Reliability of the Italian INFVo scale and correlations with objective measures and VHI scores

Affidabilità della versione italiana della scala INFVo e sua correlazione con parametri obiettivi e con i punteggi VHI

A. SCHINDLER[1], D. GINOCCHIO[2], M. ATAC[1], P. MARUZZI[1], S. MADASCHI[1], F. OTTAVIANI[3], F. MOZZANICA[1]

[1] Department of Clinical Sciences "L. Sacco", University of Milan, Italy; [2] Dipartimento di Scienze Cliniche e di Comunità, Università degli Studi di Milano, Audiology Unit, Fondazione IRCCS "Ca' Granda", Ospedale Maggiore Policlinico, Milan, Italy; [3] San Giuseppe Hospital, Ospedale Classificato, IRCCS Multimedica, Milan, Italy

SUMMARY

The objective of this study was to evaluate the reliability of the INFVo scale and its relationship with objective measures and VHI scores in 40 native Italian-speaking patients with substitution voice. The maximum phonation time (MPT), diadochokinesis (DDK) of the three syllabic sequence [pa/ta/ka], reading of a passage and a single word repetition test were recorded. Each patient completed the Italian version of the VHI. Three speech-language pathologists blindly rated the recordings using the auditory perceptual INFVo scale; one listened and assessed the voice recording twice. The INFVo intra- and inter-rater reliability reached good values. Strong to moderate correlations between the INFVo scale scores and MPT, DDK, distortions in the repetition test, speech rate during reading and the functional subscale of the VHI were found. In conclusion, the INFVo scale is a reliable tool and can be recommended for the perceptual assessment of substitution voices in Italian speaking patients.

KEY WORDS: Substitution voice • Perceptual assessment • INFVo scale • MPT • Speech rate • VHI • Total laryngectomy • Partial laryngectomy • Supracricoid laringectomy

RIASSUNTO

L'obiettivo di questo studio è stato di valutare l'affidabilità della scala INFVo e studiare la correlazione tra i suoi risultati e quelli ottenuti dalle misure obiettive e dai punteggi del questionario VHI in 40 pazienti di madre lingua italiana con voce di sostituzione. Ognuno di essi ha completato il questionario VHI, ha eseguito un test di ripetizione di una singola parola e la sua voce è stata registrata in modo da poter calcolare il massimo tempo di fonazione (MPT) e la diadococinesi (DDK) della sequenza trisillabica [pa/ta/ka]. La voce del paziente è stata anche registrata durante la lettura di un brano. Tre logopedisti hanno valutato in cieco le registrazioni utilizzando la scala di valutazione percettiva INFVo. Uno dei logopedisti ha ascoltato e valutato le registrazioni due volte. Sia l'affidabilità intra-rater che quella inter-rater hanno raggiunto valori ottimali. I punteggi ottenuti nella scala INFVo risultavano correlati con quelli ottenuti nel MPT, DDK e nella sottoscala funzionale del VHI. In conclusione la scala INFVo rappresenta uno strumento affidabile e dovrebbe pertanto essere utilizzata nella valutazione delle voci di sostituzione di pazienti italiani.

PAROLE CHIAVE: Voce di sostituzione • Valutazione percettiva • Scala INFVo • MPT • Velocità d'eloquio • VHI • Laringectomia totale • Laringectomia parziale • Laringectomia sopracricoidea

Introduction

Perceptual assessment of voice offers important information concerning the characteristics, limitations and possibilities for change of disturbed voice production[1]; therefore, perceptual assessment plays a pivotal role in functional diagnosis of voice and it is included in the multidimensional protocol of voice pathology assessment together with videostroboscopy, acoustics, aerodynamics and subjective rating by the patient[2]. The application of perceptual assessment is supported by the fact that instrumental measures cannot substitute auditory-perceptual assessment[3-9].

