

Flora Taube Manicardi¹ 
Jeniffer de Cássia Rillo Dutka² 
Thais Alves Guerra² 
Maria Inês Pegoraro-Krook² 
Eduardo Federighi Baisi Chagas³ 
Viviane Cristina de Castro Marino¹ 

Keywords

Cleft Palate
Velopharyngeal Insufficiency
Speech Disorders
Speech Perception
Speech

Descritores

Fissura Palatina
Insuficiência Velofaríngea
Distúrbios da Fala
Percepção da Fala
Fala

Correspondence address:

Flora Taube Manicardi
Programa de Pós-graduação em
Fonoaudiologia, Universidade Estadual
Paulista – UNESP
Av. Hygino Muzzi Filho, 737,
Portaria A, Mirante, Marília (SP),
Brasil, CEP: 17525-900.
E-mail: flora.tm@hotmail.com

Received: March 19, 2022

Accepted: April 01, 2023

Effect of perceptive-auditory training on the classification of speech hypernasality

Efeito do treinamento perceptivo-auditivo na classificação da hipernasalidade de fala

ABSTRACT

Purpose: To analyze the effect of auditory-perceptual training by inexperienced speech-language pathologists in the classification of hypernasality in individuals with cleft lip and palate and compare their classification of hypernasality individually, with the gold standard evaluation, before and after this training. **Methods:** Three inexperienced speech-language pathologists used a four-point scale to assess 24 high-pressure speech samples from individuals with cleft lip and palate, before and after auditory-perceptual training. The speech samples corresponded to six samples of each degree of hypernasality. The speech-language pathologists received auditory-perceptual training during the assessments. They had access to anchor samples and immediate feedback of correct answers regarding the degree of hypernasality in training. **Results:** There was no significant difference in the overall percentage of correct answers when comparing before and after the auditory-perceptual training. There was a significant association and agreement of the three evaluators with a gold standard evaluation after training, with an increase in agreement for a single evaluator for absent and mild degrees of hypernasality. The dichotomous analysis of the data showed an increase in the Kappa Index of Agreement of this evaluator. Although there was an increase in the Index of Agreement between evaluators for absent, mild, and severe hypernasality, this increase did not reach statistical significance. **Conclusion:** The auditory-perceptual training provided did not result in a significant improvement in the hypernasality classification for the inexperienced speech-language pathologists, even though the individual data analysis showed that the training favored one of the evaluators. Further studies involving gradual and more extensive auditory-perceptual training may favor the classification of hypernasality by inexperienced SLPs.

RESUMO

Objetivo: Analisar o efeito de um treinamento perceptivo-auditivo de fonoaudiólogas sem experiência na classificação da hipernasalidade de fala de indivíduos com fissura labiopalatina e comparar a classificação da presença e grau de hipernasalidade realizadas dessas fonoaudiólogas (com a avaliação padrão-ouro), antes e depois do treinamento perceptivo-auditivo. **Método:** Três fonoaudiólogas sem experiência analisaram 24 amostras de fala de alta pressão de indivíduos com fissura labiopalatina, antes e depois de treinamento perceptivo-auditivo, usando escala de quatro pontos. As amostras de fala correspondiam a seis amostras de cada grau de hipernasalidade. Entre as análises, as fonoaudiólogas receberam treinamento perceptivo-auditivo. Houve acesso às amostras de referência e feedback de respostas corretas quanto ao grau de hipernasalidade no treinamento. **Resultados:** Não houve diferença significativa na porcentagem geral de acertos entre os momentos antes e depois do treinamento perceptivo-auditivo. Houve associação e concordância significativa das três avaliadoras com avaliação padrão ouro após treinamento, com aumento da concordância para uma avaliadora (aumento de respostas corretas para os graus ausente e leve). A análise dicotômica dos dados mostrou aumento do índice de concordância Kappa dessa avaliadora. Houve aumento do índice concordância inter-avaliadores para hipernasalidade ausente, leve, e grave, porém sem significância estatística. **Conclusão:** O treinamento perceptivo-auditivo não resultou em melhora significativa da classificação da hipernasalidade de fala pelas fonoaudiólogas sem experiência, embora a análise individual dos dados tenha mostrado que o treinamento favoreceu uma dessas avaliadoras. Novos estudos envolvendo treinamento perceptivo-auditivo gradual e mais extenso poderão favorecer a classificação da hipernasalidade de fala por fonoaudiólogos sem experiência.

Study conducted at Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo – USP – Bauru (SP), Brasil.

¹ Programa de Pós-graduação em Fonoaudiologia, Universidade Estadual Paulista Júlio de Mesquita Filho – UNESP – Marília (SP), Brasil.

² Pós-graduação em Ciência da Reabilitação, Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo – USP – Bauru (SP), Brasil.

³ Departamento de Educação Física, Nutrição e Fisioterapia, Universidade de Marília – Unimar – Marília (SP), Brasil.

