

## Original articles

# Vocal quality assessment: methodological approach for a perceptive data analysis

Maria Fabiana Bonfim de Lima Silva<sup>(1)</sup>Sandra Madureira<sup>(2)</sup>Luiz Carlos Rusilo<sup>(3)</sup>Zuleica Camargo<sup>(2)</sup>

<sup>(1)</sup> Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brazil.

<sup>(2)</sup> Pontifícia Universidade Católica de São Paulo (PUCSP), São Paulo, SP, Brazil.

<sup>(3)</sup> Instituto de Ciência e Tecnologia da Universidade Federal de Alfenas (UNIFAL), Alfenas, Minas Gerais, Brazil.

Study conducted at the Speech Therapy Course of the Universidade Federal da Paraíba – UFPB – João Pessoa (PB), Brazil and at the Laboratório Integrado de Análise Acústica e Cognição (LIAAC) - PUC-SP, São Paulo, Brasil.

Conflict of interest: Nonexistent

Received on: January 30, 2017

Accepted on: September 14, 2017

**Mailing address:**

Maria Fabiana Bonfim de Lima Silva  
Departamento de Fonoaudiologia, Centro de Ciências da Saúde, Universidade Federal da Paraíba - Cidade Universitária, Campus I, Castelo Branco  
CEP: 58051-900 - João Pessoa, Paraíba, Brasil  
E-mail: fbl\_fono@yahoo.com.br

**ABSTRACT**

**Purpose:** to present a methodological approach for interpreting perceptual judgments of vocal quality by a group of evaluators who used the script *Vocal Profile Analysis Scheme*.

**Methods:** cross-sectional study based on 90 speech samples from 25 female teachers with voice disorders and/or laryngeal changes. Prior to the perceptual judgment, we performed three perceptual tasks to select samples which were then presented to five evaluators using the *Experiment script MFC 3.2* (software PRAAT). Next, we applied a sequence of tests based on successive approaches to inter- and intra-evaluator behavior. Data were treated by statistical analysis (*Cochran* and *Selenor* tests).

**Results:** with respect to the analysis of the evaluators' performance, it was possible to define those that presented the best results, in terms of reliability and proximity of analyses, compared to the most experienced evaluator, excluding one. The results of the cluster analysis also allowed designing a voice quality profile of the group of speakers studied.

**Conclusions:** the proposal of a methodological approach allowed defining evaluators whose judgments were based on phonetic knowledge, and drawing a vocal quality profile of the group of samples analyzed.

**Keywords:** Evaluation; Voice; Teachers; Voice Disorders; Speech Perception

## INTRODUCTION

Perceptual voice assessment is one of the oldest and most widely used procedures for assessing and diagnosing voice disorders<sup>1-3</sup>. The efficacy of its results depends heavily on the evaluator's experience<sup>1,4-6</sup>. Although it is considered as a gold standard for vocal evaluation, there is constant mention of possible interferences arising from the subjectivity of evaluators, lack of reliability of judgments, variety of evaluation methods, inconsistencies of instruments and lack of standardization of the terminology used<sup>1,4,7-10</sup>.

The search for overcoming such limitations is based on different strategies, such as presentation of anchor stimuli, training and calibration of evaluators, repetition of stimuli, application of *scripts* to randomize the order of presentation of speech samples (programming applicable to open source software), and a statistical approach of perceptual judgments<sup>6,11-13</sup>. In this perspective, among the evaluation instruments available for clinical use, there are few scripts and scales based on theoretical models such as phonetic theory, as is the case of the *Vocal Profile Analysis Scheme* (VPAS)<sup>14</sup>.

It is worth mentioning that the use of instruments of Phonetics in Speech-Language Pathology clinics has contributed to a detailed identification of speech structures in cases of speech disorders<sup>15,16</sup>, as well as speech control cues in the process of language acquisition of children with and without hearing impairment (HI) from early ages<sup>17-19</sup>. In addition, the use of such instruments may offer possibilities for the characterization of language sonority and linguistic variants<sup>20</sup>.

The VPAS script, and its adaptation to Brazilian Portuguese VPAS-PB<sup>21</sup>, details the occurrence of several vocal quality adjustments in phonatory, articulatory and tension areas, as well as vocal dynamics elements (pitch, loudness, use of pauses, speech rate and respiratory support) from the perspective of phonetic theory. The application of the VPAS script results in the voice quality profile of samples. An example could be a sample whose vocal quality profile is characterized by the combination of closed jaw adjustments (level 1), elevated larynx (level 2) and laryngeal hyperfunction (level 2).

