The Effect of Telephone Transmission on Voice Quality Perception

*Renata Regina Passetti, and †Ana Carolina Constantini, *†Campinas, Brazil

Summary: Objectives/Hypothesis. This study aims to analyze the effect of telephone transmission on the voice quality perception of subjects with dysphonia and seeks to contribute to the use of perceptual voice analysis in forensic context. The main hypothesis is that the auditory perception of voice quality differs according to the acoustic quality of the speech material.

Study Design. Eight male speakers were simultaneously recorded directly by a digital recorder and over a mobile phone.

Materials and Methods. A perceptual experiment containing 21 speech samples was used. The samples were analyzed via an online adaptation of the Vocal Profile Analysis Scheme (VPAS) into Brazilian Portuguese by a group of six voice-specialized professionals. Data analysis included the relative frequency of the 10 most evaluated VPAS settings in both recording conditions, and the relative frequency of the VPAS groups of features. A multidimensional scaling technique was applied to investigate the perceptual distance between stimuli pairs from both recording conditions. This was followed by a nonparametric multiple linear regression analysis that aimed to comprehend the underlying variables that motivated the spatial arrangement of the stimuli.

Results. The telephone transmission effect may have an influence on assessing correctly the supralaryngeal, laryngeal, prosodic, and temporal VPAS settings. The perceptual distance between stimuli pairs were mainly motivated by the rating of velopharyngeal system, respiratory support, laryngeal tension, speech rate, and creaky voice.

Conclusions. The auditory perception differences between the stimuli pairs with different acoustic qualities could be related to the mobile phone bandwidth and data compression, and to telephone interaction characteristics. A group of laryngeal, supralaryngeal, temporal, and respiratory settings was mainly related to the dissimilarities reported. In terms of the application of VPAS in forensic context, the protocol was considered a relevant scientific tool that enables standardizing and analysis comparison among different groups of forensic specialists.

Key Words: Voice quality−Telephone transmission effect−Perceptual analysis−VPAS−Forensic phonetics.

INTRODUCTION

The ability to recognize an individual by only assessing his vocal characteristics is a task we undergo daily even through adverse acoustical conditions, such as during telephone calls where spectral distortion affects the perception of the speech signal[1] or in crowded places where there are multiple individuals speaking simultaneously, a phenomenon known as the "cocktail party problem."[2] The analysis of an individual's voice profile by the use of auditory-perceptual approaches has become an important procedure used mainly for clinical purposes. However, the evaluation of voice quality (henceforth VQ) plays an important role in other areas, such as speech synthesis systems,[3] dialectological studies,[4] drama teaching, and forensic phonetics.[5]

For some of these areas, especially forensic phonetics, it is important to determine whether the acoustic quality of the speech material can impair the perceptual assessment of VQ and mislead practical forensic procedures. The ability to differentiate VQ in different listening contexts is related to speaker identification tasks that usually involve comparing a criminal's telephone speech sample to a suspect's direct speech sample.[6] VQ discrimination is also key to earwitnesses identifications, since VQ is likely the determining voice discrimination factor performed by naïve listeners.[7,8]

VQ is one of the most analyzed parameters among suprasegmental features examined in forensic speaker comparison practices. According to an international survey performed by Gold and French,[9] 94% of the interviewed experts analyzed VQ in their practical procedures, and 68% of them used a recognized protocol or a modified version of it. A well-known approach for VQ perceptual analysis is the Vocal Profile Analysis Scheme (VPAS) (Appendix 1). VPAS is based on the contributions of Laver et al[10] in terms of VQ description and analysis.

Laver's description of VQ was founded on an integrative concept of anatomical, acoustic, and perceptual factors. From his point of view, VQ is the product of social and organic attributes combined into unconscious phonetic adjustments of vocal tract muscles, which he named as "settings."[5,11]

The set of laryngeal and supralaryngeal settings that characterizes VQ is linked to the perceptual assessment of other groups of adjustments, such as prosodic, temporal, and respiratory factors. The conception of "setting" enables a description of an individual's VQ from its deviance from a "neutral setting" associated to a neutral vocal tract configuration as reference.[4]

Although there is some criticism about the subjectivity associated to auditory protocols,[12] the usage of these tools

Accepted for publication April 27, 2018.

From the *Department of Linguistics, State University of Campinas, Campinas, São Paulo, Brazil; and the †Department of Human Development and Rehabilitation, State University of Campinas, Campinas, São Paulo, Brazil.

Address correspondence and reprint requests to Renata Passetti, Department of Linguistics, State University of Campinas, Rua Sérgio Buarque de Holanda, 511, 13083-859, Campinas, São Paulo, Brazil. E-mail: re.passetti@gmail.com

Journal of Voice, Vol. 33, No. 5, pp. 649−658
0892-1997

https://doi.org/10.1016/j.jvoice.2018.04.018

allows a detailed evaluation of VQ, since multiple phonetic and articulatory adjustments can be simultaneously assessed. Moreover, comparative analyses can also be carried on, leading to important results for voice analysis-related areas.

The literature has shown that the telephone transmission plays an effect on speech signal, affecting the perceptual distinction between VQs such as nasality[13] and breathiness and whisper.[14] The use of VPAS in our study will allow us to compare our results to those previously reported in the literature. Therefore, the purpose of this study is to analyze the effect of the telephone transmission on the VQ perception and contribute to the use of perceptual voice analysis in forensic context. We hypothesize that VQ perception in direct recordings differs from its perception in mobile phone recordings due to telephone transmission distortions.

MATERIALS AND METHODS

The experimental procedure comprised two stages—data collection and the development and application of a perceptual experiment. The latter was aimed at verifying if a group of raters would be consistent in perceptually assessing VQ in stimuli pairs with different acoustic qualities. This research was approved by the Ethics Committee of the State University of Campinas (CAAE number: 54794016.3.0000.5404).

Subjects

Eight male speakers aged between 20 and 56 years (mean of 42 years) were recorded. The speakers were patients of the Speech Therapy Program in the Voice Ambulatory of the Rehabilitation Research and Studies Center "Gabriel Porto" – State University of Campinas. The selected participants had an otolaryngological diagnosis[a] of functional or organic-functional dysphonia.

Recording only male speakers had forensic purposes and was an attempt to get closer to the reality of the Brazilian prison system. According to the latest statistical survey released, 93% of Brazilian prisoners were men.[15] The selection of diagnosed dysphonic voices was an attempt to control the raters' VQ evaluation on perceptual test, since we could compare if their analysis were aligned with the clinical diagnosis of these voices, if necessary.

