
Interference of background experience on agreement of perceptive-auditory analysis of neutral and dysphonic voices

Interferência do tempo de experiência na concordância da análise perceptivo-auditiva de vozes neutras e disfônicas

La Interferencia del tiempo de experiencia en la concordancia del análisis perceptivo auditivo de la voz neutrales y disfónicas

Samira Byron de Oliveira*
Ana Cristina Côrtes Gama*
Cristiane Ribeiro Chaves*

Abstract

Introduction: Voice has a multidimensional evaluation, and therefore must be analyzed in its different aspects. Currently, in clinical practice, the most commonly used is the perceptive-auditory analysis, which is the golden evaluation of vocal quality. It is a subjective test that bases itself mainly on the impression of the evaluator on the patient's voice, causing the results to be heavily influenced by the evaluator's background experience in this type of evaluation. **Objective:** to analyze the interference of the evaluator's background experience on the results of the perceptive-auditory voice analysis. **Methods:** transversal analytic observational study in which six speech therapists with different levels of expertise on the test evaluated 55 emissions in connected speech and sustained vowel, utilizing the GRBASI scale. In order to analyze the agreement and discordance between the evaluators the AC1 statistical calculation was utilized. **Results:** The background experience of the speech therapists does not interfere in the agreement of the intra-rater evaluation on both speech tasks. The inter-rater agreement is greater between individuals in the most experienced group on both tasks, although regarding connected speech

*Universidade Federal de Minas Gerais – UFMG – Belo Horizonte-MG - Brazil.

Authors' contributions: SBO Development of text and revision. ACCG General orientation of the steps for the elaboration and execution of the manuscript. CRC Collection of the data present on this paper and revision.

Correspondence address: Samira Byron de Oliveira. **E-mail:** samirabyron@gmail.com

Received: 04/05/2016

Accepted: 15/07/2016

a greater agreement was observed between members of the least experienced group regarding asthenic and tension parameters than in the most experienced group. **Conclusion:** the background experience of speech therapists does not interfere in the intra-rater agreement, but it impacts positively on the inter-rater agreement, suggesting that the background experience in this analysis tends to standardize the auditory judgment process of dysphonic voices.

Keywords: Voice; Dysphonia; Auditory Perception; Speech, Language and Hearing Sciences.

Resumo

Introdução: A voz tem uma avaliação multidimensional, e por isso deve ser analisada em suas diferentes perspectivas. Atualmente na prática clínica, a mais utilizada é a análise perceptivo-auditiva, que é a avaliação ouro da qualidade vocal. É um teste subjetivo que se baseia especialmente na impressão do avaliador sobre a voz do paciente, sofrendo influência do tempo de experiência do mesmo neste tipo de avaliação. **Objetivo:** analisar a interferência do tempo de experiência do avaliador na concordância da análise perceptivo-auditiva da voz. **Método:** estudo observacional analítico transversal no qual seis fonoaudiólogos com diferentes anos de experiência avaliaram 55 emissões na fala encadeada e vogal sustentada, utilizando a escala GRBASI. Para análise da concordância intra-avaliador e interavaliador foi utilizado o cálculo estatístico AC1. **Resultados:** A experiência dos fonoaudiólogos não interfere na concordância da avaliação intra-avaliadores nas duas tarefas de fala. A concordância interavaliador é maior para o grupo mais experiente nas duas tarefas, entretanto na fala encadeada verificou-se que os inexperientes apresentaram concordância maior nos parâmetros A e S em relação aos experientes. **Conclusão:** O tempo de experiência dos fonoaudiólogos não interfere na concordância intra-avaliador, mas impacta positivamente na concordância interavaliadores, sugerindo que a experiência nesta análise tende a uniformizar o processo de julgamento auditivo de vozes disfônicas.

Palavras-chave: Voz; Disfonia; Percepção Auditiva; Fonoaudiologia.

