Journal of Voice
Vol. 15, No. 4, pp. 529–542

Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements

*†Ping Yu, *Maurice Ouaknine, *Joana Revis, and *Antoine Giovanni

**Laboratoire d'Audio-Phonologie Clinique, Centre Hospitalier Universitaire Timone, Marseille, France; †Kunming Medical University, Yunnan, People's Republic of China*

Summary: The purpose was to determine the clinical value of a multiparametric objective voice evaluation protocol including acoustic and aerodynamic parameters measured mainly on a sustained /ɑ/. This was done by comparison with perceptual analysis of continuous speech by a jury composed of 6 experienced listeners. Voice samples (continuous speech) from 63 male patients with dysphonia and 21 control subjects with normal voices were recorded and assesed by a jury of listeners. The jury was instructed to classify voice samples according to the G (overall dysphonia) component of the GRBAS score on a 4-point scale ranging from 0 for normal to 3 for severe dysphonia. Objective parameters were recorded on an EVA® workstation. As usual with this type of system, parameters were measured mainly on a sustained /a/. Measured parameters included fundamental frequency (F_0), intensity, jitter, shimmer, signal-to-noise ratio, Lyapunov coefficient (LC), oral airflow (OAF), maximum phonatory time (MPT), and vocal range (range). Estimated subglottic pressure (ESGP) was determined on a series of /pa/. Discriminant analysis was performed to detect correlation between jury classification and combinations of parameters. Results showed that a nonlinear combination of only six parameters (range, LC, ESGP, MPT, signal-to-noise ratio, and F_0) allowed 86% concordance with jury classification. Discussion deals with the relative importance of the different objective parameters for discriminant analysis. Special emphasis is placed on two measurements rarely made in routine clinical workup, i.e., estimated subglottic pressure and Lyapunov coefficient. **Key Words:** Multiparametric analysis—Voice—Dysphonia—Lyapunov coefficient—Discriminant analysis.

INTRODUCTION

There is an increasing need for a reliable technique to quantify voice disturbances and detect dysphonia in everyday clinical practice. Current methods of voice analysis based on perceptual scoring and laryngoscopic findings are well defined. However noninvasive physical measurements would provide an adjuvant approach to dysphonia and allow reliable comparison of voice samples (e.g., before and after treatment), therapeutic methods (e.g., microsurgery versus laser), or surgical groups. Despite extensive research in this domain, there is currently no single

Accepted for publication June 1, 2001.
Address correspondence and reprints requests to Antoine Giovanni, Laboratoire d'Audio Phonologie Clinique, Federation ORL, CHU Timone, 13385 Marseille cedex 5, France.
e-mail: agiovann@ap-hm.fr

widely accepted, standardized technique of objective voice evaluation.

Numerous reports using perceptual analysis as the gold standard for comparison have been performed to assess the prognostic value of acoustic measurements for objective voice evaluation. Eskenazi compared six acoustic parameters with perceptual evaluation by experienced jury listeners using the GRBAS method on a scale from 1 to 7.[1] Regression analysis showed that jitter was well correlated with grade G (overall dysphonia) and grade R (roughness) whereas signal-to-noise ratio was correlated with grade B (breathiness). Wolfe et al assessed the predictive value of four acoustic parameters, i.e., mean F_0, jitter, shimmer, and noise-to-harmony ratio in patients with three voice disorders, i.e., nodules, paralysis, and dysfunctional dysphonia.[2] Results showed that shimmer was most concordant with perceptual grading but that the degree of correlation was limited (0.54).

A major obstacle to the use of a single acoustic or aerodynamic parameter for objective assessment of dysphonia is that different disease processes affect various aspects of voice performance to different degrees. Moreover acoustic measurements cover only part of the information contained in perceptual analysis. For this reason several teams have proposed a multiparametric approach to enhance the scope of data.

Piccirillo described a multiparametric method of objective voice based on four parameters selected by regression analysis from a panel of 14 objective measurements.[3] That study was carried out on 33 dysphonic patients and 12 controls with normal voices. The gold standard was the GRBAS with a 4-point scale for each component. The correlation between the multivariate logistic voice function score and each GRBAS component was: -0.58, -0.49, -0.73, -0.66, and -0.65, respectively.

Callan used the concept of self-organizing maps (SOM) to classify subjects into four different groups, i.e., normal, spastic dysphonia (*n*=30), functional dysphonia before treatment (*n*=30), and functional dysphonia after treatment (*n*=30).[4] Discriminant analysis showed that these four groups could be differentiated using 6 of the 22 acoustic parameters included in the Multi-Dimensional Voice Program (MDVP, Kay Elemetrics, Lincoln Park, NJ).

Wuyts recently proposed a dysphonia severity index (DSI) based on the study of 319 patients with various voice disorders and 68 patients with normal voices.[5] Objective voice evaluation was performed using 13 acoustic and aerodynamic parameters. Perceptual analysis was performed by grading the G component of the GRBAS on a 4-point scale. Multivariate regression analysis demonstrated four pertinent parameters, i.e., highest frequency ($F_{0\text{-high}}$ in Hz), lowest intensity (I_{low} in dB), maximum phonation time (MPT in s), and jitter (%). Using these four parameters, Wuyts established a linear regression equation. Results showed a correlation between negative DSI and severity of dysphonia. Positive DSI was correlated with minor dysphonia. The combination of these four parameters allowed classification of subjects in the same group as perceptual analysis using the G component of the GRBAS in 50% of cases.