Data collection

This study's method was based on Künzel's study on telephone transmission effect.[16] It comprised recordings performed simultaneously under two conditions: (a) over a condenser headset microphone connected to a digital recorder and (b) over a mobile phone by connecting the calling phone to a sound card which worked as a telephone interception device. A local number was dialed for each recording.

The samples were recorded in an acoustic booth and the headset microphone was placed at a 5-cm distance from the speakers' lips. The speakers were free to hold the mobile phone in any comfortable position. The direct recordings used a Yoga HM20 condenser headset microphone (Yoga Eletronics Co., Taipei, Taiwan) and a Zoom H4N digital recorder (Zoom North America, Hauppauge, New York, USA). The mobile phone recordings used the calling phone connected to a Behringer UCA222 sound card (MUSIC Tribe Global brands, Makati City, Metro Manila, Philippines) that was connected to a Dell notebook (Dell Inc., Hopkinton, Massachusetts, USA). The recordings were set at a sample rate of 44.1 kHz and a 16-bit resolution. The software used for mobile phone recordings was *Audacity 1.2.6* (The Audacity Team, GNU General Public License, Free Software Foundation) and the mobile phones used for calls were an Apple iPhone 5s (Apple Inc., Cupertino, California, USA) (calling phone) and a Nokia 110 (Nokia, Espoo, Helsinki, Finland) (receiver phone). This experimental procedure guaranteed that all speakers' speech sample pairs (direct and mobile phone) had exactly the same content, since they were obtained simultaneously.

The speech material consisted of semi-spontaneous speech samples obtained through informal conversations on daily life activities. The conversation topics were elicited by the interviewer and consisted of a description of the speakers' work day, weekend activities, and favorite hobbies. Conversations lasted approximately 5 minutes. Finally, the *PRAAT software*[17] (By Paul Boersma and David Weenink, University of Amsterdam, Amsterdam, The Netherlands) was used to extract 20-second duration stimuli that would be used for the perceptual experiment.

Perceptual experiment

Experimental procedure

Excerpts of, approximately, 20 seconds of uninterrupted contextualized speech[18] from both acoustic means (direct and mobile phone) were selected for data collection. Twenty-one stimuli were adopted for the perceptual experiment: 16 recordings (one for each acoustic quality per speaker) plus 30% of stimuli repetitions. The stimuli's loudness level was adjusted to 70 dB.

The perceptual experiment was performed by a group of six professionals, composed by three phoneticians, two speech-language pathologists who worked in a Brazilian Government Agency for Law Enforcement and Crime Prosecution, and a forensic expert from the National Institute of Criminalistics, Brazilian Federal Police. The raters had been trained in VQ assessment and in filling out the VPAS protocol during the *Workshop on Vocal Profile Analysis* by Professor Anders Eriksson (Stockholm University, Sweden) held at the Institute of Language Studies, State University of Campinas. The workshop program comprised a day of oral presentations on VQ subjects and 4 days of training in the skills necessary for filling out the VPAS protocol.

The online platform PsyToolkit[19] was chosen to perform the perceptual experiment. The stimuli were randomly organized and authors had ensured that two stimuli from a same speaker would not be consecutively presented. A digital version of the VPAS adapted to Brazilian Portuguese (henceforth BP-VPAS)[4]

[a]The clinical diagnosis of selected voices was based on the dysphonia grade, roughness, breathiness, asthenia, and strain protocol[39,40] and is shown in Appendix 4.

(Appendix 2) was used for the raters' perceptual assessment. Raters were recommended to be in a silent environment and to use headphones. None of them reported any hearing loss or phonoarticulatory problems. They could listen to each speech sample as many times they considered necessary to perceptually evaluate a subject's VQ and fill out the BP-VPAS. The set of 21 stimuli was rated in two sessions (assessment of 10 and 11 stimuli, respectively) separated by an interval of 2 weeks.[b]

At the end of the perceptual evaluation, the raters were invited to answer a survey about application's viability of the BP-VPAS[20] that was adapted for our research purposes (Appendix 3). This survey worked as a feedback of raters' performance at perceptual experiment and of their opinions about the application of the BP-VPAS protocol in forensic contexts.

VPAS protocol

The perceptual assessment of VQ by the VPAS protocol[21] is broadly divided into three main sessions (groups of features): vocal tract (supralaryngeal) features, overall muscular tension (laryngeal and supralaryngeal), and phonation (laryngeal) features. Besides that, it is also possible to perceptually evaluate prosodic features (pitch and loudness), temporal organization, and respiratory support. Therefore, the full protocol is divided into six groups of features: vocal tract features, overall muscular tension, phonation features, prosodic features, temporal organization, and other features (respiratory support and diplophonia).

As stated earlier, the VPAS analytic unit is called "setting" and it represents the neutral activity of the vocal tract's articulations and muscles. A VQ analysis should identify non-neutral settings and quantify its deviance. Therefore, at first, raters should evaluate which are the neutral and non-neutral settings of the aforementioned features. Afterwards, when non-neutral settings are identified, raters should quantify the deviance degree of these settings from their neutral correspondent. The scale degree ranges from 1 to 6. Degrees 1 to 3 are classified as "moderate" and degrees 4 to 6 as "extreme" deviations from the neutral settings.

The BP-VPAS[4] was based on the contributions of Professors Sandra Madureira and Zuleica Camargo from the Integrated Acoustic Analysis and Cognition Laboratory (Pontifical Catholic University of São Paulo, São Paulo, Brazil), and its cultural and linguistic validation is still in process.

Data analysis

The speech material investigating the telephone transmission effect on VQ perception was analyzed via a set of statistical tests. The statistical software *R version 3.4.3*[22] (The R Foundation for Statistical Computing, Vienna, Austria) was used to perform these tests.