Judgments of the perceived quality of a voice sample are affected by several variables including listener characteristics (experience and training), the phonetic content of the sample and the rating scale[10]. Experienced listeners seem to judge voice quality more consistently than untrained raters[11], although a few hours of training may be sufficient to attain reliable scores in inexperienced listeners[12][13]. As for the voice sample, although a relatively strong relationship was found between the sustained vowels and con-

nected speech, sustained vowel sounds may not adequately reflect the dysphonic severity of continuous speech [14-16]. Finally, the role of visual analogue versus ordinal scale has been investigated; although a visual analogue scale seems to enable a finer judgment of voice quality, and it has been shown that with increased freedom of judgment the inter-rater agreement decreased considerably [17 18].

In order to be valuable, perceptual assessment should follow a standard procedure [1]. Several different frameworks for obtaining perceptual ratings have been described in the literature [19-23]; the most commonly used worldwide is the GRBAS scale, which has established reliability and validity data [4 24-26].

Until now, there has been no accurate method for perceptual assessment of substitution voicing, defined as voicing without two true vocal folds, such as after total laryngectomy, supracricoid laryngectomy and glottectomy. Even if the GRBAS scale has been used in the assessment of substitution voicing, and has presented positive correlation with acoustic, aerodynamic and voice-related quality of life (QOL) questionnaires [27-31], it does not appear to be the best tool for this type of voice. In fact, substitution voice is often scored as severely impaired [27 28 30 32 33]; furthermore, laryngeal and substitution voices are different across a variety of parameters. Moreover, some perceptual features are unique for substitution voice and are not therefore included in the GRBAS scale. Finally, the GRBAS baseline values refer to normal laryngeal voice quality, which is unattainable in substitution voice.

In order to have a more accurate method for substitution voice perceptual assessment, the INFVo scale was developed [34 35]. This scale examines the following characteristics: overall impression (I), unintended additive noise (N), fluency (F) and quality of voicing (Vo). I reflects the overall voice quality as well as the impression of intelligibility. N reflects the amount of annoyance caused by the audibility of uncontrolled noises, such as bubbly and breathy noise produced during speech; F reflects the perceived smoothness of the sound production, Vo reflects whether the voicing is supposed to be voiced or unvoiced [34]. Each of these parameters can be classified on a scale that uses a horizontal bar divided into 11 cells, of which one is scored. For each parameter, the extreme right coincides with a very good score, while the extreme left represents a very poor score. Therefore, the higher the score the better the perceived quality of the voice. The INFVo scale was studied in Dutch-speaking patients and presented good inter-subject reliability in semi-professional raters and excellent inter-reliability in professionals in Netherlands and France [35].

Voice quality is to some extent culturally conditioned and specific to a certain language community [19 36]; therefore, the reliability of the INFVo scale found in Dutch-speaking patients may not apply to Italian speakers. Acoustic analysis of substitution voices is difficult because of the extreme irregularity of the signal, and previous research found only moderate correlation between acoustic measures and perceptual ratings in substitution voices [37]. In contrast, other measures such as the maximum phonation time (MPT) and the speech rate are easier to obtain, and both are expected to have a relationship with the fluency of speech; in fact, the longer the MPT and the higher the speech rate, the longer the sentence uttered without interruption for taking a breath.

Partial, supracricoid and total laryngectomies lead to a severe voice impairment that is an obvious disability [38]; nonetheless, QOL involves many factors, including societal attitudes, environmental barriers, education, age, gender, vocation, cultural and ethnic background [39], and many individuals do not rank speech as the most important attribute that contributes to their QOL. Previous research has analyzed the correlation between perceptual measures and both QOL and voice-related QOL (V-RQOL), and only moderate correlations were found, suggesting that perceptual evaluation and V-RQOL questionnaires evaluates different aspects of voice [40-44].

To the best of our knowledge, the INFVo scale has never been applied to native Italian-speaking patients with substitution voice, and no data exist on the correlation between this scale and other measures, such as those derived from objective measures and voice-related QOL questionnaires. The aim of the study was to: a) analyze intra-subject and inter-subject reliability of the INFVo scale in native Italian-speaking patients with substitution voice; b) examine the relationship between the INFVo scale and selected objective parameters; c) analyze the relationship between the INFVo scale and voice-related QOL questionnaires.