The Laboratoire d'Audio-Phonologie Clinique at the Timone University Hospital Center in Marseille has been working in the field of objective voice analysis for several years. In conjunction with the Laboratoire Parole et Language d'Université de Provence, we developed the EVA® workstation for objective voice assessment. Since then, a major topic of research has been designing a multiparametric protocol using various acoustic and aerodynamic parameters for objective evaluation of pathological voice impairment. In a previous report we described a protocol comprising acoustic and aerodynamic parameters including jitter, signal-to-noise ratio, oral airflow, and voice onset time.[6] Results obtained using this protocol and perceptual analysis by a listening jury were compared in a population of 239 subjects with normal and pathologic voices. However, concordance[7,8] between objective and perceptual analysis was only 66.7% and, like the technique of Wuyts, this method was deemed insufficient for clinical application.

Recent study in our laboratory has been aimed at validating new parameters to improve the accuracy of our multiparametric protocol. Initial study focused on estimation of subglottic pressure using the EVA® system. Findings showed that estimated subglottic pressure (ESGP) was higher in dysphonic subjects than in controls regardless of intensity and frequency.[9] In addition our group has proposed a new acoustic parameter based on nonlinear dynamics. In the 1990s, Titze et al demonstrated the presence of nonlinear events in patients with dysphonia.[10] These events can only be explained by the interaction between vocal cords, a factor not taken into account by current vocal cord models. To specifically quantify these events, we developed a technique to calculate the Lyapunov coefficient (LC), which measures the characteristics of vibratory instability as represented in a phase portrait. A preliminary study was carried out in a population including 26 patients with unilateral vocal cord paralysis and 12 controls with normal voices. Results confirmed that LC was significantly higher in patients than in controls (0.570 versus 0.386). The clinical value and method of computation of LC have been described elsewhere.[11]

The goal of the present study was to assess the efficacy of a multiparametric protocol including LC and ESGP for objective evaluation of disease-related voice impairment.

METHODS

Study population

Dysphonic voice samples from 63 male patients were selected from our voice data bank. The mean age of the selected patients was 49 years (range, 23 to 75 years). Only male voices were used in this study since previous experience has shown that some acoustic parameters (e.g., jitter) are sensitive to fundamental frequency and that mixed-gender populations can have a confounding effect.[12–14] Laryngoscopic examination with stroboscopic illumination, when needed, was performed to detect lesions on vocal cords. The voices of 21 male volunteers working in or around the laboratory were recorded for use as controls. The mean age of controls was 38 years. All control subjects were considered to have normal voices and had no history of dysphonia or hearing loss.

Perceptual voice analysis

Voice samples were recorded in a soundproof booth on a digital audiotape (DAT) recorder (Tascam D 20). The microphone (Sennheiser) was placed at a distance of 20 cm and slightly to the side of the subject's mouth to minimize breathing noise. Gain was adjusted to avoid saturation and ensure optimal use of recording dynamics. Subjects were instructed to read a standardized text at a comfortable pitch and volume as naturally as possible without overacting. Several trials were made until the examiner judged the quality of recording and speech satisfactory. Voice samples were digitalized at a sampling rate 22.050 Hz.

Recorded voice samples were evaluated by a jury composed of six listeners including two phoniatricians and four speech therapists with experience in clinical voice evaluation as well as in studies using perceptual voice analysis. Listeners were instructed to score the G component (overall dysphonia) of the GRBAS on a 4-point scale:[15] G0: normal voice, G1: slight dysphonia, G2: moderate dysphonia, G3: severe dysphonia. Recordings were presented by a computer in random order. Final scores were assigned by consensus of four or more jury members to avoid interindividual variability. This consensus was achieved in all cases.

Objective evaluation using the EVA® system

Objective voice measurements were made using the EVA® workstation (SQ-Lab, Marseille). As described elsewhere,[6] this system allows simultaneous measurement of acoustic and aerodynamic parameters. Acoustic signals were recorded using the microphone built into the pneumotachograph used to measure oral airflow. Intraoral pressure was calculated by Phonedit software using data from a built-in pressure sensor.

The subject was instructed to pronounce a sustained /a/ three consecutive times for at least 3 seconds at a comfortable pitch and volume. Several trials were made to allow the examiner to select the one he considered as perceptually closest to the subject's natural voice. It is important to stress that all recordings were made by the same experienced examiner.

Parameters were measured during a 2-second window beginning 200 milliseconds after the onset of the signal (see Figure 1). This delay was designed to eliminate instabilities associated with the onset phase. Recorded data were as follows: F_0 (in Hz), intensity (in dB), jitter factor (in %), shimmer (in %), signal-to-noise ratio (in dB), and oral airflow (OAF in cm^3/s).

n–n–

$$\textit{Jitter factor}\ (\%) = \frac{\frac{1}{n-1} = \left[\sum_{i=1}^{n=1} \left| F_i - F_{i+1} \right.\right]}{\frac{1}{n}\sum^{n} F_i} \times 100$$