Financial support: CAPES (88882.433909/2019-01).

Conflict of interests: nothing to declare.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Hypernasality is the most representative speech symptom of velopharyngeal dysfunction (VPD) after primary palatal surgery⁽¹⁾. This excessive nasal resonance during the production of oral sounds⁽¹⁾ occurs due to abnormal coupling of the resonance cavities (oral and nasal). Although instrumental techniques (nasoendoscopy, videofluoroscopy and flow-pressure technique) are recommended to corroborate the diagnosis of VPD, the auditory-perceptual assessment is the initial tool used by clinicians to identify speech symptoms suggestive of VPD after primary palatoplasty^(2,3). Through their auditory impressions, the clinicians identify the presence and severity of hypernasality, which favor clinical decision-making and evaluation of the effectiveness of the treatment⁽⁴⁾. The auditory-perceptual evaluation must be conducted by experienced professionals to minimize variations and errors inherent to the subjective nature of this type of evaluation⁽⁵⁾.

The identification of hypernasality can be challenging even for experienced listeners^(6,7). The type of stimuli (speech or song)⁽⁸⁾, the extent of the stimulus, the phonetic context of the speech sample^(9,10) and the presence of coexisting alterations⁽¹¹⁾ can influence perceptual assessment of hypernasality, affecting its reliability⁽⁴⁾. The type of scale^(4,12-14) and the listener's familiarity with the use of a given scale can also influence the assessment of this speech symptom⁽¹³⁾. The numerical scale of equal intervals is widely used for clinical⁽¹⁵⁾ and research purposes^(6,10,16-18), including those involving auditory-perceptual training for the classification of hypernasality^(19,20).

The degree of clinical experience and the criteria adopted by each evaluator in their analyses can also affect the classification of hypernasality⁽⁴⁾. The auditory-perceptual training and the use of reference samples are strategies that can favor the consistency of the evaluators' analyses⁽⁶⁾, in which unstable internal patterns of the listeners can be replaced by references⁽⁴⁾. Although these strategies are strongly recommended to improve the reliability of the auditory-perceptual analysis of hypernasality^(6,21) only few studies documented the results of auditory-perceptual training in the assessment of this speech symptom^(6,19-21).

In a previous study⁽⁶⁾, controlled training and definition of standardized criteria for the classification of hypernasality led to an increasing agreement within and between experienced evaluators (SLPs) and, consequently, an improvement in the reliability of the auditory-perceptual analysis of this speech symptom. Other studies particularly investigated the influence of the auditory-perceptual training in hypernasality judgments performed by students in speech-language pathology^(20,22) or residents in otorhinolaryngology⁽¹⁹⁾.

Lee, Whitehill and Ciocca⁽²²⁾ studied the effect of practice and feedback on intra-judge and inter-judge reliability of hypernasality judgements performed by SLP-students. Significant differences were found between (1) the groups receiving training (practice with and without feedback) and (2) the group that had simple exposure to speech samples. Groups that had received more comprehensive training had a significantly larger range of hypernasality ratings than the exposure group. There were no significant differences in intra-judge reliability among the three listener groups (practice with feedback, practice without feedback and only exposure

to samples). The researchers attributed this result to the small speech sample size (82 speech samples that were divided in 4 training steps) and limited training length (30 minutes of the introduction and one hour in each step of the training), which may have influenced the listeners to achieve good reliability scores.

In a later study, Guerra⁽²⁰⁾ studied ratings of hypernasality performed by undergraduate SLP-students. The ratings were made at different times (before training, immediately after training, one week and one month after training) and four different conditions (without training, with training and optional access to reference samples; with training and controlled access to reference samples; and, with training, controlled access to reference samples and immediate feedback of right answer). Even controlling the possible impact of articulatory alterations in the classification of hypernasality by using only low intraoral pressure speech samples, no significant difference was not found in all hypernasality ratings among all conditions and different analysis time.

Sydney et al.⁽¹⁹⁾ determined the ability of otolaryngology residents (control group vs. training group) to rate hypernasality in patients with VPD. Although there was an improvement in agreement after training, this difference was not significant.

There is a need for better understanding of the ability of inexperienced SLPs in rating hypernasality as well as the strategies that may improve their initial ratings. Previous findings^(20,22) of auditory-perceptual judgments of hypernasality carried out by undergraduate students in SLP are inconsistent in the current literature. In addition, for otolaryngology residents, there was a trend towards improvement in the agreement of ratings after training, but with no significant difference⁽¹⁹⁾. In this sense, it is not clear whether a controlled auditory-perceptual training can influence the classification of hypernasality by listeners with limited clinical experience in the field of speech pathology.

The aims of this study were: 1) analyze the effect of auditory-perceptual training of inexperienced SLPs in the classification of hypernasality of individuals with CLP; 2) compare ratings of presence and degree of hypernasality of each inexperienced listener with the gold standard assessment, before and after the auditory-perceptual training.

