FIA 2020/22
XII CONGRESSO/CONGRESO IBEROAMERICANO DE ACÚSTICA
XXIX ENCONTRO DA SOCIEDADE BRASILEIRA DE ACÚSTICA - SOBRAC
Florianópolis, SC, Brasil

Acoustic Analysis of Five Different Voice Quality Parameters: a Pilot Study

Meireles, A. R.[1]

[1] Laboratório de Fonética, Universidade Federal do Espírito Santo, ES, Brasil, meirelesalex@gmail.com

Resumo

Este trabalho analisa cinco diferentes qualidades de voz produzidas com a nota C4 sustentada (262 Hz), a saber: a) voz soprosa (Br): voz modal com escape de ar pelas pregas vocais; b) voz fraca (W): voz modal com mínima pressão subglótica (pouco audível); c) voz modal (M): fonação semelhante à fala regular; d) voz crepitante (Cr): voz modal com adição de vibração de prega irregular; e e) voz tensa (T): voz modal com pressão subglótica máxima. Para a análise acústica foi utilizado o software VoiceSauce que extraiu automaticamente treze parâmetros de medidas de curto prazo: 1- H1: amplitude relativa do primeiro harmônico; 2- H1H2: diferença de amplitude entre o primeiro e o segundo harmônico; 3- H1A3: diferença entre a amplitude do primeiro harmônico e a amplitude do harmônico de pico na região F3; 4- CPP: proeminência do pico cepstral; 5- Energia: medida da intensidade da voz; 6-9: relação harmônico-ruído de 0-500, 0-1500, 0-2500, 0-3500 Hz; 10-11: frequência de pico do primeiro e segundo formantes; 12-13: largura de banda de F1 e F2. Os resultados estatísticos mostraram que os parâmetros diminuem da seguinte maneira: 1) H1: M > Br > T > Cr > W; 2) H1H2: W > Br > (M = Cr) > T; 3) H1A3: (W=Br) > M > Cr > T; 4) B1: T > W > Br > M > W; 5) B2: M > T > Cr > Br > W; 6) CPP: T > M > Br > Cr > W; 7) Energia: T > M > Cr > Br > W. Além disso, HNR apresentou o mesmo padrão decrescente para as quatro condições (W > M > T > Br > Cr), e os formantes F1 e F2 também apresentaram um padrão decrescente semelhante (T > M > Br > Cr > W).

Palavras-chave: qualidade de voz, qualidade da vogal, modo de fonação, f0 sustentado.

PACS: 43.70.Gr

Abstract

This paper analysis five different voice qualities produced with a sustained C4 note (262 Hz), as follows: a) breathy voice (Br): modal voice with air escape throughout the vocal folds; b) weak voice (W): modal voice with minimum subglottal pressure (barely audible); c) modal voice (M): regular speech-like phonation; d) creaky voice (Cr): modal voice with the addition of irregular fold vibration; and e) tense voice (T): modal voice with maximum subglottal pressure. For the acoustic analysis we used the software VoiceSauce that automatically extracted thirteen parameters of short-term measures: 1- H1: relative amplitude of the first harmonic; 2- H1H2: difference in amplitude between the first and the second harmonics; 3- H1A3: difference between the first harmonic amplitude and the amplitude of the peak harmonic in the F3 region; 4- CPP: cepstral peak prominence; 5- Energy: measure of voice intensity; 6-9: harmonics-to-noise ratio from 0-500, 0-1500, 0-2500, 0-3500 Hz; 10-11: peak frequency of the first and second formants; 12-13: bandwidth of F1 and F2. The statistical results have shown that the parameters decrease in the following manner: 1) H1: M > Br > T > Cr > W; 2) H1H2: W > Br > (M = Cr) > T; 3) H1A3: (W=Br) > M > Cr > T; 4) B1: T > W > Br > M > W; 5) B2: M > T > Cr > Br > W; 6) CPP: T > M > Br > Cr > W; 7) Energy: T > M > Cr > Br > W. Moreover HNR presented the same decreasing pattern for the four conditions (W > M > T > Br > Cr), and the formants F1 and F2 also presented a similar decreasing pattern (T > M > Br > Cr > W).

Keywords: voice quality, vowel quality, phonation mode, sustained f0.

1. INTRODUCTION

According to Keating and Esposito [1], the three most common voice qualities (henceforth VQs) in natural languages are related to the degree of vocal folds adduction: i) creaky voice: greater glottal and supraglottal (aryepiglottic and ventricular folds) constriction; ii) breathy voice: posterior glottal abduction; iii) modal voice: vocal folds constriction in between creaky and breathy voices. As examples of voice qualities used with phonological contrast we may cite Gujarati (modal voice x breathy voice, cf. [2]), White Mong and Jalapa Zapotek (modal x breathy x creaky, cf. [3]; [4]). Also, there are some languages in which the degree of tension in the thyroarytenoid (TA) muscle is responsible for phonological contrast such as Takhian Thong Chong with four (modal x tense x breathy x breathy-tense, cf. [5]) and Lachixo Zapotec with five (falsetto, breathy, creaky, modal, whispery, cf. [6]) Vqs.

Although DiCanio's research [5] do not define tense phonation, it is implied they are following Edmodson and Esling's definition [7]: "'tense' or 'pressed' phonation is produced by bracing the vocal folds against the ventricular folds (ventricular incursion) with a sphincteric compression of the arytenoids." [5, p.165]. Besides, Esling [8] and Edmondson and Esling [7] have shown that there is a dual-fold relationship between VQ and vowel quality, i.e., VQ is influenced by vowel quality and also the opposite. Another phenomenon that is related to vowel quality is breathiness. DiCanio [5] relates that breathy and/or breathy-tense are correlated with higher vowels.

Based on the work by Catford [9] and Laver [10] on phonatory and articulatory settings in speech, Laver and Mackenzie-Beck [11] developed a profile for voice quality analysis named VPAS. Recently, supported by the work of their predecessors, John Esling and colleagues have developed a model that proposes a laryngeal articulator and a system for transcribing voice quality ([12]; [13]). Specifically, the definition of VQ in this article will be based on the assumptions of the VPAS model, which is similar to the voice quality definition in the works by John Esling.

2. THE DEFINITION OF VOICE QUALITY

As in our previous study on singing [14], voice production is understood here as stated by Laver [10], who defines voice quality as all the vocal characteristics that are related to speech, including laryngeal, supra-laryngeal, and muscular tension and voice dynamical features (see [15] for a review). The core of VQ analysis is based on the most common voice found in the literature: modal voice.