Resumo

Introducción: La voz tiene una evaluación multidimensional, y por eso debe ser analizada en sus diferentes perspectivas. Actualmente en la práctica clínica, la más utilizada es el análisis perceptivo-auditivo, que es la evaluación oro de la cualidad vocal. Es un test subjetivo que se basa especialmente en la impresión del evaluador sobre la voz del paciente, siendo influenciado por el tiempo de-experiencia del evaluador. **Objetivo:** analizar la interferencia del tiempo de la experiencia del evaluador en la concordancia del análisis perceptivo-auditiva de la voz. **Método:** estudio observacional analítico transversal en el que seis fonoaudiólogos, con diferentes años de experiencia, evaluaron 55 emisiones en el habla encadenada y vocal sostenida, utilizando la escala GRBASI. Para el análisis de la concordancia del evaluador y entre evaluadores fue utilizado el cálculo estadístico AC1. **Resultados:** La experiencia de los fonoaudiólogos no interfiere en la concordancia de la evaluación de los evaluadores en las dos tareas de habla. La concordancia entre evaluadores es mayor para el grupo más experto en las dos tareas, sin embargo en el habla encadenada se encontró que los inexpertos presentaron concordancia mayor en los parámetros astenia y tensión en relación a los expertos. **Conclusión:** El tiempo de experiencia de los fonoaudiólogos no interfiere en la concordancia del evaluador, pero impacta positivamente en la concordancia entre evaluadores, sugiriendo que la experiencia en este análisis tiende a uniformizar el proceso de juzgamiento auditivo de voces disfónicas.

Palabras claves: Voz; Disfonia; Percepción Auditiva; Fonoaudiología.

Introduction

Voice is the main means of human communication, and its production relates to anatomical, physiological, emotional, organic, environmental and behavioral characteristics that correlate in order to generate vocal emission¹. Voice manifests through a complex muscular process, and when it occurs harmonically, the speaker emits a clean and comfortable vocal production. However, when these emissions occur with exertion and presence of noise, a case of dysphonia is observed, meaning any difficulty or alteration in vocal emission that occurs due to muscular maladjustment, impeding natural voice production, being related to various etiologies².

Voice has a multidimensional evaluation, and therefore must be analyzed in its different facets through the perceptive-auditory, acoustic and aerodynamic evaluations, as well as the patients self-evaluation^{3,4}. Currently, in clinical practice, the most commonly employed is the perceptive-auditory evaluation, being a classic vocal quality evaluation that is performed non-invasively, with a low cost and rapid execution⁵. It is considered a standard exam in vocal clinic due to its capacity to evaluate vocal characteristics that cannot be detected via instrumental analysis⁶. It is a subjective test based mainly on the evaluator's impressions on the patient's voice, being heavily influenced by the evaluator's background experience with this particular type of evaluation⁷.

The literary sources emphasize that perceptive-auditory analysis suffers from influences external to the evaluator that may alter its results⁸, the most relevant ones being the form in which the vocal stimuli was presented, the type of auditory-perceptive scale being used and the speech task^{1,9}. In regards to the evaluator's intrinsic performance, many aspects may interfere in the results of the perceptive-auditory analysis⁴, such as: background professional experience^{5,10} and qualification, auditory training¹¹, state of attention during the evaluation¹ and native language¹². Literature also emphasizes that more experienced professionals present more consistent results⁹, and connected speech tasks create greater reliability than sustained vowel speech tasks⁹.

Perceptive-auditory analysis depends on the individual perception of each evaluator, and that auditory judgment is influenced by professional and

personal background experience¹³. With adequate auditory training inexperienced individuals can greatly improve their auditory capability, consequentially improving the results of their evaluation⁸.

The goal of the current study is to analyze the interference of the evaluator's background experience on the agreement of the perceptive-auditory analysis.

Methods

Transversal analytic observational study, approved by the Comitê de Ética em Pesquisa (Research Ethics Committee), under form number 16633113.3.0000.5149, utilizing vocal samples from a database of a Speech Therapy Clinic on an Educational Institution.

For this sample, 46 connected speech (counting from one to 10) and sustained /a/ vowel voices were selected, from Brazilians of the two genders, 23 males and 23 females aged above 18 years old, with neutral voices or with different levels of vocal deviation.

