

Rafaeli Higa Scarmagnani¹
 Adriana Cristina de Almeida Santos
 Furlan de Oliveira¹
 Ana Paula Fukushiro^{1,2}
 Manoel Henrique Salgado³
 Inge Elly Kiemle Trindade^{1,2}
 Renata Paciello Yamashita¹

Keywords

Speech
 Cleft Palate
 Velopharyngeal Insufficiency
 Reproducibility of Results
 Rhinomanometry

Descritores

Fala
 Fissura Palatina
 Insuficiência Velofaríngea
 Reprodutibilidade dos Testes
 Rinomanometria

Correspondence address:

Rafaeli Higa Scarmagnani
 Rua Silvio Marchione, 3-20, Bauru (SP),
 Brasil, CEP: 17012-900.
 E-mail: rafaelihiga@usp.br

Received: 04/28/2014

Accepted: 07/28/2014

Impact of inter-judge agreement on perceptual judgment of nasality

O impacto da concordância entre avaliadores no julgamento perceptivo da nasalidade da fala

ABSTRACT

Purpose: To investigate the effect of perceptual inter-judge agreement of hypernasality on velopharyngeal (VP) closure prediction. **Methods:** Two logistic regression models were developed aiming to verify the possibility of predicting the VP closure using the following characteristics: rating of VP closure (adequate, borderline, inadequate), determined by the pressure-flow technique, degree of hypernasality (absent, mild, moderate, severe), and the presence/absence of nasal air emission and nasal rustle determined perceptually by three experienced speech language pathologists. In the first model, 100 speech samples with a moderate agreement rate of hypernasality (kappa coefficient: 0.41) were used. In the second model, 43 speech samples with a perfect agreement among judges were included. The χ^2 -test was used to compare the models ($p \leq 0.05$). **Results:** In the first model, 65 of the 100 samples were rated in the correct VP closure category, with 42 adequate and 23 inadequate. The borderline VP closure was not predicted. The second model rated 31 of the 43 samples in the correct category, with 21 adequate VP closure, 5 in the borderline VP closure, and 5 inadequate. There was no difference ($p=0.526$) between the two models. However, the second model showed a higher proportion of accuracy (7%) than the first one, and it has also predicted the borderline VP closure. **Conclusion:** These results showed the importance of high index of inter-judge agreement when using subjective parameters of speech evaluation, especially when compared to an instrumental evaluation. This suggests the need for strategies for training and calibration of judges in the perceptual judgment to improve the reliability of auditory-perceptual assessment.

RESUMO

Objetivo: Investigar o efeito da concordância entre diferentes avaliadores no julgamento perceptivo da hipernasalidade na previsão do fechamento velofaríngeo (FVF). **Métodos:** Dois modelos de regressão logística foram desenvolvidos com o objetivo de se prever o FVF utilizando as seguintes características: FVF (adequado, marginal e inadequado), determinado pela técnica fluxo-pressão, grau da hipernasalidade (ausente, leve, moderado, grave) e a presença/ausência de emissão de ar nasal e ronco nasal, determinados perceptivamente por três fonoaudiólogas experientes. O primeiro modelo foi composto de 100 amostras de fala com índice de concordância moderado para a hipernasalidade (coeficiente kappa: 0,41) e o segundo, de 43 amostras com concordância total entre os avaliadores. O teste do χ^2 foi utilizado para comparar os modelos ($p \leq 0,05$). **Resultados:** No primeiro modelo, 65 das 100 amostras foram classificadas na categoria correta de FVF, sendo 42 adequado e 23 inadequado. O FVF marginal não foi previsto. O segundo modelo classificou 31 das 43 amostras na categoria correta, sendo 21 FVF adequado, 5 marginal e 5 inadequado. Não houve diferença ($p=0,526$) entre os dois modelos. No entanto, o segundo mostrou proporção de acerto mais alta (7%) do que o primeiro modelo, além de ter conseguido prever o FVF marginal. **Conclusão:** Esses resultados mostram a importância da concordância total entre diferentes avaliadores quando se utilizam parâmetros subjetivos de avaliação da fala, especialmente quando comparados com a avaliação instrumental, e sugerem a necessidade de estratégias para o treinamento e calibração de avaliadores no julgamento perceptivo a fim de aumentar a confiabilidade da avaliação perceptivo-auditiva da fala.

Study carried out at the Laboratory of Physiology, Hospital for Rehabilitation of Craniofacial Anomalies, Universidade de São Paulo – USP – Bauru (SP) Brasil.

(1) Hospital for Rehabilitation of Craniofacial Anomalies, Universidade de São Paulo – USP – Bauru (SP), Brasil.

(2) School of Dentistry of Bauru, Universidade de São Paulo – USP – Bauru (SP) Brasil.

(3) Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP – Bauru, (SP), Brasil.

Financial support: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.

Conflict of interests: nothing to declare.

