http://dx.doi.org/10.1590/S2317-64312015000100001444

Assessment of speech nasality in individuals with cleft palate

Avaliação da nasalidade de fala na fissura labiopalatina

Edna Zakrzevski Padilha[1], Jeniffer de Cássia Rillo Dutka[2], Viviane Cristina de Castro Marino[3], José Roberto Pereira Lauris[4], Mariana Jales Felix da Silva[5], Maria Inês Pegoraro-Krook[2]

ABSTRACT

Purpose: To describe the results of speech nasality of individuals with cleft lip and palate, and to compare auditory-perceptual judgments of nasality between live ratings and multiple judges ratings of recorded speech, for two sets of speech stimuli. **Methods:** The study involved the retrospective analysis of the results of auditory-perceptual assessments of speech nasality performed live by a single speech-language pathologist and a prospective judgment of 100 recordings of speech samples obtained during production of two sets of speech stimuli: one with high pressure consonants (HPC, n=100) and another with low pressure consonants (LPC, n=100). The data belonged to patients, of both genders, with ages between 5 and 12 years, with cleft lip and palate operated by the same surgeon. **Results:** The absence of hypernasality was found for 69% of the patients during live assessment. When present, mild hypernasality was found for 23% and moderate for 8% of the patients. For judge ratings of recorded samples, 50% was identified as hypernasal during production of samples with high pressure consonants, and 62% for the samples with low pressure consonants. A statistically significant difference was found between the live perceptual judgments and judges' ratings of the recorded samples only for the stimuli with high pressure consonants. The agreement between the methods of assessment was 79% for HPC samples and 80% for LPC samples, within the moderate range. **Conclusion:** Live perceptual judgment of speech nasality revealed higher occurrence of absence of hypernasality followed by presence of mild hypernasality, when compared to multiple judges of recorded samples. The live clinical assessment of speech, however, has the disadvantage that the data may not be reproduced, quantified or shared by other team members.

Keywords: Cleft palate; Velopharyngeal insufficiency; Diagnosis; Speech; Speech disorders

RESUMO

Objetivo: Descrever os resultados da nasalidade de fala de indivíduos com fissura labiopalatina e comparar os achados de nasalidade estabelecidos por meio do julgamento perceptivo-auditivo realizado ao vivo com os achados estabelecidos por análise de gravações por juízes, em dois tipos de amostras de fala. **Métodos:** O estudo envolveu a análise retrospectiva dos resultados de avaliações perceptivo-auditivas da nasalidade de fala realizadas ao vivo por uma fonoaudióloga e o julgamento prospectivo, por consenso de juízas de 100 gravações de amostras de fala, obtidas durante a produção de dois conjuntos de estímulos de fala: um com consoantes de alta pressão (CAP, n=100) e outro com consoantes de baixa pressão (CBP, n=100). Os dados pertenciam a pacientes de ambos os gêneros, com idades entre 5 e 12 anos, que tiveram a fissura labiopalatina operada por um mesmo cirurgião. **Resultados:** A ausência de hipernasalidade foi constatada em 69% dos julgamentos ao vivo. Quando presente, a hipernasalidade leve foi constatada em 23% dos casos, enquanto a hipernasalidade moderada em 8%. Para os julgamentos das amostras gravadas, 50% foram identificadas com hipernasalidade durante a produção das amostras CAP e 62% durante a das amostras CBP. Diferença significativa foi encontrada entre o resultado do julgamento ao vivo e o julgamento pelas juízas nas amostras CAP. A concordância entre as modalidades de avaliação variou de 79% para as amostras CAP e 80% para as amostras CBP, sendo considerada moderada. **Conclusão:** O julgamento perceptivo ao vivo da nasalidade de fala pode detectar melhor a ausência de hipernasalidade, seguida pela hipernasalidade de grau leve, em comparação com o julgamento realizado por juízes múltiplos, a partir de amostras gravadas. Contudo, tem a desvantagem de os dados não poderem ser reproduzidos, nem quantificados, nem compartilhados por outros membros da equipe.