Reliability measures. Two statistical tests were performed on the perceptual experiment to estimate the reliability among raters' performance:

(1) **Cronbach's alpha:** applied to measure the internal consistency reliability among raters for filling the VPAS protocol. This measurement provides a standardized alpha (std. alpha) that, in our study, was based upon the correlations for each VPAS setting. It was also possible to access the alpha coefficient itself (α), which ranges from 0 to 1 in providing the overall VPAS internal reliability.[23,24] An interpretation of Cronbach's alpha reliability coefficient can be accessed in Gliem and Gliem.[25]

(2) **Simple matching coefficients (SMC):** applied to compute the similarity between the five repeated stimuli, obtained for randomly repeating 30% of the data, according to the experimental procedure, and the stimuli they corresponded to. This measurement was based on San Segundo and Mompean.[12] Intra- and inter-raters were performed to assess the percentage agreement. The calculation considered a "similarity" (=1) when the setting being analyzed had exactly the same configuration (degree) in both compared stimuli, otherwise the received grade was zero (=0). Therefore, the SMC consisted in summing up the similarities and dividing them by the total VPAS settings (n = 53).

Relative frequency of VPAS settings evaluation. The relative frequency of the evaluated VPAS settings was performed:

(I) to compare the 10 most evaluated VPAS settings in both recording conditions (direct and mobile);
(II) to analyze which was the most evaluated VPAS groups of features for each recording condition.

As the given scale degree (1−6) of each VPAS setting matters for this measurement, the first step was to organize data collection by calculating a weighted average (WAVG) of each evaluated VPAS setting considering the stimuli from both recording conditions. This procedure is demonstrated in Equation 1.[c]

$$WAVG_{Stimulus_x,\ Setting_y} = \frac{[(n\times 1)+(n\times 2)+(n\times 3)+(n\times 4)+(n\times 5)]}{(1+2+3+4+5)} \quad (1)$$

where x and y identified the stimulus and VPAS setting being measured, respectively, and n is the quantity of raters that gave the mentioned scalar degree to the selected stimulus and setting.

The absolute frequency was then measured, considering the WAVGs of VPAS settings from both recording conditions, according to Equation 2.

$$Absolute\ frequency = \sum WAVG_{direct} + \sum WAVG_{mobile} \quad (2)$$

[b]This time lapse was chosen so as not to overload the raters during the perceptual evaluation process of the speakers' VQ.

[c]The WAVG did not consider the sixth scale degree, since none of the raters gave it to the evaluated stimuli.

TABLE 1.
Similarity Matching Coefficients for Assessing Intrarater Agreement

	Stimuli Pairs				
Raters	S3_D × S3_D_R	S1_D × S1_D_R	S8_M × S8_M_R	S8_D × S8_D_R	S5_D × S5_D_R
R1	0.87	0.91	0.89	1.0	0.94
R2	0.93	0.93	0.94	0.92	0.94
R3	0.96	0.85	0.96	1.0	0.94
R4	0.91	0.91	0.98	0.94	0.98
R5	0.79	0.89	0.74	0.92	0.77
R6	0.94	0.92	0.98	0.94	0.98

For analysis (I), the relative frequency was measured by adding the WAVG of VPAS settings from each of recording condition at a time and dividing it by the absolute frequency considering both recording conditions, according to Equation 3.

$$Relative\ Frequency_i = \frac{\sum WAVG_i}{Absolute\ frequency} \quad (3)$$

where i identified the selected recording condition.

The same procedure was applied for analysis (II), but data were selected according to the VPAS groups of features the evaluated settings belonged to. This procedure is demonstrated in Equation 4.

$$Relative\ Frequency_{i,g} = \frac{\sum WAVG_{i,g}}{Absolute\ frequency} \quad (4)$$

where i identified the selected recording condition and g the selected VPAS group of features.

Multidimensional scaling. The perceptual distance between stimuli pairs was investigated by applying the multidimensional scaling (henceforth MDS) technique. This method detects significant underlying dimensions that can explain the (dis)similarities between the investigated objects.[26] Each object is represented by a point in a multidimensional space, and its distance from another object represents how similar they are.[27]

This technique had already been successfully applied in perceptual studies,[28,29] and it will allow us to analyze how the telephone transmission effect affected raters' assessment of VQ by investigating the stimuli pairs' spatial arrangement. The input for running the MDS technique was a similarity matrix containing the Euclidean distances among all paired stimuli from our data.

Multiple linear regression. The analysis of the two-dimensional scatter plot resulting from the MDS technique was followed by a nonparametric multiple linear regression analysis, which aimed to better understand the variables that motivated that specific spatial arrangement. The most evaluated VPAS settings (item 2.3.3.2 − first analysis) were taken as predictor variables, and the Euclidian distances between the stimuli pairs of a same subject were taken as response variables.

RESULTS

Raters' agreement

The analysis of the consistency among raters for filling out the VPAS protocol was assessed by the calculation of Cronbach's alpha coefficient. The measurements for each VPAS setting is shown in Appendix 5. Both the overall standard alpha (std. alpha) and the alpha coefficient (α) were 0.79. The lower and upper limit coefficients of the 95% confidence interval were 0.75 and 0.85, respectively.

The SMC technique results for the assessment of intra- and inter-rater agreement at the evaluated pair of stimuli, original and its repetition (followed by "_R"), are shown in Tables 1 and 2, respectively.

In general, raters obtained higher SMC for the compared stimuli. Two exact assessments were made for stimuli pair "S8_D" and the SMC was equal to 1.0, which meant perfect intra-rater agreement. Among the raters group, R5 had received the lower SMC, which was due to different evaluation degrees given for a same VPAS setting, since the technique only scored identical assessments.

On average, there was a higher agreement between stimuli pairs (SMC = 0.74). The comparison related to stimulus S8_D had the higher agreement among the raters (SMC = 0.89), and those related to stimuli S3_D and S8_M had the

TABLE 2.
Similarity Matching Coefficients for Assessing Inter-rater Agreement

	Stimuli Pairs				
	S3_D × S3_D_R	S1_D × S1_D_R	S8_M × S8_M_R	S8_D × S8_D_R	S5_D × S5_D_R
SMC	0.68	0.75	0.68	0.89	0.70

