Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2009;61:217–226
DOI: 10.1159/000227999

Published online: July 9, 2009

Perturbation Measures of Voice: A Comparative Study between Multi-Dimensional Voice Program and Praat

Youri Maryn[a] Paul Corthals[b] Marc De Bodt[d] Paul Van Cauwenberge[c] Dimitar Deliyski[e]

[a]Department of Speech-Language Pathology and Audiology, Department of Otorhinolaryngology and Head and Neck Surgery, Sint-Jan General Hospital, Bruges, [b]Faculty of Health Care 'Vesalius', Hogeschool Gent, and [c]Faculty of Medicine and Health Sciences, Department of Otorhinolaryngology and Head and Neck Surgery, University of Ghent, Ghent, and [d]Department of Communication Disorders, Department of Otorhinolaryngology and Head and Neck Surgery, University Hospital, Antwerp, Belgium; [e]Department of Communication Sciences and Disorders, University of South Carolina, Columbia, S.C., USA

Key Words

Perturbation measurement · Multi-Dimensional Voice Program · Praat · Computerized Speech Lab

Abstract

Background/Aims: Frequency and amplitude perturbations are inherent in voice acoustic signals. The assessment of voice perturbation is influenced by several factors, including the type of recording equipment used and the measurement extraction algorithm applied. In the present study, perturbation measures provided by two computer systems (a purpose-built professional voice analysis apparatus and a personal computer-based system for acoustic voice assessment) and two computer programs (Multi-Dimensional Voice Program and Praat) were compared. ***Methods:*** Correlations and inferential statistics for seven perturbation measures (absolute jitter, percent jitter, relative average perturbation, pitch perturbation quotient, shimmer in decibels, percent shimmer, and amplitude perturbation quotient) in 50 subjects with various voice disorders are presented. ***Results:*** Results indicate statistically significant differences between the two systems and programs, with the Multi-Dimensional Voice Program yielding consistently higher measures than Praat. Furthermore, correlation analyses show weak to moderate proportional relationships between the two systems and weak to strong proportional relationships between the two programs. ***Conclusion:*** Based on the literature and the proportional relationships and differences between the two systems and programs under consideration in this study, one can state that one can hardly compare frequency perturbation outcomes across systems and programs and amplitude perturbation outcomes across systems.

Introduction

Minor disturbances in the frequency and the amplitude of the voice signal, called perturbations, are unavoidably present even when one tries to produce a perfectly steady sound [1]. In patients with a voice problem, perturbation may become worse and result in a more se-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1021–7762/09/0614–0217$26.00/0

Accessible online at:
www.karger.com/fpl

Youri Maryn
Department of Speech-Language Pathology and Audiology
Department of Otorhinolaryngology and Head and Neck Surgery
Sint-Jan General Hospital, Ruddershove 10, BE–8000 Bruges (Belgium)
Tel. +32 50 452 448, Fax +32 50 452 290, E-Mail youri.maryn@azbrugge.be

vere deviation from the normal voicing pattern. Perceptually, this may be interpreted as dysphonia and described using labels like hoarse, breathy and rough. Popular acoustic metrics to assess dysphonia are jitter and shimmer, denoting short-term (cycle-to-cycle) variability in fundamental frequency (F_0) and amplitude, respectively. A comprehensive review on this topic can be found in Baken and Orlikoff [2]. Since Lieberman [3] introduced the concept of perturbation analysis in the area of voice and speech, the demand for reliable, valid and objective voice analyses has motivated acoustic voice research and perturbation measurements have undergone considerable refinement. The availability of user-friendly personal computer systems has made quantitative voice and speech analysis commonly accessible [4, 5]. A well-known commercially available computer system for voice analysis, the Computerized Speech Lab (CSL) by Kay Elemetrics Corp. (currently known as KayPentax) [6] offers several perturbation measures in its Multi-Dimensional Voice Program (MDVP) [7]. An example of freely available personal computer-based analysis software is Praat [8, 9]. It also provides perturbation measures in a voice report.

Acoustic voice analysis based on perturbation measures has long been subject to debate. A key issue is validity, in particular concurrent criterion-related validity with perceptual evaluation as the bench mark for voice quality assessment.

Several authors have found significant relationships between perceptual evaluation and acoustic perturbation. For example, Eskenazi et al. [10] point to jitter (percent) as a predictor for breathiness and hoarseness, in contrast to pitch perturbation quotient and amplitude perturbation quotient. Dejonckere et al. [11] found significant correlations between jitter (percent) and roughness, between shimmer (percent) and breathiness, and between shimmer (percent) as well as noise-to-harmonics ratio and Hirano's [12] grade index for perceptual voice assessment. Wolfe and Martin [13] revealed significant correlations between jitter (percent) and breathiness and between shimmer and hoarseness, an inclusive term the authors use for indicating glottal noise and roughness. However, such perturbation-quality relationships do not always emerge. For example, Bhuta et al. [14] reported significant multivariate correlations between MDVP noise measures (voice turbulence index, noise-to-harmonics ratio and soft phonation index) and perceptual GRBAS [12] voice evaluation, but individual perturbation measures were not observed to be significant correlates. De Bodt [15] could not find any meaningful objective acoustic correlate for perceptual GRBAS ratings. Differences in judge experience, voice samples used, type and severity of pathology, and data acquisition hardware and software often lead to inconsistent research findings. A more profound discussion on the validity of acoustic metrics for voice quality is beyond the scope of this article, and interested readers are referred to Kreiman and Gerratt [16].