Descritores: Fissura palatina; Insuficiência velofaríngea; Diagnóstico; Fala; Distúrbios da fala

The study was conducted at the Hospital for Rehabilitation of Craniofacial Anomalies, Universidade de São Paulo – USP – Bauru (SP), Brazil.
(1) Graduation Program (Master) in Sciences, Department of Speech, Language and Hearing Disorders, Bauru School of Dentistry, Universidade de São Paulo – USP – Bauru (SP), Brazil.
(2) Department of Speech, Language and Hearing Disorders, Bauru School of Dentistry, Universidade de São Paulo – USP – Bauru (SP), Brazil; Graduation Program in Rehabilitation Sciences of the Hospital for Rehabilitation of Craniofacial Anomalies, Universidade de São Paulo – USP – Bauru (SP), Brazil.
(3) Department of Speech, Language and Hearing Disorders, College of Philosophy and Sciences, Universidade Estadual Paulista "Júlio de Mesquita filho" – UNESP – Marília (SP), Brazil.
(4) Department of Orthodontics, Pediatric Dentistry and Community Heath, Bauru School of Dentistry, Universidade de São Paulo – USP – São Paulo (SP), Brazil.
(5) Graduation Program (Master) in Rehabilitation Sciences, Hospital for Rehabilitation of Craniofacial Anomalies, Universidade de São Paulo – USP – Bauru (SP), Brazil.
Funding: The first author was granted with a master scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.
Conflict of interests: No
Authors' contribution: *EZP* principal investigator, development of the study and schedule, literature review, data collection and analysis, article writing, article submission; *JCRD* development of the study, data analysis, article review; *VCC* data collection and analysis; *JRPL* data analysis; *MJFS* data collection and analysis; *MIPK* supervisor, concept and development of the study and schedule, data analysis, correction of article draft, and approval of the final version.
Correspondence address: Maria Inês Pegoraro-Krook. Departamento de Fonoaudiologia, Faculdade de Odontologia de Bauru, Universidade de São Paulo. Al. Octávio Pinheiro Brisolla, 9/75, Bauru (SP), Brazil, CEP: 17012-901. E-mail: mikrook@usp.br
Received on: 7/8/2014; **Accepted on:** 2/11/2015

INTRODUCTION

Velopharyngeal dysfunction is considered one of the major etiological factors that influence the speech production skills of individuals with cleft palate[1]. The results of speech after primary palatoplasty are commonly used as an indicator of the outcome of surgery and can reflect the effectiveness of the treatment protocols used by services that manage cleft palate[2-5].

The live auditory-perceptual judgment of speech conducted by the speech-language pathologist (SLP) is the standard procedure to evaluate speech outcome after surgical correction of cleft lip and palate[6]. A study with SLPs associated with cleft palate teams at North America found that 99% of the teams use the perceptual evaluation as the golden standard procedure to establish speech outcome during velopharyngeal evaluation[5]. Considering that the symptoms of VPD are perceptual in nature, auditory-perceptual judgments are selected as gold standard for evaluation[6,7] and should be done by trained listeners[8,9].

The identification of presence of hypernasality by craniofacial teams is often accomplished through auditory-perceptual assessment using binary scales (abnormal vs normal) or using scales with equal intervals such as the 4 point scale where 1 = normal, 2 = mild, 3 = moderate and 4 = severe hypernasality, for example. Direct magnitude estimation and paired comparisons (with or without reference samples) have also been used to identify presence of hypernasality[6]. Literature reports that descriptive categories and scale of equal intervals are the tools most often used. While using equal appearing interval scales it is assumed that the distance between two positions on the scale is the same and the intervals to register the degree of nasality may vary across the scales between 3, 4, 5 or more degrees[6,10].

In addition to the use of scales with different intervals, other methodological differences in speech evaluation of patients with cleft lip and palate can be found across teams from different parts of the world. Different stimuli have been used for capturing the speech samples recorded for later classification of hypernasality, including: samples involving production of isolated words, sentences or spontaneous speech; samples balanced according to vowel type (high vowels versus other vowels); samples balanced according to consonant type, with high pressure consonants (plosives and fricatives) and with low pressure consonants (liquids); or samples combining oral and nasal consonants. Differences in the method to elicit speech samples (such as reading, naming or repeating) and to document speech results (such as live evaluation vs. use of audio or video recordings) can also be found[10,11].

Audio recordings of speech samples are considered the main system for documenting speech outcome among craniofacial teams, particularly because it has the advantages of being easily retrieved, edited and presented for auditory-perceptual ratings by multiple judges, which allows for measurements of the intra and interjudge reliability, and also because it can be used as a tool for corroborating the findings of live evaluations[10,12,13]. Researchers from Scandinavia and the United Kingdom conducted a multicenter study (Scandcleft Project, 1997) and have standardized the recording and the analysis of the speech samples to document outcome of primary repair. The authors pointed out that audio and/or video recordings should be part of the documentation for all patients and that the recording equipment must be of good quality to allow evaluation of all speech variables[13].

