AUDITORY-PERCEPTUAL JUDGMENT OF VELAR STOPS ASSOCIATED TO CLEFT PALATE BY JUDGES WITH AND WITHOUT EXPERIENCE

Julgamento perceptivo-auditivo das oclusivas velares associadas à fissura labiopalatina por juízes com e sem experiência

Viviane Cristina de Castro Marino[1], Tayrine Sousa Marques Borges[1],

Mayalle Rocha Bonfim Jurado[1], Mariana Rodrigues Canales[1],

Aveliny Mantovan Lima-Gregio[2], Jeniffer de Cássia Rillo Dutka[3]

ABSTRACT

Purpose: to verify: a) identify the level of agreement amount judges during auditory-perceptual ratings of velar plosive sounds before and after speech therapy, b) verify if the phonetic context of the speech samples affect judges' agreement, and c) to compare the ratings among judges with and without experience in rating CA. **Methods:** speech samples of children with cleft lip and palate, 30 before and 30 after speech therapy, and 30 samples from a child without cleft lip and palate and with normal speech were rated by a group of 9 judges. Three SLPs established the gold-standard ratings used as reference for comparisons. Six other judges rated the samples for this study: three considered experienced (SLPs) and three non-experienced (students). The speech samples rated involved velar consonants /k/ and /g/ and vowels /a/, /i/ and /u/. Judges were instructed to rate presence or absence of velar consonants or presence of CA. **Results:** Kappa statistics revealed moderate agreement among experienced judges and low agreement among the judges without experience for samples recorded before speech therapy. Phonetic context had an effect on the ratings before and after speech therapy. Ratings were significantly better among experienced judges before speech therapy (p-valor <0,001). **Conclusion:** judges' experience and phonetic context of speech samples had an effect on ratings of CA.

KEYWORDS: Cleft Palate; Speech; Speech Disorders; Auditory Perception

■ INTRODUCTION

Among the speech disorders that can be found in the population with cleft lip and palate (CLP), we can highlight the compensatory articulations (CA), which are deviations in the production of sounds that are established in the early stages of phonological acquisition to functionally compensate for altered velopharyngeal mechanism. Such productions alter the articulation placements and, in most cases, are produced in the regions of the vocal tract posterior to the velopharynx[1]. When present, the CA can make the child's or the adult's speech unintelligible and bring in prejudices to their social lives, studies and even to their professional lives[2,3]. Speech therapy is essential for the establishment of standard production and different intervention approaches have been reported in the literature in order to facilitate typical speech production in this population[4,5]. It requires from the speech-language pathologist extensive knowledge about the CA in

order to identify these productions and also establish appropriate therapeutic planning [2].

The auditory perceptual evaluation is the initial procedure used by the speech-language pathologist for the identification and characterization of speech disorders associated CLP[6]. In clinical practice, it is through the auditory judgment that the speech-language pathologist characterizes the speech disorders prior to treatment and make sure if the presence of typical speech after intervention. Considering the perceptual evaluation as essential for the characterization of speech disorders associated with FLP procedure, researchers have worried about possible aspects that can influence the interpretation of the results of this assessment[6-8]. Among these aspects we can point out intra and inter judges agreement estimated through specific procedures[7,9-13], the evaluator's experience in performing auditory-perceptual judgment on these productions [6,10,12] and the selection of the speech samples used in the perceptual evaluation of the speech[6,10,14,15].

Particularly in regards to the agreement of judges, literature emphasizes the importance of verifying this aspect in order to allow comparisons of results in treatment in multicenter[7,16] longitudinal studies[13] or, also, involving different surgical techniques[17]. In an article of critical literature review on perceptual assessment of speech in patients with CLP, researchers[6] found that approximately half of the analyzed articles included measures of reliability (as agreement percent or kappa coefficient), and verified lack of information on reliability of judges in 49% of the analyzed material. In general, the literature has reported low inter judge agreement on perceptual assessment of speech[9,10,12]. One study in particular found poor agreement between professionals (with and without experience), when the phonetic transcription was used as a procedure to register the identified CA in the productions of the speech of individuals with CLP, which led the authors to conclude the need to extend training in the phonetic transcription to identify the CA[10]. Other procedures to assess articulatory productions were reported in different studies, including the identification of the presence/absence of typical speech, the calculation of the percentage of correct consonants, frequency and type of changes or description of alterations[6]. Although such procedures are referenced on them, in recent years researchers have advocated the phonetic transcription by professionals with experience in this task. Recent studies report agreements of 80-90% inter - transcriptions[11,18]. However, structured trainings prior to conducting the judgments of speech are recommended in order to facilitate judges in the proposed tasks[16], especially when samples from different languages are judged[8].

Regarding the speech samples used for the identification of CA, in general, the literature has reported variability in sample selection among different studies, including spontaneous speech, texts, sentences, words, isolated consonants and automatic speech (used in an isolated or combined form)[6]. One study, in particular, used controlled words constituted by consonant+vowel+consonant (CVC) inserted in a vehicle sentence to examine the degree of agreement of responses obtained with and without professional experience on identifying the CA[10]. The use of speech samples phonetically controlled and standardized is recommended in the literature in order to enable comparison of results, especially in multicentric studies[14].