Another issue concerns the differences in measuring outcome between computer systems and between computer programs. Since every computerized speech recording and analysis system has its own configuration for data acquisition such as microphone type and localization relative to the source [17, 18], presence or absence of external amplifying hardware (as in the case of Kay Elemetrics Corp.'s CSL versus a generic sound card, respectively), type of personal computer with its typical hardware and software settings for recording and the properties of its internal sound card [19, 20], use of external digital recording apparatus such as digital audio tape or minidisc [21], analysis and processing program [22–24], and measurement algorithms [25], etc., differences in any of these system-related items can lead to more or less *intersystem differences* in perturbation measurements. Collectively, Deliyski et al. [26] investigated the extent and the order in which gender, microphone, number of tokens, type of environmental noise, level of environmental noise, data acquisition system and software influence perturbation measures on 80,000 audio recordings. Although all of the factors were considered to be influential, Deliyski et al. [26] concluded that the most prominent effect on perturbation measures was exercised by analysis software, followed by gender and type of microphone.

When the same recording is analyzed using different software, keeping all other system-related factors invariant, the differences in results must be due to the programs (as for example between Dr. Speech, Tiger Electronics DRS Inc., Seattle, Wash., USA, and CSL [24]) and more specifically their settings such as sampling rate, method of fundamental period extraction [5, 8, 27–29], perturbation algorithm [25], etc. Especially the F_0 extraction algorithm seems to be of crucial importance in voice perturbation measures. Titze and Liang [27] investigated the performance of three event-detection F_0 extraction methods, cycle-to-cycle waveform-matching, zero-crossing, and peak-picking. They stated that peak-picking yields higher perturbation values than zero-crossing and that waveform-matching provides the lowest perturbation values. Furthermore, they concluded that waveform-matching performs best in signals with a frequency variation below 6% per cycle. Possible reasons why this is so are profoundly explored and discussed by Roark [29]. Differences in any of the program-related items can lead to *interprogram differences* in perturbation measurements. Such interprogram differences in perturbation outcomes have been investigated by Bielamowicz et al. [22], Karnell et al. [23] and Smits et al. [24]. Comparison of the F_0 measures among these three studies revealed near-perfect correlations and nonsignificant differences, illustrating very strong agreement for mean F_0. For frequency perturbation and amplitude perturbation on the other hand, there was a very poor to moderately high agreement with statistically significant differences between several computer programs. These data were more recently confirmed in the study by Deliyski and Shaw [30], who found moderate to very strong correlations between frequency and amplitude perturbation measures of three different programs. In general, these differences were attributed to the use of different F_0 extraction methods in the perturbation measurements of the various systems. These studies confirm the earlier review by Read et al. [4], who concluded that the systems generally perform quite well but differ greatly in how their operations are performed.

This study was undertaken to: (a) investigate the intersystem differences between two commonly used systems for computerized perturbation measurements (CSL with MDVP and a common desktop PC system with Praat) with dissimilar microphone type, microphone placement, external hardware, computer, and installed software; (b) examine the interprogram differences between two frequently utilized acoustic analysis programs (MDVP and Praat) for voice samples recorded with CSL.

These issues are especially interesting when clinicians, for instance, aim to relate data obtained by different systems and/or programs or when clinicians want to compare data with normative statistics. To the knowledge of the authors, a comparative study between data collected in dysphonic patients by means of these two systems or programs has not been done yet despite the fact that both are widely known and used in the clinical and scientific realm of voice disorders.

Methods

Subjects

Fifty patients participated in this study. The participants were recruited on an informed consent basis from the ENT case load of the Sint-Jan General Hospital in Bruges in the course of a 1-year period. They all presented with various voice disorders and had been referred for multidimensional voice assessment by staff otolaryngologists. There were 23 males with a mean age of 51 years and an age range from 13 to 74 years, and 27 females with a mean age of 36 years, ranging from 14 to 71. All laryngological diagnoses were made with a flexible transnasal chip-on-tip laryngoscope. Table 1 summarizes laryngoscopic findings. The scores on the Voice Handicap Index [31], as a quantification of the amount of disability caused by a voice disorder, had an average of 51 and ranged from 19 to 106. The scores on the Dysphonia Severity Index [32], as an objective and multiparametric estimate of (disordered) voice quality, ranged from –15.55 to 4.58 with a mean of –1.54. This group of subjects can be considered to be clinically representative of the population of voice-disordered patients, reflecting different age groups, different degrees of dysphonia and voice complaints, and nonorganic as well as organic laryngeal pathologies.

