VOICE QUALITY DESCRIPTION FROM A PHONETIC PERSPECTIVE: SUPRALARYNGEAL AND MUSCULAR TENSION SETTINGS

name(s) of the author(s):Zuleica Camargo 1, Perpetua Coutinho 1,Sandra Madureira 1,Luiz Carlos Rusilo2
affiliation(s) 1 Integrated Acoustic Analysis and Cognition Laboratory-LIAAC. Graduate Applied Linguistics and Language Studies Program.Pontifical Catholic University of São Paulo, Brazil 2 Engineering Center. Federal University of Alfenas, Brazil
email address(es): zcamargo@pucsp.br, perpetuacoutinho@hotmail.com, madusali@pucsp.br, luiz.rusilo@unifal-mg.edu.br

ABSTRACT

This study aims at investigating the acoustic correlates of supralaryngeal and muscular tension voice quality settings (VQS).

Speech samples from 40 speakers were perceptually analyzed by means of the Vocal Profile Analysis Scheme (VPAS). Acoustic analysis comprised formant frequency (F1, F2 and F3) and intensity measurements (I1, I2 and I3) of seven Brazilian Portuguese vowels. Multivariate statistical analysis was performed to take into account the discriminant capability of the acoustic measures to detect VQS as well as influences from gender, age and vowel type.

The results have shown the discriminant power of formant measures to predict neutral and non-neutral VQS: I1, F3, I3 and I2 correlated with supralaryngeal VQS and F1 and F1, I2 and I3 with muscular tension.The canonical correlation analysis of VQS, formant measures and gender variables have shown that neutral VQS grouped with F1 and F2 and non-neutral VQS grouped with F3, I1, I2 and I3.

Keywords: voice quality, production,perception, resonance, acoustics

1. INTRODUCTION

Detailed acoustic descriptions of voice qualities are not easy to perform, despite the importance of voice quality descriptions to vocal expressiveness [1][19], voice disorders [8]-[9] and crosslinguistic investigations [6]. In consequence, many acoustic descriptions of voice qualities have not been extensively addressed in studies focusing acoustic-perceptual correlations.

Furthermore, there are few studies addressing acoustic correlates of the supralaryngeal and the muscular tension mechanisms which can be identified perceptually with the help of the Vocal Profile Analysis Scheme (VPAS) [17].

As a phonetically-grounded evaluation of voice quality, the VPAS is based on the phonetic description of voice quality model by Laver [16]. The basic analytical unit is the setting, as a long-term muscular tendency in the vocal apparatus (supralaryngeal, laryngeal and muscular tensions domains). The settings are described as variations from a reference setting, the neutral one, in which there are no constrictive or expansive effects in the vocal tract cavities; no shortening or lengthening of the extension of the vocal tract and no extreme variations in terms of muscular tension activity in the supralaryngeal and laryngeal parts of the vocal tract.

The first step (first pass) in the perceptual evaluation using the VPAS is to identify the presence of neutral x non-neutral settings. In the second step (second pass) the judge is asked to evaluate in a scalar degree (from 1 to 6) the non-neutral settings only [17].

To perform the VPAS evaluations, judges need phonetic background and experience on the use of the profile. The phonetic descriptive model of voice quality proposed by Laver [16] follows two principles: susceptibility and compatibility. The principle of susceptibility accounts for the fact that some speech segments are more susceptible to the effects of specific voice quality settings and the principle of compatibility states that some voice quality can co-occur and others not. In this way, rounded vowels are more susceptible to the lip spreading setting of voice quality than unrounded vowels, and the lip spreading and the lip rounding settings are not compatible.

Since susceptibility is an important issue for the perception of the kinds of settings, the phonetic model proposes the use of key-speech segments to identify the effect of voice quality settings on them. So, in evaluating voice quality settings, focus on key-speech segments is important since their inherent features are modified by the long-term voice quality settings and turn out to be salient from the perceptual point of view. For the sake of describing phonetic voice quality settings, the corpus designed for this work took into account the principle of susceptibility and made use of key-speech segments [3][4].

Although perceptual and acoustic correlates for laryngeal activity have been taken into account [5][8]-[9][20]-[21]and differences related to gender and age have been found [2][14][18][20], the detailed description of the supralaryngeal (longitudinal, transversal and velopaharyngeal) and muscular (vocal tract and larynx) tension settings is not usually addressed, despite the fact that changes around 4-8% in formant frequencies are supposed to be audible and can be influenced by vocal quality settings[12][14][22].