Several authors have pointed out the importance of establishing a careful clinical protocol to evaluate speech nasality, suggesting the need for a standardization of the speech judgments[6,10,14]. Careful methodology is essential to enable comparisons of speech results and to ensure that there is no loss of data due to compromised quality of recorded speech samples[14]. For the management of cleft lip and palate the procedures for ratings of speech nasality should, at the initial stage, emphasize the identification of absence or presence of hypernasality, since this aspect of speech is a major indicator of the outcome of the surgery and it is the primary symptom of velopharyngeal dysfunction. To assess speech nasality properly it is necessary to use procedures that are efficient, accurate and reliable contributing to the identification of consistent findings. This study aimed to describe speech nasality of individuals with CLP and to compare the outcome between live perceptual judgments and listeners' judgments of audio recorded speech samples.

METHODS

The research protocol was approved by the Ethics Committee on Human Research of the Hospital for Rehabilitation of Craniofacial Anomalies, *Universidade de São Paulo* (HRAC-USP), Brazil (346/2012).

Speech samples

A total of 100 patients was identified for this study, all with unilateral cleft lip and palate with no other associated anomalies, all underwent primary palatoplasty operated consecutively by a single surgeon during a period of five years. The primary surgeries received by the patients were conducted between 9 and 18 months of age (mean = 12 months), in the same institution. One hundred recordings and the findings of auditory-perceptual evaluations were retrieved for this study, both, the recordings and the data in patient's charts, were captured at the same date and were included in this study as long as they were obtained prior to secondary intervention. Forty-one recordings were from females and 59 from males, with ages between 5 and 12 years, all acquired between August, 2006 and May, 2010.

Live auditory-perceptual assessment of nasality

The live auditory-perceptual ratings of speech nasality were retrieved retrospectively from patient´s chart. Three SLPs with more than five years of experience with the evaluation and treatment of patients with cleft lip and palate performed all live

speech assessment between 2006 and 2010, and registered the presence and the degree of hypernasality in the patient´s chart as the routine protocol established at the institution. For the classification of the degree of nasality, a 4-point equal appearing scale was used, 0 indicating absence of hypernasality and the scores of 1, 2 and 3 indicating presence of hypernasality (mild, moderate and severe, respectively). The live judgment of the occurrence and the degree of hypernasality during the clinical speech evaluation was performed using uncontrolled speech sample, which could have been established with a brief conversation or during repetition of words and phrases (which may or may not include recurrence of target sounds).

Auditory-perceptual assessment of nasality by judges

A prospective analysis of speech nasality using recorded samples of those 100 patients was conducted by three SLPs in consensus (referred from now on as judges), all of which were experienced with the evaluation of the speech of patients with cleft lip and palate. The speech recordings were retrieved from the institution's files and were used for the task of auditory-perceptual assessment as proposed for this study. The speech samples were captured during production of two sets of sentences with the predominance of oral sounds one containing phrases only with high pressure consonants (HPC) and another containing phrases only with low pressure consonants (LPC). The HPC sample included the following phrases: *Papai caiu da escada* (Father fell from the stairs), *Fábio pegou o gelo* (Fabio grabbed the ice), *O palhaço chutou a bola* (The clown kicked the ball), *Teresa fez pastel* (Teresa made pastel) and *A árvore dá frutos e flores* (The tree bears fruit and flowers). The LPC sample included: *O louro ia olhar a lua* (The parrot would look at the moon), *Laura lia ao luar* (Laura read under moonlight), *A leoa é leal* (The lioness is loyal), *Lili era loira* (Lili was blonde) and *Lulu olha a arara* (Lulu looked at the parrot).

The samples were captured using a headset condensate/unidirectional microphone (model AKG C420®), positioned in a distance of approximately 5 cm from one of the lateral labial commissure. Recordings were imported directly into an IBM-PC Intel Pentium® 4 2.8GHz computer, equipped with a Sound Blaster Audigy 2, using the Sony Sound Forge, version 7.0 (2003) program, with a sampling rate of 44100 Hz in single-channel, 16-bit, saved as wave files. All recordings were made in a silent and acoustically treated environment according to the routine for documentation of speech in the institution.