Table 1. List of laryngeal pathologies with their relative occurrence in the group of this study

	Patients	
	n	%
Nonorganic	20	40
Nodules	7	14
Polyp	5	10
Cyst	2	4
Polypoid mucosa (edema in Reinke's space)	4	8
Granuloma	1	2
Leukoplakia	2	4
Unilateral vocal fold paralysis	9	18
Total	50	100

Recordings

From every subject, a voice sample was simultaneously recorded using the two systems. Recording settings are summarized in table 2. The subjects were asked to produce sustained phonation of the vowel /a/ at a comfortable pitch and loudness. The simultaneous recording of the sustained vowel resulted in identical 3-second samples of an oscillographically steady portion of the vowel (excluding voice onset and offset). Concerning the oscillographic steadiness of the samples, decisions were made based on the presence or absence of gross signs of instability (e.g. unvoiced segments, voice breaks, etc.) while looking at the real-time waveform in MDVP (with screen width equal to 3 s). When the first trial was not sufficiently long or oscillographically too unsteady for further research, more trials were undertaken until an acceptable recording was obtained. After recording, all samples were saved in wave format on the hard disks of both computer systems. Acoustic analyses were done on these pairs of files. Recordings from the CSL system (with MDVP) and the PC system (with Praat) were utilized for investigating intersystem differences. For interprogram differences, recordings from the CSL system were analyzed in both MDVP and Praat. The ambient noise level in the labora-

Table 2. Recording and acquisition settings of the two computer systems used in this study

	System 1, CSL	System 2, PC
Microphone		
Type	AKG C420 head-mount condenser microphone with balanced output [33]	Shure Prologue 14H desktop dynamic microphone [34]
Mouth-to-microphone angle	± 45° (left)	± 45° (right)
Mouth-to-microphone distance	± 5 cm	± 15 cm
Computer		
Type	Fujitsu Siemens Scenic P300 desktop computer	Fujitsu Siemens Scenic T desktop computer with a built-in soundcard
External hardware		
	CSL 4500 [6]	–
Program		
Name	MDVP [7]	Praat [8, 9]
Model/version	Model 5105, Version 2.6.2	Version 4.4.01
Sample rate, Hz	44,100	44,100
F_0 extraction method	signum-encoded autocorrelation followed by pitch-synchronous peak detection with linear interpolation [35]	autocorrelation with sinc interpolation followed by waveform matching [36, 39]

Table 3. Descriptive statistics for the data in the two systems

Perturbation measure: name (system)	Unit	n	Mean ± SE	SD	Min	Max	Range
Absolute jitter (1)	μs	46	119.79 ± 9.57	64.92	19.97	295.60	275.63
Absolute jitter (2)	μs	45	51.52 ± 4.58	30.39	12.31	149.89	137.58
Percent jitter (1)	%	47	1.93 ± 0.16	1.13	0.38	5.08	4.70
Percent jitter (2)	%	44	0.79 ± 0.07	0.46	0.17	1.93	1.76
Relative average perturbation (1)	%	47	1.16 ± 0.10	0.70	0.22	3.02	2.80
Relative average perturbation (2)	%	45	0.45 ± 0.04	0.28	0.07	1.12	1.05
Pitch perturbation quotient (1)	%	47	1.13 ± 0.10	0.65	0.23	3.11	2.88
Pitch perturbation quotient (2)	%	46	0.48 ± 0.04	0.29	0.11	1.25	1.14
Shimmer in dB (1)	dB	45	0.38 ± 0.03	0.19	0.01	0.87	0.86
Shimmer in dB (2)	dB	47	0.33 ± 0.02	0.11	0.12	0.64	0.52
Percent shimmer (1)	%	45	4.50 ± 0.35	2.35	0.81	11.31	10.50
Percent shimmer (2)	%	47	3.69 ± 0.17	1.16	1.41	6.82	5.41
Amplitude perturbation quotient (1)	%	46	3.44 ± 0.25	1.72	1.09	8.43	7.34
Amplitude perturbation quotient (2)	%	46	2.60 ± 0.11	0.74	1.11	4.37	3.26

tory room, measured with a Larson & Davis 800B precision integrating sonometer (Larson & Davis Laboratories, Inc., Provo, Utah, USA), was 36 dB_A. The voice recordings had an intensity range from 70.08 to 85.14 dB_{SPL}, resulting in signal-to-noise ratios (SNR) ranging from 34.08 to 49.14 dB. The recommended SNR level was 42 dB [37]. For the large majority of the samples SNR was above 42 dB, while for the rest, SNR levels above 30 dB are still acceptable [26, 37], especially when measuring disordered voices characterized with higher perturbation values [37].