Materials and methods

Participants

Forty patients, 28 males and 12 females, who had undergone surgical procedures for laryngeal cancer from 2000 to 2008, were enrolled in the study. The mean age was 66 ± 9.6 years (range 46-91). None of the patients had any respiratory problems, debilitating illness or recurrence of disease. All patients had completed oncological treatment at least one year before the study was undertaken. Demographic characteristics of the participants and phonation modality are reported in Table I. Twenty-four patients were treated with total laryngectomy (TL), 7 underwent supracricoid laryngectomy (SL) with either cricohyodopexy (n = 2) or cricohyodoepiglottopexy (n = 7) and 9 patients were treated with other partial laryngectomies (3 with glottectomy and 6 with frontolateral laringectomy). Thirteen patients were exclusive oesophageal speakers, 11 were tracheo-oesophageal speakers, 9 were ventricular band speakers and 7 were arytenoid speakers, meaning that the arytenoid mucosa was the vibration source.

Speech sample

The voice signal was recorded with a microphone positioned approximately 10 cm from the patient's mouth at a 45-degree angle from the mouth axis to reduce airflow effects. Voice recordings were directly stored in the host computer using the Computerized Speech Lab (CSL) Model 4500 (Kay Elemetrics, Lincon Park, NJ). All recordings were made in a quiet room (ambient noise < 50 dB (A)). Speech samples included: a sustained /a/ at comfortable pitch and loudness, the reading aloud of a standard short passage, consisting of 5 sentences for a total of 100 syllables, a word repetition test including all phonemes of the Italian language and a diadochokinetic (DDK) test [45]. The MPT was determined by measuring the longest sustained /a/ in three productions. On the basis of the time needed to read the five sentences, the syll/s were calculated. During the repetition test, the rater utters a word and the patient is asked to repeat it. On the basis of auditory perceptual evaluation, the rater considered whether the word was uttered with or without phonetic distortions or substitutions. The test is made of 31 words and lasts about 2 minutes. Finally, the DDK test was performed by asking to each subject to utter the three syllabic sequence [pa/ta/ka] as rapidly as possible for 5 sec.

Perceptual, objective and voice-related QOL measures

The recorded material was used for auditory perceptual assessment performed using the INFVo scale. Prior to the rating session, the three clinicians involved in the study were trained in the use of the INFVo scale with female and male substitution voice samples.

The MPT (rated in s), the DDK (rated in syll/s) and the speech rate (also rated in syll/s) of the 100 syllables passage were calculated. Furthermore, the number of phonetic distortions or substitutions during the word repetition test was counted. Finally, each patient autonomously completed the Italian version of VHI [46 47] to have self-assessment data on the perceived QOL.

Reliability and correlation analysis

For the INFVo reliability analysis, each recording was blindly listened and rated by three licensed speech-language pathologists, referred to as Rater 1, Rater 2 and Rater 3. Recordings were presented in random order separately for each rater. All raters were females, with a normal hearing threshold and extensive experience in voice and substitution voice perceptual assessment; additionally, each rater had over five years of experience in voice and speech rehabilitation after partial supracricoid and total laryngectomy. In order to evaluate the intra-rater reliability of the INFVo, Rater 1 had to listen and to assess

Table I. Characteristics characteristics of study participants.

Type of laryngectomy	Number of patients	Sex (males/females)	Phonation modality (n)
Total laryngectomy	24	17/7	Oesophageal speakers (13) Tracheo-oesophageal speakers (11)
Supracricoid laryngectomy	7	4/3	Arytenoid speakers (7)
Partial laryngectomy	9	7/2	Ventricular band speakers (9)
Frontolateral laryngectomy	*6*	*5/1*	
Glottectomy	*3*	*2/1*	

the recordings twice, with a week of interval between the first and the second assessment. None of the three raters was involved in speech sample recording. The raters were unaware of the surgical procedure the patients underwent and of the phonation modality they were using.