The recordings were edited using the Sony Sound Forge Pro-10® (Sony® Media Software, 2009 Program). Two audio files (wave format) were prepared, one for HPC samples and another for LPC samples, and were saved in a folder named "samples for judgment". A number was issued for each individual sample along with an identification indicating the gender of the patient that was recorded. The information regarding the gender of the patient was used by the judges in order to retrieve the reference samples provided during the rating task. Reference samples were created to calibrate the judges to identify the 4-point intervals of the scale used in this study. These references were established after being judged with 100% agreement between listeners who indicated that the samples were representative of each degree of the scale. In order to the judges to access the reference samples during their analysis of the recordings four folders were created, representing the 4 degrees of the scale (absence of hypernasality, mild, moderate and severe hypernasality). Samples established for both genders (male and female) and for both types of speech stimuli (HPC and LPC) were established to be used as reference to establish consensus among the judges. That is, in the file with reference samples indicative of absence of hypernasality, four folders were created: one consisting of reference samples loaded with high pressure consonants for the female voice (HPC-female), one with reference samples loaded with high pressure consonants for the male voice (HPC-male); one with the reference samples loaded with low pressure consonants for the female voice (LPC-female) and one with the reference samples loaded with low pressure consonants for male voice (LPC-male). Since the same procedure was used to establish reference samples for the other three intervals of the scale (mild, moderate and severe hypernasality), 16 folders with reference samples were created, all using samples that did not belong to the patients included in this study (two genders + two types of stimuli + 4-point intervals = 16). Reference samples were used to calibrate the judges during the use of the 4-point scale and could be retrieved by the judges during the judgment task if needed. The samples rated in this study as well as the reference samples were saved on a compact disc (CD).

Before the judgment sessions, the judges underwent a brief auditory training, when they heard all the reference samples and used the 4-point scale while comparing speech samples from both genders. The auditory-perceptual judgment was conducted in a quiet room where the three judges, connected the same computer, heard the same samples for judgment, each using an AKG® K414P headset, connected to a Windows Media Player (Microsoft Windows®). Thus, the three judges heard the speech samples simultaneously and noted the outcome of their auditory-perceptual judgments in recording sheets, one for HPC samples and one for LPC samples. Judges could listen to the recordings as many times as needed to rate the sample. The judges were instructed to only rate the aspect of hypernasality in each recording, choosing from four choices: absence of hypernasality, mild hypernasality, moderate hypernasality or severe hypernasality. In case of doubt or disagreement during the judgment, the reference samples were heard again by all judges to help to achieve consensus regarding the intervals of the scale. The judges were allowed to discuss their ratings until the consensus (100% agreement) was established for each HPC and LPC sample. The classification of nasality was performed

in a single day, with the completion of a session in the morning for the HPC samples and another in the afternoon for the LPC samples. Each grading session lasted about three hours, with one interval of 10 minutes.

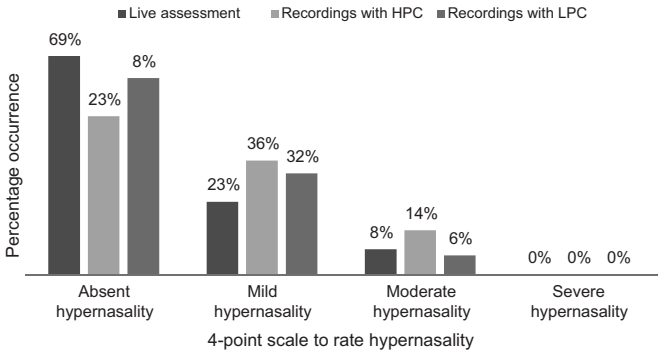
The descriptive analysis of the data included measures of the overall percentage of occurrence of hypernasality and the percentage of occurrence for each interval of the 4-point scale. For the inferential analysis of the findings, the measures obtained using the 4-point scale were transformed into in a binary scale indicating only the absence or the presence of hypernasality. The McNemar test was used to compare the findings between the two modalities of judgment of nasality considering the two possible outcomes (absence or presence of hypernasality), with the level of significance set at $p<0.05$. The level of agreement between the methods of assessment was expressed as a percentage of agreement and also with the Kappa coefficient[15].

RESULTS

Retrospective analysis of the data revealed records regarding the live assessment of hypernasality for 99 of the charts' analyzed, and indicated that 68 (69%) of the patients presented without hypernasality and 31 (31%) presented with hypernasality, and in this group, 23 (23%) were identified with mild hypernasality and 8 (8%) with moderate hypernasality. Severe hypernasality was not identified for the group of patients studied. That is, if the speech outcome of the primary palatoplasty at the research site was established based only on the identification of the presence or absence of hypernasality assessed by a single SLP during live auditory-perceptual evaluation, 31% of patients from this study presented a symptom indicative of velopharyngeal dysfunction.

The auditory-perceptual analysis of the recordings by multiple judges, was established initially with 100% consensus for 80% of the rated samples. That is, for 20% of the samples, the three judges did not agree 100% regarding the absence or presence and degree of hypernasality and had to review the ratings after accessing the reference samples to recalibrate the scale intervals. After discussing and reviewing the samples, judges reached consensus (100% agreement) for all recordings.