Garellek [16] comments that modal voice can be defined as i) a sound that it neither creaky nor breathy, or ii) the normal voice quality of a subject. Of course, in the second definition, a speaker that has a constant creaky-voice like quality, this will be his modal voice, and if he speaks a language with phonological distinction between modal and creaky, he will have to produce a sound even creakier to make the phonological distinction. According to him, "we still know very little about individual differences in voice quality" [16, p. 7].

Considering that there are different types of vocal fold adduction, we propose in this paper to investigate five VQs, as follows: i) weak voice: the minimum amount of subglottal pressure in order to move the vocal folds and produce voicing; ii) breathy voice: voicing only in the anterior part of the vocal folds with a consequently leak of air in their posterior part; iii) modal voice: full contact of the vocal folds throughout its extension with the subject's comfortable subglottal pressure; iv) strong voice: the maximum amount of subglottal pressure that the subject can produce; v) creaky voice: constriction of the supraglottal folds (aryepiglottic and ventricular folds) causing irregular vibration of the vocal folds.

3. THE RESEARCH START UP AND LITERATURE REVIEW

The motivation for the study is the lack of a full study that analyzes the mutual influence of vowel quality and the subject in VQ's spectral parameters, and, therefore, the absence of detailed information on how to interpret and differentiate the different VQ settings of vowels.

Some questions will be addressed here: Does the subject have different values of the VQ parameters in relation to different degrees of vocal fold settings? How does the vowel quality influence the VQ parameters? Although the subject has an intentional note to produce (C4=262 Hz), will there be different f0 in function of the different VQs? Some previous studies shed some light to answer these questions.

Regarding the parameters H1 (relative amplitude of the first harmonic), H1-H2 (difference in amplitude between the first and the second harmonic), H1-A3 (difference between the first harmonic amplitude and the amplitude of the peak harmonic in the third formant region), Hillenbrand and Houde [17] found that the amplitude of H1 in breathy voices, which have lesser vocal fold adduction, is higher compared to the modal voice. As a result, voices with smaller vocal fold adduction such as breathy voices have higher H1, H1-H2 and H1-A3 compared to voices with greater vocal fold adduction such as modal and creaky voices. This previous result was also found by Esposito [18] for the language Zapotec.

DiCanio [5] discusses the correlation of phonation types with certain acoustic measurements. According to him, greater adductive tension causes greater excitation of higher harmonics. Therefore, “one measures the amplitude of higher harmonics to see, albeit indirectly, how tense the vocal folds are during their vibration” [5, p. 167]. Also, he states that “low-range measures like H1-H2 have a close correlation to OQ values and are therefore good measures of the degree of glottal tension present in different phonation types" [5, p. 168]. Zhang [19] explains that “a longer open phase often leads to a more dominant first harmonic (H1) in the low-frequency portion of the resulting voice source spectrum". Also, Stevens [20] showed that the spectral tilt is correlated with the abruptness of vocal fold closure. Based on this argument, H1-H2 and H1-A3 values are expected to be smaller with greater vocal adduction.

Similarly, Ladefoged and Maddieson [21] found that lax vowels had higher H1-H2 values than tense vowels in some minority languages of China, and Esposito [22] found that H1-A3 was the best measure to distinguish breathy from modal phonation in the language Chong. In addition, Hanson et al. [23] comments that \ when vocal folds close completely but nonsimultaneously along their length, the source spectral tilt will increase", resulting, therefore in smaller values of H1-A3. This study also comments that H1-H2 is also correlated with the open quotient of the vocal folds, and the corrections in the parameters H1, H2, A3 are necessary once we want to compare vowels and speakers. In another study, Stevens [20] mentions that H1-H2 is larger for the lax vowel [ɪ] in comparison to the tense vowel [æ]. Based on this study, Hanson et al. [23] concludes that these differences suggest that the reduced vowel is produced with a greater open quotient and a larger glottal opening. Zhang [19] provides another correlation between physiology and acoustics: “for a glottal waveform of a very short open phase, the second harmonic (H2) or even the fourth harmonic (H4) may become the most dominant harmonic. Voice source with a weak H1 relative to H2 or H4 is often associated with a pressed voice quality."(p. 2619) Millgard et al. [24] found that H1-H2 decrease is correlated with increase in vocal fold adduction, and Mooshammer [25] presents higher values of H1-A3 with soft voices in comparison to loud voices.

Apart from being the most used measurements for distinguishing phonological VQs, Kreiman and colleagues gives a justification for using acoustic measurements in VQ analysis, “H1-H2, the spectral slope in the mid-frequencies (roughly 1-3 kHz), and high frequency excitation (harmonic and inharmonic) each account for more than 24% of variance in spectral shapes, suggesting that these features may be important determinants of voice quality” (26, p.608). Keating et al. [27] comments that H1-H2 with correction for the effect of formants is the most important measure of phonation contrasts across languages. See, for example, the use of H1 by Esposito [18] to show this parameter increases from creaky to modal and to breathy voice, and Garellek and Keating [28] that used H1-H2 to show this parameter increases from creak to modal and to breathy voice.

In consideration of the speech source harmonicity and broad voice intensity, v. Latoszek et al. [29] comments that the measures of cepstral

peak prominence (henceforth CPP) were the best predictors of breathiness with lower values for this VQ. Garellek and Keating [28] also shows the following pattern for the CPP in Jalap Zapotek: modal > creak > breathy. Hollien [30] found higher intensity values for modal voice than loft (falsetto-like) voice, and commented that greater thyroarytenoid muscle activity results in smaller values of HNR.

4. DATA AND METHODS

Five vowel qualities [a, ɛ, i, ɔ, u] sustained for 1000 milliseconds with a note C4 (262 Hz) were recorded with five voice qualities (modal, creaky, breathy, weak, strong), generating a total of 25000 (5 vowels x 5 VQ x 1000 ms) data for analysis of nine spectral parameters of VQ. Weak voice is considered as a sound produced with the minimum subglottal pressure necessary for vibrating the vocal folds. On the other hand, strong voice is the voice produced with the maximum subglottal pressure that the subject can support for vibrating the vocal folds.