Part of the complexity referred to by clinicians in their initial contact with the VPAS script lies in the theoretical principles behind it. The principles of compatibility and interdependence are related to relationships between vocal quality adjustments: the first deals with actions physiologically compatible or incompatible with each other; the second, in turn, focuses on actions

physiologically interdependent. A third principle, susceptibility, refers to the relationship between adjustments and segments (vowels and consonants), that is, how vocal quality adjustments affect segments along the speech chain. In this last principle, a segment (vowel and consonant) may be susceptible to the interference of an adjustment, that is, it reflects the degree of vulnerability of segments in relation to adjustments, especially of segments considered as "key" for the detection of vocal quality events<sup>22</sup>. Thus, when adjustments have characteristics not shared by the segment, the latter becomes more susceptible to the influence of the former.

Another aspect to be considered in the field of perceptual evaluation of vocal quality refers to the demand for adoption of a group of examiners/evaluators to address especially the question of subjectivity in perception tests. This issue, which applies to the universe of studies that use the script VPAS, refers to a demand for the establishment of a vocal profile based on judgments of several evaluators. That is, the final result of the evaluation of each vocal sample should be considered in light of the judgments made by evaluators individually, resulting in the definition of a vocal quality adjustment (and its degree of manifestation), which comprises the voice quality profile from a phonetic point of view.

The scarcity of studies substantiating a method for analyzing perceptual judgments of vocal quality based on phonetic models and statistical procedures, which allows estimating the most similar judgments between evaluators (in pairs) and choosing the evaluator with the greatest reliability regarding an analysis instrument that presents a scale of several dimensions, justifies the interest of this study. In addition, the discussion about principles, procedures and especially possibilities and limitations of the application of perceptual analysis to clinical care routines and research environments stimulates a fertile ground that seeks to promote thinking on the nature and the theoretical basis of evaluation protocols and perceptive descriptions of voice used in a scientific and clinical context, as well as its relation to the vocal history of the speaker.

It should also be noted that, in several studies, the description of perceptual data of vocal quality is a step of the analysis and that such findings will often be compared to acoustic and/or physiological data. Thus, the approach that allows perceptual judgment data to result in information that may be analyzed statistically is a current demand.

This study aims to present a methodological approach for the interpretation of perceptual judgments of vocal quality by a group of evaluators who used the script Vocal Profile Analysis Scheme (VPAS) adapted to Brazilian Portuguese<sup>21</sup>.

## METHODS

This research was approved by the Ethics Committee on Research with Human Beings of the Federal University of Paraíba, UFPB, under the protocol no. 298/2008. The *corpus* of the study comprised samples from a database containing 54 teachers' voices. Teachers were elected to participate in this study because they are voice professionals with the highest incidence of voice disorders, frequently seeking speech and hearing care<sup>3</sup>. The inclusion criteria were: the person had to be a female teacher with voice disorders and laryngeal changes (by perceptual information and otorhinolaryngological diagnosis). The samples must have been recorded using three speech styles. Based on such criteria, 33 teachers were selected.

The audio recordings were made in the teachers' work environment, during intervals between classes. They consisted in the following tasks: semi-spontaneous speech (interview situation), semi-spontaneous speech (lecture simulation) and reading out loud<sup>23</sup>. The choice for different speech styles (tasks) lies in variations already studied in vocal quality adjustments and vocal dynamics<sup>24</sup>.

The reading out loud task comprised reading a passage of standard text<sup>23</sup>. In addition, a semi-spontaneous speech-interview (SSI) was conducted starting with the question "What factors do you think interfere with the voice? Why?", as proposed by the authors<sup>2</sup>. Finally, lecture simulation consisted in a lecture excerpt with a topic chosen by the teacher (without a specific time limit) following the examiner's request.

Speech samples were recorded in a quiet room using a Plantronics GameCom PRO 1 headset microphone at a distance of approximately 15 cm from the right labial commissure, coupled to an HP Pavillion ZE 4920 CEL M330 1.4G notebook. The software used was *SoundForge 7.0* set at a sampling frequency of 22,050 Hz, 16 bits, extension ".wav".

All 99 samples of 33 teachers were submitted to three perception tasks, performed by different groups of evaluators. From the results of such tasks, we selected a set of samples that became the *corpus* of this study. This *corpus* is detailed below. The objectives and methods, as well as the results of each task, are summarized in Figure 1.

After analyzing the results of the three perception tasks, samples of eight teachers were excluded. Thus, the *corpus* of the perception experiment of vocal quality consisted of 90 speech samples from 25 teachers: 25 loud readings, 25 semi-spontaneous speeches (SSI), 25 semi-spontaneous lecture simulations (SLS) and 15 replications of some samples to approach the reliability of the evaluators' answers, totaling a 20% randomized sample replication of the *corpus*<sup>25</sup>. The samples were edited in extracts of approximately 20 seconds extracted from the recordings of the 3 speech tasks.

All 90 samples were labeled as statements related to reading out loud (RL), semi-spontaneous-interview (SSI) and semi-spontaneous-lecture simulation (SLS), and analyzed based on a perceptual-auditory point of view (VPAS-PB) using the software *PRAAT* and the *Experiment script MFC 3.2*, version 5143 (available at: <http://www.fon.hum.uva.nl/praat/>).