When considering all 200 ratings of recorded speech by multiple judges (100 ratings for the samples with HPC and 100 for the samples with LPC), the findings revealed that 112 samples (56%) were judged with absence of hypernasality, while 88 (44%) were judged with presence of hypernasality. Specifically for the HPC samples, 50 (50%) were judged with absence of hypernasality, 36 (36%) with mild hypernasality, 14 (14%) with moderate hypernasality, and none (0%) with severe hypernasality. For the LPC samples, 62% of the recordings were judged with absence of hypernasality, 32 (32%) with mild hypernasality, 6 (6%) with moderate hypernasality, and none (0%) with severe hypernasality (Figure 1).



Note: HPC = high pressure consonants; LPC = low pressure consonants

Figure 1. Hypernasality outcome during live assessment (retrieved from patient's charts) and during multiple judges ratings of recordings with high and low pressure consonant samples

A comparison of the live judgment retrieved from medical charts to the ratings of the recorded samples by judges, revealed that the difference in occurrence of hypernasality was statistically significant only between live (31%) and HPC (50%) ratings, with the judges identifying presence of hypernasality during production of HPC in the recordings of 19 individuals who were rated without hypernasal speech during live assessment ($p<0.001$, McNemar Test).

The percentage of agreement between the live auditory--perceptual judgments and perceptual judgments of the speech recordings by the judges was 79% for the HPC and 80% for the LPC samples with a Kappa coefficient (K) indicating moderate agreement (K=0.57 and 0.55).

DISCUSSION

In 2002, the World Health Organization (WHO)[16] released a report on research on craniofacial anomalies emphasizing the need to establish methods and measures for documenting treatment outcome. WHO recommended that the speech variable of nasality should be evaluated for their occurrence (presence/absence), and also for the classification of the degree of hypernasality.

Performing an assessment of nasality, especially hypernasality, is not an easy task, given the many variables that can influence the judgments. Among these variable, we highlight the individual characteristics of the speech stimuli used for capturing the speech samples to be judged; the procedures for the perceptual evaluation; the choice of statistical methods for comparison, among other variables[5,11,12,17,18]. In the present study, care was taken to select participants with the same type of cleft, operated by the same surgeon, within the same age range, in an attempt to minimize external variables that can impact the results of assessment of nasality after primary palatoplasty for correction of cleft lip and palate.

The difficulty in obtaining reliable judgments has been evidenced in the literature since the auditory-perceptual assessment of nasality has proven to be a challenging task[6,10,17,19,20].

In this study, there were differences regarding the identification of hypernasality between live perceptual assessment (as retrieved from patient´s charts) and auditory-perceptual judgements of recorded samples established by judges, particularly for HPC speech sample. When analyzing HPC and LPC samples, it was observed that the set of HPC samples had a higher incidence of high vowels. Researchers have reported that different vowels produce significantly different patterns of nasality, with the height of the tongue influencing the acoustic characteristics of each sample and the degrees of perception of nasality[21]. Particularly, for individuals with hypernasality and repaired cleft palate, it was emphasized by some authors[21,22] that listeners perceive high vowels as more nasal than low vowels, both in isolation and in sentences. A trend for the listeners in this study to perceive the set of HPC samples (with higher percentage of high vowels) as more nasal than the set of LPC samples with lesser percentage of high vowels therefore was expected. The literature[21,22] have also emphasized that listeners perceive low vowels as more nasal than high vowels for normal speakers. One study found that low vowels are produced by normal speakers with lesser strength of the velopharyngeal closure than high vowels[23]. One could infer that high vowels as well as high pressure consonants would require a more precise velopharyngeal closure than low vowels and low pressure consonants, and the finding from this study showing higher rate of absence of hypernasality for HPC samples may be due to the difference in phonetic content between HPC and LPC stimuli. Further studies with greater control of the phonetic context of the speech stimuli may help clarify this aspect.

Another factor that could have interfered in the perceptual judgment of the recordings in this study was the length and the phonetic context of the speech stimuli rated, with each set of recorded speech consisting of five phrases with a large variety in terms of phonetic context particularly for the HPC samples. Some authors[24] summarized findings from the literature about the effect of the length of stimulus on listener reliability. The reported findings showed that listener reliability for rating nasality was higher for sentences than for single words, and higher for single words, than isolated vowels. Data from this study, however, did not clarify whether the length of the stimuli or stimulus with lower variety of sounds may favor the perceptual judgments of recorded samples. In the presence of varied phonetic content there is a risk of listeners to pay attention to isolated parts of the speech stimuli, or even in aspects not related to the nasality of speech, such as distortion caused by audible nasal air escape, nasal turbulence, use of compensatory articulation or even dysphonia, leading to larger variation in judges reliability measures.