Acoustic Measures

The following seven perturbation measures were obtained in MDVP as well as in Praat. There were four frequency perturbation measures: absolute jitter, percent jitter, relative average per-

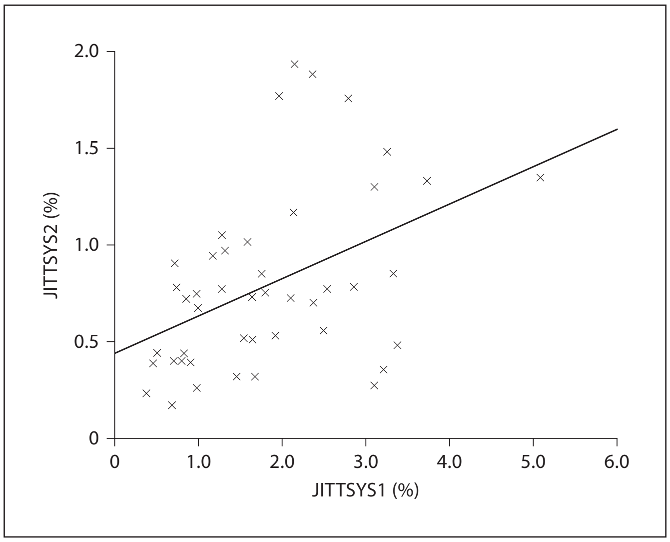


Fig. 1. Scatterplot with linear regression line to illustrate the moderate correlation of percent jitter values in the intersystem analysis between measurements in the CSL system with MDVP and the PC system with Praat (r = 0.44).

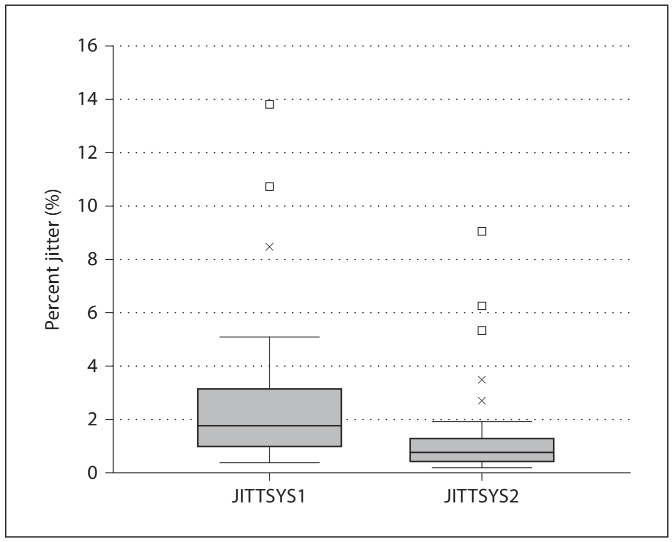


Fig. 2. Box-and-whisker diagram to illustrate the statistically significant intersystem difference in percent jitter values between measurements in the CSL system with MDVP and the PC system with Praat (× = outliers, □ = extremes).

Table 4. Pearson correlation and statistical difference values for variability between two commonly used computer systems for voice perturbation measurement

	Intersystem correlation, r	Intersystem difference, t
Absolute jitter	0.360*	7.463***
Percent jitter	0.442**	7.325***
Relative average perturbation	0.470**	7.716***
Pitch perturbation quotient	0.481**	7.653***
Shimmer, dB	0.455**	2.569*
Percent shimmer	0.332*	2.455*
Amplitude perturbation quotient	0.325*	3.469***

r = Pearson product-moment correlation coefficient; t = value of the t test for dependent samples.
* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

turbation and pitch perturbation quotient. The measures with similar order of perturbation function in Praat are: jitter local absolute, jitter local, jitter rap and jitter ppq5, respectively. There were three amplitude perturbation measures: shimmer in decibels, percent shimmer and amplitude perturbation quotient. Similar measures in Praat are: shimmer decibels, shimmer local and shimmer apq11, respectively. Profound elaboration regarding the F_0 extraction algorithms and the perturbation extraction algorithms of MDVP and Praat is provided in Deliyski [35] and Boersma [36], respectively.

Statistical Analyses

All statistical analyses were done using SPSS for Windows version 12.0 (SPSS Inc., Chicago, Ill., USA). First, all data were explored for the presence of outlying and extreme data. Outliers are defined as data with values between 1.5 and 3 times the interquartile range. Extremes are defined as data with values more than 3 times the interquartile range. Because outliers and extremes can dramatically influence and thus grossly distort the absolute value of r [38], they were omitted from the data set, excluding between 3 and 7 data points per measure of the two systems and programs. Second, for the comparison of the two systems as well as the two programs, two kinds of statistics were employed. Pearson product-moment correlation coefficients (r) were calculated in order to determine the degree of correspondence among the 7 perturbation measures produced by both systems or programs, respectively. Furthermore, as an important proportional relationship between two measures does not necessarily imply equality of the actual values produced by these programs or systems, differences were evaluated by means of the t test for two dependent samples.

Results

Comparison of the Systems

Descriptive statistics for the perturbation measurements derived from the two systems are shown in table 3. In table 4, the Pearson bivariate correlation scores for the different pairs of simultaneously recorded vowel samples are summarized. For all frequency and amplitude pertur-