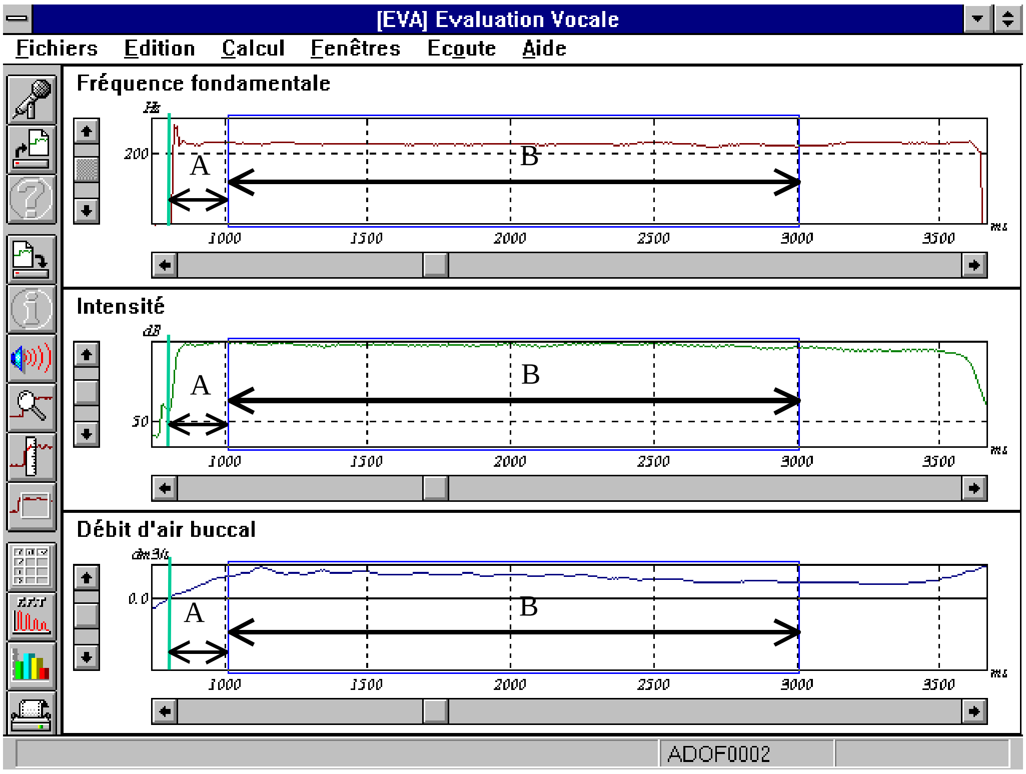

FIGURE 1. Selection of the stable part of the signal using a standardized 2000 ms window (B) starting 200 ms after the beginning of the signal to avoid instabilities during the onset phase (A).

The Lyapunov coefficient (LC) was determined on the same window using a program based on the previously described algorithm[11] and expressed in bits/s. All measurements were performed in triplicate and reported data correspond to the mean of the resulting three values.

Intraoral pressure was estimated using a PVC tube placed in the subject's mouth and connected to the pressure sensor device of the EVA® workstation. For these measurements the subject was instructed to pronounce /pa/ eight consecutive times at approximately 1 second intervals at normal pitch and volumes. Several trials were always made. Subglottic pressure was estimated as previously described[9] by placing a cursor at the top of the pressure peak corresponding to the explosive /p/ sound. These measurements allowed calculation of mean estimated subglottic pressure for each subject.

Vocal range was determined as follows. Each subject was instructed to produce the lowest-pitched and highest-pitched sounds possible. The EVA® workstation was then used to calculate the pitch of each sound and the range was calculated as the difference between the two pitch values expressed in Hertz.

Maximum phonation time was measured as follows. Each subject was instructed to pronounce /a/ for as long as possible at a comfortable pitch and volume. This sound was digitalized and its duration was determined by placing cursors at the beginning and end of the signal graph. The final value was the mean of three consecutive trials.

Statistical analysis

The goal of this study was to validate our multiparametric protocol by comparison with the results of perceptual analysis. For this purpose, a series of statistical analyses was performed using Systat® 7.0 for Windows®. Probability values less than 0.05 were considered significant.

The first stage of statistical analysis consisted of univariate analysis to identify pertinent parameters. Comparative analysis of these parameters was then carried out to detect significant differences between adjacent dysphonia grades (e.g., grade 1 versus grade 2, grade 2 versus grade 3, and grade 3 versus grade 4). Because of the non-Gaussian distribution of some variables and the disproportionate distribution of different grades, the nonparametric Mann-Whitney U test was used.

In the next stage of univariate analysis, a correlation matrix was established to determine the relationship between measured variables as well as their correlation with dysphonia grade. High correlation between variables is an indicator of redundancy and thus is important for selecting parameters for a multiparametric protocol. We used the Pearson square correlation factor (r^2), which expresses common variance between two values.

After completion of univariate analysis, multivariate analysis using the discriminant analysis method was undertaken to classify subjects into different groups using the selected combination of variables. The default *F* value settings of the Systat® software were used, i.e., *F*-to-remove less than 3.9 and *F*-to-enter greater than 4. The results of discriminant analysis were then compared to the results of perceptual analysis to evaluate the accuracy of objective evaluation.

RESULTS

Perceptual analysis

Grading of the dysphonic voice samples by the listener jury is given according to pathology in Table 1. Two of the 21 subjects in the control group were classified G1 and three patients whose voices were normal at the time of recording were classified G0. The underlying pathologies in these three patients was functional dysphonia in two cases and vocal nodules causing intermittent dysphonia in one case.

Univariate analysis

Table 2 presents the mean and standard deviation of each objective parameter as a function of jury grade. The lack of correlation between shimmer values and jury grade was probably due to inherent limitations of our method of instrumental measurement. Based on this observation, shimmer was dropped from the rest of this study. Presentation of data in the form of box plots (Figure 2) provides better visualization of the distribution of variables according to jury grade data. These graphs demonstrate the extent of overlapping values as a function of the variable.