Besides the aspects mentioned above, the experience of the evaluator in perceptually identify the CA is considered by the literature as an aspect that can influence the judgments of speech. One study in particular found greater agreement in the identification of CA by professionals with experience in assessing the CLP population, when compared to that obtained by non-experienced professionals, using the phonetic transcription to record the productions[10]. In another study, the authors found low agreement among experienced judges in identifying the parameter of nasality and moderate to good agreement in the identification of other investigated parameters (including articulatory production). Overall, the authors attributed the results of agreement obtained in their study to the need of training with reference samples, even for the experienced judges, in order to improve agreement results in assessments[12]. More recently, a study compared judgments of aspects from speech (including CA) of untrained judges (staff and first year students) to judgments of trained speech-language pathologists. The judged speech samples were produced by children with cleft palate or lip/palate and contained two sets of sentences. There was a low incidence of articulatory difficulties presented in the samples and, therefore, it was not possible to ascertain the degree of inter judge agreement regarding this parameter of speech. However, it was found that judges, with and without experience, identified the only two children with changes in articulatory production)[19].

By analyzing the various aspects presented in the literature that may influence the perceptual speech analysis, a study of literature review has brought questions about the extent to which the experience of the judge, when considered as a single factor, ensures satisfactory degree of agreement between judges[7]. Another study[16], when using a procedure specifically designed for training the speech assessment of CLP population, obtained good agreement between judges in the identification

of non-oral consonants (productions in the pharynx and glottis). From the results obtained, the authors argue in favor of judgments held after structured training sessions involving the use of pre-recorded speech samples, in contrast to the trials performed counting only on prior experience of the speech-language pathologist. In overall, the literature suggests that judgments to be made by multiple judges with experience in assessing aspects of speech presented by the CLP population[7].

Based on these, it appears that the auditory-perceptual evaluation of speech is subject to many variables that can influence the results, requiring special care for its implementation and interpretation. There is also a major concern from the researcher to recognize and discuss factors that may interfere with the identification of the CA as well as prepare speech-language pathologists for the task of identifying these changes. This task, which is performed by perceptual evaluation, is essential for clinicians and researchers, as it depends on the definition of conduct on the need for treatment and monitors the therapeutic results. Information about agreement in auditory-perceptual judgments obtained in the pre and post speech therapy, taking into account the experience or not from the judge and the phonetic composition of speech samples can contribute to a better understanding of CA speech of children with CLP. The objectives of this study were: a) to verify the degree of agreement of judges (with and without experience) in the auditory perceptual judgment of the production of velar stops before and after speech therapy in relation to judgments of reference ("gold standard"); b) to check the possible influence of the phonetic composition of speech samples in this agreement, and c) to verify whether the judgments made by judges with experience differ from those obtained by judges without experience, under the conditions investigated (pre/post- speech therapy and control).

■ **METHODS**

This prospective study included the auditory-perceptual judgment of speech samples from a child with CLP, before and after speech therapy. It also involved the trial of auditory perception speech samples from a child without CLP who presents typical speech, i.e. normal speech development (control). Consensual judgments ("gold standard") were established by three speech-language pathologists from a prestigious center specialized in craniofacial anomalies for the speech samples included in this study. Three speech-language pathologists with experience of at least five years in speech assessment of children with CLP judged

the same speech samples and were called, in the present study, "experienced judges". Three students starting the Phonoaudiology course judged the speech samples included in the study and were named "inexperienced judges".

The study was conducted in two stages, being submitted and approved by the Ethics Committee in Research by the institution of origin, under the numbers 0347/2011 and 0609/2012 and was considered riskless .

Speech samples stored in the database

Speech samples analyzed in the study are part of a database maintained by the Laboratory of Acoustic Analysis – "Laboratório de Análise Acústica" - LAAC from the institution. At this bank, speech production of the patients undergoing speech therapy is systematically recorded before and after the interventions. Particularly, the recordings used in this study were obtained from a five-year-old child girl with operated cleft palate (CP). Prior to speech therapy, this child had compromised intelligibility in the speech due to the use of CA (glottal stop) as identified in clinical evaluation (live) which consisted of repetition of words and spontaneous conversation. The child participated in a program of speech therapy directed to the establishment of the production of oral articulation points as opposed to the use of compensatory articulation, focusing on stop consonants. The speech therapy program involving stop consonants was done weekly, with two 40-minute sessions each week for a period of four months. It was also used in the study, the recordings stored in the database of the LAAC obtained from a five-year-old girl without cleft palate and with typical speech (control).

For children with CP, the repetitions were obtained in the two studied conditions, before and after speech therapy. For this study, therefore, a total of 90 samples were recorded, being 60 samples from the child with CP, considering the plosives /k/ and /g/, and the vowels /a/, /i/, /u/, the two conditions studied (pre and post speech therapy), and also 30 samples of the child with typical speech (control), considering the plosives /k/ and /g/, and the vowels /a/, /i/, /u/. So, for the child with CP: 5 (repetitions) X 3 (vowels) X 2 (occlusive) X 2 (condition) = 60 samples, and the control child: 5 (repetitions) X 3 (vowels) X 2 (occlusive) = 30 samples. In total, therefore, 90 recordings stored in the database of the LAAC were of interest to the study.

It is noteworthy that a single evaluator conducted the recordings of the child's speech with and without CP (pre and after speech therapy). Speech samples were recorded in the same acoustically treated room (LAAC), using high-fidelity digital equipment (digital recorder MARANTZ, unidirectional microphone

Shure). The microphone was positioned 20 cm from the mouth of the children. The digitized recordings were stored in a computer.