The script *Experiment MFC 3.2* was used as a tool to randomize stimuli to be presented to all five evaluators (E1 to E5), who would evaluate vocal quality with a phonetic motivation. In the first screen of the perception experiment, a test instruction was presented. In the other screens, controlled by the evaluator, 90 sound stimuli were presented.

The duration of the experiment corresponded to approximately four hours per evaluator distributed into four sessions on different days, lasting one hour each session. There were intervals (pauses) of five minutes for auditory rest after the presentation of ten samples.

The selection of the group of evaluators was based on expertise in Phonetics and experience in the application of the script VPAS-PB. We decided to select evaluators with different levels of experience and expertise in order to discuss interferences of the variables with the vocal quality evaluation, according to Figure 2.

Tasks	Objective	Method	Results
PT1	Investigate the influence of the nature of speech samples (on RL, SSI and SLS tasks) in the judgments using VPAS-PB.	Selection of 8 speech samples (4 of RL and 4 of SSI) from 4 female teachers, all with vocal complaints; 2 presented otorhinolaryngological diagnosis for laryngeal changes.  These samples were edited, randomized and submitted to 3 evaluators experienced in the use of the VPAS-PB script.	The manifestation of vocal quality adjustments and vocal dynamics elements in the script VPAS-PB varied according to speech task.  Impacts on research planning: inclusion of the 3 speech styles (RL, SSI and SLS).
PT2	Define the duration of the samples to be used in the experiment.	Selection of 4 speech samples from 2 female teachers with laryngeal changes in 2 speech tasks, which were edited into different durations (20 s, 30 s and 40 s).  Ten evaluators were asked to evaluate the vocal quality (script VPAS-PB) of those who participated in this experiment using a specific form.	The duration of samples had no impacts on the perceptual judgments of vocal quality.  Impacts on research planning: duration of speech samples set to 20 s.
PT3	Select the recordings with the best audio quality using signal-to-noise ratio (SNR).	Selection of 36 samples: 2 noise levels were added to 24 samples (at the medium and the maximum sound waves).  Samples were analyzed by 8 evaluators with varying levels of experience in the use of the script VPAS-PB and by 2 experienced evaluators, considered as reference.	It evidenced a compatibility of answers of evaluators with various degrees of experience, with a slight increase in the number of mistakes proportional to the increase in the noise level of the sample.  Impacts on research planning: exclusion of stimuli with a SNR < 2

Caption: RL = reading out loud, SLS = semi-spontaneous lecture simulation, SSI = semi-spontaneous speech-interview, PT1 = Perception task 1, PT2 = Perception task 2, PT3 = Perception task 3.

**Figure 1.** Perception tasks conducted at the planning stage of sample selection procedures for percept analysis

Evaluator	Professional training time (years)	Training	Time using the script VPAS-PB (months)	Training time on the script VPAS-PB (months)	Participation in the workshop - script VPAS-PB (prior to the application of the experiment)
E1	4	Speech therapist with a Master's degree in Speech Therapy	18	18	Yes
E2	10	Speech therapist with a Master's degree in Linguistics	36	18	Yes
E3	24	Speech Therapist with a Master's degree in Linguistics	18	18	Yes
E4	23	Speech therapist with a PhD in Linguistics	156	156	Yes
E5	28	Speech Therapist with a PhD in Linguistics	6	6	No

**Figure 2.** Characterization of evaluators participating in the experiment of vocal quality perception with phonetic motivation

In order to standardize the procedures and solve any doubts, the five evaluators were invited to participate in a workshop entitled “Roadmap VPAS-PB: auditory training”, lasting 15 hours. The Evaluator 5 did not participate in this workshop, and routinely uses perceptual voice assessment in his clinical practice, although based on other assessment tools.

The initial approach of judgments made by E1 to E5 was based on inter- and intra-evaluator concordance and reliability in the use of the script VPAS-PB and in serial tests, by which scores and a classification were gradually established between evaluators so that the analysis of the most discrepant evaluators was excluded. Then, the tests were reapplied to the

remaining group. The Cochran test was used for homogeneity of variances. The Snedecor test was used with 95% confidence levels (in an Excel worksheet). Both tests were used to define intra- and inter-group reliability (including pairs). We noted that all evaluators, but one, presented a significant congruence between them. Thus, the judgments made by the incongruous evaluator were excluded in a blind procedure (i.e., the data analyst did not know the evaluators, having access only to the judgment worksheet). Then, the evaluators were classified according to their intra-judgment reliability scores and their congruence with the other evaluators.

In addition, at the final stage, after analyzing all evaluators, the judgments were compared to the evaluator considered as a reference based on the classification generated by this analysis.

The criteria used for the selection of the reference evaluator were specific training, time of use and expertise on the VPAS script, participation in the workshop on the use of this script, and inter- and intra-judgment congruence and reliability.