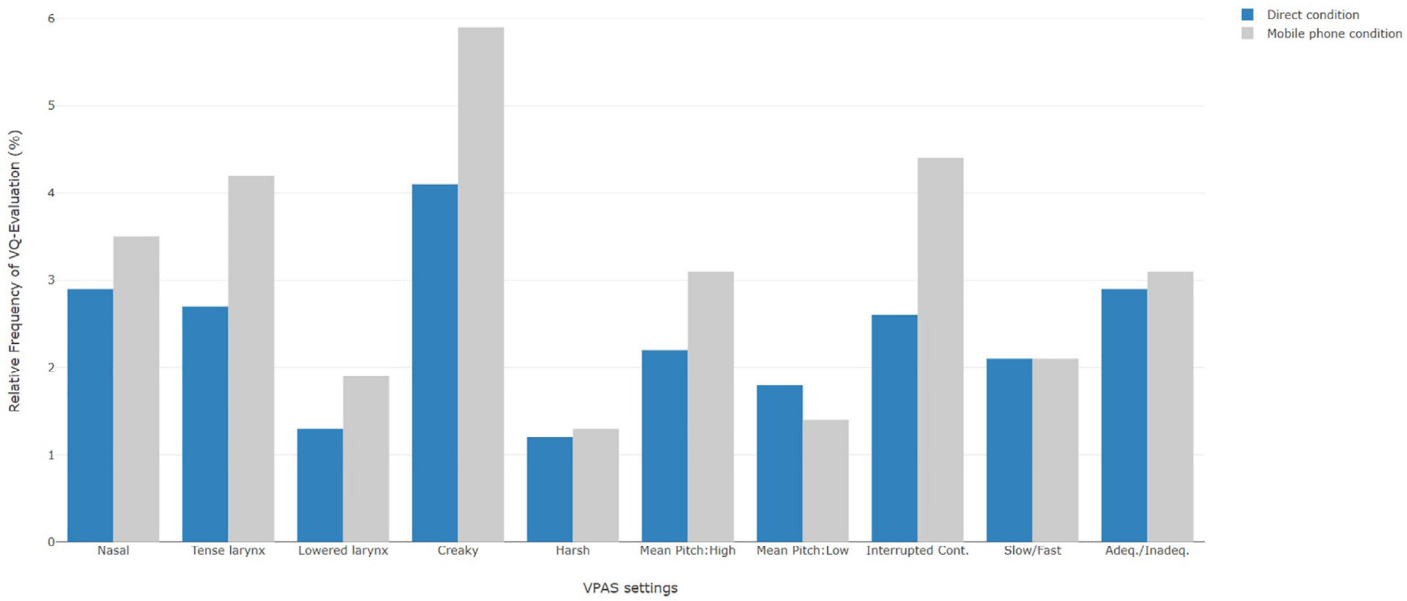

FIGURE 1. The 10 most frequent VPAS settings considering both recording conditions simultaneously.

lower agreement (both with an SMC = 0.68). The 0.68 SMC score was due to the different evaluation degrees for a same VPAS setting and not to agreement inconsistencies between the raters.

The 10 most evaluated VPAS settings in both recording conditions

The relative frequency analysis of the 10 most evaluated VPAS settings in both recording conditions (direct and mobile phone) is shown in Figure 1. The VPAS settings are presented following the order they appear in the VPAS protocol.

Most of the VPAS settings had a higher frequency of evaluation in the mobile phone recording condition. Table 3 shows the relative frequencies in each recording condition for these selected settings.

TABLE 3.
Relative Frequency Percentages for the 10 Most Evaluated VPAS Settings for Each Recording Condition

VPAS Setting		Relative Frequency (%)	
		Direct Recordings	Mobile Phone Recordings
Nasal		2.9	3.5
Tense larynx		2.7	4.2
Lowered larynx		1.3	1.9
Creaky		4.1	5.9
Harsh		1.2	1.3
Mean pitch	High	2.2	3.1
	Low	1.8	1.4
Interrupted continuity		2.6	4.4
Rate		2.1 (slow)	2.1 (fast)
Respiratory support		2.9 (adequate)	3.1 (inadequate)

It is important to highlight that the relative frequency percentages are low because their measurement considered the evaluation of the total amount of the VPAS protocol's settings (n = 53).

The VPAS settings velopharyngeal nasal, tense larynx, lowered larynx, creaky, harsh, high mean pitch, interrupted continuity, and inadequate respiratory support had higher frequency percentages in mobile phone recordings. The perception of low mean pitch was the only VPAS setting with a higher frequency in direct condition.

Considering both recording conditions together, the phonation type *creaky* was the most evaluated VPAS setting. However, an assessment inconsistency was attested for the relative frequency percentages of *rate* and *respiratory support* settings. For both recording conditions, *speech rate* constitute 2.1% of evaluations, but in direct recordings this percentage was linked to the assessment of a *slow* speech rate, while in mobile phone recordings to a *fast* speech rate. The evaluation for *respiratory support* also differed in accordance to the recording condition. For the direct condition, there was a higher perception of an *adequate respiratory support*, while for the mobile phone condition, the *inadequate* label had a higher evaluation.

An explanation for these perception differences according to the recording condition will be presented in the discussion section.

Relative frequency of VPAS groups of features

The relative frequency of VPAS groups of features is shown in Figure 2. The format presentation follows that of the VPAS protocol.

According to Figure 2, the evaluation frequency increased in mobile phone recording condition for most of the VPAS groups of features. Table 4 shows the percentages in each recording condition.

TABLE 4.
Relative Frequency Percentages for the VPAS Groups of Features for Each Recording Condition

VPAS Groups of Features	Relative Frequency (%)	
	Direct Recordings	Mobile Phone Recordings
Vocal tract	11.4	11.4
Overall muscular tension	4.4	6.0
Phonation	7.1	9.1
Prosodic	12.7	13.6
Temporal organization	6.0	8.8
Other features	3.9	5.5

The VPAS groups of features that had a higher evaluation percentage in mobile phone recording condition were: *overall muscular tension*, *phonation*, *prosodic*, *temporal organization*, and *other features* (*respiratory support*). Among them, the *prosodic* group was the most evaluated one, and was followed by the *vocal tract* feature group, which was equally evaluated in both conditions.

