite the use of the seemingly parallel acoustic measures. Apparently, acoustic measures, like those examined in our study could be useful for characterizing and quantifying voice properties, and possibly for differentiating between pathological and healthy voices [2,6]. However, these measures mostly fail to differentiate between specific laryngeal pathologies.

Table 2
Means and standard deviations (in parentheses) of mF0, jitter, shimmer, noise-to-harmonic ratio (NHR) and degree of unvoiceness (DUV) obtained for the four study groups using the MDVP and Praat programs.

Vowel	Measure	MDVP				Praat			
		Nodule	Polyp	Cyst	Functional	Nodule	Polyp	Cyst	Functional
/a/	mF0 (Hz)	197.65 (22.73)	206.40 (36.41)	225.36 (225.36)	201.62 (38.51)	198.36 (22.46)	205.96 (36.45)	217.61 (45.43)	204.64 (35.95)
	Jitter (%)	1.77 (1.68)	2.39 (1.06)	2.00 (1.47)	2.01 (1.64)	1.16 (1.37)	0.93 (0.27)	0.96 (0.77)	0.85 (0.91)
	Shimmer (%)	7.00 (9.18)	7.76 (3.46)	6.49 (3.20)	5.85 (4.30)	5.01 (5.09)	7.58 (3.75)	6.30 (4.37)	5.12 (3.91)
	NHR	0.19 (0.13)	0.20 (0.09)	0.15 (0.04)	0.17 (0.11)	0.09 (0.16)	0.09 (0.08)	0.07 (0.10)	0.06 (0.10)
	DUV	14.06 (19.46)	18.13 (18.13)	8.66 (18.13)	12.70 (19.16)	2.57 (5.49)	0.50 (0.58)	1.53 (2.46)	1.25 (2.46)
/i/	mF0 (Hz)	211.20 (25.01)	214.20 (36.16)	220.80 (34.40)	206.32 (38.68)	211.21 (25.07)	213.38 (34.08)	217.64 (35.72)	209.33 (35.51)
	Jitter (%)	1.67 (1.40)	2.49 (1.05)	3.16 (1.77)	1.94 (1.29)	1.15 (1.44)	1.09 (0.54)	1.20 (0.91)	1.20 (1.63)
	Shimmer (%)	4.57 (4.88)	5.88 (2.78)	5.93 (4.07)	4.72 (5.02)	3.01 (3.20)	5.92 (3.41)	5.25 (4.84)	3.84 (4.74)
	NHR	0.15 (0.06)	0.16 (0.04)	0.17 (0.06)	0.15 (0.08)	0.04 (0.06)	0.05 (0.04)	0.04 (0.04)	0.04 (0.07)
	DUV	10.31 (11.84)	10.21 (7.35)	11.68 (13.11)	0.93 (12.86)	0.53 (0.94)	1.86 (3.12)	1.59 (2.31)	1.43 (4.28)

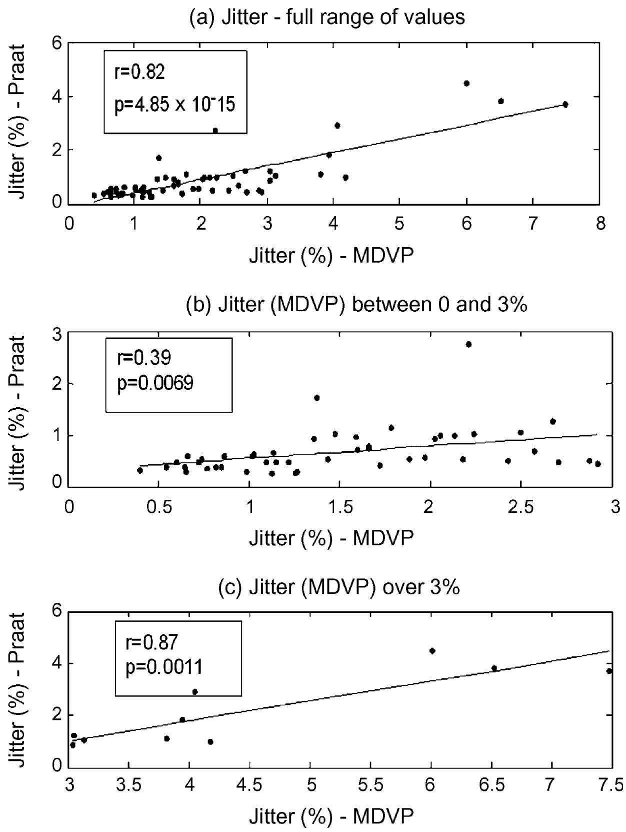

Fig. 2. Individual participants' jitter values for /a/, calculated by MDVP versus Praat, along with linear regression and correlation coefficient: (a) full range of jitter values; (b) jitter values range (MDVP) between 0% and 3%; (c) jitter values (MDVP) >3%.