Sample preparation for analysis by judges

For the study, existing recordings were edited and stored for auditory-perceptual judgments through *PRAAT software*. This material comprised a total of 90 samples randomly edited. The edited material was archived on CD-ROM and sent to the judges who established judgments considered "gold standard". Then, the same material was sent to each of the judges (with or without experience). Along with the edited material from the two children in the study, it was included on the same CD recorded reference audio samples to the judges. These reference samples were representative of each type of production (typical speech, with the omission of segment or presence of CA) and belonged to other subjects who were not included in this study.

Establishment of the "gold standard"

Three speech-language pathologists have established the reference ("gold standard") of the 90 speech samples. These professionals belong to a prestigious complexity in the treatment of craniofacial anomalies and have worked at this place for over 5 years and have extensive experience in assessing the speech of the CLP population. Speech-language pathologists reported having normal hearing, no contact with the subjects who had recorded their statements and had not received information about the objective of the study.

It was explained to the speech-language pathologists that their judgment would serve to identify the presence, absence of occlusive velar /k/ and /g/, or even the presence of CA. Prior to the judgments, instructions and reference samples (recorded audio) representative of each type of production were offered. When presenting the samples of references, it was nominated which was the kind of production that should be judged by directing judges that they should use this information as a parameter to perform their judgments. The judges were instructed to judge only the presence, the absence of velar plosives or the presence of CA in the initial position of the word, inserted into the sentence, regardless of hearing other compensation and/or hypernasality in the presented sentence. For example, upon hearing the sentence "Fala **ca**pa bem bonita" (Say **ca**se pretty well), the judge should decide between the presence of the consonant /k/ in the first syllable, the absence of occlusive /k/ in the first syllable or the presence of CA on the first syllable. The judges were not informed about the conditions under which the samples were obtained.

After the instructions, the judges had the opportunity to listen simultaneously, in the same room, the speech samples using individual headphones. Judges were allowed to listen to the samples as often as they thought necessary, and also could adjust the volume of recorded samples. After hearing (once or more) each speech sample, the judges wrote down their option (presence/absence of velar plosives or presence of CA) on the sheets of paper made for this purpose and then verified their answers. In case of disagreeing on their judgments, speech-language pathologists listened again to obtain consensual ones (one unique judgment for each of the 90 samples heard). The consensus judgment was called, in this study, "gold standard" and these judgments were reported to verify agreement with the judgments of other participants (judges with or without experience).

Auditory-perceptual judgment: experienced judges

Three other speech-language pathologists within the same center of high complexity in the treatment of craniofacial anomalies, with more than 5 years of experience in evaluating the speech of the CLP population, judged the 90 speech samples, but individually, to meet the first objective of present study (i.e., verify the agreement between their judgments and the "gold standard"). These speech-language pathologists reported having normal hearing, no contact with the subjects who had recorded their statements and had not received information about the study objective.

The three speech-language pathologists identified the presence, absence of velar plosives /k/ and /g/, or even the presence of CA in the speech samples presented, which were the same used for establishing the "gold standard". In addition, the instructions for the judgment of the three experienced speech-language pathologists followed those described for establishing the "gold standard", with one difference: every experienced speech-language pathologist heard with individual headphones the recorded material (90 samples) in a room reserved for this purpose. Thus, 90 trials were made separately and written down in a sheet similar to that used to establish the "gold standard" registration.

Auditory-perceptual judgment: judges without experience

Three students enrolled in the second year of the undergraduate course of Phonoaudiology were included in the study and considered "inexperienced" in speech evaluation of CLP population because they had not started the specific disciplines at the

Phonoaudiology course, nor had been exposed to clinical activities. These students reported having normal hearing, no contact with the subjects who had recorded their statements and had not received information about the study objective.

The three students judged individually the 90 speech samples in order to meet the first objective of the present study (i.e., verify the agreement between their judgments and the "gold standard"). These three students identified the presence, absence of velar plosives /k/ and /g/, or even the presence of CA in the speech samples presented, which were the same used for establishing the "gold standard". In addition, the instructions for the judgment of three students with no experience followed those described for establishing the "gold standard", with one difference: each student heard the recorded material (90 samples) , with individual headphones, connected to computers separated in rooms reserved for this purpose. Every student finished their judgments at the same time and 90 judgments were made separately written down in a sheet similar to that used for the other participants in this study.

Data Analysis

The obtained responses were presented as percentages of agreement for each studied condition (pre-speech therapy, post speech therapy, control), taking into account the judgments obtained for each of the velar consonants (/k/ = 45 judgments and /g/ = 45 judgments) and judgments obtained for each of the vowels (/a/ = 30 judgments, /i/ = 30 judgments and /u/ = 30 judgments). More specifically, the judgments were analyzed by consonants (/k/ or /g/) or vowels (/a/, /i/ and /u/), in order to allow further statistical analysis (Kappa agreement) of data.

The Kappa agreement index was also used to measure the degree of agreement between auditory-perceptual judgments of interest ("gold standard" x judges with experience and "gold standard" x inexperienced judges) among the studied conditions (pre speech therapy, post speech therapy, control) and the samples used (velar stop consonants and vowels). The Kappa statistic is a measure used to verify inter - judges agreement that fixes the agreement obtained by chance (distance the given observations from the expected ones made by chance, indicating how legitimate are the

interpretations). This conservative approach was also used, along with the percentage of agreement in previous studies that aimed to obtain inter - judge reliability in judgments of speech disorders presented by the CLP population[12,20]. As reported in the literature[12] critics comment that in the Kappa the expected agreement is a source of concern because the evaluators are not statistically independent and the agreement expected by chance is based on the assumption of independent evaluators.

