

The validity of some acoustic measures to predict voice quality settings: trends between acoustic and perceptual correlates of voice quality

Luiz Carlos Rusilo¹, Zuleica Antonia de Camargo², Sandra Madureira²

¹Department of Actuarial and Quantitative Methods-DAMQ-Pontifical Catholic University of São Paulo, Brazil

²Integrated Acoustic Analysis and Cognition Laboratory-LIAAC- Pontifical Catholic University of São Paulo, Brazil

<https://doi.org/10.36505/ExLing-2011/04/0028/000197>

Abstract

The present study aims at investigating the power of acoustic measures to predict voice quality settings detected perceptually. The corpus was composed by speech samples recorded by 60 subjects. The audio samples were analyzed in PRAAT by means of the SG Expression Evaluator Script (Barbosa, 2009), which extracts f0 and f0 first derivate, intensity, spectral tilt and long-term average spectrum measures. The same samples were perceptually evaluated by means of the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS). The methodological procedures made it possible to identify the clusters related to the acoustic measures group and that of the perceptual judgement results (cluster analysis) as well as to correlate these two groups (canonic correlation analysis).

Key words: speech acoustics, voice quality, perception, statistical analysis

Introduction

Many conditions that influence voice quality analysis are not addressed in studies that search for acoustic-perceptual correlations (Hammarberg, Gauffin, 1995). Furthermore, there are few studies describing the acoustic correlates of voice quality settings which can be identified perceptually with the help of the Voice Profile Analysis Scheme (VPAS) (Laver, Mackenzie-Beck, 2007).

This study aims at investigating the power of some acoustic measures (fundamental frequency-f0 and f0 first derivate, intensity, spectral tilt and long-term average spectrum-LTAS) to predict voice quality settings detected perceptually, by means of cluster analysis and canonic correlation analysis. The detailed description of acoustic correlates and its discriminatory power to predict 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.

Methods

The corpus was composed by semi-spontaneous speech samples and repetitions of 10 key-sentences samples (Camargo, Madureira, 2008), recorded in a radio studio by 60 subjects. The audio samples were analyzed by means of the SG Expression Evaluator, a revised version of the script developed by Barbosa (2009), running in the software PRAAT, which automatically extracts acoustic measures: 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).

Some of the measures were normalized during the application of the script, such as f0 and df0, SpTt and LTAS frequency. The same samples were perceptually evaluated by two expert subjects, using the Voice Profile Analysis Scheme for Brazilian Portuguese: BP-VPAS (Camargo, Madureira, 2008, 2010).

The acoustic measures and voice quality settings judgments were statistically analyzed by means of cluster analysis and canonic correlation analysis (Lattin et al, 2011) in order to investigate the validity of each acoustic parameter to predict the voice quality settings which had also been evaluated perceptually. The cluster analysis was applied intragroup and the canonic analysis intergroups. There were two groups under analysis: the acoustic measures and perceptual judgement results. To derive the statistical measures the *software* Xlstat from Addinsoft was used.

Results

The results point to some correlations between acoustic measures concerning f0, df0, intensity, SpTt and LTAS and the perceived voice quality settings.

The cluster analysis applied to perceptual data yielded four classes: Class 1 (70,1%) - modal (SD 1,3%, CI 67,6% to 72,7%), Class 2 - laryngeal hyperfunction, and harsh voice (SD 1,1%, CI 15,2% to 19,4%), Class 3 - pharyngeal expansion, lowered larynx and creaky voice (SD 0,8%, CI 7,0 to 10,1%) and Class 4 - spread lips, advanced tongue tip, pharyngeal constriction, vocal tract hyperfunction, creaky voice, whisper and harsh voice (SD 0,6%, CI 2,9 to 5,1%).

The cluster analysis applied to acoustic data yielded six classes, two of them comprising a large number of observations: Class 1 (20,80%) and Class 2 (76,11%). They were discriminated by df0-mean and skewness measures. The other four classes (3,09%) were discriminated by f0 and df0 measures. Figure 1 presents the dendrogram related to the perceptual measures and Figure 2 the correlations between acoustic and perceptual data.