Having determined the statistical parameters for the valuation of judgments, the voice quality profile of each sample was designed. The profile was established based on the mean values of data from analyses of evaluators which were congruent with each other, determining a judgment composed of expected values based on univariate statistical analysis (computation of confidence intervals).

The values of distances and relative products of inter-evaluator judgments (in pairs) for perceptual evaluation results allowed estimating the closest judgments among evaluators in pairs.

Relative distance is a measure of relative dissimilarity between evaluators defined by the difference of positions between evaluators, i.e., the difference between the ranking of each evaluator compared two

by two. The ranking, in turn, is defined as the ranking position of each evaluator according to an average of correct judgments. Since evaluators were ranked from 1 to 5, the relative distances may assume values ranging from 1 to 4. High values indicate a great dissimilarity among evaluators.

The relative product is a composite index constructed by multiplying the ranking values of the evaluators compared two by two by the relative distance between them, as defined in the previous paragraph, indicating the quality of both judgments and their relative proximity. The lower the value of this index, the better the composition formed by two evaluators.

The comparison of intra-evaluator judgment data based on congruence values (index of correct judgments) allowed us to estimate the most experienced and congruent evaluator.

## RESULTS

With respect to the analysis of the initial behavior of the evaluators (Cochran test, inter-evaluator analysis; and Snedecor test, inter- and intra-evaluator analysis), the E5 was excluded from the next step (reapplication of the test) precisely because it presented the largest intra- and inter-evaluator variance, less time of use and expertise in the script VPAS. This evaluator also did not participate in the training on the use of the script. The test was reapplied until the two smallest variances within the group, which represented the two evaluators with the greatest reliability in terms of judgment, were reached through the VPAS-PB: E2 and E4.

The correct answer indexes and confidence intervals (upper and lower limits), based on intra-evaluator analysis, are presented in Tables 1 and 2. Correct answers were considered according to comprehensiveness of results in perceptive judgments.

The correct answer indexes and confidence intervals (upper and lower limits), based on inter-evaluator

**Table 1.** Values of congruence (% of correct answers) of intra-evaluator judgments (E1 to E5) for the total results of judgments

	Lower limit	Correct Answers	Upper limit
E1	90.93	92.72	94.51
E2	92.59	94.20	95.81
E3	90.93	92.72	94.51
E4	94.14	95.56	96.97
E5	87.40	89.51	91.62

**Caption:** Total results - all adjustments in the script VPAS-PB (levels 0 to 6)

**Table 2.** Values of congruence (% of correct answers) of intra-evaluator judgments (E1 to E5) for the total non-null results of judgments

	Lower limit	Correct Answers	Upper limit
E1	48.64	56.93	65.23
E2	43.22	53.00	62.78
E3	36.52	45.87	55.23
E4	58.86	67.57	76.28
E5	57.69	63.83	69.97

**Caption:** Non-null results - adjustments evaluated in the script VPAS-PB (levels 1 to 6)

**Table 3.** Values of congruence (% of correct answers) of inter-evaluator judgments (E1 to E5) for the total results of judgments

	Lower limit	Correct Answers	Upper limit
E1-E2	81.20	82.37	83.54
E1-E3	80.29	81.48	82.68
E1-E4	79.22	80.44	81.67
E1-E5	68.31	69.73	71.14
E2-E3	84.07	85.16	86.26
E2-E4	82.79	83.93	85.06
E2-E5	70.82	72.20	73.58
E3-E4	82.26	83.41	84.55
E3-E5	69.79	71.19	72.58
E4-E5	70.57	71.95	73.33

**Caption:** Total results - all adjustments in the script VPAS-PB (levels 0 to 6)

**Table 4.** Values of congruence (% of correct answers) of inter-evaluator judgments (E1 to E5) for total non-null results of judgments

Pairs of evaluators	Lower limit	Correct Answers	Upper limit
E1-E2	8.09	10.19	12.29
E1-E3	7.63	9.64	11.65
E1-E4	8.49	10.51	12.53
E1-E5	8.58	10.18	11.79
E2-E3	8.00	10.30	12.60
E2-E4	7.43	9.58	11.73
E2-E5	8.06	9.70	11.35
E3-E4	7.44	9.56	11.67
E3-E5	6.54	8.04	9.53
E4-E5	9.32	11.04	12.76