Table 3 shows the *p* values obtained when grades were compared consecutively (i.e., G0 versus G1, G1 versus G2, G2 versus G3). All variables except intensity were significantly different between at least two contiguous grades. For this reason, no clear hier-

TABLE 1. *Jury Grading of Dysphonic Voice Samples as a Function of Pathology*

Diagnosis	G0	G1	G2	G3	Total
Controls	19	2	-	-	21
Nodules	1	1	2	-	4
Polyps	-	12	9	1	22
Reinke's edema	-	1	4	-	5
Cysts	-	1	-	-	1
Sulcus	-	1	1	1	3
Papillomatosis	-	2	4	-	6
Dysplasia	-	1	4	2	7
Dysfunctional dysphonia	2	2	-	-	4
Paralysis	-	2	3	6	11
Total	23	24	27	10	84

TABLE 2. *Means and Standard Deviation (in Parenthesis) as a Function of Severity Grade*

	G0	**G1**	**G2**	**G3**
F_0 (Hz)	119	122	128	175
	(20.5)	(29.4)	(27.7)	(18.7)
Jitter (%)	1,0	2.07	3,44	10.70
	(0.45)	(1.92)	(3.05)	(4.04)
Intensity (dB)	90,4	93,7	90,7	88,13
	(4.57)	(7.27)	(6.4)	(5.86)
Shimmer (%)	0.84	0.77	0.51	0.59
	(0.36)	(0.41)	(0.31)	(0.43)
Signal-to-noise ratio (%)	50,7	42,6	45,4	23,9
	(16,5)	(11,3)	(11,7)	(12,8)
Signal-to-noise ratio >1 kHz (%)	11,2	10,3	8,8	3,8
	(5.3)	(4.9)	(4.8)	(1.4)
OAF (cm^3/s)	157	205	295	385
	(45.5)	(109)	(110)	(202)
ESGP (hPa)	6,1	9,4	12,7	14,4
	(1,3)	(3,1)	(4,2)	(5,5)
Range (Hz)	427	194	119	179
	(168)	(129)	(74.4)	(82)
MPT (s)	21,0	13,5	9,2	4,3
	(6.8)	(5.9)	(5.4)	(2.9)
LC (bits/s)	140	190	459	1626
	(82)	(108)	(472)	(1320)

TABLE 3. P *Values Obtained When Grades Were Compared to Each Other Using the Nonparametric Mann-Whitney U Test*

	G0 /G1	**G1 /G2**	**G2 /G3**
Jitter	0.509 (ns)	0.063 (ns)	0.031
Signal-to-noise ratio	0.028	0.3931(ns)	0.001
Signal-to-noise ratio >1 kHz	0.509 (ns)	0.391 (ns)	0.007
OAF	0.038	0.001	0.374 (ns)
ESGP	0.001	0.005	0.549 (ns)
Range	0.001	0.050	0.054 (ns)
MPT	0.001	0.006	0.001
LC	0.115 (ns)	0.002	0.001
F_0	0.865 (ns)	0.322 (ns)	0.001
Intensity	0.142 (ns)	0.131 (ns)	0.304 (ns)

Abbreviations: ns, not significant.

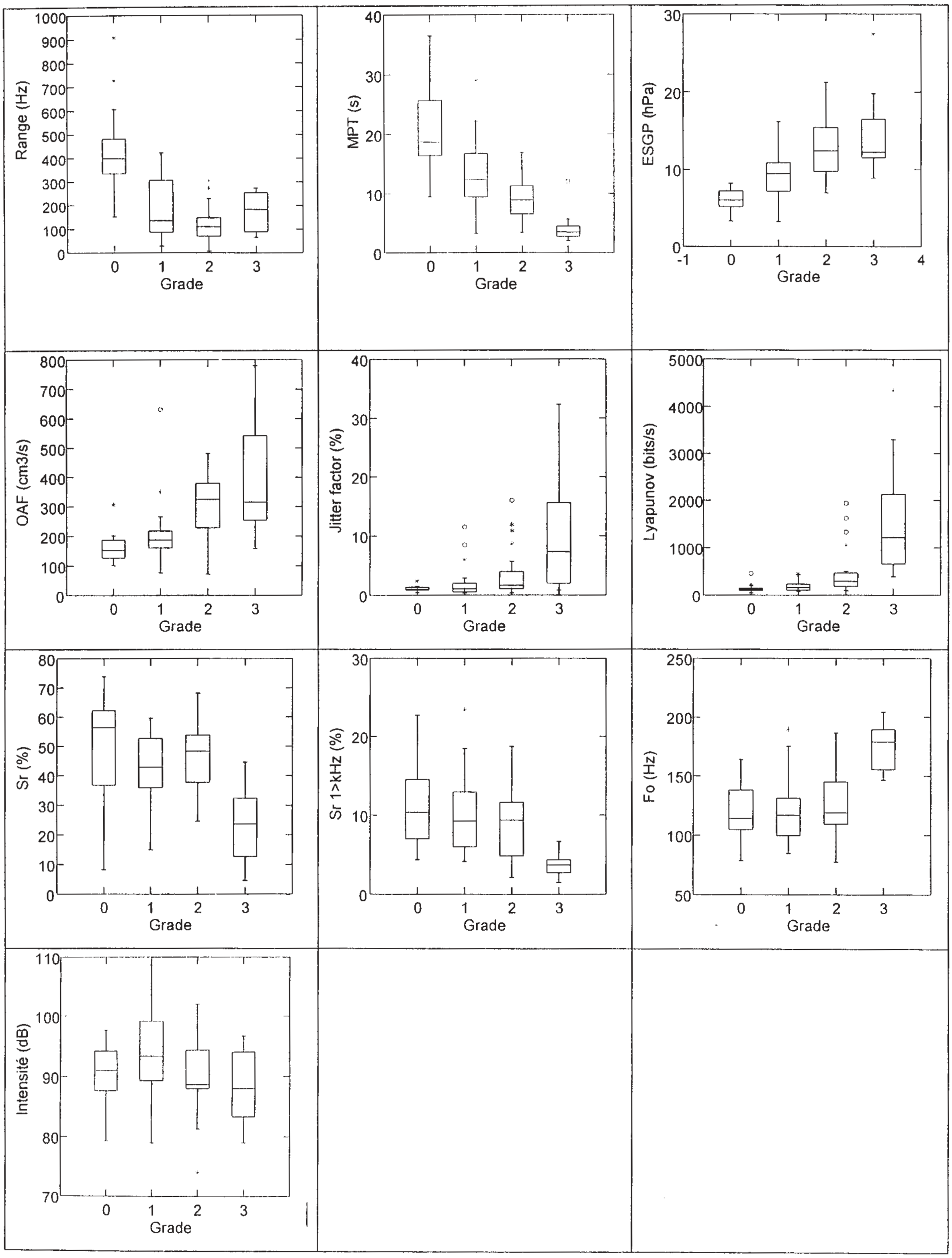

FIGURE 2. Distribution of values for 10 parameters as a function of jury grade.