Table 5. Descriptive statistics for the data in the two programs

Perturbation measure: name (program)	Unit	n	Mean ± SE	SD	Min	Max	Range
Absolute jitter (1)	μs	46	119.79 ± 9.57	64.92	19.97	295.60	275.63
Absolute jitter (2)	μs	45	43.48 ± 5.25	34.45	10.38	213.61	203.23
Percent jitter (1)	%	47	1.93 ± 0.16	1.13	0.38	5.08	4.70
Percent jitter (2)	%	44	0.62 ± 0.05	0.33	0.17	1.91	1.74
Relative average perturbation (1)	%	47	1.16 ± 0.10	0.70	0.22	3.02	2.80
Relative average perturbation (2)	%	45	0.33 ± 0.03	0.16	0.07	0.69	0.62
Pitch perturbation quotient (1)	%	47	1.13 ± 0.10	0.65	0.23	3.11	2.88
Pitch perturbation quotient (2)	%	46	0.37 ± 0.03	0.21	0.11	1.23	1.12
Shimmer, dB (1)	dB	45	0.38 ± 0.03	0.19	0.01	0.87	0.86
Shimmer, dB (2)	dB	47	0.31 ± 0.03	0.21	0.07	0.95	0.88
Percent shimmer (1)	%	45	4.50 ± 0.35	2.35	0.81	11.31	10.50
Percent shimmer (2)	%	47	3.69 ± 0.37	2.43	0.85	10.79	9.94
Amplitude perturbation quotient (1)	%	46	3.44 ± 0.25	1.72	1.09	8.43	7.34
Amplitude perturbation quotient (2)	%	46	2.81 ± 0.27	1.76	0.78	6.84	6.06

Table 6. Pearson correlation and statistical difference values for variability between two frequently utilized acoustic analysis programs for voice perturbation measurement

	Interprogram correlation, r	Interprogram difference, t
Absolute jitter	0.470**	8.669***
Percent jitter	0.366*	8.644***
Relative average perturbation	0.412**	9.059***
Pitch perturbation quotient	0.370*	8.527***
Shimmer, dB	0.542**	2.371*
Percent shimmer	0.780**	3.338**
Amplitude perturbation quotient	0.870**	4.577***

r = Pearson product-moment correlation coefficients; t = value of the t test for 2 dependent samples.
* $p \le 0.05$, ** $p \le 0.01$, *** $p \le 0.001$.

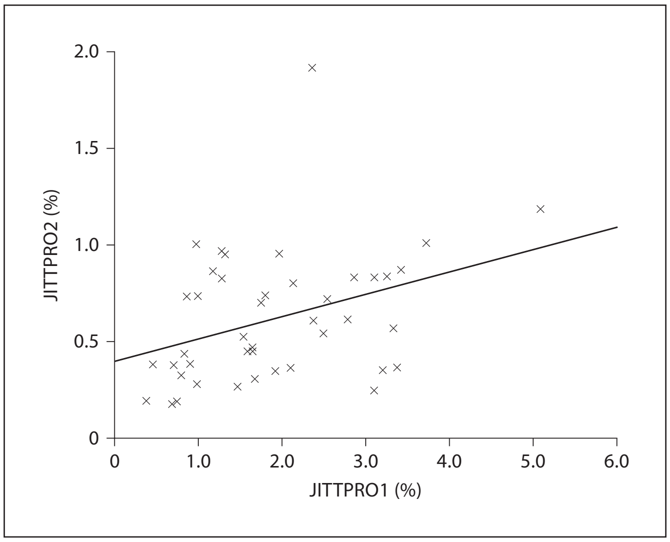
bation measures, the correlation values showed a weak to moderate relationship. As an example of the results in table 4, the regression line in the scatterplot in figure 1 illustrates the moderate correlation between the values of percent jitter obtained with the two systems.

Based on the results of the t test for two dependent samples (also in table 3), there is a statistically significant difference between the two systems for all pairs of perturbation measures. Perturbation values of the CSL system were consistently higher than those of the PC system, especially for the frequency perturbations. For percent jitter, such a difference is illustrated in the box-and-whisker plot (displaying the upper quartile, lower quartile, and interquartile ranges of a data set) in figure 2.

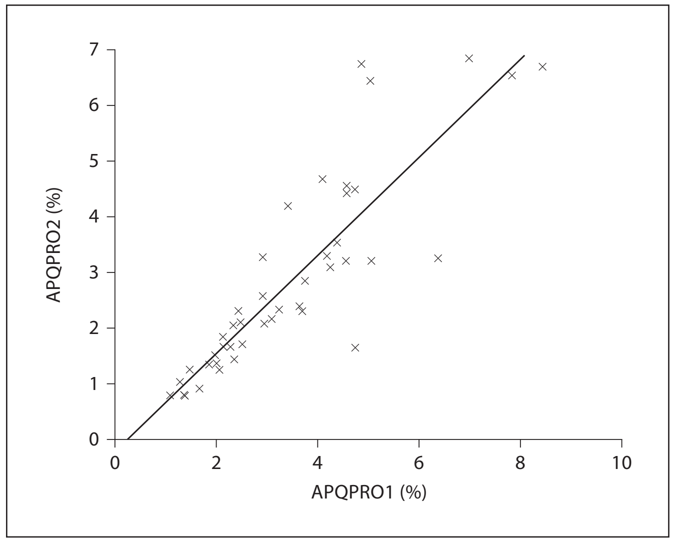
Comparison of the Programs

Table 5 represents the descriptive statistics for the perturbation measurements derived from the two programs. Table 6 summarizes the Pearson correlation coefficients for the vowel samples recorded with the CSL system and analyzed with MDVP and Praat. For all frequency perturbation measures, the results indicate a weak (for percent jitter and pitch perturbation quotient) to moderate (for absolute jitter and relative average perturbation) proportional relationship between MDVP and Praat. As an example, figure 3 illustrates the weak correlation between the values of percent jitter obtained with the two programs. Regarding the amplitude perturbations, a moderate correlation was found for shimmer in decibels and there was a strong correlation for percent shimmer and amplitude perturbation quotient (as demonstrated in the scatterplot with regression line of fig. 4).