For the sake of detecting supralaryngeal and muscular tension settings, vowels were chosen as key-speech segments in this work. Analysis of their production can reveal mobilizations of the lips, jaw, tongue (tip, body and root), pharynx, velopharyngeal mechanism, laryngeal height and in distinct muscular tension activity patterns (hypo and hyperfunctional) which differ from their inherent characteristics because are influenced by specific voice quality settings.

The following research question is pursued in this work: how do perceived long-term voice quality settings correlate with acoustic characteristics?

The present study addresses this issue for it aims at investigating the acoustic correlates of supralaryngeal and muscular tension voice quality settings in VPAS, by means of the acoustic description of vocal tract resonances (formant analysis). Gender, age and vowel type variables are discussed.

2. METHODS

The corpus was composed by semi-spontaneous speech and 03 key-sentences samples (ks) (including key-speech segments), recorded in a radio studio by 40 subjects (from 20 to 58 years old, 17 males and 23 females).

These samples were perceptually evaluated in consensus by two expert subjects according to the VPAS.

For the sake of describing the acoustic correlates of perceived voice quality settings, formantfrequencies (F1, F2, F3) and intensities (I1, I2, I3) were measured by means of the Akustyk (BartekPlichta's add-on to Praat software, available in: http://bartus.org/akustyk/akustyk_features.html). From the 03 key sentences aforementioned, seven oral vowel sounds in the stressed syllable position were selected or analysis: [a] in *cidade*, [ɛ] in *detesto*, [e] in *dele*, [i] in *turista*, [ɔ] in *poderosa*, [o] in *todos*and [u] in *estudo*.

The voice quality settings judgments and the acoustic measures were statistically analyzed by means of several multivariate analysis [13]-[15] procedures. To derive the statistical measures the s*oftware* Xlstat (Addinsoft) was used.

The agglomerative hierarchical cluster analysis was applied to the perceptual data. Discriminant analysis was applied to search for the influence of perceptual data to predict gender, age and vowel types analyzed. Discriminant analysis was also applied to search for the discriminant capability of acoustic data (frequency and intensity formant measures) to predict VQS, and also gender and age. The canonical correlation analysis was applied to search for the correlations between perceptual and acoustic data, male and female speakers considered apart. Moreover, logistic regression was applied to analyze the influences of each oral vowel in detecting neutral and non-neutral VQS.

This project was approved by Ethics Committee (number 101/11).

3. RESULTS AND DISCUSSION

3.1. The perceptual data - supralaryngeal and muscular tension voice quality settings

The agglomerative hierarchical cluster analysis yielded seven classes: Class 1 (46,4% of total occurrences) - supralaryngeal-transversal (rounded lips, minimized and extended range of lips, open and minimized range jaw, retracted tongue tip and tongue body, pharyngeal expansion), supralaryngeal-velopharyngeal (denasal) and muscular tension (vocal tract hypofunction and laryngeal hypofunction) settings; Class 2 (21,4%) - supralaryngeal-transversal (spread lips, closed jaw, advanced tongue tip and pharyngeal constriction), supralaryngeal - longitudinal (raised larynx) and muscular tension (vocal tract hyperfunction) settings; Class 3 (7,1%) – supralaryngeal-transversal (extended range of lips and jaw) settings; Class 4 (7,1%) - supralaryngeal-transversal (protruded jaw and advanced tongue body) settings; Class 5 (7,1%) - supralaryngeal-velopharyngeal (audible nasal air escape and nasal) settings; Class 6 (7,1%) – supralaryngeal-transversal (lowered tongue body) and supralaryngeal - longitudinal (lowered larynx) settings; and Class 7 (3,5%) – muscular tension (laryngeal hyperfunction) setting.

The discriminant analysis showed that muscular tension settings (neutral and non-neutral groups) were influential to distinguish speaker´s gender (86,1% of congruence). Specifically in the supralaryngeal group, nasal (36,9%), labiodentalization (30,3%) and lips extended range (29,8%) settings were influent. This gender related distribution will be considered on the next topic, when addressing the acoustic-perceptual correlations.

The discriminant analysis for the groups of voice quality settings (supralaryngeal and muscular tension) and age revealed discriminant capability to distinguish between neutral and non-neutral muscular tension settings (99,5% of congruence). This result can be interpreted in relation to changes in the vocal tract structure as a consequence of bone, muscles, cartilages and membranous tissues adaptations to the aging process, specially lowered tongue root and larynx [2][14][18]. It is important to point out that the lower the larynx, the more lax is the vocal folds contact [7] and that directly affects the muscular tension settings and reveals the interdependence between voice quality settings.