The sample calculation was defined for the sample size and number of evaluators, considering a sampling error of 10%, significance level of 5% and a power of 80%. The samples from the database were randomized until the 46 necessary voices were obtained. For the intra-rater agreement analysis, 20% of the samples were doubled randomly, resulting in 55 voices total; 55 sustained /a/ vowel and 55 connected speech (counting from one to 10) emissions were utilized for the vocal analysis.

In the auditory-perceptive analysis, the Japanese GRBASI scale (Hirano, 1991 in conjunction with Dejonckere, Remacle & Fresnel-Elbaz, 1996) was utilized. It is comprised of six standards: (G) general level of dysphonia, (R) roughness, (B) soprosity, (A) asthenia, (S) tension, and (I) instability. For each parameter, a degree of deviation is defined on a 4 point Likert scale, being 0 unaltered, 1 low deviation, 2 moderate deviation and 3 intense deviation.

Six evaluators were observed in the study, speech therapists aged between 25 and 50 years old, divided in three groups according to their self proclaimed auditory-perceptive analysis background. In order to define the degree of the evaluator's experience, they were asked to perform a self-evaluation, by choosing the concept that best suited their current state regarding back-

ground experience in auditory perceptive analysis of dysphonic voices: 1.Beginner: evaluators who had just started performing the tests, having only the basic notions; 2.Experienced: evaluators with considerable experience; 3.Senior: a professional with great experience in a certain occupation¹⁴. Based on the self-evaluations of the speech therapists, three evaluator groups were created, and their background experience was defined.

The beginner evaluator group (GI) was represented by two female speech therapists with up to two years of auditory training (average of 1,75 years). The experienced evaluator group (GE) was comprised of two female speech therapists with up to 10 years of auditory training (average of 6,5 years). The senior evaluator group (GS) was composed of two female speech therapists with up to 20 years of auditory training (average of 19 years). After signing the Term of Consent, the evaluators received the voices stored on Compact Disk, along with the GRBASI protocols in print, in order to evaluate the sustained vowel and connected

speech tasks. The auditory-perceptive analysis was performed independently and individually by each evaluator using a model RP-DJ120PP-K-Panasonic headphone. The evaluators could repeat the voices as many times as necessary to complete the auditory-perceptive evaluation.

In order to analyze intra-rater and inter-rater agreement the AC1 statistic was used, with the following classifying criteria: below zero – no agreement; 0 to 0,20 – small agreement; 0,21 to 0,40 – weak agreement; 0,41 to 0,60 – moderate agreement; 0,61 to 0,80 – good agreement; 0,81 to 1,00 – almost perfect agreement¹⁵. Software R was the computer program utilized, in the S language, S-Plus.

Results

The background experience of the speech therapists does not interfere in the agreement of the intra-rater evaluation on sustained vowel and connected speech tasks.

Table 1. Results of intra-rater agreement on the three observed groups

Evaluators	Parameters	Sustained Vowel	Connected Speech
GI	G	66,56 [28,75;104,37]	65,66 [24,60;106,71]
	R	44,43 [6,76;82,10]	59,83 [16,89;102,76]
	B	43,92 [6,88;80,95]	50,48 [9,15;91,81]
	A	55,46 [14,83;96,09]	74,68 [26,78;122,59]
	S	43,67 [6,41;80,92]	82,73 [38,91;126,55]
	I	44,13 [7,29;80,98]	54,31 [10,29;98,33]
GE	G	60,83 [23,61;98,06]	58,09 [16,90;99,28]
	R	59,86 [22,93;96,78]	54,79 [13,23;96,35]
	B	60,62 [23,11;98,12]	60,02 [17,34;102,69]
	A	54,80 [14,89;94,71]	69,72 [25,06;114,38]
	S	41,96 [4,29;79,63]	63,28 [18,52;108,04]
	I	45,44 [8,68;82,20]	62,04 [17,80;106,28]
GS	G	69,30 [31,91;106,69]	73,54 [30,20;116,88]
	R	54,99 [18,07;91,92]	72,91 [28,25;117,57]
	B	56,21 [19,28;93,13]	73,97 [28,16;119,79]
	A	69,40 [27,98;110,82]	76,36 [28,43;124,29]
	S	59,62 [20,88;98,37]	81,17 [32,45;129,89]
	I	56,73 [19,00;94,46]	87,67 [37,26;138,07]

The AC1 agreement coefficient was utilized for the statistical analysis.
Legend: GI = Beginners; GE = Medium experience; GS = Sênior¹⁴

The inter-rater agreement is greater between more experienced evaluators on the auditory-perceptive evaluation. On the sustained vowel task the G, A, S and I parameters presented greater agreement on the most experienced group (GS), and GE presented better results on parameters G and I when compared to the least experienced group

(GI). The R and B parameters didn't present any variations related to evaluator experience.

