The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures

Zuleica Antonia de Camargo[1,2], Sandra Madureira[1]

[1]Integrated Acoustic Analysis and Cognition Laboratory – LIAAC, Pontifical Catholic University of Sao Paulo, Brazil
[2]Voice and Speech Laboratory - CEFAC Institute, São Paulo, São Paulo, Brazil

Abstract

The present study aims at describing voice quality settings by means of long-term, f0 and intensity analysis of samples which integrate the material developed with the purpose of introducing the Voice Profile Analysis Scheme-VPAS in the Brazilian Portuguese context (BP-VPAS). The voice quality database was recorded in a radio studio. The audio samples were analyzed by means of the SG Expression Evaluator, a script developed by Barbosa (2009), which automatically extracts fundamental frequency (f0) and its first derivate (df0), intensity, spectral tilt (SpTt) and Long-Term Average Spectrum (LTAS). These measures have been statistically analyzed by means of multiple and non-linear regression tests. The results contemplate the correlations between perceived voice quality settings and f0 and intensity long-term acoustic measures.

Key words: Voice Quality; Phonetics; Speech Acoustics; Auditory Perception; Speech Production Measures

Introduction

Studies focusing voice quality are rather incipient in the prosodic field in spite of its importance to the analysis of both vocal expressiveness and voice disorders.

Furthermore, there are few studies describing the long-term acoustic correlates of the laryngeal, the supralaryngeal and the tension settings which can be identified perceptually with the help of the Voice Profile Analysis Scheme (VPAS) developed by Laver et al (1981), as a phonetically-grounded evaluation of voice quality. The present study addresses these two issues for it aims at describing voice quality settings by means of long-term, f0 and intensity analysis of the audio samples which integrate the instructional material developed by a group of researchers with the purpose of introducing the Voice Profile Analysis Scheme-VPAS in the Brazilian Portuguese context: BP-VPAS (Camargo, Madureira, 2008) .

Due to the importance of the susceptibility of the speech segments to the effects of voice quality settings (Laver, 1980), a team of researchers from Queen Margareth University College (QMUC- Edinburgh)

engaged in elaborating an English corpus specially designed for auditory voice quality evaluation by means of VPAS. This procedure has also been adopted by the group of researchers at LIAAC-PUCSP (Integrated Acoustic Analysis and Cognition Laboratory of the Pontifical Catholic University of São Paulo) concerned with the adaptation of VPAS to the Brazilian Portuguese context so as to improve the material aimed at instructional purposes.

The description of voice quality settings by means of long-term acoustic analysis techniques have been proved relevant (Hammarberg, Gauffin, 1991; Cukier-Blaj et al, 2008).

The detailed description of long-term acoustic correlates of the perceived voice quality settings is not an easy task, but it can provide a detailed description of events related to voice quality settings in laryngeal, supralaryngeal and tension domains, as well as for voice dynamics description. This paper addresses this demand, applying a method for speech expressiveness change detection developed by Barbosa (2009) in order to describe samples from a group of adults without any complaints about their voices.

Methods

The building of the corpus took into account the principle of susceptibility proposed by Laver (1980) and made use of the keyspeech segments (Camargo, Madureira, 2008). The sentences that constitute the corpus were recorded in a radio studio and retrieved in the voice database of LIAAC. The building of the corpus took into account the principle of susceptibility proposed by Laver (1980) and made use of the keyspeech segments (Camargo, Madureira, 2008). It was recorded by 38 subjects, aged from 20 to 58 years, being 10 male and 28 female. Besides the keyspeech segment corpus, a semi-spontaneous corpus was also recorded. The 38 subjects were asked to talk about the city where they were born. Their recordings were evaluated by two phoneticians and integrated into the instructive material of the BP-VPAS, as samples illustrating voice quality settings (laryngeal, supralaryngeal and tension settings).

For the present analysis, 509 utterances (52 semi-spontaneous speech samples and 457 repetitions of the key-sentences).The total duration of the audio samples analyzed was 114 minutes, mean of 03 minutes per speaker.