MDS and multiple linear regression

The two-dimensional space obtained as a result of the MDS technique is shown in Figure 3. Each point in the space corresponds to a stimulus. The abbreviations identify the subjects (S1, S2,. . .) and the recording condition (direct [D] and

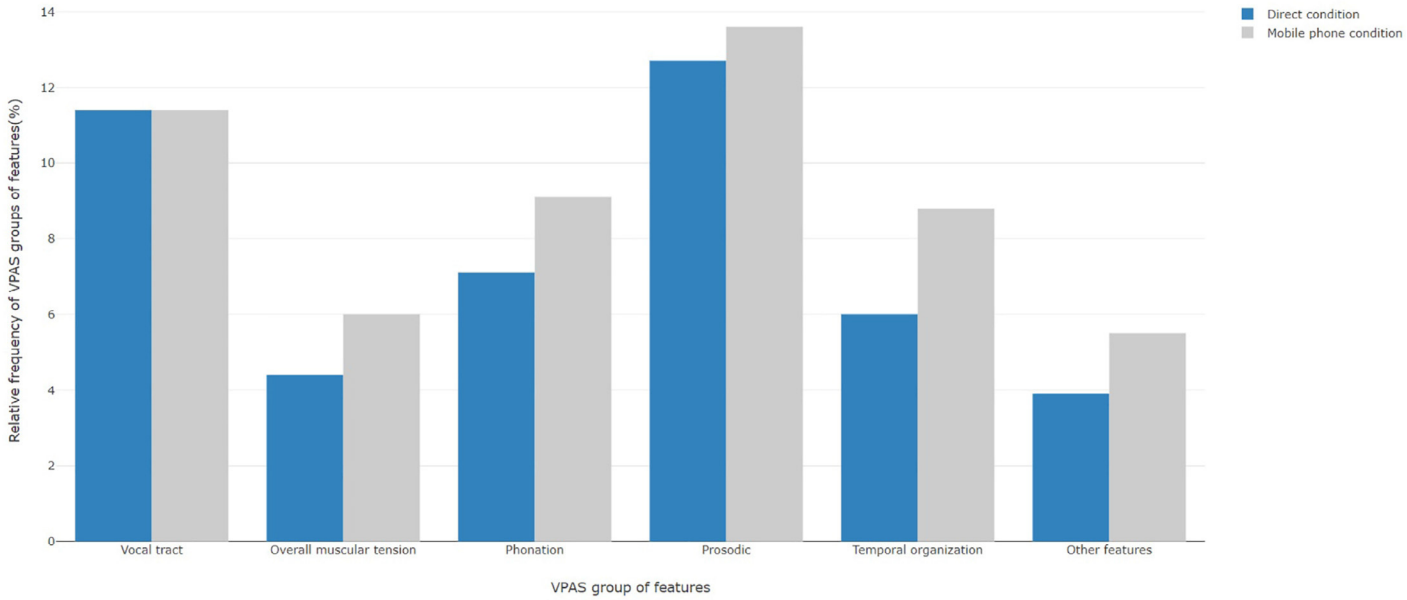

FIGURE 2. Evaluation frequency of VPAS groups of features considering both recording conditions simultaneously.

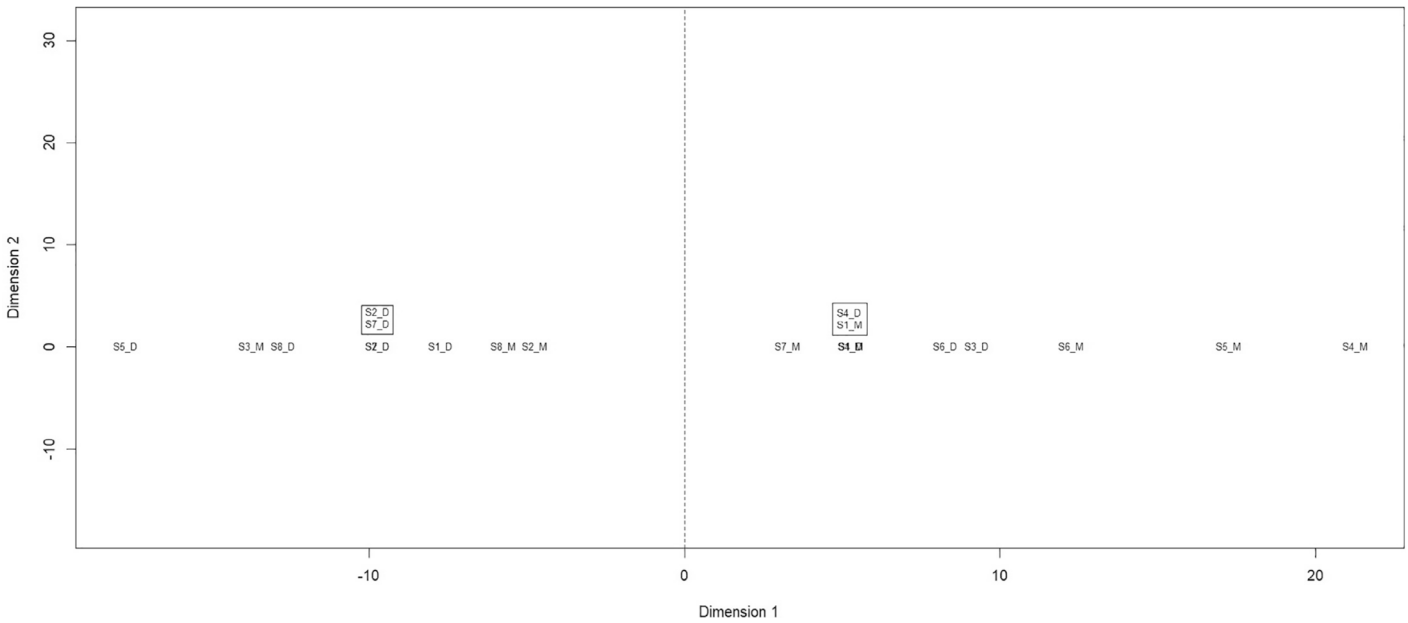

FIGURE 3. Representation of two perceptual dimensions produced by MDS technique.

mobile phone [M]). The point *S8_D*, for instance, corresponds to the stimulus of speaker 8 in the direct recording condition. There were two situations of overlapping stimuli (S2_D and S7_D, and S4_D and S1_M). These stimuli were rewritten in a text box above the overlapping place to facilitate their visualization.

The visual inspection of Figure 3 shows that the highest amount of variance is displayed in MDS dimension 1, and that dimension 2 accounts very little to data variance. Therefore, data are divided by a dotted line crossing the zero on the x-axis.

The amount of stimuli was equally spaced along the dimension 1 sides, eight for each side. However, the left side had more stimuli from the direct condition (n = 5), and the right side had more stimuli from the mobile phone condition (n = 5). None of the stimuli pairs overlapped, which indicated that their VQs had not been perceived equally in both recording conditions. Moreover, the stimuli that were equally assessed (S2_D and S7_D, and S4_D and S1_M) were not pairs of a same speaker. Only three pairs of stimuli were placed in the same dimension side: S2_D and S2_M, S4_D and S4_M, and S6_D and S6_M.

According to the MDS results, acoustic quality likely played an important role on the stimuli arrangement. Therefore, a regression analysis was performed to investigate which VPAS settings were significantly correlated to this spatial disposition.