It is interesting to observe that the strong correlations between the values calculated by the two programs initially suggested that values from one program can be linearly transformed to approximate the values calculated by another program. Maryn et al. [11], for example, have attempted to compare values of perturbation measures that were calculated using MDVP with those calculated using Praat. They, too, found a strong correlation between the values obtained by the two programs. Further, they suggested a conversion factor between MDVP and Praat that would allow for combining results of perturbation measures calculated by both programs. In contrast to this view, our preliminary data implies that this might lead to inaccurate results and conclusions. As shown in Fig. 2, examining jitter values between 0% and 3% only, revealed far lower correlation coefficients between MDVP and Praat values, than when overall jitter values were examined. This observation was consistent for all measures and vowels. This result is of special interest in light of two considerations. First, in the present data set jitter values of 0–3% covered the greater part of cases. Second, voices with relatively low perturbation values are expected to be analyzed more reliably [13]. Therefore we expected that stronger correlations between the two programs would be found within this range of perturbation value. The fact that the correlation between the results obtained by the two programs depended on the level of periodicity of the signal implies that a simple conversion factor between the two programs is problematic. Hence, this further suggests that results obtained from both programs are not comparable.

Based on these preliminary findings, it should be noted that the use of the reported thresholds for "normal" voice, as presented by MDVP, for example, should be restricted to measures calculated by a specific program, and could not be used for analyses made with other programs. This is especially pertinent when examining measures that are based on cycle-to-cycle variation. Future research should evaluate these findings in comparison with listener judgment. In addition, the F0 extraction performed in these two programs could also be evaluated in conjunction with manual F0 tracking.

Acknowledgement

The authors would like to thank Ms. Esther Shabtai, M.Sc. for her assistance in the statistical analyses.

Appendix A. Further considerations

Aside from the purely clinical aspects of MDVP and Praat, it may be of interest to the reader to point out some additional practical differences between them, most of them being to the advantage of Praat, in our opinion. Praat has a scripting language that can be used to automate the execution of largely repetitive tasks. It also has the ability to create easily used Graphical User Interfaces for these scripts. Praat is open-source, enabling the knowledgeable user to examine the exact code the program is based on. In addition, an executable named "Praatcon" is available in conjunction with Praat. In essence it is the same software without the interface. This executable can be called from other languages (e.g. Matlab), allowing Praat capabilities to be incorporated in user written programs. Finally, the F0 algorithms in Praat have several tunable parameters (aside from range, available also in MDVP), which can be used to fine tune them in specific cases, e.g. in the presence of background noise.

References

[1] M.M. Hakkesteegt, M.P. Brocaar, M.H. Wieringa, L. Feenstra, The relationship between perceptual evaluation and objective multiparametric evaluation of dysphonia severity, J. Voice 22 (2008) 138–145.
[2] I.R. Titze, Principles of Voice Production, 2nd edn., National Center for Voice and Speech, Iowa City, IA, 2000.
[3] J.I. Godino-Llorente, V. Osma-Ruiz, N. Saenz-Lechon, et al., Acoustic analysis of voice using WPCVox: a comparative study with Multi Dimensional Voice Program, Eur. Arch. Otorhinolaryngol. 265 (2008) 465–476.
[4] D.D. Deliyski, H.S. Shaw, M.K. Evans, Influence of sampling rate on accuracy and reliability of acoustic voice analysis, Logop. Phoniatr. Vocol. 30 (2005) 55–62.
[5] D.D. Deliyski, H.S. Shaw, M.K. Evans, Regression tree approach to studying factos influencing acoustic voice analysis, Folia. Phoniatr. Logop. 58 (2006) 274–288.
[6] R.J. Baken, Clinical Measurement of Speech and Voice, Allyn and Bacon, Needham Heights, MA, 1987.
[7] I. Smith, P. Ceuppens, M.S. De Bodt., A comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab, J. Voice 19 (2005) 187–196.
[8] M.P. Karnell, K.D. Hall, K.L. Landahl, Comparison of fundamental frequency and perturbation measures among three analysis systems, J. Voice 9 (1995) 383–393.
[9] D.D. Deliyski. MDVP software instruction manual, Kay Elemetrics Corp., Appendix E.
[10] P. Boersma, Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, in: IFA Proceedings, 1993, p. 17.
[11] Y. Maryn, P. Corthals, J. Vanderbeke, E. Vanwynsberge, M. De Bodt, P. van Cauwenberge, Intersystem and interprogram difference in acoustic perturbation measurement: a comparison between MDVP and Praat, in: Paper presented in the 7th Pan European Voice Conference (PEVOC7), Groningen, The Netherlands, 29th August–1st September, 2007.
[12] G.V. Glass, K.D. Hopkins, Statistical Methods in Education and Psychology, 2nd edn., Needham Heights, Massachusetts, 1984.
[13] I.R. Titze, Workshop on Acoustic Voice Analysis: Summary Statement, National Center for Voice and Speech, Iowa City, 1995.