Finally, discriminant analysis for perceptual data and vowels revealed the capability for distinguishing between neutral and non-neutral muscular tension settings (degree ofcongruence: 97,5%).

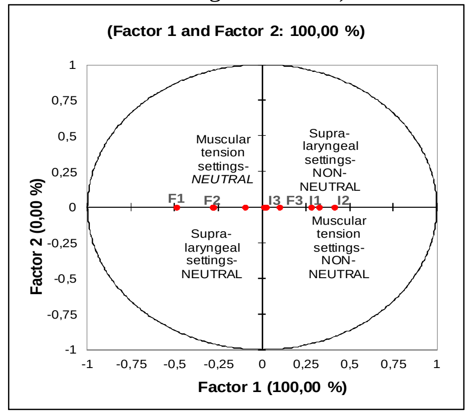
3.2. Correlations between perceptual and acoustic voice quality data: some insights on vocal tract resonances

Discriminant analysis revealed that formant measures, especially F3, I2 and I3 were able to distinguish between neutral and non-neutral supralaryngeal and muscular tension VQS. The relevant measures for the detection of supralaryngeal voice quality settings were I1 (72,9%), F3 (63,7%), I3 (36,2%) and I2 (35,3%) and for muscular tension voice quality settings were F1 (49,7%), I2 (41,6%) and I3 (35,2%).

The discriminant capability of acoustic variable for distinguishing gender was 94,41% for female and for male was 93, 28%.

When considering the discriminant analysis of the groups of variables (supralaryngeal and muscular tension variables, formant measures (F1,F2, F3; I1, I2, I3) and gender), neutral voice quality settings grouped with F1 and F2 measures and non-neutral voice quality settings grouped with F3, I1, I2 and I3 (Figure 1). The relevant variables for this distribution were: F1 (49,1%), I2 (41,3%), I1 (32,5%) and F2 (28,5%). Shifts in the lowest three or four formants are generally associated with individual changes in vowel segment productions and higher formant frequencies are usually associated to voice quality [14], but the findings in the present study (Figure 1) showed relevance for F3 and voice quality settings.

Figure 1: Diagram from discriminant analysis (voice quality settings and formant measures related to gender) (factor 1: I2, I1, F1, F2 and muscular tension settings influences).



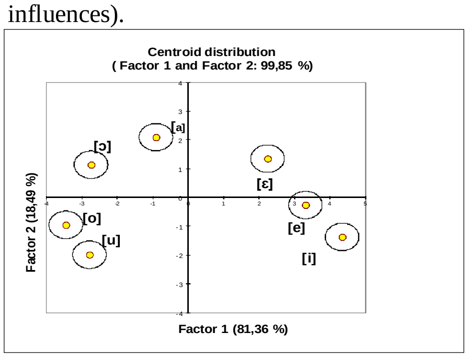
Concerning both male and female data, the canonical correlation analysis showed distinct levels of correlations between voice quality settings and formants measures. For the male group, F3 correlated to vocal tract hypofunction (32,1% of correlation), pharyngeal constriction (27,2%), closed (27,2%) and protruded (24,7%) jaw settings, as well as I3 correlated with jaw-minimized range (31,4%) and retracted tongue body (28,4%). For the female group, F3 correlated to raised tongue body (25% of correlation) and I2 showed correlations with advanced tongue body (24,7%) and raised larynx (23,6%) settings. These findings rely on the principle of compatibility (in auditory and physiological terms) of voice quality settings [3][16], in which some mechanisms for reducing the dimensions of supraglottal cavities and associating muscular tension mechanisms seemed to be influential[5]. For females, the association of raised larynx and raised and advanced tongue body settings can be thought of as related to a mechanism of hyoid-larynx complex concerning a specific-to-gender glottal configuration[10].

Discriminant analysis taking into account acoustic and age variables revealed that formant frequencies and intensities exhibited a high discriminant capability for the 21 years-old (92,86% of congruence for frequency and 94,64% for formant intensity measures). Other ages were underscored.

Considering the perceptual judgments of voice qualities (neutral and non-neutral groups) and formant measures (frequencies and intensities) in relation to vowels, discriminant analysis showed the discriminant capability of F1 (86,4%), I2 (70,8%), F2 (53,5%) and I1 (51,4%). These findings indicate that the methodological approach in this study differentiated the events related to segmental (vowels as key speech-segments) and prosodic (voice quality settings) levels [3][11]. They are illustrated by the centroid distributions of formant frequencies presented at Figures 2 and 3, respectively for male and female speakers.