In the present study, the Kappa values were interpreted according to literature[21] in which: 0.00 does not indicate agreement; 0.00 to 0.20 indicates poor agreement; from 0.21 to 0.40 indicates fair agreement; from 0.41 to 0.60 indicates moderate agreement; 0.61 to 0.80 indicates substantial agreement, and 0.81 to 1.00 indicates perfect agreement (or almost perfect). Confidence intervals were constructed with 95% statistical confidence and adopted a significance level of 5% ($p < 0.05$). The Kappa coefficient analysis was presented unifying the judgments from the three judges with experience and, similarly, unifying the trials of the three judges without experience, resulting in a single value of Kappa for experienced and inexperienced judges.

The test of equality of two proportions (nonparametric) was used to compare the proportion of responses from two specific variables (judges with and without experience) and/or their levels were statistically significant.

■ RESULTS

The results refer to the agreement in auditory-perceptual judgments of experienced judges and "gold standard" ("experienced/gold") and judges without experience and "gold standard" ("no experience/gold"), at three distinct conditions: pre, post speech therapy and control.

Tables 1 and 2 show the percentage of "experienced/gold" inter-judge agreement obtained in the pre-speech therapy condition. More specifically, Table 1 shows the percentage of agreement for the 45 judgments relating to the consonant /k/ and 45 judgments concerning the consonant /g/. Table 2 on the other hand shows the percentage of agreement for the 30 judgment concerning each of the studied vowels (/a/, /u/, /i/).

Table 1 – Percentage of inter judges agreement (with experience/gold) among the judgments made (omission-O, compensatory articulation-CA or typical speech (T) for velar (/k/ and /g/), in the pre-speech therapy. Occurrences of O, CA and T in regards to each word involving /k/ or /g/ are presented in parentheses

Velar Occlusive	Agreement				Total
	O	CA	T	N (%)	
k (capa/cuca/quibe)	11 (11/0/0)	25 (0/14/11)	0 (0/0/0)	36 (80%)	45
g (gato/gula/guizo)	14 (11/3/0)	23 (0/8/15)	0 (0/0/0)	37 (82%)	45
Total				73 (81%)	90

Table 2 – Percentage of inter judges agreement (with experience/gold) among the judgments made (omission-O, compensatory articulation-CA or typical speech (T) with vowel (/a/, /u/ /i/), in the pre-speech therapy condition. Results O, CA and T on every word involving /k/ or /g/ are presented in parentheses

Vowels	Agreement				Total
	O	CA	T - %	N - %	Total
a (capa/gato)	22 (11/11)	0	0 (0/0/0/)	22 (74%)	30
u (cuca/gula)	3 (0/3)	22 (14/8)	0 (0/0/0)	25 (84%)	30
i (quibe/guizo)	0	26 (11/15)	0 (0/0/0)	26 (87%)	30
Total				73 (81%)	90

Table 3 summarizes the percentage of inter-judge "experienced/gold" agreement obtained for judgments (usually with a consonant and with a vowel) in the pre- speech therapy condition, besides presenting the Kappa agreement index (usually with a consonant and with a vowel).

The percentage of inter-judge "experienced/gold" agreement obtained for all judgments (generally with a consonant and with a vowel) in the post-speech therapy condition and in the control condition was 100% with perfect Kappa agreement.

Tables 4 and 5 show the percentage of inter-judge "no experience/gold" agreement obtained in the pre-speech therapy condition. More specifically, Table 4 shows the percentage of agreement for the 45 judgments regarding the consonant /k/ to 45 judgments regarding the consonant /g/. Table 5 on the other hand shows the percentage of agreement for the 30 judgments for each of the studied vowels (/a/, /u/, /i/).

Table 6 summarizes the percentage of inter-judge "no experience/gold" agreement obtained for judgments (usually with a consonant and with a vowel) in the pre- speech therapy condition, besides presenting the Kappa agreement index (usually with a consonant and with a vowel).

Table 7 summarizes the percentage of inter-judge "no experience/gold" agreement obtained for judgments (usually with a consonant and with a vowel) in the post-speech therapy condition, besides presenting the Kappa agreement index (usually with a consonant and with a vowel). It is noteworthy that in two, from 90 productions, there was disagreement, and these occurred for "gula" speech sample.

The percentage of inter-judge "no experience/ gold" agreement obtained for all judgments (generally with a consonant and with a vowel) in the control condition was 100% with perfect Kappa agreement.

Table 8 shows the (absolute and relative) distribution of correct answers by the judges (with and without experience), taking into account the three conditions under investigation (control, pre and post speech therapy), and indicate the p values found for the groups of judges and the three investigated conditions.