**Caption:** Non-null results - adjustments evaluated in the script VPAS-PB (levels 1 to 6)

**Table 5.** Distances and relative products of inter-evaluator judgments

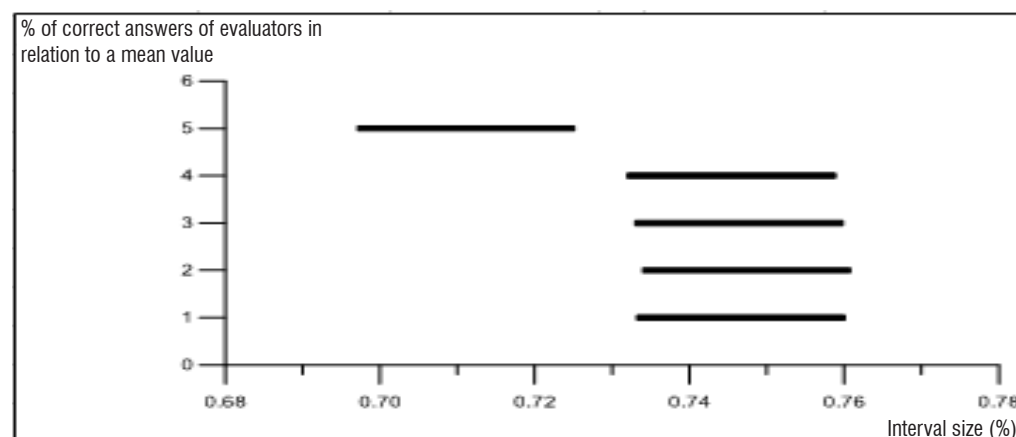
Pairs of evaluators	Relative distances	Related products
E1-E2	1	6
E1-E3	1	12
E1-E4	2	6
E1-E5	2	30
E2-E3	2	16
E2-E4	1	2
E2-E5	3	30
E3-E4	3	12
E3-E5	1	20
E4-E5	4	20

analysis, are presented in Tables 3 and 4. Correct answers were considered according to comprehensiveness of results in perceptive judgments.

Table 5 shows the distances and the relative products of inter-evaluator judgments (in pairs) for the results of perceptual judgments.

The approach of distances and relative products allowed estimating which judgments were closer between the evaluators; the result was E2-E4, E1-E2, E1-E4 and E3-E5. From the comparison of intra-evaluator judgment data, we noted that the E4

presented the highest correct answers ratio in total, followed by the E2 and the pair E1-E3. In turn, the E4 was considered the most experienced and congruent evaluator. It is noteworthy that the group of evaluators did not reveal an absolutely similar behavior at all steps when considering the inter-evaluator approach. However, they were consistent in their judgments in repeated samples in intra-evaluator analysis. Therefore, their contributions to judgments could be considered in the composition of vocal quality profile. In this respect, the results of the answers to the repetition of stimuli

**Figure 3.** Distribution of confidence intervals in judgments based on the means of each evaluator (E1 to E5)

revealed a homogeneity among the four evaluators (E1 to E4), and the segregation of the E5 (Figure 3). In view of the data presented, the evaluators were classified according to their experience in the VPAS-PB script in the following descending order: E4, E2, E1, E3 and E5.

The profile of vocal quality and elements of vocal dynamics traced for this study contemplated a set of

analyses of four evaluators (from the inter- and intra-evaluator approaches), which reached a level of distribution of variances and confidence intervals of similar judgments made between them. The vocal quality profile of the studied group was characterized by decreasing order of occurrence: laryngeal hyperfunction, rough voice, elevated larynx, vocal tract hyperfunction, closed

mandible, pharyngeal constriction, raised tongue body and breathiness. As for vocal dynamics aspects, in descending order, the following stood out: inadequate respiratory support, decreased variability of pitch, usual high pitch, high habitual loudness, fast elocution rate and increased loudness variability.

## DISCUSSION

In speech-language practice, the perceptual evaluation of vocal quality is considered the gold standard<sup>26</sup>. Although some researchers classify it as a subjective, inconstant method, with a great terminological variability, we emphasize that perceptual evaluation depends on the expertise and the experience of the evaluator, as well as on its attention throughout the procedure<sup>27-29</sup>.

There are few studies presenting a methodological approach for the analysis of perceptual judgments of vocal quality based on a phonetic model<sup>14</sup>, as well as on statistical treatment procedures for the consideration of judgments of several evaluators together. This study aimed to present a methodological approach to develop an experiment of perceptive evaluation of vocal quality with samples of teachers with voice disorders and/or laryngeal changes using the script *Vocal Profile Analysis Scheme (VPAS-PB)*, and also aimed to evaluate the performance of a group of evaluators. The choice to compose a group of evaluators with a varied experience in terms of time of exposure to the script, history of training and participation in training prior to the application of the perception task of vocal quality had the objective of discussing precisely the evaluator training demands for the method under analysis.

The issue of the experience of evaluators has been widely debated in the literature, especially as for a possible subjectivity on analyses<sup>30</sup>. In this study, it was possible to redeem the time of training and the performance of evaluators according to the classification established in terms of their experience by statistical analysis data of the results of the evaluators' judgments by using the script *VPAS-PB*. The degree of experience was related to the time spent using the *VPAS-PB* in a descending order: thirteen years (E4), three years (E2), one year and six months (E1 and E3), and six months (E5).