METHODS

This study was approved by the Research Ethics Committee at the Hospital de Reabilitação de Anomalias Craniofaciais - Universidade de São Paulo, in Bauru, SP Brazil (No. 3.131.704.3.131.704). All participants who agreed to participate signed informed consent.

Pre-existing speech recordings from individuals with a history of cleft lip and palate (CLP), both genders, were selected for this study. The pre-existing speech recordings selected were further evaluated for three participants under two conditions: before and after auditory-perceptual training.

2.1 Speech samples

The speech recordings were retrieved from a pre-existing set of recordings saved into computer files. These recordings were originally obtained in an acoustically treated room and had

good audio quality. The speech material was recorded directly on a computer with a Sound Blaster Audigy 2 sound card and Sony® Sound Forge v8.0, with a sampling rate of 44100 Hz, in single channel, 16 Bit. The audio signal was captured using a headset microphone (AKG C420®). All pre-existing recordings had a standardized interval of one second between each phrase. The speech stimuli of the study consisted of a set of 12 high-pressure consonant phrases, in which each phrase comprised a single recurring pressure consonant target sound. This stimuli was selected considering findings of a previous study, in which perceptual ratings of hypernasality were more favorable for extended oral stimuli, even with the presence of other speech disorders.

The speech samples selected for this study comprised forty-eight samples (in audio) stored in computer files that represent the hypernasality degrees: absent (A), mild (MI), moderate (MO) and severe (S), in male and female voices. These samples consisted of a set of 12 oral high-pressure phrases. For selection purposes, all the speech samples of the databases were listened to and the first 12 samples of these bases that represent each of the 4 hypernasality degrees were included, in total 48 speech samples. The representativeness of the recorded speech samples regarding the four degrees of the scale (A=25%; MI=25%; MO=25%; and S=25%) was established in a previous study⁽²³⁾.

In Silva-Mori's previous study⁽²³⁾, three Brazilian certified-SLPs with at least 10 years' experience working in a large craniofacial center and with daily clinical routine in identifying speech disorders in individuals with CLP (Specialists SLP) achieved 100% agreement in their ratings. Therefore, for this study, the specialists' ratings of 48 speech samples were used as the gold standard for comparisons with the participants' ratings and also to select speech samples for auditory perceptual training. The samples included in this study were controlled for dysphonia. Other speech symptoms (nasal air emission, compensatory articulation) were not controlled.

The 48 speech samples representing the four degrees of hypernasality initially selected for the study were subsequently distributed into rating samples, reference, training and retraining samples. Thus, the 48 samples were distributed into 24 classification samples (evaluators' analysis, before and after training), 8 reference samples (anchor), 8 auditory-perceptual initial training samples and 8 retraining samples. In turn, the 24 classification samples consisted of 6 representative samples of each 4 degrees of hypernasality (6 A, 6 MI, 6 MO, 6 S) paired by gender. Finally, the 8 reference samples, 8 training samples and 8 retraining samples consisted of 2 samples of each degree of hypernasality (2A, 2MI, 2MO, 2S) also paired by gender.

Participants

Three female certified-SLPs, mean age of 24 years, native speakers of Brazilian Portuguese were invited as participants for this study. All of them were starting the residence ship in the craniofacial area. All participants self-reported normal hearing and no previous experience in classifying hypernasality degrees in individuals with CLP.

Procedures

The participants assessed the 24 classification speech samples (6A, 6MI, 6MO, 6S) using a 4-point scale, in two moments: before and after perceptual-auditory training. Samples were presented simultaneously by the first author (FTM) through a Powerplay PRO-8 HA8000 signal divider and individual high-quality headphones (AKG Harman, model K414P). The answers were inserted on an Excel® spreadsheet.

Samples could be heard as many times as the participant deemed necessary to carry out their judgment. When necessary, the participants could ask for the repetition of the sample to the first author and the other participants could choose if they would listen to the sample again or not. Other speech symptoms (nasal air emission, compensatory articulation) were not controlled. Only samples with normal voice were included in this study.

Auditory-perceptual training

The auditory-perceptual training was performed by a Powerpoint presentation, which consisted of a calibration session, training and retraining. In the calibration session, the participants were introduced to the concept of hypernasality and were instructed to use the 4-point scale to rate the degrees of hypernasality. Three recorded audio speech samples (which were not included in any other stage of the study) were presented to the participants so that they could become familiar with the task.

In person auditory-perceptual training and retraining sessions were offered by the first author, for each participant individually, using a high-quality headphone (AKG Harman, model K414P) and a computer to indicate their answers on an Excel® spreadsheet.

For the auditory-perceptual training itself eight references (anchor samples), were four voice males and four voice females, representing each degree of hypernasality (A, MI, MO, S) and presented to each participant. Participants were instructed to rate hypernasality using a 4-point scale. After each answer, the evaluator provided feedback regarding the degree of hypernasality. If the participant rated the sample incorrectly, then the evaluator re-presented two reference samples selected for this study. One reference sample corresponded to the correct rating degree of hypernasality. The other corresponded to the previous evaluator's rating, that is, the sample rated incorrectly. Then, the evaluator presented the target sample again only for participant's comparison with reference samples. After every 20 minutes of training, a five-minute break was performed in order to avoid fatigue.