Table 3 – Percentage of agreement and Kappa agreement index to inter-judges (with experience/gold) to the judgments made by consonant (/k/ and /g/) and vowel (/a/, /u/, /i/) in the pre-speech therapy condition

Pre- speech therapy	Percentage	Kappa	p value
K	80%	0,56	<0,001*
G	82%	0,64	<0,001*
A	74%	0,15	0,112
U	84%	0,44	0,014*
I	87%	0,30	0,02*
General	81%	0,60	<0,001*

* Kappa Coefficient p<0.05

Table 4 – Percentage of inter-judges (no experience/gold) in judgments performed (omission-O, compensatory articulation-CA or typical (T) for velar (/k/ and /g/), on pre-speech therapy condition. Occurrences O, CA and T related to each word involving /k/ or /g/ are presented in parentheses

Velar Occlusive	Agreement				Total
	O	AC	T	N - %	
k (capa/cuca/quibe)	(14/0/0)	(0/5/0)	0 (0/0/0)	19 (42%)	45
g (gato/gula/guizo)	(13/6/0)	(0/0/5)	0 (0/0/0)	24 (54%)	45
Total				43 (48%)	90

Table 5 – Percentage of inter-judges (no experience/gold) in performed judgments (omission-O, compensatory articulation-CA or typical (T) with vowel (/a/, /u/ /i/), in the pre-speech therapy condition. Occurrences O, CA and T related to each word involving /k/ or /g/ are presented in parentheses

Vowels	Agreement				Total
	O	CA	T	N - %	
a (capa/gato)	(14/13)	(0/0)	0 (0/0/0)	27 (90%)	30
u (cuca/gula)	(0/6)	(5/0)	0 (0/0/00)	11 (37%)	30
i (quibe/guizo)	(0/0)	(0/5)	0 (0/0/0)	5 (17%)	30
Total				43 (48%)	90

Table 6 – Percentage of agreement and Kappa agreement index to inter-judges (no experience/gold) for the judgments made by consonant (/k/ and /g/) and vowel (/a/, /u/, /i/) in the pre-speech therapy condition

Pre-speech therapy	Percentage	Kappa	P value
K	42%	0,13	0,002*
G	54%	0,11	0,296
A	90%	0,37	0,008*
U	37%	0,15	0,002*
I	17%	0,6	0,003*
General	48%	0,12	<0,001*

* Kappa Coefficient p<0.05

Table 7 – Percentage of agreement and Kappa agreement index to inter-judges (no experience/gold) to the judgments made by consonant (/k/ and /g/) and vowel (/a/, /u/, /i/) in the post-speech therapy condition

Post-Speech Therapy	Percentage	Kappa	P value
K	100%	1,00	<0,001*
G	95,5%	0,65	<0,001*
A	100%	1,00	<0,001*
U	94%	0,64	<0,001*
I	100%	1,00	<0,001*
General	98%	0,66	<0,001*

* Kappa Coefficient p<0.05

Table 8 - Distribution of correct answers by judges (with and without experience) for the investigated conditions (control, pre and post speech therapy)

	Correct Answers	N	%	P value
Control	No experience	90	100%	1,000
	Experience	90	100%	
Pre	No experience	43	48%	0,001*
	Experience	73	81%	
Post	No experience	88	98%	0,155
	Experience	90	100%	

Equality Test of two Proportions (p<0,05)

■ DISCUSSION

The perceptual assessment is essential for the assessment of speech characteristics of subjects with CLP[6] and the importance of this type of assessment was previously emphasized[7,22]. Professionals (and future professionals) involved in perceptual speech assessment of the CLP population need to be prepared to identify, among other changes, atypical productions (CA) in order to determine the appropriate treatment plan. Thus, it is of interest to investigate the variables that can affect the judgments of these professionals, as the degree of agreement of judges in relation to judgments of reference ("gold standard"), as well as the influence of the phonetic composition of speech samples in this assessment.

In this study, in regards to the agreement between the judgments of underlined experienced judges and "gold standard" for pre-speech therapy condition, there was a percentage of agreement in 81% of the total judgments performed with moderate Kappa agreement, statistically significant. Further analysis of the data revealed that when the judges disagreed in their judgments such disagreement occurred only for the type of change (CA or omission) identified in judged productions. The differences in the two types

of analyzes (percentage of agreement, 81 % and Kappa agreement index, 0.60), can be explained by the fact that Kappa corrects the obtained agreement by chance (separating the made observations from the expected ones by chance). Previous studies have reported differences between these two types of analyzes, with lower Kappa values compared to the percentage of agreement[12,20], probably due to Kappa be a conservative approach that assumes that any agreement that could be obtained by chance was indeed random[12]. According to literature, the option to use both the percentage of agreement as the Kappa index is justified when seeking a more complete description of the data[12].

When considering the results obtained through the two analyzes performed jointly, it was found that although the participants ("gold standard " and experienced judges) have experience in evaluating the speech of the CLP population, the requested task (judging typical speech, omission or presence of CA) may have influenced the judgments. The literature describes the use of graduated scales as the most common method used to evaluate the speech of individuals with CLP[6].

However, a tendency in the literature to present results of speeches from the use of phonetic transcription[11,18] have been observed. Even though

the use of phonetic transcription is desirable[7] it is rarely used in clinical practice involving speeches from the CLP population and often in scientific research[6]. Overall, at the clinic, the atypical consonant productions are described using models proposed in the literature[23]. A study, in particular, to verify the agreement between professionals with and without experience in the identification of CA, obtained poor agreement even for trained judges, when using the phonetic transcription to record found productions, leading the authors to conclude on the need for specific training in phonetic transcription for evaluators who work with individuals who have CA[10]. In the present study, moderate agreement obtained between experienced professionals and "gold standard" (judgments of multiple judges with consensus) can be explained, at least in part, by the fact that the judges had to identify not only the presence or absence of CA, but also the omission of the velar segment. If the task was only to identify the presence or absence of CA, the agreement could have been higher (or even perfect). Higher agreement was found in a study in which the task was to judge the presence or absence of hypernasality when compared to the agreement obtained for the proposed categorizations for the judgment of resonance changes[20]. This information suggests that the requested task can influence the results obtained in the classification of speech disorders associated with CLP, including CA. The auditory-perceptual speech assessment therefore is subject to many variables that can influence the results, requiring special care for its implementation and interpretation. The influence of methodological issues involved in the auditory-perceptual assessment has been emphasized by several authors[6,7]. The need to standardize the task to be performed by speech-language pathologists in clinical and research contexts has also been reported[14], since the selection of samples may affect the obtained results[6], disadvantaging comparisons of results in multicenter studies.