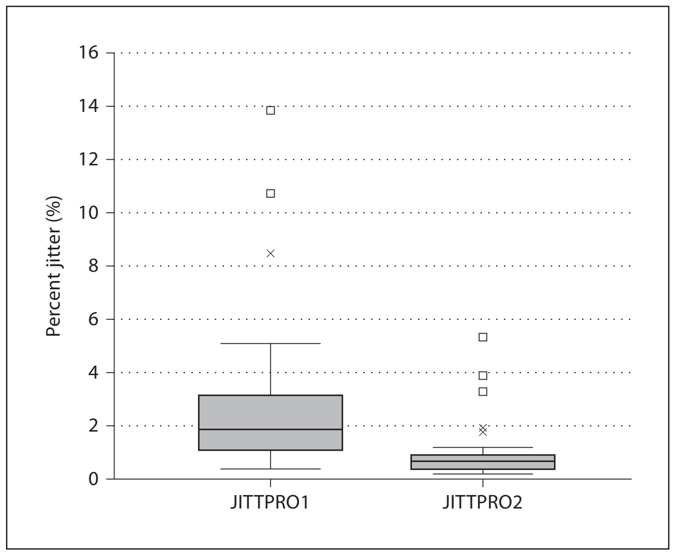
For all pairs of perturbation measures, a statistically significant difference between the two programs was found. Looking at the box-and-whisker diagrams of figure 5, where the results for percent jitter serve as an example for all the other frequency perturbation measures,



3



4



5

Fig. 3. Scatterplot with linear regression line to illustrate the weak correlation of percent jitter values in the interprogram analysis between measurements in MDVP and Praat, both acquired with the CSL system (r = 0.37).

Fig. 4. Scatterplot with linear regression line to illustrate the strong correlation of amplitude perturbation quotient values in the interprogram analysis between measurements in MDVP and Praat, both acquired with the CSL system (r = 0.87).

Fig. 5. Box-and-whisker diagram to illustrate the statistically significant interprogram difference in percent jitter values between measurements in the CSL system with MDVP and the PC system with Praat (× = outliers, □ = extremes).

there is almost no overlap of the interquartile ranges between the two programs. The MDVP measures are consistently higher than the Praat measures. For the amplitude perturbations there is more overlap, and thus there is less difference between similar measures, as evidenced by the lower t values in table 6.

Discussion

This study reports on the differences and similarities of perturbation measures obtained by two computer-based acoustic analysis programs (MDVP and Praat) and systems (CSL with MDVP and a personal computer with Praat), when examining a corpus of 3-second segments of sustained vowel /a/ obtained from 50 patients with various voice disorders.

Before discussing the results of this investigation, attention is to be drawn to the data that were excluded from the data set. In this study, statistical exploration was chosen to be the basis upon which data (outliers and extremes) were excluded. Another method for excluding perturbation data (expressed in percentage) from further analyses is the implementation of the threshold of 5%, since perturbation measures less than about 5% have been found to be reliable [27, 28]. Practised on the frequency perturbation data from the CSL system, both methods exclude almost the same data. For percent jitter, three values (8.494, 10.738, and 13.835%) were omitted based on statistical exploration. Only one value higher (5.075%) than 5% remained in the data set. However, this is a very laminar value. For relative average perturbation, statistical exploration also excluded three values (4.748, 6.366, and 7.849%) and there were no other values above 5%. For pitch perturbation quotient, also three values (5.582, 7.170, and 8.639%) were excluded on the basis of statistical explorations and again there were no other values above 5%. The three values that were excluded across all frequency perturbation measures originate from the same three voice recordings: recording 38 (unilateral vocal fold paralysis), recording 40 (hyperfunctional dysphonia with ventricular hyperadduction) and recording 44 (unilateral vocal fold paralysis). Visual investigation of the narrowband spectrograms revealed type 3 signals in all three recordings (with near-absence of harmonics), according to the classification of Titze [38]. There was 98, 98 and 100% agreement in exclusion of data between these two methods for percent jitter, relative average perturbation and pitch perturbation quotient, respectively. The threshold of 5% cannot be utilized for absolute jitter and shimmer in decibels, since both are not expressed as a percentage.