INTRODUCTION

The auditory-perceptual assessment of the speech is an important instrument for the velopharyngeal function, because the speech characteristics identified in this kind of assessment provide clues on the extent of the velopharyngeal failure⁽¹⁻³⁾. Among the speech symptoms directly related to the velopharyngeal dysfunction (VPD), there is the hypernasality, defined as a perceptive phenomenon and, therefore, a subjective one, fact that interferes in the reliability of the auditory-perceptual of the speech and that has been targeted by researches for many years, according to the review recently conducted⁽⁴⁾.

To ensure greater reliability of results in researches concerning the speech of individuals with the VPD, the literature recommends the use of different evaluators for the perceptual judgment of speech symptoms^(5,6). However, a high level of agreement may be difficult to be achieved, because of several variables.

Considering the possibility that the level of agreement among different evaluators on the judgment of speech symptoms may influence the prediction of velopharyngeal closure (VPC), the proposal of this study was defined.

This study investigated the effect of the agreement level between the evaluators in the perceptual judgment of hypernasality for the prediction of the VPC.

METHODS

This study was approved by the research ethics committee of the Hospital for Rehabilitation of Craniofacial Anomalies, University of São Paulo (document No. 360/2010 and No. 254/2012 SVAPEPE-CEP), and all the participants signed the informed consent form.

The inter-evaluator agreement on the judgment of hypernasality was obtained by the kappa coefficient.

Two models of logistic regression were developed with the objective of verifying possibilities of predicting the VPC using the following speech characteristics: VPC (adequate, borderline, and inadequate closure) determined by the pressure-flow technique, hypernasality degree (absent, mild, moderate, severe), and the presence or absence of nasal air emission and assessed nasal snoring, perceptually, by three speech language pathologists with 12 years of experience, on average, in the auditory-perceptual assessment of the speech in patients with cleft lip and palate. In the first model, 100 samples of speech with moderate agreement level to hypernasality (kappa coefficient: 0.41) were used. In the second model, 43 of the 100 samples of speech obtained full agreement among the evaluators. Later on, the proportion of correctness referring to each model was compared through the χ^2 -test at a significance level of 5%.

RESULTS

Table 1 shows the number of samples, provided according to the logistic model in relation to the real classification of VPC determined by the pressure-flow technique.

Table 1. Distribution of the 100 samples of speech predicting the velopharyngeal closure, from the hypernasality degree and the presence or absence of nasal air emission and nasal snoring

Real VPC	Prediction of VPC according to the model		Total
	Adequate VPC	Inadequate VPC	
Adequate VPC	42*	3	45
Borderline VPC	7	13	20
Inadequate VPC	12	23*	35
Total	61	39	100

*Model's correctness

Caption: VPC = velopharyngeal closure

According to this model, elaborated from the 100 samples that obtained moderate inter-judge agreement level in the perceptual judgment of the degree of hypernasality, 65 were predicted in the correct category. However, the borderline VPC was not predicted.

Table 2 shows many predictions according to the second logistic model in relation to the real VPC classification, using 43 samples that obtained perfect agreement level among the evaluators during the perceptual judgment of the hypernasality degree.

Table 2. Distribution of the 43 samples of speech predicting the velopharyngeal closure, in relation to the hypernasality, nasal air emission, and nasal snoring

Real VPC	Prediction of VPC according to the model			Total
	Adequate VPC	Borderline VPC	Inadequate VPC	
Adequate VPC	21*	1	0	22
Borderline VPC	4	5*	2	11
Inadequate VPC	2	3	5*	10
Total	27	9	7	43

*Model's correctness

Caption: VPC = velopharyngeal closure

According to this model, 31 of the 43 samples were classified correctly. This model predicted borderline VPC.

Although the comparison between the two models did not show any difference ($p=0.526$), it was verified that the second model, consisting of the full inter-judge agreement as for the degree of hypernasality, showed correctness proportion 7% higher than the first model, besides having predicted the borderline VPC.

DISCUSSION

There is a consensus in the literature that the auditory-perceptual assessment of speech is the means through which the speech language pathologist may identify the changes in speech, classifying its severity and, thus, defining the conduct and assessing the treatments performed⁽⁷⁻¹¹⁾. However, it is a subjective method and, therefore, subject to mistakes and influences of various factors. The main effect is related to the internal standards of each listener, that is, the individual

references of each evaluator that differ from one another. This led clinicians and researchers into searching strategies to enhance the perceptual assessment. The main change was the use of recordings of speech samples, which allowed the assessing of the speech by more than one listener. From there, several studies showed the importance of the presence of the analysis of speech symptoms by different evaluators and of obtaining agreement among them.