In this study, we also aimed to analyze the data separately taking into account either just the consonants or the vowels. When considering just the consonants, there was a agreement percentage of 80% for the consonant /k/ and 82% for /g/, with Kappa agreement (0.56) for moderate /k/ and substantial (0.64) to /g/, both statistically significant. Jointly, these results suggest that the consonant /g/ showed a tendency to a higher agreement (for omission of segment) among judgments (Table 1). The "gula" speech sample seemed to contribute to the differentiation of agreement between the velar consonants. The fact that this speech sample is constituted by consonantal segment /l/ (requiring

low intraoral pressure) may, in perceptual terms, have generated questions to the listeners, resulting in judgments of both omission of segments (N=3) as CA (N=8), which increased the final percentage of agreement of the consonant [g].

When considering the vowels, the percentage of agreement was 74% for vowel /a/, 84% for /u/ and 87% for /i/, with little Kappa agreement (0.15 not significant) for /a/, regular (0.30 significant) for / i / and moderate (0.44, significant) for /u/, being the differences between the two forms of analysis (percentage agreement and Kappa When considering the vowels, the percentage of agreement was 74% for vowel /a/, 84% for /u/ and 87% for /i/, with little Kappa agreement (0.15 not significant) for /a/, regular (0.30 significant) for / i / and moderate (0.44, significant) for / u /, being the differences between the two forms of analysis (percentage agreement and Kappa agreement index) probably due to the Kappa fixing the agreement obtained by chance (separating the observations made from those expected due to chance). Further analysis of the data revealed that, regardless of the consonant, the correlation (smaller and not significant) for the vowel /a/ always occurred in relation to the judgment of default, suggesting that this vowel in perceptual terms, favors the identification of the absence of occlusive ensure, as shown in Table 2.

As for vowel /u/, the observed correlation was greater in regards to the presence of CA (Table 2). It is noteworthy that out of the 30 judgments involving the vowel /u/ 15 succeeded the consonant /k/. In these, there was agreement regarding the presence of CA in 14 judgments, suggesting that, in perceptual terms, the vowel /u/ may have favored the identification of the presence of CA, in addiction to the fact that the word "cuca" has a high pressure consonant /k/, which may have influenced the identification of CA. Yet, in regards to /g/, in 15 judgments, 8 showed agreement for the presence of CA and 3 for omission of the segment, suggesting that the vowel /u/ with /g/ generated questions for the listener as to the classification of altered speech. It is noteworthy that the word "gula" consists of the consonantal segment /l/, which requires low intraoral pressure, which may have being a disadvantage for the identification of CA. Regarding the vowel /i/, it was observed that she favored the agreement for the presence of CA both before /k/ as to /g/. It is noteworthy that out of the 30 judgments involving the vowel /i/, 15 involved the consonant /k/ (with 11 concordant judgments, indicating the presence of CA) and 15 involved the consonant /g/ (15, i.e., 100 % of the concordant judgments, indicating the presence of CA) (Table 2). These data indicate that

the vowel /i/ favors the identification of the presence of CA, especially before /g/.

Jointly, the results for judgments from experienced judges in regards to "gold standard", in pre-speech therapy condition, suggest that there is variability which could be due to the phonetic composition of the samples, besides the requested task. This was not observed in the post-speech therapy and control conditions, as in both conditions, there was 100 % agreement (Kappa = perfect agreement) among the judgments, suggesting that on typical speech (post-speech therapy or control), neither requested task nor the phonetic-context samples raise questions for listeners. The literature indicates the difficulty of evaluators to identify phonemic categories that do not exist in the native language of the speaker[9], as occurs, for example, in the presence of glottal occlusive CA (also known as a "glottal stop") for the Brazilian Portuguese. Taking into account the influence of the phonetic composition of judgments from the multiple judges, in the last years scholars[14] have recommended the use of speech samples consisted by isolated words or particularly phonetically elaborated sentences, but with recurrence of the target sound. The recurrence of the same sound in the speech samples may facilitate the comparison of results in multicenter studies[14] and, therefore, should be considered in clinical practice and research. The development of standardized protocols for Brazilian Portuguese that meet international recommendations[14] can contribute a lot in studies that seek to verify the agreement among judges, especially if the judgments are carried out by multiple judges.