The scores obtained in the four parameters of INFVo scale were correlated with MPT, DDK, speech rate and number of distortions detected with the word repetition test. Finally, INFVo scores were correlated with VHI scores.

Statistical Analysis

Statistical tests were performed using SPSS 18.0 statistical software (SPSS, Inc., Chicago, IL). A student's t-test was used to compare INFVo scores of the three raters. The test-retest reliability and the inter-rater reliability of the INFVo were evaluated with Pearson product-moment correlation test; a value greater than 0.5 was considered 'strong' and values between 0.3 and 0.5 were considered 'moderate'. Values below 0.3 were considered 'low'. Two-way mixed-effects model (consistency definition) intraclass correlation coefficients (ICCs) was also used for reliability analysis. In order to analyze correlations between INFVo scores, objective measures and VHI data, a Pearson product-moment correlation test was also used. The research was reviewed and approved by the Institutional Review Board of the Sacco Hospital of Milan.

Results

The mean scores, standard deviation and ranges of INFVo parameters obtained by each rater (Rater 1, re-test Rater 1, Rater 2, Rater 3) are reported in Table II. For Rater 1, a small increase of the mean values of some parameters was apparent in the re-test condition. A difference between the 3 raters was seen: Rater 2 assigned the lowest mean scores, while Rater 3 gave higher mean scores compared to Rater 1. The differences never reached statistical significance by Student's t-test (p > 0.05).

Reliability analysis

The intra- and inter-rater reliability scores obtained through the Pearson test and the ICC analysis for each parameter of the INFVo are reported in Table III. Concerning intra-rater reliability, the correlation was strong

in all four parameters of the IN-FVo. The inter-rater reliability also showed strong correlation; in particular the parameter that best correlated among raters was over-all impression (I) (r = 0.88), while the lowest correlation was found for the additive noise parameter (N) (r = 0.63).

Objective measures
The objective measures scored as follows: 5.4 ± 5.2 s (range 1-24 s) for MPT, 1.6 ± 0.6 syll/s (range 1-4 syll/s) for DDK, 2.7 ± 0.8 syll/s (range 1-5 syll/s) for the speech rate and 3.7 ± 4.8 (range 0-16) for the number of distortions in the word repetition test.

Voice-related QOL questionnaire - VHI
The VHI total score was 34.9 ± 8.2 (range 6-81). The emotional subscale of the VHI scored the lowest values, with a mean of 8.1 ± 7.2 (range 0-28); the functional and the physical subscales scored 13.5 ± 7.5 (range 2-33) and 13.6 ± 7.6 (range 1-35), respectively.

Correlation analysis
The correlation between the four parameters of INFVo and MPT, DDK, speech rate and number of distortions are reported in Table IV. The correlations between the parameters of INFVo scale were strong with MPT, ranged between low and strong with DDK and were strong with speech rate. Strong correlation was found between the parameter F of the INFVo scale and the speech rate (r = 0.83), while low but significant correlation was found between the parameter N of the INFVo and the DDK (r = 0.45). Moreover, the number of distortions was inversely correlated with all the INFVo parameters. In particular, the strongest correlation was found with the parameter I (r = -0.82), while the lowest was found with the F one (r = -0.64).
The correlation between INFVo parameters scores and VHI is reported in Table V. Only the functional subscale of the VHI was negatively correlated with the parameters I (r = -0.51), N (r = -0.62), and F (r = -0.41) of the INFVo scale.