In order to assure the right note tuning, before recording each vowel, the note C4 was played in a Yamaha PSR E36 keyboard, and then the five VQs for each vowel were recorded (eg.: note C4 recorded with creaky [a], modal [a], breathy [a], weak [a], strong [a]). As we can notice in tables 2, 4, 6, 8, and 10, this procedure was effective for ensuring a perfect pitch in all VQs. Based on his self-perception, the subject claimed that five voice qualities were created with five different laryngeal configurations, and that the note C4 (262 Hz) was effectively produced in all cases. The investigated subject was a vocalist with a bachelor degree in Music, with more than 20 years of experience in choral and pop singing, phonetician with specialization in voice quality in speech and singing, and with VPAS training (the author A.R.M.1).

The audios were recorded using a Shure Beta 58a microphone at a sampling rate of 96 kHz, converted to WAV (mono) in Praat [34], and then the audio was down-sampled to 16 kHz in VoiceSauce. To avoid errors related to the recording methods, before calculating the acoustic measures, the audios were normalized to the maximum amplitude in the software Audacity.

For the acoustic analysis we used the software VoiceSauce [35, 36] that automatically extracted thirteen parameters of short-term measures (H1*2: relative amplitude of the first harmonic; H1*H2*: difference in amplitude between the first and the second harmonic; H1*A3*:difference between the first harmonic amplitude and the amplitude of the peak harmonic in the F3 region; CPP: cepstral peak prominence; Energy: measure of voice intensity; HNR5, HNR15, HNR25, and HNR35: harmonics-to-noise ratios from 0-500 Hz, 0-1500 Hz, 0-2500 Hz, and 0-3500Hz; F1: peak frequency of the first formant; F2: peak frequency of the second formant; B1: bandwidth of F1; B2: bandwidth of F2). For f0 extraction we used VoiceSauce' Straight measure (strF0), and for formant extraction we used the Snack method [37]. The settings for calculating the formants and bandwidths were: window length of 25 ms, steps of 1 ms, LPC order of 12, and pre-emphasis of 0.98.

5. RESULTS

All statistical analyses were run in RStudio (version 1.0.153). First we checked the histograms of the five different voice qualities for every vowel to observe whether they exhibit a normal distribution, and then ran a Shapiro test for statistical significance. As no data passed the normality test, we ran a Kruskal-Wallis test with the spectral parameters as a function of the vowel for each voice quality (eg. H1* ~ breathy vowel), and then a Dunn's test for multiple comparisons with Bonferroni correction. Afterwards wer ran this same statistical procedures for the spectral parameters as a function of the voice quality for each vowel (eg. H1* ~ voice quality, vowel [a]). Statistical significance was found for all parameters regarding the Kruskal-Wallis test. Moreover, considering the fundamental frequency (see tables 2, 4, 6, 8, and 10) all the voice qualities were produced around the note C4 (=262 Hz).

5.1 The influence of the vowel quality on spectral parameters

Flanagan [38] and Titze [39] have discussed the non-linear relation between source and filter. Therefore, it is expected that the vowel

quality may influence the voice quality. Nevertheless, it is very rare to find articles that predicts how this influence is manifested on the voice quality parameters. Esling [8] and Edmondson and Esling [7], for example, have shown that there is a dual-fold relationship between VQ and vowel quality, i.e., VQ is influenced by vowel quality and also the opposite. In another study, DiCanio [5] relates that breathy and/or breathy-tense are correlated with higher vowels.

In a previous paper [40], we have proposed the following hypotheses regarding the influence of the vowel quality on voice quality:

H1. H1* has greater values in high vowels, intermediate in mid vowels, and smaller values in low vowels.

H2: H1*H2* has greater values in high vowels, intermediate in mid vowels, and smaller values in low vowels.

H3: H1*A3* has greater values in high vowels, intermediate in mid vowels, and smaller values in low vowels.

H4. CPP has greater values in low vowels, intermediate in mid vowels, and smaller values in high vowels.

H5. Energy has greater values in low vowels, intermediate in mid vowels, and smaller values in high vowels.

H6. HNR (HNR5, HNR15, HNR25, and HNR35) have greater values in low vowels, intermediate in mid vowels, and smaller values in high vowels.

5.1.1 Breathy voice

Considering only the breathy voice, Tables 1, 2 and figures 1 to 9 show that: H1: partially corroborated since [a̤]=[ɔ̤] < [i̤] < [ɛ̤] < [ṳ] for H1*; H2: not corroborated since we observed somewhat the reverse pattern: [i̤] < [ṳ] = [ɔ̤] < [a̤] < [ɛ̤] for H1*H2*; H3: not corroborated since [ɔ̤] < [ṳ] < [i̤] < [a̤] < [ɛ̤] for H1*A3*; H4: not corroborated since [a̤] < [ɛ̤]=[ɔ̤]= [i̤] < [ṳ] for CPP; H5: not corroborated since [a̤] < [i̤] < [ɛ̤] < [ɔ̤] < [ṳ] for Energy; H6: not corroborated since [a̤] = [ɛ̤] = [i̤] = [ɔ̤] = [ṳ] for HNR5, [a̤] < [ɔ̤] < [ɛ̤] < [ṳ] < [i̤] for HNR15, [a̤] < [ɛ̤] < [ɔ̤] < [i̤] < [ṳ] for HNR25 and HNR35.

Table 1: Spectral parameters for all vowels with breathy voice quality.

V	H1*	H1H2*	H1A3*	CPP	Energy
a̤	15.5(1.8)	7.4(2.4)	15.6(3.0)	16.4(2.0)	5.5(2.2)
ɛ̤	18.0(3.3)	8.0(3.2)	17.3(3.9)	19.5(2.1)	7.4(3.8)
i̤	16.8(4.0)	3.5(4.5)	7.0(4.3)	19.1(2.8)	6.7(5.0)
ɔ̤	15.3(3.3)	6.8(4.9)	-8.8(4.5)	19.8(1.8)	11.8(7.9)
ṳ	19.4(4.1)	5.9(4.6)	0.9(4.8)	20.6(1.7)	17.5(11.2)

Table 2: Spectral parameters for all vowels with breathy voice quality (continuation).