The literature also indicates the use of anchor stimuli (reference samples) either to calibrate the judges or to improve the reliability of their judgments[25] in attempts to improve the task of rating nasality. The present study used reference samples during both, the training/calibration of the judges and also during the judgments when no consensus was achieved. The reference samples were used particularly when the recordings were rated with mild hypernasality. The threshold between normality and a speech disorder is considered of clinical importance because after identifying hypernasality, particularly, the clinician usually refers the patient for further assessments and/or monitoring and/or more intervention (secondary surgery, for example). Usually instrumental evaluations are obtained from more invasive techniques which allow visualization of the structures of the velopharyngeal mechanism such as nasoendoscopy and videofluoroscopy. Instruments that provide the possibility of analyzing acoustic aspects (as nasometry) or aerodynamic aspects of speech (such as pressure-flow technique) can used only to corroborate perceptual findings, since they allow clinicians only to infer the adequacy or inadequacy of the velopharyngeal function[5,6,8,26]. As with prior literature, the variation of outcome as identified with different modalities of assessment with this study also point towards the importance of combining different methods for assessment of speech and velopharyngeal function in clinical practice. Special care is needed during the process of identifying the best approach to treat the detected disorders particularly when the perceptual findings are within the range of mild speech disorder or marginal velopharyngeal dysfunction. The findings of this study, therefore, agree with those reported in the literature[14] that suggest that it is difficult for the human ear to identify aspects of disordered speech that are close to the normal threshold.

The present study revealed that more than half of the samples were judged with absence of hypernasality (69% identified during live auditory-perceptual assessment, 50% identified by multiple judges during ratings of samples produced with HPC stimuli and 62% identified by multiple judges during ratings of samples produced with LPC stimuli). While looking into de samples rated with presence of hypernasality, most were identified as representative of mild hypernasality (23% identified during live auditory-perceptual assessment, 36% identified by multiple judges during ratings of samples produced with HPC stimuli and 32% identified by multiple judges during ratings of samples produced with LPC stimuli). Important to consider, however, that while doing the ratings of recorded samples the judge had to choose between the absence or presence of hypernasality in a situation distinct from the ratings established during the live clinical assessment. That is, since the live ratings were not established simultaneously with the recordings, the speakers with borderline nasality may have presented with minor changes in nasality which were not equally detectable under both rating conditions (live vs recorded). The level of agreement between the live ratings during clinical assessment and the multiple judges' ratings of recorded samples, therefore, may be influenced by the difficulty that the human ear has for distinguishing between the intervals of the scale used to rate nasality, especially with the samples within the threshold

between normal and disordered speech. Kappa statistics, particularly, requires an equal distribution of samples across all the intervals of the scale, what was not observed in this study and is not possible (nor warranted) during clinical practice. Most recordings were representative of speech without hypernasality, followed by the group with mild hypernasality, and only few cases with moderate hypernasality. Therefore, even though there was a high percentage of agreement between live ratings and judgment performed by judges for both HPC and LPC stimuli (HPC=79%, LPC=80%), Kappa statistics showed agreement of 0.57 and 0.55 between live ratings and HPC and LPC stimuli, respectively. Interestingly, when comparing live ratings with the judgment of samples recorded during production of HPC stimuli, there was a significant difference in the occurrence of hypernasality. The difference between live ratings and LPC stimuli, however, was not significant, even though Kappa statistics were very similar for both stimuli. This finding was also reported in literature[27].

The choice of judges, for both modalities of assessment, is also an important variable while establishing speech nasality outcome. In the present study, the live auditory-perceptual judgment was established during the clinical evaluation of the patient and was performed by one of three possible SLPs that worked at the institution. While experienced with the speech evaluation of patients with CLP each SLP presented with their own internal pattern for rating nasality. When listeners classify characteristics of speech or voice to some criterion of quality, they compare the stimulus presented to an internal standard or scale (like a "personal scale"). These internal standards are developed and maintained within each judges' memory and may be different from listener to listener. Moreover, the internal standard that one judge use while performing perceptual rating are inherently unstable and can be influenced by internal factors such as lapses in memory and attention, and by external variables such as acoustic context and listening tasks[25,28]. Some studies recommend that the analysis of speech data in subjects with CLP should be made by independent speech pathologists[12] while others recommend that the perceptual judgment of hypernasality has more credibility and reliability when it is done by speech pathologists in consensus, using speech samples audio or video recorded[11,13].