Discussion

Auditory-perceptual evaluation is the most commonly used methods for clinical assessment of voice, and it is often considered as the gold standard for documentation of voice disorders [1]. However, perceptual evaluation has been heavily criticized because it is subjective and listener reliability is not always adequate [7-9]. This is even truer for substitution voices; while their irregularity makes instrumental measures unreliable and increases the need of adequate perceptual assessment, little agreement exists on the terms to use in the evaluation of this particular type of voice. In the present study, the intra-rater and inter-rater reliability as the relationship of the INFVo auditory perceptual scale with objective measures and VHI scores in native Italian-speaking patients with substitution voice were studied. The need for a study with Italian speakers and raters lies in the fact that social and cultural aspects have a great importance for the perceptual judgment of voice [19]. Reliability was overall good and correlation between the INFVo scale and MPT, DDK, speech rate, number of distortions and VHI scores were found. This is the first report on the application of the INFVo scale of Italian-speaking patients with substitution voice; these results add further support to previous studies on the application of the INFVo rating scale as a perceptual method in a multidimensional assessment protocol for substitution voicing.
Intra-rater reliability scores were excellent with correlation values ranging between r = 0.95 and r = 0.97; these values are considered optimal for individual measurements over time [48]. This is the first report in the literature on INFVo intra-rater reliability, since in the original

Table II. Mean ± standard deviation and ranges of the INFVo scores by the three raters in all patients.

	I	N	F	Vo
Rater 1	4.7 ± 3.3 (0-10)	4.8 ± 2.9 (0-10)	4.0 ± 3.5 (0-10)	2.9 ± 3.5 (0-10)
Retest Rater 1	4.8 ± 3.1 (0-10)	4.7 ± 2.6 (1-10)	4.2 ± 3.5 (0-10)	3.1 ± 3.4 (0-10)
Rater 2	4.6 ± 3.5 (0-10)	4.5 ± 3.1 (0-10)	3.5 ± 3.3 (0-9)	3.5 ± 3.5 (0-10)
Rater 3	4.7 ± 3.4 (0-9)	5 ± 2.7 (1-10)	4.1 ± 3.4 (0-10)	3.1 ± 3.1 (0-10)

I: overall impression; N: additive noise; F: fluency; Vo: quality of voicing.

Table III. Inter-rater reliability analysis of the INFVo using Pearson test (r) and ICC analysis.

		I	N	F	Vo
Test-retest Rater 1	R	0.97	0.95	0.97	0.95
	ICC	0.97	0.95	0.97	0.95
Rater 1 vs 2	R	0.88	0.63	0.80	0.73
	ICC	0.88	0.63	0.79	0.73
Rater 1 vs 3	R	0.87	0.83	0.87	0.82
	ICC	0.86	0.82	0.87	0.82
Rater 2 vs 3	R	0.86	0.68	0.79	0.87
	ICC	0.86	0.68	0.79	0.86

I: overall impression; N: additive noise; F: fluency; Vo: quality of voicing.

studies of Moerman et al. [34-35] no attempts had been made to measure it. While the data is encouraging for the reliability of the INFVo scale, it should be considered with caution, since only one rater was recruited for judging twice the same group of voices. In a reliability study with another perceptual scale for tracheo-oesophageal voice, intra-rater scores of 12 expert speech and language pathologists were also very high [49]. Intra-rater reliability for so-called common dysphonia also appeared rather high; in the first study on GRBAS scale reliability, the intra-rater correlation scores between two judges ranged between r = 0.70 and r = 0.46 [50], while more recent studies on CAPE-V intra-rater reliability report correlation scores up to r = 0.93 [51] [52]. Thus, it appears that the

Table IV. correlation between INFVo scores and MPT, DDK, speech rate and number of distortions.

	I	N	F	Vo
MPT (s)	0.59*	0.57*	0.73*	0.57*
DDK (syll/s)	0.47*	0.45†	0.67*	0.47*
Speech rate (syll/s)	0.71*	0.63*	0.83*	0.64*
Number of distortions	- 0.82*	- 0.81*	- 0.64*	- 0.71*

* p < 0.01; † p < 0.05.
I: overall impression; N: additive noise; F: fluency; Vo: quality of voicing.

Table V. Correlation between INFVo parameters and self-assessment of the voice quality measured by VHI.