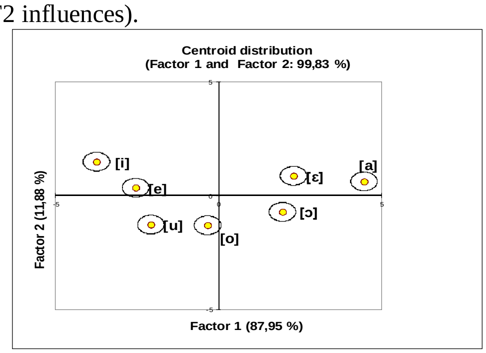
The discriminant capability of vowels by means of F1, F2 and F3 measures for male subjects (figure 2) varied from 70,5% of congruence (for [e]) to 88,24% (for [ɛ], [u] and [i]). The influential variables were F1 (99,3%) and F2 (95,3%).

Figure 2: Centroid distribution for vowels from discriminant analysis (formant frequencies) for male speakers (factor 1: F2 influences; factor 2: F1 influences).



The discriminant capability of vowels by means of F1, F2 and F3 measures for female subjects (figure 3) varied from 43,4% of congruence (for [e]) to 91,3% (for [i]). The influent variables were F1 (99,5%) and F3 (48,7%).

Figure 3: Centroid distribution for vowels from discriminant analysis (formant frequencies) for female speakers (factor 1: F1 influences; factor 2: F2 influences).



Centroid distributions of formants intensities are presented at Figures 4 and 5, respectively for male and female speakers. Lower percentage values were obtained in relation to intensity formant measures. These measures showed discrimination power.

For male (Figure 4), the scores ranged from 17,65% of congruence for [e] to 76,46% for [ɛ]. For female speakers (figure 5), values ranged from 26,09% for [i] to 69,57% for [ɛ].

Figure 4: Centroid distribution for vowels from discriminant analysis (formant intensities) for male speakers (factor 1: I2 and I1 influences; factor 2: I3 influences).

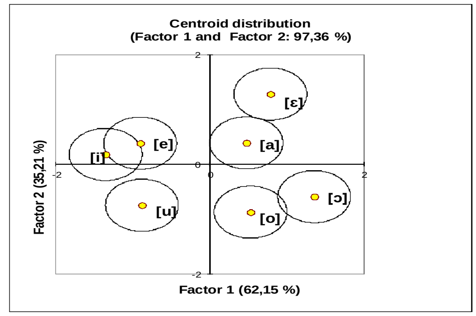
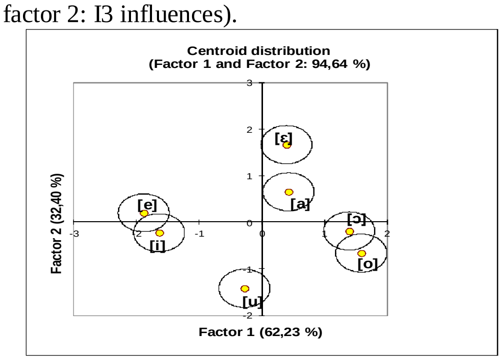


Figure 5: Centroid distribution for vowels from discriminant analysis (formant intensities) for female speakers(factor 1: I2 and I1 influences; factor 2: I3 influences).



Logistic regression applied to formant frequencies measures for each studied vowel showed the relevance of vowels in detecting non-neutral supralaryngeal and non-neutral muscular tension settings. For the group of supralaryngeal settings, for all Brazilian Portuguese oral vowels, no discriminant capability for the neutral settings was found. On the contrary, vowel [e] revealed 97,3% of discriminant capability for non-neutral supralaryngeal settings and the other six vowels showed total discriminant capability (100%). When tension settings were considered, non-neutral settings showed 100% of relevance for [o], 96% for [e], 91,67% for [ɔ], 84% for [i], 83,3%% for [u] and 80% for [ɛ] and [a]. For the neutral tension settings, lower values were obtained (0% for [o], 2% for [i],18,75% for [ɔ], 20% for [e] and [a] and 25% for [u]).

4. CONCLUSIONS

Formant frequency and intensity measures, mainly F3, I2 and I3, were found to be relevant to distinguish between neutral and non-neutral VQS. The relevant measures for the detection of supralaryngeal VQS were I1, F3 , I3 and I2 and for muscular tension VQS were F1, I2 and I3.