The audio samples were analyzed by means of the SG Expression Evaluator, a script developed by Barbosa (2009), running in the software PRAAT, which automatically extracts acoustic parameters: fundamental frequency (f0- median, inter-quartile semi-amplitude, skewness and 0,995 quantile) and its first derivate (df0- mean, standard-deviation and skewness), intensity (skewness), spectral tilt (SpTt–mean, standard-deviation and skewness) and Long-Term Average Spectrum

(LTAS frequency standard-deviation). Data related to intensity measures were normalized taking into account the distance between lips and microphone and the intensity level at a reference tone of 1kHz one. Some of the measures were normalized during the application of the script, such as f0 and df0, SpTt and LTAS frequency.

These measures were statistically analyzed by means multiple and non-linear regression tests, in order to investigate the acoustic parameters as related to distinct perceived voice quality settings.

Results

The results contemplate the correlations, detected by means of statistical analysis, between perceived voice quality settings and acoustic measures, specially for retracted tongue tip (66,7% SD 27,2%), creaky voice (57,5% SD 7,8%), limited lips extension (28,5% SD 12,1%), laryngeal hyperfunction (23,1% SD 11,7%) and limited jaw extension (12,5% SD 11,7%).

F0 (median, skewness and 99,5% quantil) and f0 first derivate (mean, standard deviation and skewness) measures have correlated with laryngeal tension (both hyper and hypofunction) and laryngeal (mainly harsh voice, combined or not with whisper, breathy voice and creaky voice) settings. The 99,5% quartile measure made it possible to distinguish utterances characterized by periodicity form the ones characterized by aperiodicity.

Spectral Tilt measures have also correlated with laryngeal tension settings and the laryngeal settings which are characterized by aperiodicity. High values were related to laryngeal hypofunction and low values to hyperfunction. Besides Spectral Tilt, Intensity Skewness has also been helpful to characterize laryngeal tension settings.

Among the findings has been that of the association between supralaryngeal and the laryngeal tension settings: hyperfunction associated with tongue body settings, pharyngeal constriction and raised larynx and hypofunction with pharyngeal expansion and lowered larynx.

Another finding was related to the identification of modal voice settings. If there is little variation of f0 and df0 values across speech samples of the same speaker, voices are identified as modal. Furthermore, F0 and df0 measures were found to be useful to detect intermittent occurrences of voice quality settings such as laryngeal settings with irregular vocal fold vibration: breathy voice, harsh voice and creaky voice.

The investigation of voice quality settings based on the measures extracted by means of the SG Expressive Evaluator made it possible to correlate acoustic to perceptual dimensions and to identify instances of co-occurrence of settings. Those aspects are relevant not only to consider compensatory strategies to voice disturbances in the clinical context but also but also to foster research on speech expressivity (Madureira, 2008).

Conclusions

The results reinforce the importance of the description of voice quality based on long-term acoustic and perceptual correlations and the role of distinct keyspeech segments in the corpus designed for voice quality evaluation.

Acknowledgements

References

Barbosa, P.A. 2009. Detecting changes in speech expressiveness in participants of a radio program. In Proc. of Interspeech. v.1,2155-2158. Brighton, Reino Unido.

Camargo, Z., Madureira, S. 2008. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In Proc. of the 4th Conf on Speech Prosody v.1, 57–60. Campinas, Brazil.

Cukier-Blaj, S.C., Camargo, Z., Madureira, S. 2008. Long-term average spectrum loudness variation in speakers with asthma, paradoxical vocal fold motion and without breathing problems. In Proc.of the 4th Conf on Speech Prosody v.1, 41-44. Campinas, Brazil.

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.). 1995. Vocal fold physiology, 283-303. San Diego, Singular Publishing.

Laver, J. 1980. The phonetic description of voice quality. Cambridge University Press.

Laver, J., Wirs, S., Mackenzie, J., Hiller, S.M. 1981. A perceptual protocol for the analysis of vocal profiles. Edinburgh University Department of Linguistics Work in Progress 14,139-155.

Madureira, S. 2008. Reciting a sonnet: production strategies and perceptual. In Proc.of the 4th Conf. on Speech Prosody v.1, 697-700. Campinas, Brazil.