	I	N	F	Vo
VHI tot	- 0.29	- 0.46*	- 0.20	- 0.13
VHI e	- 0.18	- 0.25	- 0.02	- 0.07
VHI f	- 0.51†	- 0.62†	- 0.41*	- 0.26
VHI p	- 0.06	- 0.25	- 0.04	- 0.02

* p < 0.05; † p < 0.01.
I: overall impression; N: additive noise; F: fluency; Vo: quality of voicing; tot: total; e: emotional; f: functional; p: physical.

intra-rater reliability found in this study is in agreement with previous research on both alaryngeal voice and common dysphonia.

Inter-rater reliability was good, with correlation values ranging between r = 0.73 and r = 0.88 for all the parameters of the INFVo scale, with the exception of the N parameter, where correlation scores ranged between r = 0.63 and r = 0.83. Moerman et al. [34] also reported significant high inter-rater reliability; moreover, in the two previous studies on the INFVo as well as in the present, the I and F parameters seem to have stronger reliability compared to the other two. Nonetheless, the inter-rater reliability data should be considered with caution because of the small number of raters in the present study.

A significant correlation between the INFVo scale and MPT, DDK, speech rate and number of distortions was found in the present study; in particular, the correlations between the F parameter and the MPT and the speech rate were very good. These findings are particularly interesting, since aerodynamic alterations in tracheo-oesophageal speakers have been demonstrated [53] [54]; furthermore, it seems reasonable that the fluency of speech increases as the MPT and speech rate increase. A very good correlation between the number of distortions and the I and N parameters was found. Since the parameter I is strongly correlated with intelligibility [35], it is not surprising that there is reciprocal impact between articulation distortions on one side and I and N on the other.

A fair to good inverse correlation between I, N and F parameters of the INFVo scale and the functional sub-scale of the VHI were found; the N parameter also

correlated with the VHI total score. In two studies on so-called common dysphonia, a moderate correlation between self-assessment measures and GRBAS scale were found [41] [42]. More interestingly, the VHI has been studied in patients with substitution voice after supracricoid laryngectomy, showing positive correlation between the G parameter of the GRBAS scale and total VHI [43]. These studies are in agreement with our findings. Furthermore, the fact that correlations between INFVo parameters and VHI were limited to the functional sub-scale suggests that the INFVo parameters are particularly informative on perceived functional communication rather than on an overall perceived voice handicap, which reflects many different factors, including the patient's psychosocial traits and cultural and ethnic backgrounds.

Conclusions

In conclusion, the INFVo scale appears to be a reliable tool and can be recommended for the perceptual assessment of substitution voices in Italian-speaking patients. A moderate to strong relationship between MPT, DDK and speech rate and INFVo scores was also found, although further studies on the relationship between objective measures and INFVo scores are warranted. Furthermore, the functional sub-scale of VHI correlated with the parameters of INFVo supporting the hypothesis that the INFVo parameters are informative on the perceived communication rather than on the overall voice handicap.

References

1 Oates J. *Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions*. Folia Phoniatr Logop 2009;61:49-56.

2 Dejonckere PH, Bradley P, Clemente P. *A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)*. Eur Arch Otorhynolaryngol 2001;258:77-82.

3 Dejonckere PH. *Perceptual and laboratory assessment of dysphonia*. Otolaryngol Clin North Am 2000;33:731-50.

4 Dejonckere PH, Remacle M, Fresnel-Elbaz E, et al. *Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements*. Rev Laryngol Otol Rhinol (Bord) 1996;117:219-24.

5 Bhuta T, Patrick L, Garnett D. *Perceptual evaluation of voice quality and its correlation with acoustic measurements*. J Voice 2004;18:299-304.

6 Eadie TL, Doyle PC. *Classification of dysphonic voice: acoustic and auditory-perceptual measures*. J Voice 2005;19:1-14.

7 Kent RD. *Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders*. Am J Speech Lang Pathol 1996;5:7-23.

8 Yiu EM. *Limitations of perturbation measures in clinical acoustic voice analysis*. Asia Pacific J Speech Lang Hear 1994;4:155-66.