The perceptual assessment, while the "gold standard" measure, has the advantage of not requiring special instruments/equipment and therefor is a low cost procedure. As disadvantage, however, it does not directly evaluate the function and structures of the velopharyngeal mechanism, and also is subjective and susceptible to poor reliability among examiners[26]. Audio recordings of speech samples are the most frequent means for the documentation of speech outcome, and it has the advantage of being easily retrieved, edited and presented for auditory-perceptual judgment by judges, allowing for measures of intra and interjudge reliability, providing means to corroborate the findings of live evaluation. The literature, however, still points out the need for standardized protocols for the collection of speech samples while indicating the importance of using recording equipment of good quality[10,12,13]. While using audio recordings instead of live judgments to establish speech outcome, it is possible to obtain information about speech production that is captured without the influence of the expectations of the clinician, free of visual information or other information about each individual patient. The live judgment of nasality, conversely, considers all the data available to make a full clinical judgment, however, without the ability to reproduce the situation to verify the reliability[12,14].

The moderate agreement between the two methods of assessment, therefore, also could have been affected by the fact that recorded samples are more limited regarding the information available to the listener compared to a live evaluation, that is, although the recordings could be heard many times as necessary for the judgment, there was no possibility of obtaining visual or individual patient information which affect the clinician while doing a live assessment. The SLP who performed the live ratings had access to medical records, to the results of prior assessments and also other information, such as facial and body expression. Studies providing greater control of live ratings and including access to the medical records and also to prior assessment during ratings of recorded samples are needed to clarify these aspects. Another important issue to be considered in future studies is the time when the judgments of the speech nasality were performed. The judgement of recorded speech was performed by the multiple judges within a single day while the live clinical rating was retrieved from patient's charts and performed during clinical evaluation of the patients within a period of five years (2006-2010). In general, when we consider the nasality findings obtained with the two methods of assessment studied, there is great similarity between the findings, even when we take into account the different variables discussed above. Additionally, during the judgment of the recorded samples, the judges were able to compare the recordings with reference samples during their attempts to classify hypernasality, and, therefore, had their answers "anchored" which establishing consensus as reported by other authors[25]. Gerrat et al.[25] compared the perceptual judgments of normal and rough voices using a 5-point equal-appearing interval scale and a scale with explicit anchor stimuli for voices, the authors found that the ratings made using the anchored scale were significantly more reliable than those gathered using the unanchored traditional scale.

Finally, in order to increase the understanding of hypernasality and its evaluation, different aspects should be further investigated and controlled in future studies, including the effect of auditory training, the use of reference samples as anchor stimuli, the type of scales and also the variation among the intervals of the scale. Studies addressing the perception of nasality in the presence of associated speech errors such as

the use of compensatory articulation, audible nasal emission with and without nasal turbulence, intraoral weak pressure and speech distortions related to dental-occlusion conditions are also needed.

CONCLUSION

Percentage absence of hypernasality was higher while presence of mild hypernasaliity of lower when identified during live auditory-perceptual rating when compared to the judgement of recorded samples by multiple judges. Speech outcome established by live ratings, however, have the disadvantage of not being able to be reproduced making comparisons among different cleft palate team more difficult. Further studies with control of the phonetic context of the speech samples used for both, live and recorded ratings, are warranted to help improving the reliability of the auditory-perceptual judgment of speech nasality.