In the sample of five evaluators, it is possible to identify important factors in the definition of the evaluators' experience: time spent with the instrument *VPAS-PB*, specific training in a phonetic approach to vocal quality, participation in the workshop on the use

of this script, and inter- and intra-judgment congruence and reliability. The statistical procedures adopted, being the subjective characteristics of the subjects (evaluators) unknown when applied, allowed establishing a scale of experience of evaluators congruent with the aspects of training and time of activity in phonetic evaluation of vocal quality (*VPAS-PB*), as well as the participation of the evaluators in the training on the *VPAS-PB*. It is worth mentioning that the application of the script to evaluators was made after the training (workshop), aiming to reach a level of calibration using anchor-stimuli, a procedure also defended by the authors, which explore the complexity of experiments of perception of vocal quality<sup>12</sup>. Thus, the Evaluator 5 was excluded since it had the shortest time of training and use of the *VPAS-PB* script, did not participate in the workshop and presented the greatest variance in intra- and inter-judgment analysis. It was therefore considered an incongruent evaluator.

The choice for the interpretation method of findings of the initial group of five evaluators was also challenging, especially as regards the complexity of considering the uniqueness of individual analyses in search for a "consensus", or one analysis that reflected the opinion of the group. In order to discuss the specificities and, in particular, the complexity of perceptual analyses by groups of evaluators, we decided not to work with consensus analyses or reliability assessments of the answers of evaluators which could lead to a choice for one of the evaluators. In view of the demand for a discussion on the advantages and disadvantages of adopting a phonetic model for the description of vocal quality<sup>14</sup>, we decided to study in a more detailed way the set of perceptual judgments of the five initial evaluators until it was possible to define a set of tests which allowed the definition of a vocal quality profile of a group of voiced samples.

After the global analysis of judgments and the analysis of the general behavior of evaluators, it was possible to develop a sequence of tests which resulted in the choice of the evaluators whose judgments were based on the principles of the phonetic model.

It is worth emphasizing that this evaluation was not intended to qualify evaluators in terms of their perceptual skills, but to qualify and estimate their performance in terms of the proposed task, considering their consistency of answers for the same stimuli at different moments of the analysis, which characterizes an intra-evaluator analysis.



Another important point is that the four evaluators whose analyses comprised the average profile of vocal quality judgments of the set of samples studied did not have a similar behavior when analyzed using an inter-evaluator approach. However, they were consistent as for their judgments for repeated samples. Thus, although the group is not absolutely homogeneous, their judgments are consistent at different moments. Such findings were similar to those found by another study involving a group of students of Speech-Language Therapy. There was a concordance of intra-evaluator answers in relation to the analyses of evaluators (speech-language therapists with expertise on voice)<sup>6</sup>.

The information from intra-evaluator analyses was also interesting since it provided a comparison of a group of evaluators with the judgments of an evaluator with more experience with the instrument, who could be considered as a reference evaluator, a procedure used at other stages of studies based on VPAS-PB judgments of two evaluators experienced in this script<sup>11</sup>. In one of such studies, the authors<sup>11</sup> reported relevant results arising from the training of a group of 16 evaluators (14 first-year students in a voice specialization course; the two other evaluators were Speech-Language Therapy and Linguistics professors with experience in VPAS) upon investigating the validity and the consensus in the use of VPAS among examiners. One of the highlights was the lack of consensus among the participating evaluators regarding the group of phonatory adjustments and their possible combinations, which, according to the authors<sup>11</sup>, revealed a lack of systematization of auditory-based methods of vocal evaluation and familiarity with the mentioned model. These data were compatible with this study because, by the inter-evaluator analysis, there were some discrepancies in judgments, which reinforce that the extension of the training period or, more precisely, the constant updating and the continuous work with the evaluators becomes essential to create a cohesively qualified group to conduct phonetic analyses of vocal quality.

The care in the procedures of this study for the perceptive analysis of vocal quality, regarding the training and experience in the script VPAS of the group of evaluators and the comparison between the judgments of this group and the judgments of an evaluator with more experience, sheds light on a complexity inherent to studies focusing on the answers of evaluators according to several modes of perception. As for the

perception of vocal quality, we highlight the criticism of the way by which the statistical analysis was conducted in many studies, in which some correlations may be effects of test artifacts and specificities of samples<sup>12</sup>. In this study, we considered the several steps of the study of the evaluators' behavior and of successive approaches in intrinsic terms (intra-evaluator approach: consistency between task repetitions) and extrinsic terms (intra-evaluator approach: in relation to other evaluators, or more precisely, to each evaluator). The new proposal of statistical approach presented in this study may be a contribution to the continuity of the exploration of studies of auditory perception of vocal quality<sup>31</sup>.