The nonparametric model that accounted for the higher explained variance was the one that considered the interaction among the settings *nasal velopharyngeal*, *respiratory support*, *tense larynx*, *slow rate*, and *creaky voice*. The explained variance percentage (R^2) of this model was 73% (standard deviation = 0.16, $p = 7.24 \times 10^{-6}$).

DISCUSSION

The techniques applied to assess the raters' agreement have shown a reliable agreement within and between raters. In general, the obtained results for all analyses reinforced the hypothesis about the telephone transmission effect on the perceptual assessment of VQ. The following topics will discuss the main findings of each conducted analysis.

Relative frequency of evaluated VPAS settings in both recording conditions

The results of the evaluation frequency of VPAS settings in both recording conditions showed that VQ was not equally assessed in all stimuli pairs. The telephone quality had increased the evaluation frequencies of most VPAS settings in this recording condition. Therefore, by perceptually analyzing VQ in speech samples from mobile phone recordings, raters tended to overestimate the degree of a perceived non-neutral setting. This overestimation can be mostly related to the mobile phone bandwidth (300−3400 Hz),[16,30] and to the dynamic change of the data compression rate in Global System for Mobile Communications (GSM) system,[31] which affect the acoustic quality of the speech material by adding noises that can impair the adequate perception of VQ characteristics.

For now, we will provide some explanation about the assessment of the 10 most evaluated VPAS settings. *Nasality* is an acoustic-auditory phenomenon expressed in speech by the velopharyngeal movement, which causes the coupling of oral and nasal cavities. During the production of nasal segments, the resonance regions formed by the coupling of these cavities can be amplified, creating nasal formants, or blocked, producing spectral valleys and anti-formants. Although the lower cut-off frequency of mobile phone band-pass filter can eliminate some acoustic information of nasal segments,[13,32,33] its higher perception in mobile phone condition could be explained in terms of the spectral distortion effect that shifts frequencies near the cut-off limits.[30] This effect can narrow the telephone dynamic range and artificially change spectral information, which could result in a different perception of this setting.

The perception of *tense larynx* setting was aligned to the clinical diagnosis for most evaluated dysphonic voices, since it had already attested the presence of vocal strain. However, this evaluation does not match the perception of a *lowered larynx height*, a setting related to the decrease of formant frequencies values as a response to the action of the infrahyoid group of muscles that cause the expansion of the pharyngeal cavity and, consequently, of the fundamental frequency (henceforth F_0), which is auditorily expressed by a lowered pitch.[21]

However, the assessment of *lowered larynx* can be related to the adoption of a specific body posture that may have contributed to its setting perception.[21] Such body posture consists in rotating both chin and head slightly down, which decreases the angle between the neck and under the surface of the chin, and anatomically enables the process of lowering the larynx. This explanation associated to the experimental design of our data collection (a corpus of simultaneous direct and mobile phone recordings) is justifiable, since part of the speakers may have decided, consciously or not, to keep their hands free and to hold the mobile phone by pressing their shoulder towards their ear, a similar position to that described by Laver as motivating the lowered larynx.

The assessment of the voicing types *creaky* and *harsh* was also aligned to the clinical diagnosis of all evaluated voices, since it had already attested the presence of an associated phonation type (*roughness*). In Brazilian Portuguese, as in other languages, the voicing type *creaky* can signal both terminal and nonterminal prosodic boundaries in speech.[34−36] Therefore, the perception of this setting may also be associated with this prosodic function manifestation in the speakers' speech. However, the percentage increase in the assessment of these settings in the mobile phone condition can be related to the aforementioned mobile phone effect. The poor acoustic quality of mobile phone recordings may have affected the correct perception of this VPAS setting.

The higher evaluation percentage of a *high pitch* in mobile phone condition can be associated to the mobile phone transmission effect on the F_0, which is the acoustic correlate of intonation. Previous studies on this subject have shown

that, in mobile phone speech samples, F_0 values are subjected to an upward shift when compared to samples directly obtained.[1,37] Thus, the perceptual assessment of this setting could have played a role on the acoustic modifications caused by the spectral distortion of the mobile device. Therefore, the higher evaluation percentage of a *low pitch* in direct recordings was also justifiable, as this recording condition did not have the filtering restrictions imposed by the telephone channel, allowing this setting to be properly assessed in this condition.

Concerning the analysis of *continuity* setting, most of the analyzed speech samples were considered as *interrupted*. However, it is necessary to observe that, in both recording conditions, the stimuli were excerpts from a telephone conversation and, hence, typical characteristics of this interaction type may have been manifested. Telephone interactions do not have any of the paralinguistic components that characterize face-to-face interactions. The lack of visual information can compromise the expected fluency in turn-talking. In this case, stimuli considered as noncontinuous could be signaling the listeners' attention to their interlocutor and to the message interpretation.[38] Moreover, a noncontextualized speech excerpt taken out of a telephone interaction can be mistakenly interpreted as noncontinuous.

The same perceptual assessment of *speech rate* (*slow* and *fast*) signals a conflict in this setting evaluation. This VPAS setting was mostly perceived as *slow* in stimuli from direct condition, and as *fast* in stimuli from mobile phone condition. Since all data were simultaneously recorded in direct and mobile phone conditions, there is no reason to believe that there was a change in the speakers' speech rate during the interaction. Therefore, our hypothesis is that raters have a mental representation of speech rate characteristics in telephone interactions and have made their evaluations based in this mental representation. Another possible explanation would be related to the low acoustic quality of mobile phone recordings which could mask the reliable perception of the speech rate.

A similar explanation can be applied to the *respiratory support* evaluation. Raters have evaluated this setting as *adequate* in direct condition and as *inadequate* in mobile phone condition. Laver's[38] explanation about this setting states that its evaluation should consider a continual flow of air being provided as the basis for whole utterances of normal pulmonic egressive speech. The dynamic change of data compression rate in GSM system may interfere in the acoustic quality of the signal. So, this process combined with the mobile phone band-pass filter characteristics can result in noises in the acoustic signal, which can certainly have affected the adequate perception of speakers' respiratory support in this condition.

Perceptual assessment of VPAS groups of features and the viability of VPAS application in forensic context

We decided to discuss some of the results of the perceptual assessment of VPAS groups of features in relation to the raters' answers to the survey about the viability of VPAS application in forensic context.