In this study, it was also of interest to verify the agreement among the judgments of judges without experienced and "gold standard" ("no experience/ gold"). For pre-speech therapy condition, there was a low percentage of agreement (48%) in judgments conducted with small but statistically significant Kappa correlation. Further analysis of the data revealed that while there was agreement, this occurred for samples in which the "gold standard" also judged as a omission and, when there were disagreements, these were among the types of changes identified as the judges without experience tended to assess productions as "omission" while the "gold standard" judged the same productions as "CA". Only two out of the total (N=90) of samples were considered as typical speech by graduate students, when compared to the "gold standard", suggesting that undergraduate students in Phonoaudiology perceive the presence of altered speech, although they are still unable to distinguish between omission of the velar stops (/k/, /g/) and presence of CA. A previous study[10] investigated agreement between

judgments of judges with and without experience in identifying CA in relation to the "gold standard ". Results indicated greater agreement in the identification of CA by professionals with experience in the assessment of subjects with CLP when compared to that obtained by professionals without experience, when the phonetic transcription was used. Another study found poor agreement, even among experienced judges in identifying atypical productions, attributing this finding to the lack of sufficient training of the evaluators[12]. Overall, the data from these studies highlight the need for specific training for the identification of compensatory productions, which should be started since graduate school. It is known that the period during the undergraduate course of Phonoaudiology represents an important opportunity for professional training to prepare future professionals for specialized clinical assessment. The use of materials that favor such experience is recommended for these future professionals.

When analyzing the data separately and taking into account only the consonants, there was agreement percentage of 42% for the consonant /k/ and 54% for /g/, with little Kappa agreement, but statistically significant for /k/. A more detailed analysis of the data suggested that while there was agreement among judges without experience and "gold standard", this occurred more times as for the omission of the segment, for both consonants (Table 4). When considering the vowels, the percentage of agreement was 90% for /a/ , 37% for /u/ and 17% for /i/, with regular Kappa agreement (0.37, significant) for /a/ and small for /u/ (0.15, significant) and /i/ (0.6, significant) .

These data suggest that while the vowel /a/ favored the agreement for omission of the segment, the vowels /i/ and /u/ favored the agreement (albeit low) for the presence of CA. Jointly the results suggest that even with low agreement between judges (no experience and "gold standard"), it seems to have been influenced by the phonetic context of the speech samples. Thus, for clinical and research purposes, it is suggested to take into account the influence of phonetic context on the results obtained. Furthermore, it is suggested the use of speech assessment based on global standards protocols as recommended internationally in 2008[14]. Whether used in classes or supervised training graduation, these samples may contribute to the development of the skills to identify CA for future professionals, empowering them including the judgment of other speech samples, such as those in vehicle sentences or spontaneous speech. In general, recent studies using phonetic transcriptions indicated agreement of 80 and 90% among judgments[11,18], contradicting previous studies that found lower agreement[7,9,10,12].

These data suggest that the agreement among evaluators can be high with proper [16].

For post-speech therapy condition, there was agreement percentage of 98% (88/90), with Kappa agreement index of 65.2% for /g/ and 64.4% for /u/, both statistically significant. Further analysis of the data showed that the vowel /u/ has raised questions for the judges with no experience for this condition, and judges without experience judged two samples of speech "gula" as altered speech (one as omission and another as CA), which suggests that for judges without experience, the production of speech after speech therapy may not yet be enough to be rescued aurally as typical in all productions. This was not observed in the control condition, as there was 100% agreement (Kappa=perfect agreement) among the judgements, suggesting that under-graduates at initial series are able to identify the speech samples consisting of velar consonants in subjects with typical speech (no change history), regardless of the phonetic context in which they find themselves .

Finally, the present study aimed to determine whether there are differences in judgments made by judges with and without experience in the three conditions (pre-speech therapy, post-speech therapy and control). The results showed that there is a statistical difference between groups, with a higher percentage of correct answers for experienced judges (81%), which is statistically different from the 48% observed for undergraduates (p - value <0.001). This analysis confirms that, in this study, both judges with experience (speech-language pathologists) and inexperienced judges (undergraduate students) had difficulties in identi-fying the CA when the requested task involved not only the identification of the presence of CA or typical speech, but also to differentiate between CA and omission of segments. The use of phonetic transcription, even though it is recommended, could also have hampered the task of judges with no experience in judging the presented speech samples. In this study, the judgments made were made using words inserted in vehicle sentences approaching the speech samples used in the previous study[10]. The agreement obtained could be higher if the speech samples were constituted by sentences with recurrence of sounds which were presented to the judges, with or without experience. A study, now underway, is intended to determine the effect of using different speech samples in regards to the judgments of evaluators experience regarding the "gold standard" (multiple judges)

■ CONCLUSION

The data obtained in this study suggested that the requested task, the phonetic composition of speech samples and the experience of the judges influenced perceptual judgments, particularly for pre-speech therapy condition. Overall, the study data indicate the need to select the speech samples that can better respond to questions in relation to atypical productions associated with CLP. For example, when investigating acoustic facts, the use of words inserted in vehicle sentences is recommended to control possible variables that could impact on measures of interest. Yet, when investigating speech through auditory-perceptual assessment, the use of standardized protocols involving speech samples with recurrence of sounds can encourage speech analysis by different judges.

RESUMO

Objetivo: verificar: a) o grau de concordância de juízes no julgamento perceptivo-auditivo da produção de oclusivas velares, antes e depois da fonoterapia; b) a possível influência da composição fonética das amostras de fala nesta concordância e c) se os julgamentos obtidos por juízes com experiência diferem daqueles obtidos por juízes sem experiência, nas condições investigadas. **Métodos:** 60 amostras de fala de uma criança com fissura labiopalatina (30 previamente e 30 posteriormente a fonoterapia) e 30 amostras de fala de uma criança com fala típica (normal) foram julgadas por um grupo de 9 juízes. Três fonoaudiólogos estabeleceram os julgamentos consensuais "padrão ouro" para o estudo. Seis outros juízes julgaram as amostras: três considerados com experiência (fonoaudiólogos) e três sem experiência (alunos de graduação). As amostras de fala incluíram palavras constituídas pelas oclusivas velares /k/ e /g/ combinada com as vogais /a/, /i/ e /u/. Os juízes foram instruídos a julgar a presença, a ausência das oclusivas velares ou a presença de AC nestas amostras. **Resultados:** verificou-se diferenças no grau de concordância de juízes com experiência (Kappa moderada) e sem experiência (Kappa baixo) para os julgamentos realizados na condição pré-fonoterapia. O contexto fonético das amostras de fala influenciaram os julgamentos nas condições pré e pós-fonoterapia. Houve maior percentual de acerto para os juízes experientes condição pré-fonoterapia (p-valor <0,001). **Conclusão:** a experiência dos juízes e composição fonética das amostras de fala influenciam os julgamentos perceptivo-auditivos das AC.