V	HNR5	HNR15	HNR25	HNR35	sF0
a̤	29.9(3.9)	19.1(3.5)	19.9(3.6)	19.7(3.3)	263.3(3)
ɛ̤	30.3(4.8)	22.6(4.6)	21.0(4.2)	21.5(4.1)	262(4)
i̤	29.0(6.5)	31.4(6.5)	29.0(6.1)	26.8(5.8)	262(5)
ɔ̤	30.3(4.8)	20.3(5.1)	23.4(5.1)	23.0(4.8)	262(3)
ṳ	28.3(5.3)	26.0(5.5)	32.3(6.1)	31.4(5.6)	265(3)

5.1.2 Weak voice

Considering only the weak voice, Tables 3, 4, and figures 1 to 9 show that: H1: corroborated since [a_w] < [$ɔ_w$] < [$ɛ_w$] < [i_w] < [u_w] for H1*; H2: partially corroborated since [$ɔ_w$] < [a_w] < [$ɛ_w$] < [i_w] = [u_w] for H1*H2*; H3: partially corroborated since [$ɔ_w$] < [u_w] < [a_w] < [$ɛ_w$] < [i_w] for H1*A3*; H4: not corroborated since [a_w] < [u_w] < [$ɛ_w$] < [i_w] < [$ɔ_w$] for CPP; H5: not corroborated since [a_w] < [i_w] < [$ɛ_w$] < [u_w] < [$ɔ_w$] for Energy; H6: partially corroborated since [u_w] < [$ɛ_w$] = [i_w] < [a_w] < [$ɔ_w$] for HNR5, [a_w] < [$ɛ_w$] = [$ɔ_w$] < [u_w] < [i_w] for HNR15, [$ɛ_w$] < [a_w] < [$ɔ_w$] < [u_w] = [i_w] for HNR25 and HNR35.

5.1.3 Modal voice

Considering only the modal voice, Tables 5, 6, and figures 1 to 9 show that: H1: corroborated since [a] < [ɛ] < [ɔ] = [i] = [u] for H1*; H2:

partially corroborated since [ɛ] < [ɔ] < [a] < [u] < [i] for H1*H2*; H3: not corroborated since [u] < [a] = [i] = [ɛ] < [ɔ] for H1*A3*; H4: not corroborated since [u] = [a] < [i] < [ɛ] < [ɔ] for CPP; H5: not corroborated since [i] < [a] < [u] < [ɛ] < [ɔ] for Energy; H6: partially corroborated since [u] < [ɛ] = [ɔ] < [i] < [a] for HNR5, [ɛ] = [ɔ] < [u] < [a] < [i] for HNR15, [ɛ] < [ɔ] < [a] < [u] = [i] for HNR25, and [ɛ] < [ɔ] < [a] = [i] < [u] for HNR35.

Table 3 Spectral parameters for all vowels with weak voice quality.

V	H1*	H1*H2*	H1*A3*	CPP	Energy
a_w	6.1(4.1)	5.3(4.9)	9.3(9.0)	15.4(1.5)	0.6(0.2)
$ɛ_w$	9.1(3.3)	6.4(7.1)	12.2(3.9)	17.4(1.7)	1.1(0.4)
i_w	10.1(3.7)	12.3(6.1)	15.9(6.5)	17.9(2.0)	1.0(0.9)
$ɔ_w$	7.6(3.7)	2.4(4.8)	-10.8(3.7)	18.4(1.5)	2.6(1.4)
u_w	11.0(2.6)	11.5(5.2)	6.4(9.4)	16.3(1.9)	2.0(0.8)

Table 4 Spectral parameters for all vowels with weak voice quality (continuation).

V	HNR5	HNR15	HNR25	HNR35	sF0
a_w	37.9(3.6)	29.3(3.0)	33.5(3.4)	32.8(3.6)	259(2)
$ɛ_w$	36.2(6.0)	32.6(5.4)	31.5(4.9)	32.7(4.4)	262(2)
i_w	37.3(3.9)	43.4(3.7)	41.8(3.6)	40.8(3.6)	263(5)
$ɔ_w$	40.0(4.3)	33.1(4.0)	38.4(3.4)	37.4(2.8)	260(4)
u_w	34.1(5.5)	34.9(4.8)	42.3(4.8)	41.4(4.2)	263(4)

Table 5: Spectral parameters for all vowels with modal voice quality.

V	H1*	H1H2*	H1A3*	CPP	Energy
a	15.8(4.8)	0.9(4.7)	6.1(10.7)	23.0(1.7)	23.0(17.9)
ɛ	16.0(3.3)	-5.9(2.1)	6.4(4.0)	24.8(2.8)	60.8(20.1)
i	18.7(2.4)	12.4(1.5)	7.5(4.2)	23.7(2.6)	12.2(3.5)
ɔ	18.6(3.0)	-3.5(3.0)	9.6(7.4)	26.5(2.5)	75.7(24.8)
u	18.6(2.8)	9.3(3.2)	-4.7(4.5)	22.5(2.8)	33.0(12.6)

Table 6: Spectral parameters for all vowels with modal voice quality (continuation).

V	HNR5	HNR15	HNR25	HNR35	sF0
a	42.5(4.7)	30.7(4.4)	31.4(4.5)	30.0(4.4)	258(2)
ɛ	33.0(4.8)	24.3(4.5)	23.6(4.6)	23.6(4.3)	262(2)
i	34.1(5.0)	35.0(4.6)	33.0(4.6)	30.0(4.3)	263(3)
ɔ	32.6(4.7)	24.5(4.3)	26.1(4.3)	26.6(4.3)	261(3)
u	31.7(4.4)	28.4(4.3)	34.1(4.4)	33.3(4.6)	263(4)

5.1.4 Strong voice

Considering only the strong voice, Tables 7, 8, and figures 1 to 9 show that: H1: not corroborated since [as] = [ɛs] = [is] = [us] < [ɔs] for H1*; H2: partially corroborated since [ɛs] = [ɔs] = [as] = [us] < [is] for H1*H2*; H3: not corroborated since [is] < [ɛs] < [us] = [as] = [ɔs] for H1*A3*; H4: corroborated since [is] < [us] < [ɔs] < [ɛs] = [as] for CPP; H5: corroborated since [is] < [us] = [ɛs] < [ɔs] < [as] for Energy; H6: partially corroborated since [us] = [is] < [ɛs] < [ɔs] < [as] for HNR5, not corroborated since [as] < [is] = [ɛs] = [ɔs] < [us] for HNR15, not corroborated since [as] = [is] = [us] < [ɛs] < [ɔs] for HNR25, and not corroborated since [is] = [ɔs] < [as] = [us] < [ɛs] for HNR35.

Table 7: Spectral parameters for all vowels with strong voice quality.