9 Ma EP, Yiun EM. *Multiparametric evaluation of dysphonic severity*. J Voice 2006;20:380-90.

10 De Bodt MS, Van de Heyning PH, Wuyts FL, et al. *The perceptual evaluation of voice disorders*. Acta Otorhinolaryngol Belg 1996;50:283-91.

11 Gelfer MP. *Perceptual attributes of voice: development and use of rating scales*. J Voice 1988;2:320-6.

12 Bassich CJ, Ludlow C. *The use of perceptual methods by new clinicians for assessing voice quality*. J Speech Hear Dirsord 1986;51:125-33.

13 Eadie TL, Baylor CR. *The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice*. J Voice 2006;20:527-44.

14 Revis J, Giovanni A, Wuyts F, et al. *Comparison of different voice samples for perceptual analysis*. Folia Phoniatr Logop 1999;51:108-16.

15 Wolfe V, Cornell R, Fitch J. *Sentence/vowel correlation in the evaluation of dysphonia*. J Voice 1995;9:297-303.

16 Zraick RI, Wendel K, Smith-Olinde L. *The effect of speaking task on perceptual judgment of the severity of dysphonic voice*. J Voice 2005;19:574-81.

17 Wuyts FL, De Bodt MS, Van de Heyning PH. *Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia*. J Voice 1999;13:508-17.

18 Yu P, Revis J, Wuyts FL, et al. *Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analog scale*. Folia Phoniatr Logop 2002;54:271-81.

19 Hammarberg B. *Pathological voice qualities. Perceptual and acoustic characteristics of a set of Swedish "reference" voices*. Bull Audiophonol 1992;8:39-52.

20 Wilson DK. *Voice problems of children.* 3rd ed. Baltimore: Williams & Wilkins; 1987.

21 Laver J. *The phonetic description of voice quality*. London: Cambridge University Press; 1987.

22 Hirano M. *Clinical examination of voice*. New York: Springer; 1981.

23 Zraick RI, Kempster GB, Connor NP, et al. *Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)*. Am J Speech Lang Pathol 2011;20:14-22.

24 Carding PN, Carlson E, Epstein R, et al. *Formal perceptual evaluation of voice quality in the United Kingdom*. Logop Phon Vocol 2000;25:133-8.

25 Webb AL, Carding PN, Deary IJ, et al. *The reliability of three perceptual evaluations scales for dsyphonia*. Eur Arch Otorhinolaryngol 2004;261:429-34.

26 Ghidini A, Trebbi M, Piccinini A, et al. *Functional evaluation protocol*. Acta Otorhinolaryngol Ital 2010;30:254-6.

27 Schindler A, Favero E, Nudo S, et al. *Long-term voice and swallowing modifications after supracricoid laryngectomy: objective, subjective, and self-assessment data*. Am J Otolaryngol 2006;27:378-83.

28 Schindler A, Cuccarini V, Bottero A, et al. *Long-term vocal functional result after glottectomy: a multidimensional analysis*. Eur Arch Otorhinolaryngol 2007;264:1039-44.

29 Kazi R, Kiverniti E, Prasad V, et al. *Multidimensional assessment of female tracheoesophageal prosthetic speech*. Clin Otolaryngol 2006;31:511-7.

30 Makeieff M, Barbotte E, Giovanni A, et al. *Acoustic and aerodynamic measurement of speech after supracricoid partial laryngectomy*. Laryngoscope 2005;115:546-51.

31 Makeieff M, de la Breteque A, Guerrier B, et al. *Voice handicap evaluation after supracricoid partial laryngectomy*. Laryngoscope 2009;119:746-50.

32 Bron L, Pasche P, Brossard E, et al. *Functional analysis after supracricoid partial laryngectomy with cricohyoidoepiglottopexy*. Laryngoscope 2002;112:1289-93.

33 Laudadio P, Presutti L, Dall'Olio D, et al. *Supracricoid laryngectomies: long-term oncological and functional results*. Acta Otolaryngol 2006;126:640-9.