We emphasized that the analysis of an agreement inter-evaluator and intra-evaluator is a fundamental factor to provide reliability to the perceptual evaluation of voice<sup>32</sup>. Such an agreement may increase according to the experience and training in analyses of vocal changes, and is influenced by factors such as fatigue, attention lapses and misunderstandings during the evaluation<sup>27,33,34</sup>, in addition to the very conception and structuring of the perception experiment.

At this point, we may state that the data collected reinforce that time of training in the method is fundamental. Inter-evaluator data, in which there were some discrepancies, reinforce that the lengthening of the training period or, more precisely, the constant updating and the continuous work with the evaluators becomes essential in order to create a cohesively qualified group to conduct vocal quality analyses, which, although it may be considered subjective, that is, without an objective and extrinsic standard, may be replicated by training. Another point to be taken for the design of future studies in this subject refers to a higher number of the group of evaluators.

The findings reinforce the multidimensional character of vocal quality and the complexity involved in the perceptive judgments of this phenomenon, as well as the demand for training and perception experiments to select evaluators.

The issue of voice multidimensionality lies in the fact that vocal quality emerges from a combination of actions, so that it is not possible to analyze vocal quality based on only one parameter. The VPAS script is presented as an alternative to address phonatory, muscular tension and supralaryngeal activity aspects. To do so, it requires, as in other modalities of analysis scripts, training and familiarity in the use of the instrument. Such a situation makes researchers adopt

the mentioned script in order to adopt several steps of composition and selection of their examiners<sup>12,30,35-37</sup>. With a proper phonetic training, the evaluator becomes able to evaluate the prominent sound quality in the speech of an individual. This was the path taken, by which the adoption of references from Phonological Sciences provided conditions for detailing events related to vocal quality in a group of teachers with voice disorders.

## CONCLUSION

Based on the perceptual analysis of the four congruent evaluators, the mean vocal quality profile of the group (female teachers of the public education network with voice disorders and/or laryngeal changes), was studied. The most frequent adjustments in this group, in a descending order, were adjustments to the laryngeal hyperfunction, rough voice, elevated larynx, vocal tract hyperfunction, closed mandible, pharyngeal constriction, raised tongue body and air leak.

As for vocal dynamics, in a descending order, the following aspects were seen: inadequate respiratory support, decreased variability of pitch, usual high pitch, high habitual loudness, fast elocution rate and decreased loudness variability.

The proposal of a methodological approach to evaluate the performance of a group of evaluators for voice quality assessment was adequate, since the proposed set of tests allowed defining evaluators whose judgments were based on phonetic principles, as well as designing the mean vocal quality profile of a group of voiced samples.

## REFERENCES

1. De Bodt MS, Van De Heyning PH, Wuyts FL, Lambrechts L. The perceptual evaluation of voice disorders. *Acta Otorhinolaryngol. Belg.* 1996;50(4):283-91.
2. Simões-Zenari M, Latorre MRDO. Changes in behavior associated to the use of voice after a speech therapy intervention with professionals of child day care centers. *Pró-Fono R Atual Cient.* 2008;20(1):61-6.
3. Lima-Silva MFB, Ferreira LP, Oliveira IB, Andrada e Silva MA, Ghirardi ACAM. Voice disorders in teachers: self-report, auditory-perceptive assessment of voice and vocal fold assessment. *Rev Soc Bras Fonoaudiol.* 2012;17(4):391-7.
4. Blaustein S, Bar A. Reliability of perceptual voice assessment. *J Commun Disord.* 1983;16(2):157-61.
5. Webb A, Carding PN, Deary IJ, Mackenzie K, Steen N, Wilson JA. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol.* 2004;261(8):429-34.
6. Silva RSA, Simões-Zenari M, Nemr NK. Impact of auditory training for perceptual assessment of voice executed by undergraduate students in Speech Language Pathology. *J Soc Bras Fonoaudiol.* 2012;24(1):19-25.
7. Kreiman J, Gerratt BR, Precoda K. Listener experience and perception of voice quality. *Speech Hear Res.* 1990;33(1):103-15.
8. Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception. *J Speech Hear Res.* 1992;35(3):512-20.
9. Kreiman J, Gerratt B, Kempster G, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and framework for future research. *J Speech Hear Res.* 1993;36(1):21-40.
10. Behlau M, Hogikyan ND, Gasparini G. Quality of life and voice: study of a Brazilian population using the voice-related quality of life measure. *Folia Phoniatr Logop.* 2007;59(6):286-96.
11. Camargo ZA, Madureira S. Avaliação vocal sob a perspectiva fonética: investigação preliminar. *Distúrb Comun.* 2008b;20(1):77-96.
12. Kreiman J, Gerratt B. Measuring voice quality. In: Kent R, Ball M. (Org). *Voice quality measurement.* San Diego: Singular Publishing; 2000. p. 73-99.
13. Valentim AF, Côrtes MG, Gama AC. Spectrographic analysis of the voice: effect of visual training on the reliability of evaluation. *Rev Soc Bras Fonoaudiol.* 2010;15(3):335-42.
14. Laver J. Phonetic evaluation of voice quality. In: Kent MJ, Martin JB (org). *Voice Quality Measurement.* San Diego: Singular, 2000.
15. Benninguer MS. Quality of the voice literature: what is there and what is missing. *J Voice.* 2011;25(6):647-52.
16. Pessoa-Almeida NA, Novaes BCC, Camargo Z. Dados perceptivo-auditivos e acústicos como indicadores prosódicos da fala em criança com deficiência auditiva. In: Camargo ZA (org). *Fonética Clínica: vinte anos de LIAAC.* São Paulo: Pulso, 2016. p.81-104.
17. Bailly G. Learning to speak. Sensori-motor control of speech movements. *Speech Communication.* 1997;22(1):251-67.