On the connected speech task, the most experienced group (GS) presented greater agreement on the analysis of the R, B, A, S and I parameters. The G parameter didn't present any variations related to evaluator experience. When comparing the medium

experience (GE) and least experienced (GI) groups, it was observed that greater experience impacted positively on the agreement of parameter R and negatively on parameters A and S.

Table 2. Results of inter-rater agreement on the three observed groups

Evaluators	Parameters	Sustained Vowel	Connected Speech
GI	G	57.00 [48.84 ; 59.11]	66.63 [58.39 ; 64.91]
	R	43.75 [35.51 ; 51.65]	32.30 [23.98 ; 44.22]
	B	58.30 [50.09 ; 60.18]	42.01 [33.77 ; 49.67]
	A	63.76 [56.12 ; 65.52]	57.94 [49.93 ; 61.65]
	S	40.35 [32.11 ; 49.52]	65.37 [57.03 ; 64.92]
	I	26.16 [18.01 ; 39.93]	31.93 [23.60 ; 44.22]
GE	G	71.96 [63.72 ; 80.20]	68.32 [60.07 ; 76.58]
	R	58.36 [50.15 ; 66.56]	64.94 [56.71 ; 73.18]
	B	59.08 [50.82 ; 67.33]	7.68 [39.48 ; 55.87]
	A	56.29 [48.12 ; 64.47]	40.02 [31.75 ; 48.29]
	S	49.05 [40.81 ; 57.29]	38.24 [29.90 ; 46.58]
	I	48.43 [40.27 ; 56.59]	34.76 [26.49 ; 43.03]
GS	G	69.10 [60.95 ; 77.24]	66.59 [58.34 ; 74.84]
	R	53.60 [45.45 ; 61.76]	59.08 [50.78 ; 67.39]
	B	60.24 [52.05 ; 68.42]	70.21 [61.97 ; 78.44]
	A	73.56 [65.76 ; 81.36]	73.34 [65.28 ; 81.41]
	S	76.30 [68.13 ; 84.47]	72.35 [64.91 ; 79.80]
	I	68.52 [60.29 ; 76.76]	73.12 [65.16 ; 81.08]

The AC1 agreement coefficient was utilized for the statistical analysis.
Legend: GI = Beginners; GE = Medium experience; GS = Sênior¹⁴

Discussion

The auditory-perceptive evaluation is a sovereign procedure in the global scenario of vocal quality analysis^{8,9}, in spite of being partially based on intuition, subjective, affected by the evaluator's background experience and state of attention, the scale utilized^{6,9} and comparisons with other voices heard throughout the years¹⁶. These aspects will deeply affect the outcome of the evaluation.

The intra-rater agreement is analyzed statistically by the repeated response of the same evaluator, meaning how often a speech therapist gave the same voices the same results on the evaluation, in different occasions¹⁷. The inter-rater agreement is the comparison of the response of two or more evaluators¹⁷. The results of this research indicate that the evaluator's background experience does not interfere in the intra-rater agreement, only improving the inter-rater agreement. It is valid to infer that increased background experience in auditory-perceptive evaluation of voices tends to cause greater consistency in the analysis's results, since this type of analysis is impressionistic, in which the evaluators compares the sample voices with an internal reference system², formed via their own auditory experiences. The results of this re-

search allow us to conclude that more experienced evaluators tend to judge voices with a greater agreement than their vocal clinic peers. Therefore, the auditory perceptive analysis suffers reduced individual interferences when executed by more experienced evaluators.