Several studies have already investigated the *interprogram* differences in acoustic vocal perturbation measurements [22–24, 26, 30]. Although Bielamowicz et al. [22], Karnell et al. [23] and Smits et al. [24] found a very strong interprogram agreement in the F_0 measurements, the analysis of voice perturbation measures yielded much less significant correlations. Furthermore, the correlations between the programs were higher for amplitude perturbation measures than for frequency perturbation measures [24, 30]. Bielamowicz et al. [22] explained this difference in frequency and amplitude perturbation by the fact that jitter is far more dependent on the exact placement of cycle boundaries than shimmer. Whereas minimal errors in placing these boundaries (e.g. due to F_0 tracking dissimilarities) markedly adds noise to frequency perturbations measurements, the effect of such errors is less detrimental to amplitude perturbations because they generally lack sufficient magnitude to eliminate an entire peak from a cycle. Smits et al. [24] compared the measurement of absolute jitter, relative (percent) jitter and relative (percent) shimmer between CSL and Dr. Speech software. They found Pearson correlation coefficients ranging from 0.26 (absolute jitter) and 0.31 (relative jitter) to 0.69 (relative shimmer). Deliyski and Shaw [30] compared frequency and amplitude perturbation between MDVP, TF32 (formerly known as CSpeech, by Paul Milenkovic, Madison, Wisc., USA) and Praat. For the frequency perturbation, they found moderate to very strong correlations (0.40, 0.44 and 0.90) and for the amplitude perturbation there were strong to very strong correlations (0.75 and 0.98). Their interprogram comparison between MDVP and Praat yielded correlations of 0.44 and 0.98 for relative average perturbation and percent shimmer, respectively. We found similar correlations (0.41 and 0.78) in our interprogram comparison of the same measures. In general, these results in the literature corroborate with the findings of the interprogram comparison in the present study: weak to moderate correlations for frequency perturbation measures and moderate to strong correlations for the amplitude perturbation measures (table 4). It should be noted that the different programs utilized different F_0 tracking methods. A profound tutorial on F_0 extraction methods and the effects of discrepancies in F_0 extraction is given by Roark [29].

Next to comparing two programs for perturbation measurement, this study also investigated the differences and similarities between two commonly used data acquisition systems: CSL with MDVP and a personal computer with Praat. The *intersystem* comparison for frequency perturbation measures yielded weak to moderate correlations and was therefore similar to the interprogram comparison. For the amplitude perturbation measures, on the other hand, the moderate to strong correlations from the interprogram comparison dropped to weak to moderate correlations in the intersystem comparison. This suggests that the amplitude perturbation measures are more susceptible for differences in the data acquisition and harmonize with the results of Deliyski et al. [19], who found a statistically significant impact of data acquisition environment and microphone on amplitude perturbation but not on frequency perturbation.

The present study also revealed differences between the perturbation measures stemming from both analysis programs/systems. While all differences were statistically significant for all perturbation measures (with MDVP values being consistently higher than Praat values), the interquartile ranges in the box-and-whisker plots are clearly less overlapping for the frequency perturbations than for the amplitude perturbations. In the case of the comparison between the two programs (MDVP and Praat), the recording hardware and acquisition were identical. Furthermore, the perturbation measures across the two programs were rather similar regarding the order of the perturbation function. Statistical differences between the actual values, on the other hand, can be explained by the dissimilarities between the two systems/programs: the pitch extraction algorithm was different. Praat utilizes an autocorrelation method with sinc interpolation followed by a cycle-to-cycle waveform-matching period detection [36, 39], while MDVP uses a combination of a signum-encoded autocorrelation method followed by pitch-synchronous peak detection with linear interpolation [35]. This important difference causes Praat measuring smaller perturbation values than MDVP.

As for the intersystem comparison, very similar results arose and analogous explanations can be given. The strong correlations could be attributed to similarities in computer apparatus, noise conditions and computation algorithms used in both systems. Both systems differed in the presence/absence of external preamplifying hardware, microphone type, and mouth-to-microphone angle and distance. Although earlier research states that perturbation measures depend on microphone and hardware characteristics [17–19], these dissimilarities did not have a drastic impact on the correlation coefficients in the present study, according to interprogram correlation results. Statistical differences are slightly smaller in the intersystem than in the interprogram variability study.

An additional comment is warranted regarding the number of subjects included in this study. In order to be representative for the population of voice-disordered patients, the inclusion of more than 50 subjects can empower the results of this study. However, Karnell et al. [23] and Deliyski and Shaw [30] included only 20 pathologic and normal subjects, respectively, and Smits et al. [24] included 120 normophonic subjects. Bielamowicz et al. [22] included a selection of 50 pathologic subjects, a number similar to the number of subjects in this study.

Conclusion

Based on the available literature and on the proportional relationships and differences between the two systems and programs under consideration in this study, one can state that one can hardly compare frequency perturbation outcomes across systems and programs and amplitude perturbation outcomes across systems. It is therefore important to have system-specific or program-specific normative data. The normative data for MDVP are given in the MDVP manual [7]. For Praat, however, there are no such data available, inducing a direction for future research.