V	H1*	H1H2*	H1A3*	CPP	Energy
a_s	16.5(3.6)	0.13(5.3)	1.2(6.9)	27.4 (3.3)	123.3 (37.5)
$ɛ_s$	16.1(2.6)	-0.18(3.9)	-4.9(8.9)	27.5(3.6)	106.6(27.7)
i_s	16.0(5.6)	7.5(4.4)	-6.7(7.7)	24.4(3.2)	62.4(41.6)
$ɔ_s$	10.7(8.5)	-5.9(2.7)	1.9(8.8)	26.3(5.2)	105.6(52.6)
u_s	15.3(5.0)	-0.21(3.8)	0.7(10.3)	25.5(5.3)	90.4(51.6)

Table 8: Spectral parameters for all vowels with strong voice quality (continuation).

V	HNR5	HNR15	HNR25	HNR35	sF0
aₛ	31.3(9.1)	22.4(8.8)	23.8(9.2)	24.4(9.4)	257(4)
ɛₛ	29.9(7.8)	26.3(9.5)	25.6(10.3)	25.7(10.4)	265(4)
iₛ	26.0(6.6)	27.9(6.8)	25.0(6.3)	22.0(5.9)	266(7)
ɔₛ	33.7(9.7)	25.9(10.4)	27.9(10.4)	27.1(10.5)	269(6)
uₛ	25.6(7.5)	23.1(7.6)	25.0(7.5)	24.8(7.4)	272(3)

5.1.5 Creaky voice

Considering only the creaky voice, Tables 7, 8, and figures 1 to 9 show that: H1: not corroborated since [ɔ̰] < [a̰] = [ḭ] < [ɛ̰] = [ṵ] for H1*; H2: partially corroborated since [ɔ̰] < [ɛ̰] < [a̰] < [ṵ] < [ḭ] for H1*H2*; H3: not corroborated since [ɔ̰] < [ṵ] < [a̰] < [ɛ̰] < [ḭ] for H1*A3*; H4: not corroborated since [ṵ] < [ɔ̰] = [a̰] < [ɛ̰] = [ḭ] for CPP; H5: not corroborated since [ḭ] < [a̰] < [ɛ̰] < [ɔ̰] = [ṵ] for Energy; H6: not corroborated for HNR 5 since [ṵ] = [ɔ̰] < [a̰] = [ɛ̰] < [ḭ], not corroborated for HNR15 since [ɔ̰] < [a̰] < [ṵ] < [ɛ̰] < [ḭ], partially corroborated for HNR25 [ɔ̰] < [a̰] = [ɛ̰] < [ṵ] < [ḭ], and partially corroborated [ɔ̰] < [a̰] = [ɛ̰] < [ṵ] < [ḭ] for HNR35.

Table 9: Spectral parameters for all vowels with creaky voice quality.

V	H1*	H1H2*	H1A3*	CPP	Energy
a̰	13.0(4.7)	1.96(5.1)	-0.9(5.8)	17.3(3.4)	23.0(9.3)
ɛ̰	14.9(3.5)	-1.11(5.8)	2.8(5.2)	18.9(3.1)	38.2(13.3)
ḭ	13.0(2.9)	9.6(1.5)	5.4(3.7)	19.0(3.0)	9.2(3.6)
ɔ̰	11.5(3.1)	-3.8(2.2)	-7.2(7.2)	16.9(3.5)	42.7(9.0)
ṵ	14.3(2.4)	3.5(4.1)	-3.8(5.5)	15.4(2.5)	42.1(18.0)

Table 10: Spectral parameters for all vowels with creaky voice quality.

V	HNR5	HNR15	HNR25	HNR35	sF0
a̰	16.2(12.3)	4.7(12.5)	8.6(11.6)	9.0(10.8)	252(10)
ɛ̰	16.7(9.3)	9.2(8.4)	7.1(7.8)	8.9(7.1)	254(6)
ḭ	23.4(11.5)	26.6(10.9)	23.6(10.4)	21.0(9.5)	267(6)
ɔ̰	9.3(7.8)	-1.8(8.1)	3.1(8.1)	3.7(7.4)	260(9)
ṵ	8.3(6.6)	6.4(6.6)	11.3(6.9)	12.4(6.6)	256(7)

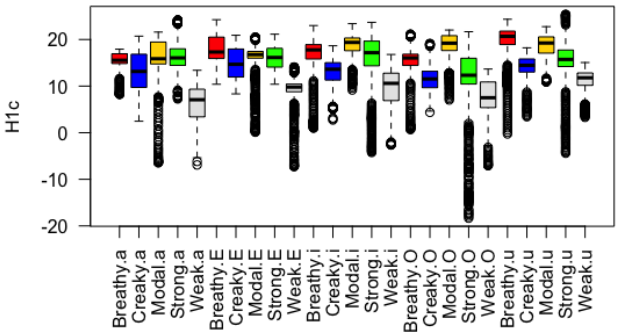

Figure 1: H1* as function of the vowels and the voice qualities.

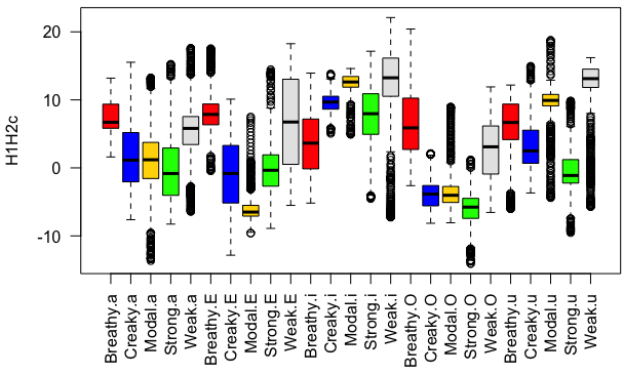

Figure 2: H1*H2 as a function of the vowels and the voice qualities.

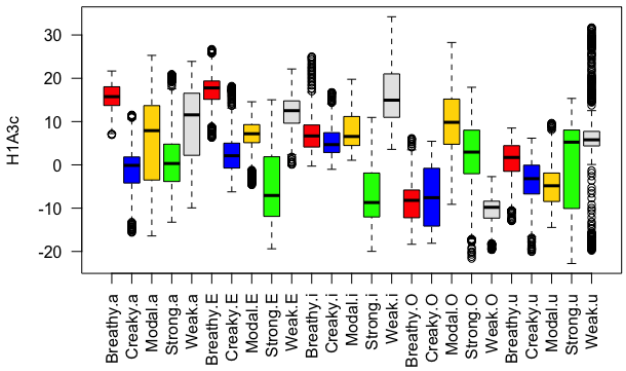

Figure 3: H1*A3* as a function of the vowels and the voice qualities.