18. Meier RP, McGarvin G, Zakia RA, Willerman R. Silent mandibular oscillations in vocal babbling. *Phonetica*. 1997;54(4):153-71.
19. Buder EH, Choma LB, Oller DK, Robinson RB. Vibratory regime classification of infant phonation. *J Voice*. 2007;22(5):553-64.
20. Lima MFB, Camargo ZA, Ferreira LP, Madureira S. Qualidade vocal e formantes das vogais de falantes adultos da cidade de João Pessoa. *Rev. CEFAC*. 2007;9(1):99-109.
21. Camargo ZA, Madureira S. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In: Fourth Conference on Speech Prosody; 2008; Campinas, BR. São Paulo: Capes, Fapesp, CNPq, 2008a, v.1: 14.
22. Mackenzie-Beck J. Perceptual analysis of voice quality: the place of vocal profile analysis. In: Hardcastle WJ, Mackenzie-Beck J (orgs). *A figure of speech: a festschrift for John Laver*. Lawrence Erlbrum Associates: Mahwah; 2005. p. 285-322.
23. Camargo ZA, Madureira S, Tsuji DH. Analysis of dysphonic voices based on the interpretation of acoustic, physiological and perceptual data. In: 16th International Seminar on Speech Production Proceedings; 2003; Sidney, Austrália. Sidney: Speech Production; 2003.
24. Lima MFB, Madureira S, Camargo ZA. Avaliação fonética de qualidade vocal em diferentes estilos de fala (semi-espontânea e leitura). In: Anais do 17<sup>a</sup> Congresso Brasileiro de Fonoaudiologia e 1<sup>o</sup> Congresso Ibero-Americano de Fonoaudiologia; 2009; Bahia, Brasil. São Paulo: Sociedade Brasileira de Fonoaudiologia; 2009. p. 1814.
25. Guirardello EB. Adaptação cultural e validação do instrumento Demandas de Atenção Dirigida. *Rev Esc Enferm USP*. 2005;39(1):77-84.
26. Köhle J, Camargo Z, Nembr K. Análise perceptivo-auditiva da qualidade vocal de indivíduos submetidos a laringectomias parciais verticais pela auto-avaliação dos indivíduos e pela avaliação fonoaudiológica. *Rev. CEFAC*. 2004;6(1):67-76.
27. Bele IV. Reliability in perceptual analysis of voice quality. *J Voice*. 2005;19(4):555-73.
28. Sellars C, Stanton AE, McConnachie A, Dunnet CP, Chapman LM, Bucknall CE et al. Reliability of perceptions of voice quality: evidence from a problem asthma clinic population. *J Laryngol Otol*. 2009;123(7):755-63.
29. Oates J. Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. *Folia Phoniatr Logop*. 2009;61(1):49-56.
30. Kreiman J, Gerratt B. The perceptual structure of pathologic voice quality. *Journal of the Acoustical Society of America*. 1996;100(3):1787-95.
31. Kreiman J, Sidtis D. *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Wiley-Blackwell: Malden, 2011.
32. Gama AC, Santos LL, Sanches NA, Côrtes MG, Bassi IB. Studying the effect of spectrogram visual support of in the auditory-perceptive voice evaluation reliability. *Rev. CEFAC*. 2011;13(2):314-21.
33. Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice*. 2006;20(4):527-44.
34. Carding PN, Wilson JA, Mackenzie K, Deary IJ. Measuring voice outcomes: state of the science review. *J Laryngol Otol*. 2009;123(8):823-9.
35. Kreiman J. Listening to voices: theory and practice in voice perception research. In: Johnson K, Mullennix JW (orgs). *Talker variability in speech processing*. San Diego; 1997. p. 85-108.
36. Kreiman J, Gerratt B. Categorical judgments of vocal quality. Presented at the 134th Meeting of the Acoustical Society of America. San Diego; 1997.
37. Camargo ZA, Madureira S. The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long term f0 and intensity measures. In: Proceedings of the third ISCA Tutorial and research workshop on Experimental Linguistics; 2010; Athens, Greece. Athens: International Speech Communication Association; 2010: 33-6.