The results reinforce the relevance of analysing both supralaryngeal and muscular tension voice quality settings in clustering groups by gender and muscular tension in clustering groups by age.

5. REFERENCES

[1] Barbosa, P.A. 2009. Detecting changes in speech expressiveness in participants of a radio program. *Proc. Interspeech*Brighton, 2155-2158.

[2]Beck, J.M. 1999. Organic variation of the vocal apparatus. In: Hardcastle, W.J., Laver,J. (eds), *The handbook of phonetic sciences*. Malden: Blackwell, 256-297.

[3] Beck, J.M. 2005. Perceptual analysis of voice quality: the place of vocal profile analysis. In: Hardcastle,W.J., Mackenzie-Beck, J. (eds), *A figure of speech: a festschrift for John Laver*. Mahwah: Lawrence Erlbrum Associates, 285-322.

[4] Camargo, Z., Madureira, S. 2008. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). *Proc. 4th Int.Conf Speech Prosody* Campinas, 57-60.

[5] Camargo, Z., Madureira, S., Pessoa, A.N., Rusilo, L.C. 2012. Voice quality and gender: some insights on correlation between perceptual and acoustic dimensions. *Proc. 6th Int. Conf. Speech Prosody* Shanghai, 115-118.

[6] Esling, J.H. 2000. Crosslinguistic aspects of voice quality. In: Kent, R.D., Ball, M.J (eds), *Voice quality measurement*. San Diego: Singular Publishing, 25-35.

[7]Fink, B.R., Basek, M., Epanchin, V. 1956. The mechanism of opening of the human larynx. *Larygoscope* 66(4), 410-425.

[8]Gillespie, A.I., Dastolfo, C., Magid, N., Gartner-Schmidt, J. 2014. Acoustic analysis of four common voice diagnoses: moving toward disorder-specific assessment. *J. Voice* 28(5), 582–588.

[9]Hammarberg, B., Gauffin, J. 1995. Perceptual and acoustics characteristics of quality differences in pathological voices as related to physiological aspects. In: Fujimura, O., Hirano, M. (eds), *Vocal fold physiology*.San Diego: Singular Publishing, 283-303.

[10]Hammer, M.J., Krueger, M.A. 2014.Voice-related modulation of mechanosensory detection thresholds in the human larynx. *Exp. Brain Research* 231(1), 13-20.

[11]Hewlett, N., Beck, J.M. 2006. *An introduction to the science of phonetics* London: Lawrence Erlbaum Associates.

[12] Ives, D.T., Smit, D.R.R., Patterson, R.D. 2005. Discrimination of speaker size from syllables phrases. *J. Acoustic. Soc. Am.* 118(6), 3816-3822.

[13] Kreiman, J., Gerrat, B. 2000. Measuring vocal quality. In: Kent,R.D., Ball, M.J. (eds), *Voice quality measurement.* San Diego: Singular Publishing, 73-101.

[14] Kreiman, J., Sidts, D. 2011. *Foundations of voice studies: an interdisciplinary approach to voice production and perception* Malden: Wiley-Blackwell.

[15] Lattin, J., Carroland, D.J.D., Green, P.E. 2003. *Analyzing Multivariate Data*Independence: Thomson Brooks/Cole.

[16] Laver, J. 1980, Reprinted 2009. *The phonetic description of voice quality* Cambridge: Cambridge University Press.

[17] Laver. J., Wirz, S., Mackenzie, J., Hiller, S. 1981. A perceptual protocol for the analysis of vocal profiles. *Edinburg University Department of Linguistics Work in Progress* 14, 139-155.

[18]Linville, S.E. 2000. The aging voice. In: Kent, R.D., Ball, M.J. (eds), *Voice quality measurement*. San Diego: Singular Publishing, 359-395.

[19] Madureira, S. 2008. Reciting a sonnet: production strategies and perceptual. *Proc. 4th Int. Conf. Speech Prosody* Campinas, 697-700.

[20] Mendoza, E.,Valencia, N., Muñoz J., Trujillo, H.1996. Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *J. Voice* 10(1), 59-66.

[21] Rusilo, L.C., Madureira S., Camargo, Z. 2011. The validity of some acoustic measures to predict voice quality settings: trends between acoustic and perceptual correlates of voice quality. *Proc.4th ISCA Tutorial Research Workshop Experimental Linguistics* Paris, 115-118.

[22] Smith, D.R.R., Patterson, R. D., Turner, R. 2005. The processing and perception of size information in speech sounds. *J. Acoustic. Soc. Am.* 117(1), 305-318.