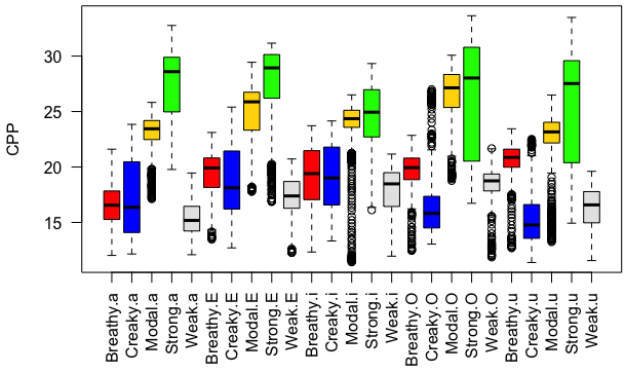

Figure 4: CPP as a function of the vowels and the voice qualities.

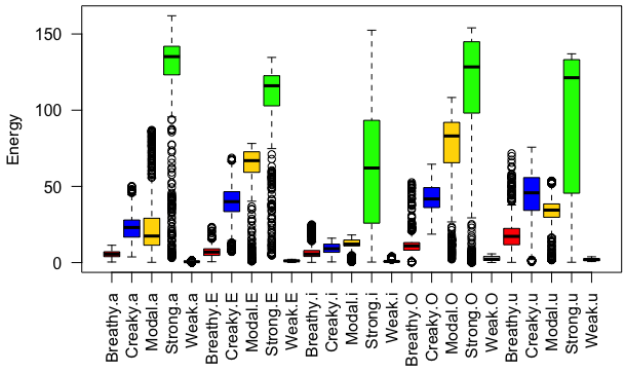

Figure 5: Energy as a function of the vowels and the voice qualities.

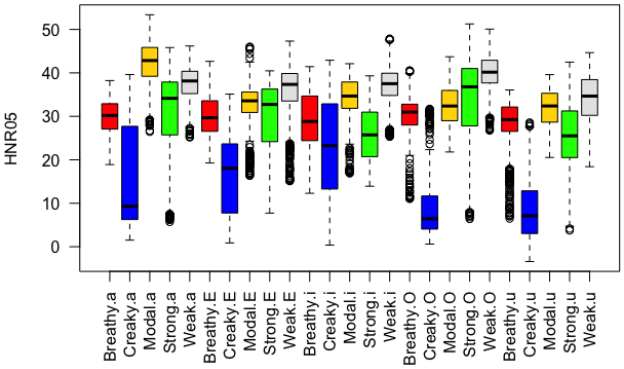

Figure 6: HNR5 as a function of the vowels and the voice qualities.

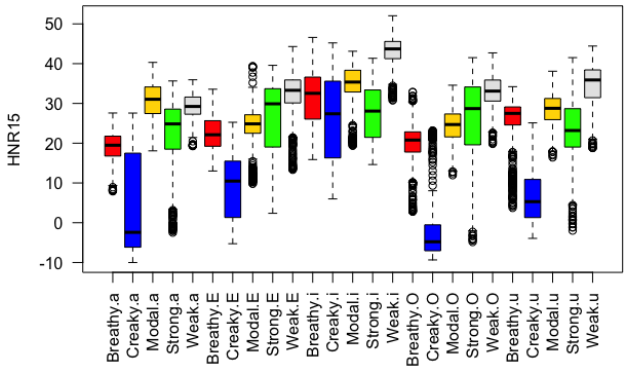

Figure 7: HNR15 as a function of the vowels and the voice qualities.

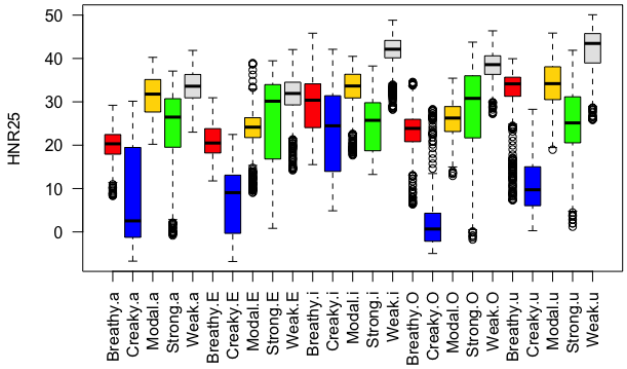

Figure 8: HNR25 as a function of the vowels and the voice qualities.

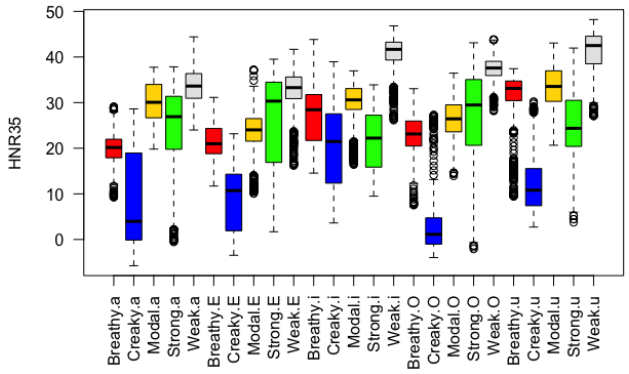

Figure 9: HNR35 as a function of the vowels and the voice qualities.

5.2 The influence of the voice quality on spectral parameters

After displaying the influence of the vowel quality on voice quality parameters, we are going to show how the voice quality changes the voice quality parameters for each individual vowel.

Higher values of H1* are usually associated in the literature with breathy voice [17, 18], i.e., the presence of a certain air escape (breath) through the vocal folds during phonation (modal voice). Thus, considering that a higher H1 is correlated with greater air escape and lesser vocal adduction, we hypothesize the following order in our data: breathy voice > weak voice > modal voice > creaky voice > strong voice.

According to Keating and colleagues [27], higher values of H1*H2* are associated with breathy and lax phonations, and lower values with creaky and tense phonations. We hypothesize the following order for this parameter: breathy voice > weak voice > modal voice > creaky voice > strong voice.