archy between measurements could be established and all variables were retained for further study to avoid any loss of information. The matrix in Table 4 shows the degree of intercorrelation between objective parameters and jury grade. All variables except intensity were significantly correlated with jury grade.

Discriminant analysis

As shown in Table 3, discriminant analysis was carried out using the nine variables correlated with dysphonia. Results showed six variables with relatively higher *F* values, i.e., range, LC, MPT, ESGP, total signal-to-noise ratio, and F_0. Since covariance between groups was not identical, quadratic analysis was carried out using these six pertinent variables. Severity grading using this combination of variables was the same as perceptual analysis by the jury in 86% of cases (Table 5).

Table 6 shows the *F* values of the variables, i.e., their relative contribution to discrimination as detailed in Table 5. The second most pertinent parameter after range was LC. This finding confirms the importance of LC for objective voice evaluation.

Using Systat® for Windows®, the coefficients of discriminant functions were determined and a discrimination equation was established for each grade (Figure 3). These equations were applied to each patient so as to obtain four values per patient. Patients were classified in the grade with the highest value and the respective probability of classification was calculated for each grade. This approach allows more precise asessment of the degree of accuracy of objective evaluation. Table 7 illustrates results obtained for each subject.

DISCUSSION

A number of studies have been devoted to development of acoustic measurements for objective voice evaluation and validation of their correlation with per-

TABLE 4. *Matrix Showing Pearson Square Correlation Coefficient (r^2) Between Objective Parameters and Jury Grade*

	Grade	F_0	Range	MRT	Jitter	Intensity	OAF	ESGP	Signal-to-noise ratio	Signal-to-noise ratio >1 kHz	LC
F_0	0.224 (0.002)	1.000									
Range	0.354 (0.001)	0.002 (1.000)	1.000								
MPT	0.531 (0.001)	0.097 (0.220)	0.227 (0.001)	1.000							
Jitter	0.220 (0.001)	0.069 (0.873)	0.050 (1.000)	0.142 (0.022)	1.000						
Intensity	0.012 (1.000)	0.109 (0.120)	0.003 (1.000)	0.004 (1.000)	0.089 (0.323)	1.000					
OAF	0.320 (0.001)	0.158 (0.010)	0.052 (1.000)	0.349 (0.001)	0.116 (0.085)	0.000 (1.000)	1.000				
ESGP	0.417 (0.001)	0.135 (0.001)	0.155 (0.011)	0.257 (0.001)	0.078 (0.557)	0.047 (1.000)	0.191 (0.002)	1.000			
Signal-to-noise ratio	0.160 (0.009)	0.272 (0.001)	0.073 (0.697)	0.032 (0.176)	0.249 (0.001)	0.018 (1.000)	0.013 (1.000)	0.153 (0.013)	1.000		
Signal-to-noise ratio >1 kHz	0.153 (0.013)	0.054 (1.000)	0.092 (0.273)	0.097 (0.115)	0.085 (0.394)	0.071 (0.794)	0.061 (1.000)	0.085 (0.402)	0.292 (0.001)	1.000	
LC	0.297 (0.001)	0.138 (0.028)	0.055 (1.000)	0.203 (0.001)	0.578 (0.001)	0.192 (0.002)	0.134 (0.035)	0.076 (0.628)	0.318 (0.001)	0.189 (0.002)	1.000

Note: Probability (*p*) values are indicated in parentheses.

TABLE 5. *Grading of Subjects Using Six Pertinent Objective Variables*

	Group 0	Group 1	Group 2	Group 3	% correct
Grade 0	**22**	1	0	0	96
Grade 1	2	**20**	2	0	83
Grade 2	0	5	**20**	2	74
Grade 3	0	0	0	**10**	100
Total	24	26	22	12	**86**

Note: Bold numbers correspond to the number of cases that were classified in the same grade as by perceptual analysis. The validity of statistical analysis was confirmed by Wilk's lambda test ($p < 0.001$).

TABLE 6. *F Values for Discriminant Analysis of Grade G*

Variables	F
Range	10.44
LC	8.54
MPT	5.39
ESGP	5.36
F_0	3.31
Signal-to-noise ratio	2.07

$$f_n = \frac{1}{2}\begin{bmatrix}1 & V_1 & V_2 & V_3 & V_4 & V_5 & V_6\end{bmatrix}\begin{bmatrix}2A_0 & B_1 & B_2 & B_3 & B_4 & B_5 & B_6 \\ & 2C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ & & 2C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ & & & 2C_{33} & C_{34} & C_{35} & C_{36} \\ & & & & 2C_{44} & C_{45} & C_{46} \\ & & & & & 2C_{55} & C_{56} \\ & & & & & & 2C_{66}\end{bmatrix}\begin{bmatrix}1 \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6\end{bmatrix}$$

FIGURE 3. Matrix showing the coefficients of discriminating functions where $n = 0, 1, 2, 3$; V: variable; A: constant; B and C: coefficients of discriminating function.