A study which compared inter-rater agreement in auditory-perceptive evaluations of vocal quality performed by speech therapists specialized in voice, singing masters and inexperienced listeners, resulted in a high agreement between speech therapists, a medium agreement between singing masters and a low agreement between inexperienced listeners¹⁸. These results confirm the current studies' findings, highlighting that the background experience in auditory-perceptive analysis influenced the results.

Literature indicates that a greater background experience has a positive influence in the reliability of the auditory-perceptive analysis, and has been imperative in the documenting of vocal disorders^{5,6,9,13,18,19,20,21}. A study that evaluated the influence of knowledge on laryngeal diagnosis on the results of auditory-perceptive analysis when performed by speech therapists with varying degrees of professional background experience and students on their last year of their speech therapy course, concluded that the inexperienced evalu-

ators presented inconsistent results and had low agreement on the evaluation of voices with low and medium degrees of dysphonia²⁰.

Research also indicates that individuals who underwent auditory training presented greater agreement in the analysis of vocal parameters^{5,19}, therefore, background experience and training are factors that have a positive impact on the agreement between evaluators on the auditory-perceptive evaluation.

In regards to the speech task employed on the auditory-perceptive analysis, it is noted that the sustained vowel is easier to produce, but tends to suffer deviations on evaluation, while the connected speech is a more natural emission that promotes higher reliability on evaluation^{9,22}.

On both speech tasks, the GS group presented greater agreement. The most concordant parameters in the inter-rater evaluation for the sustained vowel task were: G, A, S, I and for the connected speech task, R, B, A, S and I were the most concordant. On groups GI and GE, background experience was observed to be a detrimental factor on evaluation of parameters A and S, being the least concordant. These results allow us to infer that since parameters A and S are less frequent on vocal clinic, the evaluator's background experience didn't contribute on the agreement of the analysis²³.

Studies that analyze the results of auditory-perceptive evaluation of voices are still few, and many aspects must be considered in the analysis of the results of this evaluation. It must be mentioned that this study analyzed the degree of intra and inter-rater agreement, a standard that analyzes the frequency with which the answers of evaluators attribute the same classification. Reliability, on the other hand, is the measurement of consistency between evaluators, not depending on the total value of each evaluator's classification²⁴. The evaluations of the different evaluators may be concordant, but that does not determine if they are being specific in their answers causing greater reliability. Reliability analysis, alongside the comparison of other vocal evaluation measurements is of great importance in order to greater advance the study of auditory-perceptive analysis.

The statistical analysis of this study was executed via the AC1 calculation, which is the best estimator for the evaluation of agreement, since it considers error rates, as opposed to other statistical calculations that do not consider it, meaning that

when the evaluator sets a certain value on a first evaluation and a different one on a second, the AC1 will take that difference into account, while other statistical calculations would not account for it. The AC1 also weights consistency, in case the evaluator gives the same answer on both evaluations, adding a positive weight to its answer¹⁵. Literature suggests that this particular statistical calculation is the most adequate for studies that analyze the degree of agreement of answers²⁵.

Other studies on the impact of the evaluator's background experience on agreement of auditory-perceptive analysis exist^{5, 6, 11, 13, 18, 19, 20, 21, 26}, but apply different methodologies. Future studies are important in order to validate the findings of this study.

Conclusion

Background experience of speech-therapists does not interfere on intra-rater agreement, but impacts inter-rater agreement positively, suggesting that experience in this particular type of analysis tends to improve consistency of results from the auditory evaluation of dysphonic voices.

All the auditory-perceptive parameters presented greater agreement in evaluators with greater background experience, with the exception of parameters A and S, in which the least experienced evaluators presented greater agreement than the more experienced ones.