According to Gordon and Ladefoged [41], spectral tilt (H1*A3*) is "the degree to which intensity drops off as frequency increases" (p. 15) and "characteristically most steeply positive for creaky vowels and most steeply negative for breathy vowels" (p. 15). Higher H1*A3 values, therefore, are found in phonations with greater air escape, i.e., lesser activation of the muscles thyroarytenoid and lateral cricoarytenoid muscles. Then, we hypothesize the following pattern for this parameter: breathy voice > weak voice > modal voice > creaky voice > strong voice.

According to Hillenbrand and colleagues [17, p. 772], "the idea behind the CPP measure is that a highly periodic signal should show a well defined harmonic structure and, consequently, a more prominent cepstral peak than a less periodic signal". Also, Shue [35] states that it "should be larger for modal phonations and smaller for breathy phonations", and "smaller for creaky voices if the phonation is aperiodic". We hypothesize the following order for this parameter: strong > modal > weak > breathy > creaky.

Energy, according to Shue [35, p. 61-62], may be correlated with vocal effort. So, we hypothesize the following order for this parameter: strong > modal > creaky > breathy > weak.

Yumoto and Gould [42] found that the HNR for a healthy group varied between 7.0 and 17.0 dB with a mean of 11.9 dB. They also considered values below 7.4 dB to be pathological. As HNR can be correlated to noise in the spectrum, we hypothesize the following order for this parameter: strong > modal > weak > creaky > breathy.

According to these previous findings in the literature, we have proposed the following hypotheses:

H7. H1 will decrease in the following order: breathy > weak > modal > creaky > strong.
H8: H1*H2* will decrease in the following order: breathy > weak > modal > creaky > strong.
H9: H1*A3* will decrease in the following order: breathy > weak > modal > creaky > strong.
H10. CPP will increase in the following order: breathy < creaky < weak < modal < strong.
H11. Energy will increase in the following order: weak < breathy < creaky < modal < strong.

H12. HNR (HNR5, HNR15, HNR25, and HNR35) will increase in the following order: breathy < creaky < weak < modal < strong.

5.2.1 Vowel [a]

Considering only the vowel [a], Tables 1 to 10, and figures 1 to 9 show that: H7: not corroborated since breathy = modal = strong > creaky > weak for H1*; H8: corroborated since breathy > weak > modal > creaky > strong for H1*H2*; H9: partially corroborated since breathy > weak > modal > strong > creaky for H1*A3*; H10: not corroborated since weak < breathy = creaky < modal < strong for CPP; H11: corroborated since weak < breathy < creaky < modal < strong for Energy; H12: not corroborated for HNR 5 since creaky < breathy < strong < weak < modal, not corroborated for HNR15 since creaky < breathy < strong < weak < modal, not corroborated for HNR25 creaky < breathy < strong < modal < weak, and not corroborated for HNR35 since creaky < breathy < strong < modal < weak.

5.2.2 Vowel [ɛ]

Considering only the vowel [ɛ], Tables 1 to 10, and figures 1 to 9 show that: H7: not corroborated since breathy > modal = strong > creaky > weak for H1*; H8: not corroborated since breathy > weak > strong > creaky > modal for H1*H2*; H9: corroborated since breathy > weak > modal > creaky > strong for H1*A3*; H10: partially corroborated since weak < creaky < breathy < modal < strong for CPP; H11: corroborated since weak < breathy < creaky < modal < strong for Energy; H12: not corroborated for HNR 5 since creaky < strong < breathy < modal < weak, not corroborated for HNR15 since creaky < breathy < modal < strong < weak, not corroborated for HNR25 since creaky < breathy < modal < strong < weak, and not corroborated for HNR35 since creaky < breathy < strong < modal < weak.

5.2.3 Vowel [i]

Considering only the vowel [i], Tables 1 to 10, and figures 1 to 9 show that: H7: not corroborated since modal > breathy = strong > creaky > weak for H1*; H8: not corroborated since weak = modal > creaky > strong > breathy for H1*H2*; H9: partially corroborated since weak > breathy = modal > creaky > strong for H1*A3*; H10: not corroborated since weak < breathy = creaky < modal = strong for CPP; H11: corroborated since weak < breathy < creaky < modal < strong for Energy; H12: not corroborated for HNR 5 since creaky = strong < breathy < modal < weak, not corroborated for HNR15 since creaky = strong < breathy < modal < weak, not corroborated for HNR25 since creaky = strong < breathy < modal < weak, and not corroborated for HNR35 since creaky = strong < breathy < modal < weak.

5.2.4. Vowel [ɔ]

Considering only the vowel [ɔ], Tables 1 to 10, and figures 1 to 9 show that: H7: not corroborated since modal > breathy > creaky > strong > weak for H1*; H8: corroborated since breathy > weak = modal > creaky > strong for H1*H2*; H9: not corroborated since modal > strong > creaky = breathy > weak for H1*A3*; H10: not corroborated since creaky < weak < breathy < strong < modal for CPP; H11: corroborated since weak < breathy < creaky < modal < strong for Energy; H12: not corroborated for HNR5, HNR15, HNR25, and HNR35 since creaky < breathy < modal < strong < weak.

5.2.5. Vowel [u]

Considering only the vowel [u], Tables 1 to 10, and figures 1 to 9 show that: H7: not corroborated since breathy > modal > strong > creaky > weak for H1*; H8: not corroborated since weak > modal > breathy > creaky > strong for H1*H2*; H9: not corroborated since weak > breathy > strong > creaky > modal for H1*A3*; H10: not corroborated since creaky < weak < breathy < modal < strong for CPP; H11: partially corroborated since weak < breathy < modal < creaky < strong for Energy; H12: not corroborated for HNR5, HNR15, HNR25, and HNR35 since creaky < strong < breathy < modal < weak.

6. CONCLUSION

Although preliminary, our acoustic results support the idea that the action of different vowels and different voice qualities result in change of the spectral source parameters. The main results have shown that 9 spectral parameters (H1*, H1*H2*, H1*A3*, CPP, Energy, HNR5, HNR15, HNR25, HNR25) varied as a function of 5 vowels and 5 voice qualities.

The next steps of research are to gather physiological and perceptual data to examine more deeply how the voice qualities and vowels influence the spectral parameters.

REFERENCES

[1] Keating, P., Esposito, C. (2006). Linguistic voice quality. UCLA Working Papers in Phonetics 105, 85-91.

[2] Esposito, C.M. (2011). The perception of pathologically-disordered phonation by gujarati, english, and spanish listeners. Language and Speech 54 (3), 415430.

[3] Esposito, C.M. (2011). The perception of pathologically-disordered phonation bygujarati, english, and spanish listeners. Language and Speech 54 (3), 415-430.