TABLE 7. *Classification Obtained by Discriminant Analysis Using a Model with Six Objective Parameters*

Patient	Perceptual Grade	Probability				Objective Grade	Classification
		Grade 0 (%)	Grade 1 (%)	Grade 2 (%)	Grade 3 (%)		
1	0	0.19	0.65	0.16	0.00	1	W
2	0	0.92	0.08	0.00	0.00	0	R
3	0	1.00	0.00	0.00	0.00	0	R
4	0	0.82	0.18	0.00	0.00	0	R
5	0	0.93	0.07	0.00	0.00	0	R
6	0	0.94	0.06	0.00	0.00	0	R
7	0	0.88	0.12	0.00	0.00	0	R
8	0	0.70	0.30	0.00	0.00	0	R
9	0	0.95	0.05	0.00	0.00	0	R
10	0	0.97	0.03	0.00	0.00	0	R
11	0	0.98	0.02	0.00	0.00	0	R
12	0	1.00	0.00	0.00	0.00	0	R

(continues)

TABLE 7. *(continued)*

Patient	Perceptual Grade	Probability				Objective Grade	Classification
		Grade 0 (%)	Grade 1 (%)	Grade 2 (%)	Grade 3 (%)		
13	0	0.52	0.48	0.00	0.00	0	R
14	0	0.97	0.03	0.00	0.00	0	R
15	0	1.00	0.00	0.00	0.00	0	R
16	0	0.90	0.10	0.00	0.00	0	R
17	0	1.00	0.00	0.00	0.00	0	R
18	0	0.99	0.01	0.00	0.00	0	R
19	0	0.99	0.01	0.00	0.00	0	R
20	0	1.00	0.00	0.00	0.00	0	R
21	0	0.99	0.01	0.00	0.00	0	R
22	0	1.00	0.00	0.00	0.00	0	R
23	0	1.00	0.00	0.00	0.00	0	R
24	1	0.14	0.80	0.06	0.00	1	R
25	1	0.01	0.82	0.18	0.00	1	R
26	1	0.45	0.49	0.06	0.00	1	R
27	1	0.00	1.00	0.00	0.00	1	R
28	1	0.97	0.03	0.00	0.00	0	W
29	1	0.15	0.51	0.34	0.00	1	R
30	1	0.06	0.94	0.00	0.00	1	R
31	1	0.55	0.45	0.00	0.00	0	W
32	1	0.01	0.61	0.38	0.00	1	R
33	1	0.00	1.00	0.00	0.00	1	R
34	1	0.05	0.83	0.12	0.00	1	R
35	1	0.00	0.86	0.14	0.00	1	R
36	1	0.00	1.00	0.00	0.00	1	R
37	1	0.00	0.36	0.64	0.00	2	W
38	1	0.00	0.16	0.84	0.00	2	W
39	1	0.39	0.61	0.00	0.00	1	R
40	1	0.02	0.93	0.05	0.00	1	R
41	1	0.01	0.57	0.42	0.00	1	R
42	1	0.03	0.66	0.31	0.00	1	R
43	1	0.00	0.90	0.10	0.00	1	R
44	1	0.00	1.00	0.00	0.00	1	R
45	1	0.38	0.62	0.00	0.00	1	R
46	1	0.00	0.79	0.00	0.21	1	R
47	1	0.00	0.69	0.31	0.00	1	R
48	2	0.01	0.37	0.62	0.00	2	R
49	2	0.00	0.83	0.17	0.00	1	W

(continues)

TABLE 7. *(continued)*

Patient	Perceptual Grade	Probability Grade 0 (%)	Probability Grade 1 (%)	Probability Grade 2 (%)	Probability Grade 3 (%)	Objective Grade	Classification
50	2	0.00	0.06	0.94	0.00	2	R
51	2	0.00	0.00	1.00	0.00	2	R
52	2	0.00	0.00	1.00	0.00	2	R
53	2	0.00	0.28	0.72	0.00	2	R
54	2	0.00	0.26	0.74	0.00	2	R
55	2	0.00	0.14	0.86	0.00	2	R
56	2	0.00	0.09	0.91	0.00	2	R
57	2	0.00	0.01	0.99	0.00	2	R
58	2	0.09	0.78	0.13	0.00	1	W
59	2	0.00	0.01	0.94	0.06	2	R
60	2	0.00	0.01	0.99	0.00	2	R
61	2	0.06	0.66	0.28	0.00	1	W
62	2	0.00	0.14	0.86	0.00	2	R
63	2	0.30	0.63	0.07	0.00	1	W
64	2	0.00	0.33	0.67	0.00	2	R
65	2	0.00	0.00	0.07	0.93	3	W
66	2	0.00	0.00	0.23	0.77	3	W
67	2	0.00	0.01	0.99	0.00	2	R
68	2	0.01	0.82	0.17	0.00	1	W
69	2	0.00	0.08	0.92	0.00	2	R
70	2	0.00	0.00	1.00	0.00	2	R
71	2	0.00	0.00	1.00	0.00	2	R
72	2	0.00	0.45	0.55	0.00	2	R
73	2	0.00	0.00	1.00	0.00	2	R
74	2	0.00	0.02	0.98	0.00	2	R
75	3	0.00	0.00	0.14	0.86	3	R
76	3	0.00	0.00	0.02	0.98	3	R
77	3	0.00	0.00	0.00	1.00	3	R
78	3	0.00	0.00	0.00	1.00	3	R
79	3	0.00	0.00	0.09	0.91	3	R
80	3	0.00	0.00	0.00	1.00	3	R
81	3	0.00	0.00	0.00	1.00	3	R
82	3	0.00	0.00	0.00	1.00	3	R
83	3	0.00	0.00	0.01	0.99	3	R
84	3	0.00	0.00	0.00	1.00	3	R