Retraining was carried out one week after the auditory-perceptual training, using 8 different speech samples from those used in the training. Reference samples (N=8) used in the auditory-perceptual training were also used as references during ratings of the retraining samples. The same procedures used in the auditory-perceptual training were adopted in the retraining with a single difference: the samples were presented at random in order to prepare the participants for the final classification step.

Data analysis

Quantitative variables were described by the mean and standard deviation (SD). The normality distribution of the percentage

of correct answers was verified by the Kolmogorov-Smirnov test. The Wilcoxon signed-rank test was used to compare the mean percentage of participants' correct answers in relation to the gold standard evaluation, before and after the auditory-perceptual training.

Qualitative variables were described by the absolute frequency distribution. Chi-squared test was used to assess the association between qualitative variables (answers from each participant versus gold standard evaluation), taking into account the degrees of hypernasality. Chi-squared test was also used to further assess the association between qualitative variables for presence and absence of hypernasality (dichotomous analysis), both before and after the auditory-perceptual training.

The Cohen's kappa coefficient (k) was used to analyze the agreement between each participant and the gold standard evaluation and the moment. For the strength of agreement associated with Kappa statistics, the following labels were assigned to the corresponding Kappa intervals: <0.00 No agreement (N); 0.00 - 0.19 Slight (S); 0.20 - 0.39 Fair (F); 0.40 - 0.59 Moderate (M); 0.60-0.79 Substantial (SB) and 0.80-1.00 Almost perfect (AP). The difference between moments (before and after training)

for k was analyzed by 95% confidence intervals. The SPSS v19.0 for Windows was used for all statistical analyses, adopting a significance level of 5% ($p < 0.05$).

RESULTS

The mean percentage of correct answers in rating the degree of hypernasality of the three participants (evaluators), before and after the auditory-perceptual training, was 65.3% and 62.5%, respectively. There was no significant difference in the overall percentage of correct answers when comparing training moments (p -value=0.972; Wilcoxon signed-rank test).

Additionally, before training, a significant association was observed between the three evaluators (EV1, EV2 and EV3) with the gold standard evaluation. EV1 showed moderate agreement, EV2 substantial agreement and EV3 regular agreement (Table 1). After training, there was a significant association and agreement of the three evaluators with the gold standard evaluation. EV1 continued to show moderate agreement, EV2 had a reduction in agreement in relation to the pre-training moment and EV3 showed improvement in agreement (Table 2).

Table 1. Analysis of the agreement and association of the absolute frequency distribution of the classification of the hypernasality of each evaluator with the gold standard rating, before the auditory-perceptual training, by degree of hypernasality⁽¹⁾

Before training		Gold standard rating				p-value	χ^2 Kappa
		Absent	Mild	Moderate	Severe		
EV 1	Absent	6	2	0	0	<0.001*	0.556 [‡]
	Mild	0	4	3	0		
	Moderate	0	0	3	3		
	Severe	0	0	0	3		
EV 2	Absent	6	2	0	0	<0.001*	0.667 [‡]
	Mild	0	3	1	1		
	Moderate	0	0	5	1		
	Severe	0	1	0	4		
EV 3	Absent	3	2	0	0	<0.001*	0.389
	Mild	3	2	1	1		
	Moderate	0	2	5	2		
	Severe	0	0	0	3		

*significant association by the Chi-square test for p -value ≤ 0.05 ; ‡Significant kappa coefficient for p -value ≤ 0.05 . Findings in bold signify results that agree with the gold standard rating

Caption: EV=evaluator

Table 2. Analysis of the agreement and association of the absolute frequency distribution of the evaluation of the hypernasality classification of each evaluator with the gold standard evaluation, after auditory-perceptual training, by degree of hypernasality⁽²⁾

Before training		Gold standard evaluation				p-value	χ^2 K	Kappa
		Absent	Mild	Moderate	Severe			
EV 1	Absent	6	1	0	1	<0.001*	0.500 [‡]	
	Mild	0	5	2	0			
	Moderate	0	0	2	3			
	Severe	0	0	2	2			
EV 2	Absent	6	2	0	0	<0.001*	0.333 [‡]	
	Mild	0	3	3	2			
	Moderate	0	1	1	2			
	Severe	0	0	2	2			
EV 3	Absent	6	1	0	0	<0.001*	0.667 [‡]	
	Mild	0	5	1	1			
	Moderate	0	0	4	2			
	Severe	0	0	1	3			

*significant association by the Chi-square test for p -value ≤ 0.05 ; ‡Significant kappa coefficient for p -value ≤ 0.05 . Findings in bold signify results that agree with the gold standard evaluation