References

1. Lopes LW, Cavalcante DP, Costa PO. Severity of voice disorders: integration of perceptual and acoustic data in dysphonic patients. *Codas*. 2014; 26(5): 382-8.
2. Behlau M, Madazio G, Feijó D, Pontes P. Avaliação da voz. In: Behlau M. *Voz: o livro do especialista I*. Rio de Janeiro: Revinter; 2001.p.85-180.
3. Kreiman J, Vanlancker-Sidtis D, Gerratt BR. Defining and measuring voice quality. Geneva: Voqual'03; 2003.p.115-20.
4. Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx*. 2015; 3(42):183-8.
5. Costa FP, Yamasaki R, Behlau M. Influência da escuta contextualizada na percepção da intensidade do desvio vocal. *Audiol Commun Res*. 2014; 19(1): 69-74.
6. Oates J. Auditory-Perceptual Evaluation of disordered voice quality: pros, cons and future directions. *Folia Phoniatr Logop*. 2009; 61(1): 49-56.

7. Nemr K, Simões-Zenari M, Cordeiro GF, Tsuji D, Ogawa AI, Ubrig MT, et al. GRBAS and Cape-V Scales: high reliability and consensus when applied at different times. *J Voice*. 2012; 26(6): 812-22.
8. Silva RS, Simões-Zenari M, Nemr NK. Impact of auditory training for perceptual assessment of voice executed by undergraduate students in SpeechLanguage Pathology. *Rev Soc Bras Fonoaudiol*. 2012; 24(1): 19-25.
9. Bele IV - Reliability in perceptual analysis of voice quality. *J Voice*. 2005; 19(4): 555-73.
10. Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception. *J Speech Hear Res*. 1992; 35: 512-20.
11. Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. *J Speech Lang Hear Res*. 2011; 54: 430-47.
12. Ghio A, Weisz F, Baracca G, Cantarella G, Robert D, Woisard V, et al. Is the perception of voice quality language-dependant? A comparison of French and Italian listeners and dysphonic speakers. Florence, Italy: INTERSPEECH; 2011. p.525-8.
13. Iwarsson J, Reinholt Petersen N. Effects of Consensus Training on the reliability of Auditory Perceptual ratings of voice quality. *J Voice*. 2012; 26(3): 304-12.
14. Ferreira ABH. Míni Aurélio: O Dicionário da língua portuguesa. 7ª ed. Rio de Janeiro: Positivo; 2009.
15. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008; 61(1): 29-48.
16. Ghio A, Dufour S, Wengler A, Pouchoulin G, Revis J, Giovanni A. Perceptual Evaluation of Dysphonic Voices: Can a Training Protocol Lead to the Development of Perceptual Categories?. *J Voice*. 2014; 29: 304-11.
17. Gouveia VHO, Araújo AGF, Maciel SS, Ferreira JJA, Santos HH. Confiabilidade das medidas inter e intra-avaliadores com goniômetro universal e flexímetro. *Fisioter Pesq*. 2014; 21(3): 229-35.
18. Sofranko JL, Prosek RA. The effect of experience on classification of voice quality. *J Voice*. 2012; 26(3): 299-303.
19. Kreiman J, Gerratt BR. Comparing two methods for reducing variability in voice quality measurements. *J Speech Lang Hear Res*. 2011; 54(3): 803-12.
20. Eadie T, Sroka A, Wright DR, Merati A. Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia?. *J Voice*. 2011; 25(4): 420-9.
21. Eadie TL, Van Boven L, Stubbs K, Giannini E. The effect of musical background on judgments of dysphonia. *J Voice*. 2010; 24(1): 93-101.
22. Law T, Kim JH, Lee KY, Eric C, Tang EC, Joffe HL, Andrew C, et al. Comparison of Rater's Reliability on Perceptual Evaluation of Different Types of Voice Sample. *J Voice*. 2012; 26(5): 535-676.
23. De Bodt MS, Wuyts FL, Van de Heyning PH, Croux C. Test-retest study of the GRBASI scales: influence of experience and professional background on perceptual ratings of voice quality. *J voice*. 1997; 11: 74-80.
24. Matos DAS. Confiabilidade e concordância entre juízes: Aplicações na área educacional. *Est. Aval. Educ*. 2014; 25(59): 298-324.
25. Lu FL, Matteson S. Speech Tasks and Interrater Reliability. *J Voice*. 2014; 28(6): 725-32.
26. Eadie TL, Kapsner M, Rosenzweig J, Waugh P, Hillel A, Merati A. The role of experience on judgments of dysphonia. *J Voice*. 2010; 24(5): 564-73.