Considering the evaluation of VPAS groups of features by recording condition, this analysis reinforces the aforementioned results for the relative frequency of VPAS settings: the evaluation percentage from most of the VPAS groups of features was higher in mobile phone condition. The most evaluated VPAS group was the *prosodic* features followed by *phonation* features. This result reflected the raters' personal experience, since three of them were phoneticians who worked in speech prosody subjects. The others worked with forensic phonetics, an area that involves acoustic-auditory analysis and knowledge in experimental acoustic phonetics. Therefore, the perceptual assessment of VQ was expected to involve not only the specific abilities and knowledge for filling out the VPAS protocol, but also the raters' personal experience in the speech sciences area.

The third most evaluated VPAS group was the *vocal tract* features group, which was equally assessed in both recording conditions. On the other hand, this group was mentioned as the hardest group of features to identify (83.3% of the raters answered *vocal tract* and 16.7% answered *prosodic* features). Then, the relative frequency of this group can be related to its amount of VPAS settings, 25 of 53 settings, which corresponds to 47% of the totality of VPAS settings. Among this group of features, the analysis of mandibular, lingual body, and pharyngeal settings were pointed out as the hardest ones.

When asked about the advantages of using the VPAS in forensic context, raters highlighted three main VPAS strengths: (1) standardize analytical practices in forensic tasks, (2) compare the analysis of a same material among different experts, and (3) identify and quantify idiosyncratic characteristics of an individual's voice. The mentioned disadvantages were related to the extensive training to filling out the protocol.

MDS and multiple linear regression

The MDS analysis showed that the stimuli were arranged mainly in one dimension. Their division along the first dimension seemed to be motivated by the acoustic quality of the recordings, since most of the direct and mobile phone stimuli were placed in opposite sides. None of the stimuli pairs were overlapped, which could indicate that they were equally assessed. Therefore, the distance between stimuli pairs (direct and mobile phone) signaled the telephone effect on VQ evaluation.

Different VPAS settings contributed to this spatial arrangement. Four stimuli (grouped in two pairs) were equally assessed but they were not pairs of a same speaker. Caution must be taken when different stimuli are grouped incorrectly due to the different acoustic qualities present in the auditory analysis of the speech samples, as this may mislead a forensic investigation.

STUDY LIMITATIONS

As we face some of the outcomes obtained in this research, some procedures need clarification at the experimental

design level. First, the relative frequency of *lowered larynx* evaluation, a VPAS setting associated to larynx height, requires such clarification. The assessment of this setting could be related to data collection. Since our corpus consisted of stimuli obtained from telephone interactions, and although we did not video record these interactions, it is possible that some of the speakers held the mobile phone by pressing it between their shoulder and ear, which may have contributed to this setting manifestation. Speakers only were advised to hold the mobile phone in the opposite side of the headset microphone to avoid reverberations, but no recommendation was made about an appropriate way to hold it. Facing this result, it is recommended to observe this in future studies or procedures involving the use of mobile phones for VQ analysis.

Second, the assessment of opposite labels of *speech rate* setting for simultaneously obtained stimuli allows for reconsideration of the VPAS applying procedures. For contrastive settings, as is the case of *speech rate*, it would be appropriate to present a speech sample of a *neutral* speech rate setting, which should work as a standard model in which the raters should compare their assessments for other rated stimuli. This procedure would guarantee that all raters assess this VPAS setting in a similar way.

CONCLUSIONS

This research aimed at investigating the telephone transmission effect on the perceptual assessment of VQ. The results reinforced this study's hypothesis that mobile phone acoustic quality affects adequate perception and evaluation of VQ.

The main finding is associated to the overestimation of non-neutral VPAS setting degrees in telephone speech samples. Since there was reliable agreement among raters, results have shown that auditory dissimilarities in the VQ assessment of stimuli pairs were due to the acoustic quality of mobile phone recordings. The telephone transmission effect could be observed in the assessment of *supralaryngeal* settings, such as nasality and larynx tension; *laryngeal* settings, as for the voicing types *creaky* and *harsh*; and *suprasegmental* settings.

The analysis of the perceptual distance between stimuli pairs reinforced the outcomes of the relative frequencies of VPAS settings, since the stimuli spatial arrangement was motivated mainly by the mistaken assessment of the aforementioned settings. The spectral distortion caused by the mobile phone bandwidth and the background noises motivated by data compression of GSM system are considered the main causes of such perceptual inconsistencies in VQ analysis.

Some study limitations due to experimental design were also noted. One limitation was related to the assessment of the *lowered larynx* setting, which could be explained by how some speakers held the mobile phone. The other limitation was related to having a proper assessment of *speech rate* setting by comparison to a *neutral* model for standardizing its evaluation.

The current study also provided assessment of the raters' feedback on the perception experiment, and also their opinion on the application viability of the VPAS in forensic context. Their answers to the survey have shown that their performance on the perception test reflected their personal experience in the acoustic phonetics and voice analysis areas. The raters pointed out that the main contribution of the VPAS application on forensic tasks was its ability of standardizing the forensic analysis of VQ, which allowed the comparison of the same speech material among different experts. It also allowed identifying and quantifying idiosyncratic characteristics of a subject's voice. The disadvantages of the VPAS were related to the extensive training to fill out the protocol.

VQ auditory protocols, such as VPAS or simplified versions of them,[12] could improve practical tasks performed not only in forensic, but also in clinical contexts. However, VQ analysis should be led with caution when involving speech materials with different acoustic qualities. In such cases, as shown in this study (direct vs. telephone quality), the acoustic quality effect should be considered so as not to mislead the auditory analysis of VQ.

Acknowledgments

The first author thanks grant #2015/12174-9, São Paulo Research Foundation (FAPESP), São Paulo, Brazil. The opinions, hypothesis, and conclusions expressed in this article are the authors' own and do not necessarily reflect the view of FAPESP.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi:10.1016/j.jvoice.2018.04.018.