Caption: EV=evaluator

Table 3. Comparison of the results of the participants (evaluators) with the gold standard rating, before and after the auditory-perceptual training, according to the dichotomous analysis^(2,4)

		Gold standard rating		χ^2 <i>p</i> value	Kappa
		Present	Absent		
Before training					
EV 1	Present	16	0	<0.001*	0.800 [‡]
	Absent	2	6		
EV 2	Present	16	0	<0.001*	0.800 [‡]
	Absent	2	6		
EV 3	Present	16	3	0.047*	0.412 [‡]
	Absent	2	3		
After training					
EV1	Present	16	0	<0.001*	0.800 [‡]
	Absent	2	6		
EV 2	Present	16	0	<0.001*	0.800 [‡]
	Absent	2	6		
EV 3	Present	17	0	<0.001*	0.895 [‡]
	Absent	1	6		

*significant association by the Chi-square test for p -value ≤ 0.05 ; [‡] Significant kappa coefficient for p -value ≤ 0.05

Caption: EV=evaluator

Table 4. Agreement between raters (Kappa coefficient) before and after auditory-perceptual training by response category and for total agreement⁽³⁾

Training	Categories	Kappa	IC 95%		<i>p</i> -value
			LI	LS	
Before	absent hypernasality	0.597	0.366	0.828	<0.001*
	mild hypernasality	-0.072	-0.303	0.158	0.538
	moderate hypernasality	0.261	0.030	0.491	0.027*
	severe hypernasality	0.356	0.125	0.587	0.003*
	Total	0.284	0.147	0.420	<0.001*
After	absent hypernasality	0.681	0.450	0.912	<0.001*
	mild hypernasality	0.280	0.049	0.511	0.0018*
	moderate hypernasality	0.074	-0.157	0.305	0.532
	severe hypernasality	0.600	0.369	0.831	<0.001*
	Total	0.413	0.276	0.550	<0.001*

*significant kappa coefficient for p -value ≤ 0.05

Caption: IC = Confidence Interval; LI = Upper Limit; LS = Inferior Limit

Dichotomous analysis (presence and absence of hypernasality) further showed that all the evaluators presented a significant association and significant agreement with the gold standard evaluation before the auditory-perceptual training. However, EV3 showed an increase in the Kappa Index of Agreement with the gold standard evaluation after the auditory-perceptual training (Table 3).

Table 4 shows the results of the analysis of agreement between the evaluators, before and after the auditory-perceptual training, by each degree of hypernasality and total ratings. Although there was significant agreement in the Cohen's kappa coefficient before the auditory-perceptual training among the three evaluators for the absent, moderate and severe degrees and total ratings, there was no agreement among the evaluators for mild hypernasality. Additionally, there was significant agreement in the Cohen's kappa coefficient after the auditory-perceptual training among the three evaluators for the absent, mild and severe degrees and in total ratings, but there was no agreement among the evaluators

for moderate hypernasality. When comparing the findings obtained before and after the auditory-perceptual training, an increase in the Index of Agreement was observed for absent, mild, and severe hypernasality and total ratings. However, this increase was not statistically significant (analysis of confidence interval of 95%).

DISCUSSION

This study investigated the effect of auditory-perceptual training of SLPs with no experience in rating speech hypernasality in individuals with CLP. The results obtained in the present study agree with those reported by the study of Sydney et al.⁽¹⁹⁾, that showed that training did not affect the rating task by participants.

Findings of this study are contrary to the findings obtained by Lee et al.⁽²²⁾. These authors found a significant difference in hypernasality ratings for the groups that received training (with or without feedback) and the group that was only exposed to

speech samples. Although the present study offered training and retraining, both with feedback for correct answers, the focus was only the rating of hypernasality degrees, regardless of whether this symptom occurred in coexistence with other speech disorders. Lee et al.⁽²²⁾, on the other hand, offered a gradual training, in a hierarchy from simpler to more difficult tasks, which may have favored the assessment of hypernasality even in the coexistence of other speech disorders.

The findings of the present study agreed with the results obtained previously by Guerra⁽²⁰⁾. Some researchers argue that rating the degree of hypernasality can be even more challenging with the presence of multiple speech disorders due to the need to accurately isolate individual auditory perceptual characteristics⁽²⁵⁾. Guerra⁽²⁰⁾ controlled the possible impact of compensatory articulations on hypernasality ratings performed by SLP-students by including only low intraoral pressure sounds in the samples rated. The present study controlled only the voice disorders and suggested new training's proposal. The amplified listener's training is recommended for future studies to make them capable of identifying other speech's disorders with or without hypernasality. This study also analyzed the association between qualitative variables (answers from each participant versus gold standard ratings), considering the degrees of hypernasality. It was expected improvement in percentage agreement with predetermined severity ratings for all evaluators, particularly for the extreme of resonances. Results showed that for the offered training, the agreement of one evaluator remained moderate after training, while there was a reduction in agreement in relation to the pre-training moment for two other evaluators, with fewer correct answers for the moderate and severe degrees. However, the two evaluators agreed with the gold standard evaluation in all speech samples with no hypernasality, in both moments, before and after training. Previous study also found better accuracy ratings for the extreme of resonance, with more stable judgments of normal resonance⁽¹⁹⁾. In a previous study, it was also observed that extended oral stimuli (set of oral phrases) would result in better scores of perceptive assessment reliability by expert raters, therefore, these stimuli were selected for this study⁽¹⁰⁾.