References

1 Titze IR: Principles of Voice Production. Englewood Cliffs, Prentice-Hall, 1994.
2 Baken RJ, Orlikoff RF: Clinical Measurement of Speech and Voice. San Diego, Singular Publishing Group, 2000.
3 Lieberman P: Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J Acoust Soc Am 1963; 35:344–353.
4 Read C, Buder EH, Kent RD: Speech analysis systems: an evaluation. J Speech Hear Res 1992;35:314–332.
5 Howard DM: The real and non-real in speech measurements. Proc 2nd Int Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications MAVEBA, Florence, 2001.
6 Kay Elemetrics Corp: Multi-Speech and CSL Software: Software Instruction Manual. Lincoln Park, Kay Elemetrics, 2004.
7 Kay Elemetrics Corp: Multi-Dimensional Voice Program (MDVP) Model 5105: Software Instruction Manual. Lincoln Park, Kay Elemetrics, 2003.
8 Boersma P: Praat, a system for doing phonetics by computer. Glot Int 2001;5:341–345.
9 Boersma P, Weenink D: Praat: doing phonetics by computer (version 4.3.14); computer program. Amsterdam, Institute of Phonetic Sciences. http://www.praat.org (accessed May 5, 2005).
10 Eskenazi L, Childers DG, Hicks DM: Acoustic correlates of vocal quality. J Speech Hear Res 1990;33:298–306.
11 Dejonckere PH, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier-Buchman L, Millet B: Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. Rev Laryngol Otol Rhinol 1996;117:219–224.
12 Hirano M: Clinical Examination of Voice. Vienna, Springer, 1981.
13 Wolfe V, Martin D: Acoustic correlates of dysphonia: type and severity. J Commun Disord 1997;30:403–416.
14 Bhuta T, Patrick L, Garnett JD: Perceptual evaluation of voice quality and its correlation with acoustic measurements. J Voice 2004;18:299–304.
15 De Bodt M: A Framework of Voice Assessment: The Relation between Subjective and Objective Parameters in the Judgement of Normal and Pathological Voice; doct diss University of Antwerp, Antwerp, 1997.
16 Kreiman J, Gerratt B: Measuring vocal quality; in Kent RD, Ball MJ (eds): Voice Quality Measurement. San Diego, Singular Publishing Group, 2000, pp 73–101.
17 Titze IR, Winholtz WS: Effect of microphone type and placement on voice perturbation measurements. J Speech Hear Res 1993;36: 1177–1190.
18 Winholtz WS, Titze IR: Miniature head-mounted microphone for voice perturbation analysis. J Speech Lang Hear Res 1997;40: 894–899.
19 Deliyski DD, Evans MK, Shaw HS: Influence of data acquisition environment on accuracy of acoustic voice quality measurements. J Voice 2005;19:176–186.
20 Deliyski DD, Shaw HS, Evans MK: Influence of sampling rate on accuracy and reliability of acoustic voice analysis. Logop Phoniatr Vocol 2005;30:55–62.

21 Winholtz WS, Titze IR: Suitability of minidisc (MD) recordings for voice perturbation analysis. J Voice 1998;12:138–142.
22 Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS: Comparison of voice analysis systems for perturbation measurement. J Speech Hear Res 1993;39:126–134.
23 Karnell MP, Hall KD, Landahl KL: Comparison of fundamental frequency and perturbation measurements among three analysis systems. J Voice 1995;9:383–393.
24 Smits I, Ceuppens P, De Bodt MS: Comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab. J Voice 2005;19:187–196.
25 Rabinov CR, Kreiman J, Gerratt BR, Bielamowicz S: Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. J Speech Hear Res 1995;38:26–32.
26 Deliyski DD, Shaw HS, Evans MK, Vesselinov R: Regression tree approach to studying factors influencing acoustic voice analysis. Folia Phoniatr Logop 2006;58:274–288.
27 Titze IR, Liang H: Comparison of F0 extraction methods for high-precision voice perturbation measurements. J Speech Hear Res 1993;36:1120–1133.
28 Awan SN, Scarpino SE: Measures of vocal F0 from continuous speech samples: an interprogram comparison. J Speech Lang Pathol Audiol 2004;28:122–131.
29 Roark RM: Frequency and voice: perspectives in the time domain. J Voice 2006;20: 325–354.
30 Deliyski D, Shaw E: Acoustic measurement of jitter and shimmer: inter- and intrasystem relationships. Proc ASHA Convention, Miami Beach, November 2006.
31 Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS: The Voice Handicap Index (VHI): development and validation. Am J Speech Lang Pathol 1997;6:66–70.
32 Wuyts FL, De Bodt MS, Molenberghs G, Remacle M, Heylen L, Millet B, Van Lierde K, Raes J, Van de Heyning PH: The Dysphonia Severity Index: an objective measure of vocal quality based on a multiparameter approach. J Speech Lang Hear Res 2000;43: 796–809.
33 AKG Acoustics: C420: User Instruction. MicroMic series II. München, AKG Acoustics Harman Pro, 2000.
34 Shure Inc: Model Prologue 14H User Guide. Evanston, Shure, 2003.
35 Deliyski DD: Acoustic model and evaluation of pathological voice production. Proc Eurospeech, Berlin, September 1993.
36 Boersma P: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proc Inst Phonet Sci, Amsterdam, 1993, vol 17, pp 97–110.
37 Deliyski DD, Shaw HS, Evans MK: Adverse effects of environmental noise on acoustic voice quality measurements. J Voice 2005;19: 15–28.
38 Titze IR: Workshop on acoustic voice analysis: summary statement. Iowa City, National Center for Voice and Speech, 1995.
39 Boersma P: Stemmen meten met Praat. Stem-Spraak-Taalpathol 2004;12:237–251.