34 Moerman M, Martens JP, Van der Borgt MJ, et al. *Perceptual evaluation of substitution voices: development and evaluation of the (I)INFVo rating scale*. Eur Arch Otorhinolaryngol 2006;263:183-7.

35 Moerman M, Martens JP, Crevier-Buchman L, et al. *The INFVo perceptual rating scale for substitution voicing: development and reliability*. Eur Arch Otorhinolaryngol 2006;263:435-9.

36 Yamagichi H, Shrivastav R, Andrews ML. *A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale*. Folia Phoniatr Logop 2003;55:147-57.

37 Moerman M, Pieters G, Martens JP, et al. *Objective evaluation of the quality of substitution voices*. Eur Arch Otorhinolaryngol 2004;261:541-7.

38 Luppi MP, Nizzoli F, Bergamini G, et al. *Speech therapy rehabilitation*. Acta Otorhinolaryngol Ital 2010;30:244-7.

39 Frattali CM. *Measuring modality-specific behaviours, functional abilities and quality of life*. In: Frattali CM, editor. *Measuring outcomes in speech and language pathology*. New York: Thieme; 1997.

40 Eadie TL, Doyle PC. *Auditory-perceptual scaling and quality of life in tracheoesophageal speakers*. Laryngoscope 2004;114:753-9.

41 Murry T, Medrado R, Hogikyan ND, et al. *The relationship between ratings of voice quality and quality of life measures*. J Voice 2004;18:183-92.

42 Jones SM, Carding PN, Drinnan MJ. *Exploring the relationship between severity of dysphonia and voice-related quality of life*. Clin Otolaryngol 2006;31:411-7.

43 Makeieff M, de la Breteque A, Guerrier B, et al. *Voice handicap evaluation after supracricoid partial laryngectomy*. Laryngoscope 2009;119:746-50.

44 Lundström E, Hammarberg B. *Speech and voice after laryngectomy: perceptual and acoustical analyses of tracheoesophageal speech related to voice handicap index*. Folia Phoniatr Logop 2011;63:98-108.

45 Schindler O. *Propedeutica*. Turin: Omega; 1977.

46 Jacobson BH, Johnson A, Grywalski C. *The Voice Handicap Index (VHI): development and validation*. Am J Speech Lang Pathol 1997;6:66-70.

47 Schindler A, Ottaviani F, Mozzanica F, et al. *Cross-cultural adaptation and validation of the Voice Handicap Index into Italian*. J Voice 2010;24:708-14.

48 Lohr K. *Assessing health status and quality-of-life instruments: attributes and review criteria*. Qual Life Res 2002;11:193-205.

49 Hurren A, Hildreth AJ, Carding PN. *Can we perceptually rate alaryngel voice? Developing the Sunderland tracheoesophageal voice perceptual scale*. Clin Otolaryngol 2009;34:533-8.

50 Dejonckere PH, Obbens C, de Moore GM, et al. *Perceptual evaluation of dysphonia: reliability and relevance*. Folia Phoniatr 1993;45:76-83.

51 Karnell M, Melton S, Childes J, et al. *Reliability of clinical-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders*. J Voice 2007;21:576-90.

52 Kelchner LN, Brehm SB, Weinrich B, et al. *Perceptual evaluation of severe pediatric voice disorders: rater reliability using the Consensus Auditory Perceptual Evaluation of Voice*. J Voice 2010;24:441-9.

53 Searl J. *Bilabial contact pressure and oral air pressure during tracheoesophageal speech*. Ann Otol Rhinol Laryngol 2007;116:304-11.

54 Kotby MN, Hegazi MA, Kamal I, et al. *Aerodynamics of the pseudo-glottis*. Folia Phoniatr Logop 2009;61:24-8.

Address for correspondence: Francesco Mozzanica, U.O. Otorinolaringoiatria, Ospedale "L. Sacco", via G.B. Grassi 74, 20154 Milan, Italy. Fax +39 02 50319855. E-mail: francesco.mozzanica@gmail.com