DESCRITORES: Fissura Palatina; Fala; Distúrbios de Fala; Percepção Auditiva

■ **REFERENCES**

1. Hanayama EM. Distúrbios da comunicação nos pacientes com sequela de fissura labiopalatina. Rev Bras Cir Craniomaxilofac. 2009;12(3):118-24.

2. Marino VCC, Dutka, JCR, Pegoraro-Krook MI, Lima-Gregio, AM. Articulação compensatória associada à fissura de palato ou disfunção velofaríngea: revisão de literatura. Rev CEFAC. 2012;14(3):528-43.

3. Palandi BBN, Guedes ZCF. Aspectos da fala de indivíduos com fissura palatina e labial, corrigida em diferentes idades. Rev CEFAC. 2011;13(1):8-16.

4. Pamplona MC, Ysunza A, Chavelas K, Arámburu E, Patiño C, Martí F et al. A study of strategies for treating compensatory articulation in patients with cleft palate. J Maxillofac Oral Surg. 2012;11(2):144-51.

5. Bessell A, Sell D, Whiting P, Roulstone S, Albery L, Persson M et al. Speech and language therapy interventions for children with cleft palate: a systematic review. Cleft Palate Craniofac J. 2013;50(1):e1-e17.

6. 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.

7. Sell D. Issues in perceptual speech analysis in cleft palate and related disorders: a review. Int J Lang Com Dis. 2005;40(2):103-21.

8. 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 Craniofacial J. 2009;46(4):347-62.

9. Santelman L, Sussman J, Chapman K. Perception of middorsum palatal stops from the speech of three children with repaired cleft palate. The Cleft Palate Craniofac J. 1999;36:233-42.

10. Gooch J, Hardin-Jones M, Chapman K, Trost-Cardamone J, Sussman J. Reliability of listener judgments of compensatory articulations. Cleft Palate Craniofac J. 2001;38:59-67.

11. Chapman KL, Hrdin-Jones MA, Goldstein JA, Halter KA, Havlik RJ, Schulte J. Timing of Palatal Surgery and Speech Outcome. The Cleft Palate Craniofac J. 2008;45(3):297-308.

12. Brunnegard K, Lohmander A. A cross-sectional study in 10-year-old children with cleft palate: results and issues of rater reliability. The Cleft Palate Craniofac J. 2007;44:33-44.

13. Lohmander A, Friede H, Lilja J. Long-Term, longitudinal follow-up of individuals with unilateral cleft lip and palate after the Gothernburg primary early veloplasty and delayed hard palate closure protocol: speech outcome. Cleft Palate Craniofac J. 2012;49(6):657-71.

14. Hennigsson 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.

15. Klinto K, Salameh EK, Svensson H, Lohmander A. The impact of speech material on speech judgment in children with and without cleft palate. Int J Lang Commun Disord. 2011;46(3):348-60.

16. Sell D, John A, Harding-Bell A, Sweeney T, Hegarty F., Freeman J. Cleft audit protocol for speech (CAPS-A): a comprehensive training package for speech analysis. Int J Lang Comm Dis. 2009;44:529-48.

17. Nyberg J, Westberg LR, Neovius E, Larson O., Henningsson G. Speech results after one-stage palatoplasty with or without muscle reconstruction for isolated cleft palate. Cleft Palate Craniofacial J. 2010;47(1):92-103.

18. Lohmander A, Persson,C. A longitudinal study of speech production in Swedish children with unilateral cleft lip and palate and two-stage palatal repair. Cleft Palate Craniofacial J. 2008;45(1):32-41.

19. Brunnegård K, Lohmander A, van Doorn J. Untrained listeners' ratings of speech disorders in a group with cleft palate: a comparison with speech and language pathologists' ratings. Int J Lang Commun Disord. 2009;44(5):656-74.

20. Prathanee B, Lorwatanapongsa P, Makarabhirom K, Suphawatjariyakul R, Manochiopining S, Thinnaithorn R. Reliability of perceptual assessment for resonance disorders in patients with cleft palate. J Med Assoc Thai. 2012;95 Suppl 11:S73-9.

21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

22. 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.

23. Bzoch KR. A battery of clinical perceptual tests, techniques, and observations for the reliable clinical assessment, evaluation, and management of 11 categoriasl aspects of cleft palate speech disorders. In: Bzoch KR (ed). Communicative disorders related to cleft lip and palate. 5th ed. Austin: Pro-Ed; 2004. p. 375-462.

Mailing address:
Viviane Cristina de Castro Marino
Rua Santa Helena, 1967 casa 07
Marília – SP - Brasil
CEP: 17514-410
E-mail: vivianemarino2@yahoo.com.br