In this study, a single evaluator showed improvement in agreement with predetermined ratings (gold standard evaluation), with greater agreement for mild degree and absence of hypernasality. Some authors⁽²⁶⁾ argued that the use of auditory anchors during rating tasks resulted in a significant improvement in the accuracy of severity ratings of speech samples in the normal and mild categories. The use of anchor samples during the training stages of this study may have also favored this single evaluator's ratings for the lower extremes of the scale, when hypernasality was absent and even mild, in which speech can be socially accepted by lay listeners⁽¹⁵⁾.

External variables in the speech samples may have influenced the classification of hypernasality by all evaluators for moderate and severe degrees. In this study, the presence of speech disorder characteristics (atypical articulatory patterns and nasal airflow/nasal turbulence) was not a controlled variable and, therefore, may have impaired the classification of moderate and severe hypernasality by inexperienced speech-language pathologists.

Presence of dysphonia, which may affect perceptual ratings, was a controlled variable in this study.

In this study, a 4-point equal-interval scale (absent, mild, moderate, and severe) was used to rate hypernasality. This scale is widely used for clinical and research purposes^(6,10,17,19,20,27). In this scale, the evaluator assigns an index to the speech aspect assessed, grading the level of severity based on the assumption that the different degrees measured are equivalent to the human ear. Although the study used a 4-point scale to classify hypernasality, a dichotomous analysis of the results was further performed in order to investigate how the data would behave if the response options were only the presence or absence of hypernasality. The results showed an increase in the index of agreement with the gold standard evaluation (from significant to almost perfect) for one of the evaluators, suggesting that deciding between the presence or absence of the alteration is easier than grading the alteration, when present. A binary scale to identify hypernasality by experienced listeners was used^(28,29), but not to test the auditory-perceptual training effect. In further studies, the identification of the presence and absence of hypernasality could be carried out in the initial stage of a training, in order to favor the analysis of non-experienced listeners. Previous study⁽²²⁾ offered a training that included four parts, presented in a hierarchy from simpler to more difficult tasks. This hierarchy may have favored the results, in which significant differences in the hypernasality rating scores were found between the groups receiving training (practice with and without feedback) and the group that had only exposure to speech samples.

In this study, the agreement between evaluators was also obtained in two moments: before and after training. There was significant agreement in the Cohen's kappa coefficient among the three evaluators at both times for the absent and severe degrees, suggesting that inexperienced speech-language pathologists were more consistent in their analyses, regardless of training, when hypernasality was absent or severe. The findings of this study agree with a previous study⁽¹⁹⁾ that also found more accurate analyzes with expertise ratings for normal and severe resonance. Although the comparison of the findings obtained before and after the auditory-perceptual training has shown an increase in the index of agreement between evaluators for absence and mild and severe hypernasality, this increase was not statistically significant. Difficulty in achieving high agreement ratings of hypernasality degree is reported in the literature, even for experienced listeners^(6,21).

The present study made an important contribution to a better understanding of the variables (amount of speech samples, to use anchor samples, training duration) and all of this can influence the degree's rating of nasality made by non-expert listeners. Although it is strongly recommended, only a few studies have investigated the influence of perceptual training on rating's agreement of non-expert listeners. The existing studies show the variability of results.

As pointed out in a recent study⁽³⁰⁾, there is a need to keep searching well elaborated perceptual training for non-experts (speech language pathologist's undergraduate students) that would enable them to perform hypernasality assessments⁽³⁰⁾. This argument is based on findings of a recent study which

found little or no effect of perceptual training (two hours) for identification of speech disorders presented by individuals with cleft lip and palate disorders, after offering this training for 31 students⁽³⁰⁾. Findings of our study also show the need for training improvements and suggest how important it is to take care of the training offered to the future speech-language professionals.

The data found in this study shows the need for new investigation that could overcome the possible limitations. In the new studies, it is suggested to include extended training, and as often as possible, with a bigger amount of speech samples. Also, gradual increase of the severity levels of the analyses is recommended (for example, start with binary scales, and afterwards, include more amplified scales). Moreover, it is recommended to incorporate strategies that could favor identifying speech errors that may coexist with hypernasality. Offered in the beginning stages of the perceptual training, these strategies could favor posterior hypernasality ratings by the listeners. In a previous⁽²²⁾ perceptual training strategies involving hierarchy from simpler to more difficult tasks resulted in favorable ratings by non-expert listeners. This result was not observed in other studies that did not offer easy-to-hard training procedures, as well as in the present study.