Note: In the classification column, *R* indicates that the patient was correctly classified and *W* indicates that the patient was misclassified.

ceptual analysis and/or pathological changes. The study design used here was established as a function of the conclusions of previous reports in the literature. Since voice is mainly a perceptual phenomenon, perceptual voice analysis was chosen as the gold standard, as in most studies, aimed at validating objective voice evaluation. The GRBAS scale was chosen for perceptual analysis because its efficacy has been validated in numerous previous studies.[6,16–18]

This study confirms the clinical value of a multiparametric objective voice evaluation protocol including acoustic and aerodynamic parameters measured mainly on a sustained /a/. Our findings showed that the acoustic and aerodynamic parameters selected for the model vary in the same direction as perceptual parameters and thus can be expected to provide the same conclusion as jury analysis. We performed statistical analysis to determine if each parameter of the protocol was in agreement with perceptual analysis. In addition, we compared the results of various combinations of parameters with the jury classification.

A frequent objection to objective voice analysis protocol is the use of a token voice sample such as the sustained /a/ rather than continuous speech for perceptual analysis. This issue has been addressed in several previous studies.[7,8,29] Results of a study by Revis et al[8] on the impact of voice material on perceptual analysis showed that judgments made on sustained vowel samples were comparable to judgments made on continuous speech samples. The use of continuous speech for perceptual analysis and sustained vowels for objective analysis is methodologically inevitable insofar as each material is a prerequisite for optimal performance of the analysis technique.

A total of 11 acoustic and aerodynamic parameters were measured for each patient or control. Univariate analysis showed nine of these parameters to be relatively pertinent, i.e., consistent with overall jury grading and significantly different between at least two contiguous grades ($p < 0.05$). The pertinence of these parameters was also confirmed by a correlation matrix showing significant correlation ($p < 0.01$). These nine parameters were submitted to quadratic discriminant analysis to define the best model for multiparametric assessment. Selection of parameters was performed step by step according to the *F* value. In discriminant analysis, the *F* value represents the relative contribution of each variable to discrimination and its relative importance in the model. Two types of discriminant analysis are feasible using a combination of selected variables, i.e., linear discriminant analysis and quadratic discriminant analysis. Unlike linear discriminant analysis, which requires identical covariance between groups, quadratic discriminant analysis has no prerequisite for validity and thus may optimize results.

Discriminant analysis can select different combinations of acoustic and aerodynamic parameters. In this regard, it is important that the model exhibit a good trade-off between statistical requirements and pathological features. A model that is statistically sound may not be pathologically relevant. In this study, six of the original nine parameters were selected by discriminant analysis. They were range, LC, ESGP, MPT, total signal-to-noise ratio, and F_0. This combination allowed 86% concordance with perceptual analysis. Inclusion of LC in our model led to a sharp improvement in the accuracy of objective analysis (Table 6). With a 14% error rate, the accuracy of the protocol described herein is almost reliable enough for clinical use.

In another study with the same goal as this one, Piccirillo performed objective voice evaluation on 97 dysphonic patients and 35 control subjects using a statistical approach similar to the one described here.[19] A total of 14 parameters were measured, i.e., subglottic pressure, oral airflow, intensity (low, middle, and high), frequency (low, middle, and high), range, laryngeal resistance, glottic efficacy, MPT, transglottic airflow, and maximum rate of change of contact area. Four of these parameters were identified as pertinent, namely, subglottic pressure, OAF, MPT, and voice range. Three of these parameters are the same as in our study.

In the literature OAF and jitter have been widely used for evaluation of the pathological voice and have been correlated with the intensity of dysphonia.[19–22] Although these parameters were correlated with perceptual grade in our study, they were not selected by discriminant analysis and their inclusion did not improve the accuracy of our model. Closer analysis of the relationship between parameters (Table 4) provides a possible explanation for this finding. OAF was strongly correlated with MPT, and jitter with LC. This suggests that OAF and jitter may have been redundant in this study.

Another novel aspect of this study is assessment of the probability of classification in each grade. In our clinical experience, we have observed audible differences in the degree of dysphonia between voices classified within the same jury grade. The most likely explanation for this observation is that the 4-point scale of grading does not fully reflect the continuum of disease-related voice deterioration. Since it can be assumed to be related to degree of dysphonia, probability of classification in each grade could provide a means for more accurate voice evaluation. In Table 7, patient 48 classified in objective grade 2 with a probability of 0.62 for grade 2 and 0.37 for grade 1 could probably be assumed to be less severe than patient 50 classified in grade 2 but with a probability 0.94 for grade 2 and 0.06 for grade 1.

In previous studies, we noted that it was often difficult to discriminate between grade 1 and grade 2. In this study all but one of the misclassified cases were grade 1 and grade 2. The remaining case was grade 0. As in most previous studies in the literature,[2,6,7,12,23,24] perceptual analysis was based on a 4-point scale. The advantage of this method is to reduce jury variablility but it also limits discriminatory ability by restricting choice of degree of dysphonia. Use of the 4-point scale in this study could have increased the risk of misclassification between grades 1 and 2. Several authors have reported greater jury variability for grades 1 and 2 probably due to internal standards which are by definition unstable.[17,24–27] However this is not the case for grade 0 (normal voice) and grade 3 (severe dysphonic) since listeners are highly experienced and their judgments are quite similar in these cases.[24–26] Further study is needed to improve interlistener reliability for intermediate forms of dysphonia.