REFERENCES

1. Passetti RR. *O efeito do telefone celular no sinal da fala: uma análise fonético-acústica com implicações para a verificação de locutor em português brasileiro*. Dissertação de Mestrado (Linguística)—Universidade Estadual de Campinas. 2015.
2. Cherry EC. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am*. 1953;25:975–979.
3. Keller E. The analysis of voice quality in speech processing. In: Chollet G, Esposito A, Faundez-Zanuy M, eds. *Nonlinear Speech Modeling and Applications*. Lecture Notes in Artificial Intelligence, Vol. 3445, Berlin/Heidelberg, Germany: Springer-Verlag; 2005:54–73.
4. Camargo Z, Madureira S. *Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS)*. Proceedings of the 4th Conference on Speech Prosody, Campinas, Brazil. Vol. 1. 20082008:57–60.
5. Mackenzie Beck J. Perceptual analysis of voice quality: the place of vocal profile analysis. In: Hardcastle WJ, Mackenzie Beck J, eds. *A Figure of Speech: A Festschrift for John Laver*. Mahwah, NJ: Lawrence Erlbaum Associates; 2005:285–322.
6. Hollien HF. *Forensic Voice Identification*. London: Academic Press; 2002.
7. Rose P. *Forensic Speaker Identification*. London: Taylor & Francis; 2002:289.
8. Watt D, Burns J. Verbal descriptions of voice quality differences among untrained listeners. *York Papers in Linguistics Series*. 2012;2:1–28.
9. Gold E, French P. International practices in forensic speaker comparison. *Int J Speech Lang Law*. 2011;18:293–307.

10. Laver J, Wirz S, Mackenzie Beck J, et al. *A perceptual protocol for the analysis of vocal profiles [work in progress]*. Edinburgh: University of Edinburgh; 1981:139–155.
11. Laver J. *The Gift of Speech*. Edinburgh: Edinburgh University Press; 1991.
12. San Segundo E, Mompean JA. A simplified vocal profile analysis protocol for the assessment of voice quality and speaker similarity. *J Voice*. 2017;31:644 e11.
13. Nolan F. Forensic speaker identification and the phonetic description of voice quality. In: Hardcastle WJ, Mackenzie Beck J, eds. *A Figure of Speech: A Festschrift for John Laver*. Mahwah, NJ: Lawrence Erlbaum Associates; 2005:385–411.
14. Jessen M. Speaker-specific information in voice quality parameters. *Forensic Linguist*. 2007;4:84–103.
15. BRASIL. *Ministério da Justiça. Departamento Penitenciário Nacional*. Levantamento Nacional de Informações Penitenciárias. INFOPEN. 2014.
16. Künzel H. Beware of the "telephone effect": the influence of the telephone transmission on the measurement of formant frequencies. *Forensic Linguist*. 2001;8:80–99.
17. Boersma P, Weenink D. PRAAT: doing phonetics by computer (version 5.1. 05) [Computer program]. Available at: http://www.fon.hum.uva.nl/praat/. Accessed May 1, 2009.
18. Broeders APA, Van Amelsvoort AG. *Lineup construction for forensic eyewitness identification: a practical approach*. Proceedings of the 7th International Congress of Phonetic Sciences, San Francisco. 19991999:1373–1376.
19. Stoet G. PsyToolkit: a software package for programming psychological experiments using Linux. *Behav Res Methods*. 2010;42:1096–1104.
20. Camargo Z, Madureira S. Avaliação vocal sob a perspectiva fonética: investigação preliminar. *Revista Distúrbios da Comunicação*. Vol. 20, 200877–98 1; São Paulo.
21. Laver J. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press; 1980 29; 31; 122.
22. R Development Core Team, version 3.4.3. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017.3-900051-07-0. Available at: https://www.r-project.org/. Accessed February 10, 2017.
23. Jackson S. Five ways to calculate internal consistency. R Bloggers: R news and tutorials contributed by (750) R bloggers. Available at: https://www.r-bloggers.com/five-ways-to-calculate-internal-consistency/. Accessed 26 February 2018.
24. Goforth C. Using and interpreting Cronbach's alpha. Research data services + sciences. University of Virginia Library. Available at: http://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/. Accessed 26 February 2018.
25. Gliem JA, Gliem RR. *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales*. Presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, The Ohio State University, Columbus, OH, October 8–10. 2003.
26. Multidimensional Scaling. Technical documents: statistics—textbook. Available at: https://documents.software.dell.com/statistics/textbook/multidimensional-scaling. Accessed 13 February 2017.
27. Young FW. Multidimensional scaling. Kotz S, Johnson NL, eds. *Encyclopedia of Statistical Sciences*. Vol. 5, New York: John Wiley & Sons; 1985.
28. McDougall K. Assessing perceived voice similarity using multidimensional scaling for the construction of voice parades. *Int J Speech Lang Law*. 2013;20:168.
29. Nolan F, McDougall K, Hudson T. Effects of the telephone on perceived voice similarity: implications for voice line-ups. *Int J Speech Lang Law*. 2013;20.
30. Rose P. Effect of telephone transmission. In: Selby H, Freckelton I, eds. *Expert Evidence*. Sydney: Thomson Lawbook Company; 2003.
31. Guillemin BJ, Watson C. Impact of the GSM mobile phone network on the speech signal: some preliminary findings. *Int J Speech Lang Law*. 2008;15:193–218.
32. Stevens KN, Fant G, Hawkins S. Some acoustical and perceptual correlates of nasal vowels. In: Channon R, Shockey L, eds. *In honor of Ilse Lehiste*. Dordrecht, Holland: Foris Publications; 1987.
33. Barbosa PA, Madureira S. *Manual de Fonética Acústica Experimental: Aplicações a Dados do Português*. São Paulo: Editora Cortez; 2015.
34. Dilley L, Shattuck-Hufnagel S, Ostendorf M. Glottalization of word-initial vowels as a function of prosodic structure. *J Phon*. 1996;24:423–444.
35. Gordon M, Ladefoged P. Phonation types: a cross-linguistic overview. *J Phon*. 2001;29:383–406.
36. Lima-Gregio AM. *Oclusiva glotal e laringalização em sujeitos com fissura palatina: um estudo segundo abordagem dinamicista. Tese (doutorado)*. Universidade Estadual de Campinas. 2011.
37. Hirson A, French P, Howard D. Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics. In: Lewis JW, ed. *Studies in General and English Phonetics: Essays in Honour of Professor J.D. O'Connor*. London: Routledge; 1995:230–240.
38. Laver J. *Principles of Phonetics*. Cambridge: Cambridge University Press; 1994 142; 154; 538.
39. Hirano M. *Clinical Examination of Voice*. Vienna/NewYork: Springer-Verlag; 1981.
40. Behlau M, Madazio G, Feijó D, et al. Avaliação de voz. Behlau M, ed. *Voz: O livro do Especialista*. Vol. I, São Paulo: Revinter; 2001:85–180.