In future studies, short and long-term assessments are recommended to verify the possibility to keep the abilities learned in a long period of time⁽³⁰⁾ with the offered training.

As previously proposed in the literature⁽¹⁹⁾, an online training can also help in the adhesion of more extensive auditory-perceptual training, by the listeners, and thus, the online modality must also be incorporated in these trainings.

CONCLUSION

The auditory-perceptual training with controlled access to the reference samples and the immediate feedback did not result in a significant improvement in the overall percentage of correct answers obtained by the inexperienced speech-language pathologists. However, the individual data analysis has shown that the training favored one of the evaluators. Incorporating gradual and more extensive auditory-perceptual training may favor the rating task of speech hypernasality by inexperienced SLPs, particularly for mild and moderate degrees of hypernasality.

REFERENCES

1. Zajac DJ, Vallino LD. Evaluation and management of cleft lip and palate: a developmental perspective. San Diego: Plural Publishing; 2017.
2. Kummer AW. Speech evaluation for patients with cleft palate. *Clin Plast Surg.* 2014;41(2):241-51. <http://dx.doi.org/10.1016/j.cps.2013.12.004>. PMID:24607192.
3. Baylis A, Chapman K, Whitehill TR, The Americleft Speech Group. Validity and reability of visual scaling for assessment of hypernasality and audible nasal emission in children with repaired cleft palate. *Cleft Palate Craniofac J.* 2015;52(6):660-70. <http://dx.doi.org/10.1597/14-040>. PMID:25322442.
4. Bettens K, De Bodt M, Maryn Y, Luyten A, Wuyts FL, Van Lierde KM. The relationship between the Nasality Severity Index 2.0 and perceptual judgments of hypernasality. *J Commun Disord.* 2016;62:67-81. <http://dx.doi.org/10.1016/j.jcomdis.2016.05.011>. PMID:27310727.

5. Kent RD. Hearing and believing some limits to the auditory-perceptual assessment of speech and voice disorders. *Am J Speech Lang Pathol.* 1996;5(3):7-23. <http://dx.doi.org/10.1044/1058-0360.0503.07>.
6. Oliveira ACASF, Scarmagnani RH, Fukushiro AP, Yamashita RP. The influence of listener training on the perceptual assessment of hypernasality. *CoDAS.* 2016;28(2):141-8. <http://dx.doi.org/10.1590/2317-1782/20162015163>. PMID:27191877.
7. Ahl R, Harding-Bell A. Comparing methodologies in a series of speech outcome studies: challenges and lessons learned. *Cleft Palate Craniofac J.* 2018;55(1):35-44. <http://dx.doi.org/10.1177/1055665617718546>. PMID:34162055.
8. Peter S, Abdul Rahman ZA, Pillai S. Hypernasality in singing among children with cleft palate: a preliminar study. *Int J Oral Maxillofac Implants.* 2019;48(10):1317-22. <http://dx.doi.org/10.1016/j.ijom.2019.03.896>. PMID:31014926.
9. Yamashita RP, Borg E, Granqvist S, Lohmander A. Reliability of hypernasality rating: comparison of 3 different methods for perceptual assessment. *Cleft Palate Craniofac J.* 2018;55(8):1060-71. <http://dx.doi.org/10.1177/1055665618767116>. PMID:29634363.
10. Marino VCC, Dutka JCR, Manicardi FT, Gifalli G, Silva PP, Pegoraro-Krook MI. Influence of speech stimuli in the auditory perceptual identification of hypernasality in individuals with cleft lip and palate. *CoDAS.* 2020;32(6):e20190269. <http://dx.doi.org/10.1590/2317-1782/20202019269>. PMID:33331425.
11. Lee A, Potts S, Bressmann T. Speech-language therapy students' auditory-perceptual judgements of simulated concurrent hypernasality and articulation disorders. *Clin Linguist Phon.* 2019;1-14. <http://dx.doi.org/10.1080/02699206.2019.1655666>. PMID:31429313.
12. Castick S, Knight RA, Sell D. Perceptual judgments of resonance, nasal airflow, understandability, and acceptability in speakers with cleft palate: ordinal versus visual analogue scaling. *Cleft Palate Craniofac J.* 2017;54(1):19-31. <http://dx.doi.org/10.1597/15-164>. PMID:28067575.
13. Bettens K, Bruneel L, Maryn Y, Bodt M, Luyten A, Van Lierde KM. Perceptual evaluation of hypernasality, audible nasal airflow and speech understandability using ordinal and visual analogue scaling and their relation with nasalalance scores. *J Commun Disord.* 2018;76:11-20. <http://dx.doi.org/10.1016/j.jcomdis.2018.07.002>. PMID:30071470.
14. Ramos-Favaretto FS, Fukushiro AP, Scarmagnani RH, Yamashita RP. Borg scale: a new method for hypernasality rating. *CoDAS.* 2019;31(6):e20180296. <http://dx.doi.org/10.1590/2317-1782/20192018296>. PMID:31800882.
15. Henningson G, Kuehn DP, Sell D, Sweeney T, Trost-Cardamone JE, Whitehill TL. Universal parameters for reporting speech outcomes in individuals with cleft palate. *Cleft Palate Craniofac.* 2008;45(1):1-17. <http://dx.doi.org/10.1597/06-086.1>. PMID:18215095.
16. Tak HR, Waknis AP, Kulkarni SP. Perceptual and instrumental analysis of hypernasality in children with repaired cleft palate. *J Cleft Lip Palate Craniofacial Anomalies.* 2016;3(2):67-72. <http://dx.doi.org/10.4103/2348-2125.187508>.
17. Ferreira GZ, Bressmann T, Dutka JCR, Whitaker ME, De Boer G, Marino VCC, et al. Analysis of oral-nasal balance after intensive speech therapy combined with speech bulb in speakers with cleft palate and hypernasality. *J Commun Disord.* 2020;85:105945. <http://dx.doi.org/10.1016/j.jcomdis.2019.105945>. PMID:31607437.
18. Girish KS, Pushpavathi M, Satish HV. Influence of native language on Nasalalance values in Kannada and speakers. *J Cleft Lip Palate Craniofacial Anomalies.* 2021;8(2):111-8. http://dx.doi.org/10.4103/jclpca.jclpca_33_20.
19. Sydney CB, Truong A, Forde C, Stefanov DG, Marrinan E. Perceptual assessment of velopharyngeal dysfunction by otolaryngology residents. *Otolaryngol Head Neck Surg.* 2016;155(6):1034-9. <http://dx.doi.org/10.1177/0194599816662247>. PMID:27484234.
20. Guerra TA. Treinamento de avaliadores para identificação da hipernasalidade [tese]. Bauru: Faculdade de Odontologia de Bauru da Universidade de São Paulo; 2019. <http://dx.doi.org/10.11606/T.25.2019.tde-13112019-191450>.
21. 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. <http://dx.doi.org/10.1597/13-310>. PMID:24446923.