Another possible factor in misclassification is that objective analysis was performed on the stable part of a sustained /a/. Numerous studies have stressed that study of the unsteady parts of the signal, such as the onset, could provide valuable information for objective voice evaluation and allow finer discrimination of the severity of dysphonia.[6,28,29] However a relatively stable signal is a prerequisite for the use of the EVA workstation as is the case for most other available objective voice analysis systems. Irregularities in the onset phase currently prevent objective measurements from being made even using techniques issued from nonlinear mathematics.[30,31]

Because this study was specifically designed to assess the accuracy of the multiparametric approach to objective voice assessment, the scope of its conclusions is limited in several respects. Since no posttherapeutic data were included (e.g., after phonosurgery), no conclusion can be drawn as to the efficacy of the protocol in evaluating therapeutic modalities. Since the patient population was exclusively male, no conclusion can be drawn as to the efficacy of the protocol for female voices. Since the number of patients classified in grade 3 was low, there is a possibility of bias. Further study will be necessary to resolve these unanswered questions and to develop a clinically applicable protocol for evaluation of voice quality, expecially for intermediate-grade dysphonia.

REFERENCES

1. Eskenazi L, Childers DG, Hicks DM. Acoustic correlates of vocal quality. *J Speech Hear Res.* 1990;33:298–306.
2. Wolfe V, Fitch J, Cornell R. Acoustic prediction of severity in commonly occurring voice problems. *J Speech Hear Res.* 1995;38:273–279.
3. Piccirillo J, Colin P, Dennis F. Assessment of two objective voice function indices. *Ann Otol Rhinol Laryngol.* 1998; 107:396–400.
4. Callan DE, Kent RD, Roy N, Tasko SM. Self-organizing map for the classification of normal and discordered female voices. *J Speech Hear Res.* 1999;42:355–366.
5. Wuyts FL, De Bodt MS, Molenberghs G, Remacle M, Heylen L, Millet B, Lierde KV, Raes J, Van de Heyning PH. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approch. *J Speech Hear Res.* 2000;43:796–809.
6. Giovanni A, Robert D, Estublier N, Teston B, Zanaret M, Cannoni M. Objective evaluation of dysphonia: preliminary results of a device allowing simultaneous acosutic and aerodynamic measurements. *Folia Phoniatr Logop*. 1996;48:175–185.
7. De Krom G. Consistency and reliability of voice quality ratings for different types of speech fragments. *J Speech Hear Res.* 1994;37:985–1000.
8. Revis J, Giovanni A, Wuyts FL, Triglia JM. Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop.* 1999;51:108–116.
9. Giovanni A, Heim C, Demolin D, Triglia JM. Estimated subglottic pressure in normal and dysphonic subjects. *Ann Otol Rhinol Laryngol.* 2000;109:500–504.
10. Titze I, Baken R, Hertzel H. Evidence of chaos in vocal fold vibration. In: Titze I, ed. *Frontiers in Basic Science.* San Diego, Calif.: Singular Publishing Group; 1993:143–188.
11. Giovanni A, Ouaknine M, Triglia JM. Determination of largest Lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis. *J Voice*. 1999;13:341–354.

12. Martin D, Fitch J, Wolfe V. Pathologic voice type and the acoustic prediction of severity. *J Speech Hear Res.* 1995; 38:765–771.
13. Kojima H, Gould W, Lambiase A, Isshiki N. Computer analysis of hoarseness. *Acta Otolaryngol.* 1980;89:547–554.
14. Yumoto E, Sasaki Y, Okamura H. Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *J Speech Hear Res*. 1984;27:2–6.
15. Hirano M. Psycho-acoustic Evaluation of Voice: GRBAS Scale for Evaluating the Hoarse Voice. *Clinical Examination of Voice* . Wien, Springer Verlag: 1981.
16. Hirano M, Hibi S, Terasawa R, Fujiu M. Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia. *J Phonetics*. 1986;14:445–456.
17. De Bodt M, Wuyts F, Van de Heyning P, Croux C. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice.* 1997;11:74–80.
18. Dejonckere PH, Obbens C, De Moor GM, Wieneke GH. Perceptual evaluation of dysphonia: Reliability and relevance. *Folia Phoniatr*. 1993;45:76–83.
19. Piccirillo J, Colin P, Dennis F, Frederickson J. Multivariate analysis of objective vocal function. *Ann Otol Rhinol Laryngol.* 1998;107:107–112.
20. Hirano M, Hibi S, Yoshida T, Hirade Y, Kasuya H, Kikuchi Y. Acoustic analysis of pathological voice. Some results of clinical application. *Acta Otolaryngol*. 1988;105:432–438.
21. Dejonckere P. Bruit de turbulence et apériodicité dans la voix pathologique. Une approche multifactorielle. *Rev Laryngol.* 1990;111:353–357.
22. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res.* 1994;37:769–778.
23. Hammarberg B, Tritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol.* 1980;90:441–451.
24. Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception*. J Speech Hear Res.* 1992;35:512–520.
25. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial and a framework for future research. *J Speech Hear Res.* 1993;36:21–40.
26. Anders L, Holline H, Hurme P, Sonninnen A, Wendler J. Perceptual evaluation of hoarseness by several classes of listeners. *Folia Phoniatr.* 1988;40:91–100.
27. Askenfelt AG, Hammarberg B. Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. *J Speech Hear Res.* 1986;29:50–64.
28. Dejonckere PH, Obbens C, De Moor GM, Wieneke GH. Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatr*. 1993;45:76–83.
29. De Krom G. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J Speech Hear Res.* 1995;38:794–811.
30. Giovanni A, Ouaknine M, Guelfucci B, Yu P, Zanaret M, Triglia JM. Non-linear behavior of vocal fold vibration: the role of coupling between the vold fold. *J Voice*. 1999;13:465–476.
31. Revis J, Barberis S, Giovanni A, Triglia JM. Définition d'une mesure temporelle de l'attaque vocale. *Rev Laryngol Otol Rhinol.* In press.