[4] Esposito, C.M., ud Dowla Khan, S. (2012). Contrastive breathiness across consonants and vowels: A comparative study of gujarati and white hmong. Journal of the IPA 42 (2), 123-143.

[5] DiCanio, C.T. (2009). The phonetics of register in takhian thong chong. Journal of the IPA 39(2), 162-188.

[6] Sicoli, M.A. (2010). Shifting voices with participant roles: Voice qualities and speech registers in mesoamerica. Language in Society 39, 521-553..

[7] Edmondson, J.A., Esling, J.H. (2006). The valves of the throat and their functioning in tone, vocal register, and stress: Laryngoscopic case studies. Phonology 520 23(2), 157-191.

[8] Esling, J.H. (2005). There are no back vowels: The laryngeal articulator model. Canadian Journal of Linguistics 50, 13-44.

[9] Catford, J.C. (1977). Fundamental Problems in Phonetics. Edinburgh University Press, Edinburgh.

[10] Laver, J. (1980). The phonetic description of voice quality. Cambridge University Press, Cambridge.

[11] Laver, J., Mackenzie-Beck, J. (2007). Vocal profile analysis scheme: A user's manual. Speech Science Research Centre, Queen Margaret University College-QMUC.

[12] Ball, M.J., Esling, J., Dickson, C. (1995). The voqs system for the transcription of voice quality. Journal of the IPA 25(2), 71-80. doi:10.1017/S0025100300005181.

[13] Esling, J., Moisik, S.R., Benner, A., Crevier-Buchman, L. (2019). Voice Quality: The Laryngeal Articulator Model. Cambridge University Press, Cambridge.

[14] Meireles, A., Mixdorff, H. (2020). Voice quality in low and high registers in two different styles of singing, in: Proc. Speech Prosody 2020, Tokyo, Japan. p. 5 pages.

[15] Meireles, A., de Medeiros, B.R. (2018). Acoustic analysis of voice quality in Iron Maiden's songs, in: Proc. Speech Prosody 2018, Poznan, Poland. pp. 20-24.

[16] Garellek, M. (2019). The routledge handbook of phonetics, in: Katz, W.F., Assmann, P.F. (Eds.), The phonetics of voice. Routledge, London, pp. Part II, Chapter 4.

[17] Hillenbrand, J., Houde, R.A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. Journal of Speech, Language, and Hearing Research 39, 311-321.

[18] Esposito, C.M. (2010). Variation in contrastive phonation in santa ana del valle zapotec. Journal of the IPA 40(2), 181-198. doi:10.1017/S0025100310000046.

[19] Zhang, Z. (2016). Mechanics of human voice production and control. JASA 140, 650 2614-2635.

[20] Stevens, K.N. (1994). Prosodic influences on glottal waveform: preliminary data, in: Proceedings of the International Symposium on Prosody, pp. 53-64.

[21] Ladefoged, P., Maddieson, I. (1985). Tense and lax in four minority languages of china. Journal of Phonetics 13, 433-454.

[22] Esposito, C. (2006). The effects of linguistic experience on the perception of phonation. Ph.D. thesis. UCLA.

[23] Hanson, H.M., Stevens, K.N., Kuo, H.K.J., Chen, M.Y., Slifka, J. (2001). Towards models of phonation. Journal of Phonetics 29, 451-480.

[24] Millgård, M., Fors, T., Sundberg, J. (2016). Flow glottogram characteristics and perceived degree of phonatory pressedness. Journal of Voice 30(3), 287-292.

[25] Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in german. JASA 127(2), 605 1047-1058.

[26] Kreiman, J., Gerratt, B.R., Antoñanzas-Barroso, N. (2007). Measures of the glottal source spectrum. Journal of Speech, Language, and Hearing Research 50, 595-610.

[27] Keating, P., Esposito, C., Garellek, M., Khan, S., Kuang, J. (2010). Phonation contrasts across languages. UCLAWorking Papers in Phonetics 108, 188-202.

[28] Garellek, M., Keating, P. (2011). The acoustic consequences of phonation and tone interactions in jalapa mazatec. Journal of the IPA 41 (2), 185-205.

[29] v. Latoszek, B.B., Maryn, Y., Gerrits, E., Bodt, M.D. (2018). A meta-analysis: Acoustic measurement of roughness and breathiness. Journal of Speech, Language, and Hearing Research 61, 298-323.

[30] Hollien, H. (1974). On vocal registers. Journal of Phonetics 2, 125-143.

[31] Herbst, C. (2004). Evaluation of Various Methods to Calculate the EGG Contact Quotient. Ph.D. thesis. Kungliga Tekniska Högskolan.

[32] Herbst, C., Ternström, S., Švec, J. G. (2009). Investigation of four distinct glottal configurations in classical singing—A pilot study. JASA 125 (3), March 2009.

[33] Herbst, C., Hess, M, Müller, F., Švec, J. G., Sundberg, J. (2015). Glottal Adduction and Subglottal Pressure in Singing. Journal of Voice, Vol. 29, No. 4, pp. 39-402.

[34] Boersma, P., 2001. Praat: a system for doing phonetics by computer. Glot International, 341-345.

[35] Shue, Y.-L. (2010). The Voice Source in Speech Production: Data, Analysis and Models, PhD Thesis, UCLA.

[36] Shue, Y.-L., Keating, P., Vicenik, C, Yu. (2011). VoiceSauce: A program for voice analysis, Proceedings of the ICPhS XVII, 1846-1849.

[37] Sjölander, K. (2004). “Snack Sound Toolkit.” KTH Stockholm, Sweden, http://www.speech.kth.se/snack/ (last viewed Sep. 2019).

[38] Flanagan, J. L. (1968). Source-system interaction in the vocal tract, Annals of the New York Academy of Sciences, vol. 155(1), pp. 9–17.

[39] Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory, J Acoust Soc Am., vol. 123(5), p. 2733–49, 2008.

[40] Meireles, A., Mixdorff, H. (submitted). Acoustic Study of the Voice Quality of Brazilian Portuguese Stressed Vowels, Proceedings of Speech Prosody 2022.

[41] Gordon, M. and P. Ladefoged, P. (2001). Phonation types: a crosslinguistic overview, J. of Phonetics, 29, 383-406, 2001.

[42] Yumoto, E., Gould, W., Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness, J. Acoust. Soc. Am. 71, 1544–1550.