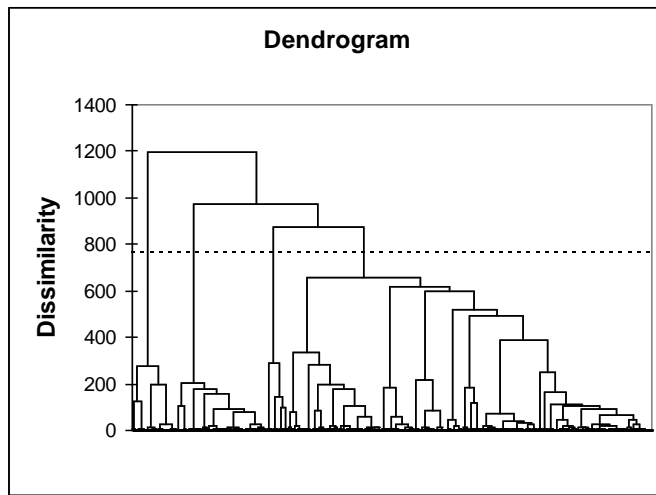


Figure 1. Dendrogram of perceptual data related to voice quality judgements (BP-VPAS).

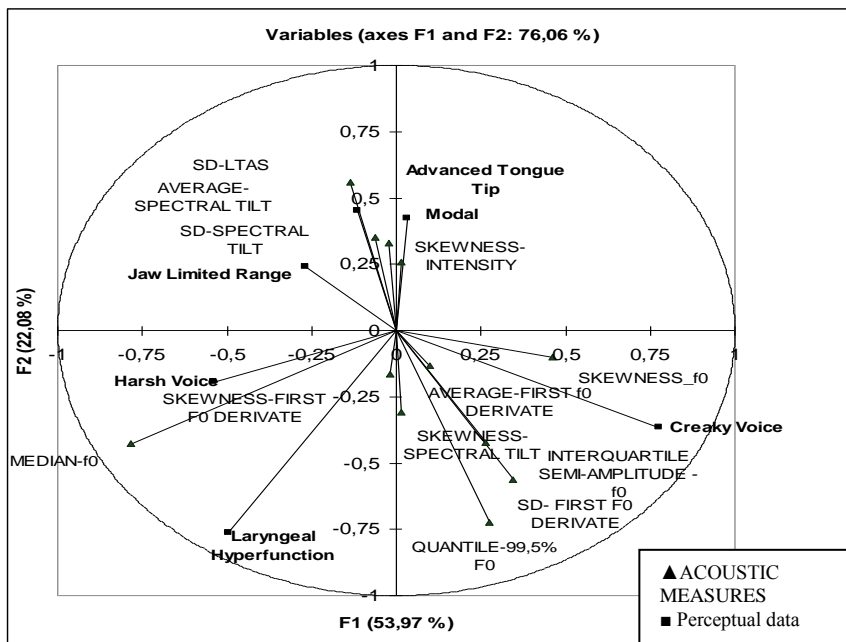


Figure 2. Correlations between acoustic and perceptual data.

The correlations shown in Figure 2 concern the most frequent voice quality settings (jaw minimized range, advanced tongue tip, laryngeal hyperfunction, modal, creaky and harsh voice) and acoustic measures.

Conclusions

The methodological procedures made it possible to identify the clusters related to the acoustic measures group and that of the perceptual judgement results (cluster analysis) as well as correlate these two groups (canonic correlation analysis).

Acknowledgements

We acknowledge Plínio Barbosa from UNICAMP for the revised version of the SG Expression Evaluator Script (2009).

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.
- Camargo, Z., Madureira, S. 2010. The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures In Proc. of the 3rd ISCA Workshop on Experimental Linguistics ExLing 2010, 33-36. Athens, Greece.
- 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*, 283-303. San Diego, Singular Publishing.
- Lattin, J., Carrol, Douglas J.D., Green, P.E. *Análise de dados multivariados*. São Paulo: Cengage Learning, 2011.
- Laver, J; Mackenzie-Beck J. 2007. *Vocal Profile Analysis Scheme-VPAS*. Queen Margareth University College-QMUC, Speech Science Research Centre, Edinburgh.