Specifically whit regard to hipernasality, a high level of agreement between evaluators is difficult to be obtained. This is because it is a perceptual phenomenon affected by many factors, for example, the listener's experience that determines the internal standards of each listener⁽¹²⁾. Although the evaluators consulted had 12 years of experience, on average, on speech assessment in the presence of cleft palate and, also, similar academic background as for the speech assessment process, having been trained by the same reference center, though differences related to the internal standards of each evaluator may be expected. It is speculated that this is the reason for obtaining moderate agreement levels as to the degree of hypernasality verified in the first model, a result also obtained in other studies⁽¹³⁻¹⁵⁾. This, probably, led to the reduced correctness percentage shown in this model, making the reliability of the VPC prediction questionable. Thus, a second logistic model was developed, based on the 43 samples whose inter-judge agreement level in the judging of hypernasality was perfect. In this case, there was an increase in the percentage of correct predictions of VPC based on the perceptive speech characteristics. Besides, in this second model, the borderline VPC was predicted, which did not occur in the previous model. This result was due to the full agreement of the evaluators as for the hypernasality degree, showing, therefore, the effect of agreement levels among different evaluators when subjective parameters of speech are being assessed, especially those concerning listener's perception.

CONCLUSION

The high level of inter-judge agreement as for the hypernasality degree positively affected the prediction of VPC. This means that, besides the perceptual judgment made by more than one listener, in the auditory-perceptual assessment of the speech characteristics, it is essential to use strategies that ensure high levels of agreement among them, to improve the reliability of results.

**RHS was responsible for the study's original idea, data collection, data analysis and the writing of the article; ACASFO collaborated with data collection and tabulation; APF collaborated in the collection and analysis of the data and the writing of the article; MHS participated in the statistical analysis of the data and the writing of the article; IEKT participated on the writing of the article; RPY was responsible for the project and outlining of the study and overall orientation on the stages of execution and elaboration of the manuscript.*

REFERENCES

1. Warren DW, Dalston RM, Mayo R. Hypernasality and velopharyngeal impairment. *Cleft Palate Craniofac J.* 1994;31(4):257-62.
2. Kummer AW, Briggs M, Lee L. The relationship between the characteristics of speech and velopharyngeal gap size. *Cleft Palate Craniofac J.* 2003;40(6):590-6.
3. Ma L, Shi B, Li Y, Zheng Q. Velopharyngeal function assessment in patients with cleft palate: perceptual speech assessment versus nasopharyngoscopy. *J Craniofac Surg.* 2013;24(4):1229-31.
4. Bressmann T, Sell D. Plus ça change: selected papers on speech research from the 1964 issue of the cleft palate journal. *Cleft Palate Craniofac J.* 2014;51(2):124-8.
5. Lohmander A, Olsson M. Methodology for perceptual assessment of speech in patients with cleft palate: a critical review of the literature. *Cleft Palate Craniofac J.* 2004;41(1):64-70.
6. Henningsson G, Kuehn DP, Sell D, Sweeney T, Trost-Cardamone JE, Whitehill TL. Speech parameters group. Universal parameters for reporting speech outcomes in individuals with cleft palate. *Cleft Palate Craniofac J.* 2008;45(1):2-17.
7. Dotevall H, Lohmander-Agerskov A, Ejnell H, Bake B. Perceptual evaluation of speech and velopharyngeal function in children with and without cleft palate and the relationship to nasal airflow patterns. *Cleft Palate Craniofac J.* 2002;39(4):409-24.
8. Whitehill TL, Lee ASY, Chun JC. Direct magnitude estimation and Interval scaling of hypernasality. *J Speech Lang Hear Res.* 2002;45(1):80-8.
9. Smith BE, Kuehn DP. Speech evaluation of velopharyngeal dysfunction. *J Craniofac Surg.* 2007;18(2):251-61.
10. Kummer AW. Resonance disorders and velopharyngeal dysfunction (VPD). In: Kummer AW. *Cleft palate and craniofacial anomalies: the effects on speech and resonance.* San Diego: Singular; 2008. p. 176-213.
11. Baylis AL, Munson B, Moller KT. Perceptions of audible nasal emission in speakers with cleft palate: a comparative study of listener judgments. *Cleft Palate Craniofac J.* 2011;48(4):399-411.
12. Lee A, Whitehill TL, Ciocca V. Effect of listener training on perceptual judgement of hypernasality. *Clin Linguist Phon.* 2009;23(5):319-34.
13. John A, Sell D, Sweeney T, Harding-Bell A, Williams A. The cleft audit protocol for speech-augmented: a validated and reliable measure for auditing cleft speech. *Cleft Palate Craniofac J.* 2006;43(3):272-88.
14. Scarmagnani RH. Correlação entre as dimensões do orifício velofaríngeo, hipernasalidade, emissão de ar nasal audível e ronco nasal em indivíduos com fissura de palato reparada [dissertação]. Bauru: Universidade de São Paulo; 2013.
15. Barbosa DA, Scarmagnani RH, Fukushima AP, Trindade IE, Yamashita RP. Surgical outcome of pharyngeal flap surgery and intravelar veloplasty on the velopharyngeal function. *CoDAS.* 2013;25(5):451-5.