22. Lee A, Whitehill TL, Ciocca V. Effect of listener training on perceptual judgement of hypernasality. *Clin Linguist Phon.* 2009;23(5):319-34. <http://dx.doi.org/10.1080/02699200802688596>. PMID:19399664.
23. Silva-Mori MJF. Base de dados de amostras de referência da nasalidade de fala e sua influência na avaliação perceptivo-auditiva na fissura labiopalatina [tese]. Bauru: Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo; 2018. <http://dx.doi.org/10.11606/T.61.2018.tde-06102020-123820>.
24. Kummer AW. Management of velopharyngeal insufficiency: the evolution of care and the current state of the art. *J Cleft Lip Palate Craniofacial Anomalies.* 2019;6(2):65-72. http://dx.doi.org/10.4103/jclpca.jclpca_10_19.
25. Kreiman J, Gerratt BR. Measuring vocal quality. In: Kent JF, Ball MJ, Voice quality measurement. San Diego: Singular, 2000. p. 73–101.
26. Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity rating. *J Voice.* 2009;23(3):341-52. <http://dx.doi.org/10.1016/j.jvoice.2007.10.006>. PMID:18346869.
27. 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. <http://dx.doi.org/10.1597/02-136>. PMID:14697067.
28. Prado-Oliveira R, Marques IL, Souza L, Souza-Brosco TV, Dutka JCR. Assesment of speech nasality in children with Robin Sequence. *CoDAS.* 2015;27(1):51-7. <http://dx.doi.org/10.1590/2317-1782/20152014055>. PMID:25885197.
29. Pegoraro-Krook MI, Rosa RR, Aferri HC, Andrade LKF, Dutka JCR. Pharyngeal bulb prothesis and speech outcome in patients with cleft palate. *Braz J Otorhinolaryngol.* 2022;88(2):187-93. <http://dx.doi.org/10.1016/j.bjorl.2020.05.028>.
30. Bruneel L, Danhieux A, Van Lierde K. Training speech pathology students in the perceptual evaluation of speech in patients with cleft palate: reliability results and the students' perspective. *IJOPRL.* 2022;157:111145. <http://dx.doi.org/10.1016/j.ijporl.2022.111145>.

Authors' contributions

FTM participated as principal investigator, research elaboration, schedule elaboration, literature survey, data collection and analysis, article writing, article submission and procedures; JCRD participated in the collaboration in the elaboration of the research, in the analysis and interpretation of the data and in the final writing of the article; TAG participated in the collaboration in the analysis of the study design and data; MIPK participated in the collaboration in the analysis and interpretation of data and correction of the article's wording; EFBC participated in collaboration in reviewing the study design and interpreting the data. VCCM participated in research design, data analysis and interpretation, research writing, correction of article writing and approval of the final version.