REFERENCES

1. D'Antonio L, Scherer N. Communication disorders associated with cleft palate. In: Losee JE, Kirschner RE, organizers. Comprehensive cleft care. New York: Mc Graw Hill Professional; 2008. p.114-35.
2. Bae YC, Kim JH, Lee J, Hwang SM, Kim SS. Comparative study of the extent of palatal lengthening by different methods. Ann Plast Surg. 2002;48(4):359-62. http://dx.doi.org/10.1097/00000637-200204000-00004
3. Farzaneh F, Becker M, Peterson AM, Svensson H. Speech results in adult Swedish patients born with unilateral complete cleft lip and palate. Scand J Plast Reconstr Surg Hand Surg. 2008;42(1):7-13. http://dx.doi.org/10.1080/02844310701694522
4. Koh KS, Kang BS, Seo DW. Speech evaluation after repair of unilateral complete cleft palate using modified 2-flap palatoplasty. J Craniofac Surg. 2009;20(1):111-4. http://dx.doi.org/10.1097/SCS.0b013e318195ab0a
5. Kummer AW, Clark SL, Redle EE, Thomsen LL, Billmire DA. Current practice in assessing and reporting speech outcomes of cleft palate and velopharyngeal surgery: a survey of cleft palate/craniofacial professionals. Cleft Palate Craniofac J. 2012;49(2):146-52. http://dx.doi.org/10.1597/10-285
6. Kuehn D, Moller K. Speech and language issues in the cleft palate population: the state of the art. Cleft Palate Craniofac J. 2000;37(4):348-83. http://dx.doi.org/10.1597/1545-1569(2000)037<0348:SALIIT>2.3.CO;2
7. Brunnegård K, Lohmander A. A cross-sectional study of speech in 10-year-old children with cleft palate: results and issues of rater reliability. Cleft Palate Craniofac J. 2007;44(1):33-44. http://dx.doi.org/10.1597/05-164
8. Smith B, Guyette TW. Evaluation of cleft palate speech. Clin Plast Surg. 2004;31(2):251-60. http://dx.doi.org/10.1016/S0094-1298(03)00123-8
9. Kummer AW. Perceptual assessment of resonance and velopharyngeal function. Semin Speech Lang. 2011;32(2):159-67. http://dx.doi.org/10.1055/s-0031-1277718
10. 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
11. Henningsson 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 J. 2008;45(1):1-17. http://dx.doi.org/10.1597/06-086.1
12. Sell D. Issues in perceptual speech analysis in cleft palate and related disorders: a review. Int J Lang Commun Disord. 2005;40(2):103-21. http://dx.doi.org/10.1080/13682820400016522
13. Lohmander A, Willadsen E, Persson C, Henningsson G, Bowden M, Hutters B. Methodology for speech assessment in the Scandcleft project-an international randomized clinical trial on palatal surgery: experiences from a pilot study. Cleft Palate Craniofac J. 2009;46(4):347-62. http://dx.doi.org/10.1597/08-039.1
14. Sweeney T, Sell D. Relationship between perceptual ratings of nasality and nasometry in children/adolescents with cleft palate and/or velopharyngeal dysfunction. Int J Lang Commun Disord. 2008;43(3):265-82. http://dx.doi.org/10.1080/13682820701438177
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74. http://dx.doi.org/10.2307/2529310
16. World Health Organization. Global strategies to reduce the health-care burden of craniofacial anomalies: report of WHO meetings on International Collaborative Research on Craniofacial Anomalies; Geneva; 2000 May 24-26. Geneva: World Health Organization; 2002.
17. Wyatt R, Sell D, Russell J, Harding A, Harland K, Albery E. Cleft palate speech dissected: a review of current knowledge and analysis. Br J Plast Surg. 1996; 49(3):143-9. http://dx.doi.org/10.1016/S0007-1226(96)90216-7
18. Kent R. 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
19. Keuning KH, Wieneke GH, Dejonckere PH. The intrajudge reliability of the perceptual rating of cleft palate speech before and after pharyngeal flap surgery: the effect of judges and speech samples. Cleft Palate Craniofac J. 1999;36(4):328-33. http://dx.doi.org/10.1597/1545-1569(1999)036<0328:TIROTP>2.3.CO;2
20. Persson C, Lohmander A, Elander A. Speech in children with an isolated cleft palate: A longitudinal perspective. Cleft Palate Craniofac J. 2006;43(3):295-309. http://dx.doi.org/10.1597/04-071.1
21. Bae Y, Kuehn DP, Ha S. Validity of the nasometer measuring the temporal characteristics of nasalization. Cleft Palate Craniofac J. 2007;44(5):506-17. http://dx.doi.org/10.1597/06-128.1
22. Lewis KE, Watterson T, Quint T. The effect of vowels on nasalance scores. Cleft Palate Craniofac J. 2000;37(6):584-9. http://dx.doi.org/10.1597/1545-1569(2000)037<0584:TEOVON>2.0.CO;2
23. Kuehn DP, Moon JB. Velopharyngeal closure force and levator veli

palatini activation levels in varying phonetic contexts. J Speech Lang Hear Res. 1998;41(1):51-62. http://dx.doi.org/10.1044/jslhr.4101.51

24. Watterson T, Lewis K, Allord M, Sulprizio S, O'Neill P. Effect of vowel type on reliability of nasality ratings. J Commun Disord. 2007;40(6):503-12. http://dx.doi.org/10.1016/j.jcomdis.2007.02.002
25. Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. J Speech Hear Res. 1993;36(1):14-20. http://dx.doi.org/10.1044/jshr.3601.14
26. Smith BE, Kuehn DP. Speech evaluation of velopharyngeal dysfunction. J Craniofac Surg. 2007;18(2):251-61. http://dx.doi.org/10.1097/SCS.0b013e31803ecf3b
27. Watterson T, Lewis KE, Deutsch C. Nasalance and nasality in low pressure and high pressure speech. Cleft Palate Craniofac J. 1998;35(4):293-8. http://dx.doi.org/10.1597/1545-1569(1998)035<0293:NANILP>2.3.CO;2
28. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. J Speech Hear Res. 1993;36(1):21-40. http://dx.doi.org/10.1044/jshr.3601.21