Low band spectral tilt analysis for pathological voice discrimination

Hugo Cordeiro
Department of Electronic, Telecommunications and Computers
High Institute of Engineering of Lisbon (ISEL)
Polytechnic Institute of Lisbon
Lisbon, Portugal
hcordeiro@deetc.isel.ipl.pt

Carlos Meneses
Department of Electronic, Telecommunications and Computers
High Institute of Engineering of Lisbon (ISEL)
Polytechnic Institute of Lisbon
Lisbon, Portugal
cmeneses@deetc.isel.ipl.pt

***Abstract*— This paper presents a new method for discriminating between subjects with healthy voices and subjects with diseases in the vocal folds. This method uses speech signals and spectral analysis of the sustained vowel /a/. The slope between a first band of the signal defined in the first two harmonics and a second band defined in the zone of the /a/ first formant contains information that allows to correctly classify the database of pathological voices of the University of São Paulo. The presented method can be applied in the direct analysis of spectra or implemented in high-level classifiers as a complement to other parameters.**

Keywords—Pathological Voices, Spectral Tilt, Spectral Band Analysis

I. INTRODUCTION

Pathological voice identification is the process of discriminate subjects with and without voice pathologies. The use of speech in these tasks, as a non-invasive, easy and quick method, is useful in screening situations or as a complementary method for the diagnosis.

Speech analysis for identifying pathological voices typically has the acoustic or spectral approach. Acoustic analysis methods provide perturbation measures such as jitter and shimmer [1], [2]. Spectral parameters such as energy spectrum [3], [4], mel-frequency cepstral coefficients (MFCC) [5]–[7] and formant analysis [8], [9] achieve an accuracy rate over 90%. One of the advantages of this approach is that pitch estimation, a difficult task when dealing with voice disorders [6], is not required. Typically, the sample used for analysis is the sustained vowel /a/, since it is produced with the vocal tract completely open and is correlated with the electroglottograph [10].

The method of characterizing pathological voices using formant analysis proposed in [9], explained in detail in the next section, contains some flaws: the first spectral peak that models the first two harmonics can be detected in healthy subjects, although as higher bandwidth; in some unhealthy subjects, with predominance of energy in the first two harmonics, the first peak is not detected due to the smaller slope of the signal.

This article presents a new method of spectral analysis to discriminate pathological voices that allows, on one hand, to find simple, reliable and intelligible parameters through visual analysis of a spectrum, and on the other hand can also be used in systems with high complexity, along with other characteristics (parameters), for the automatic tracking of pathological voices. The proposed parameter is the spectral tilt between the first two harmonics and the first formant, denominated Low Band Spectral Tilt (LBST).

The remainder of this article is organised as followed: section II presents the related work about voice pathologies identification based on spectral analysis. Section III describes the database and section IV the proposed method. Section V presents the results and discussion. Conclusions are presented in section VI.

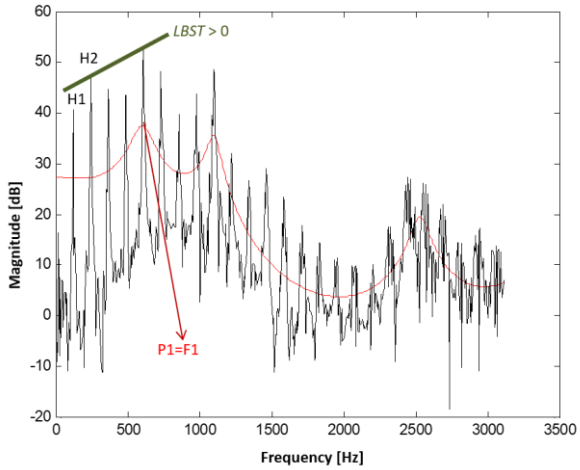
II. RELATED WORK

For healthy voices, the spectral envelope peaks with small bandwidth have formants information. In [9], to verify if there was a change in the formants for unhealthy voices, spectral envelope was estimated by spectral analysis with a 30^{th} order linear prediction filter.

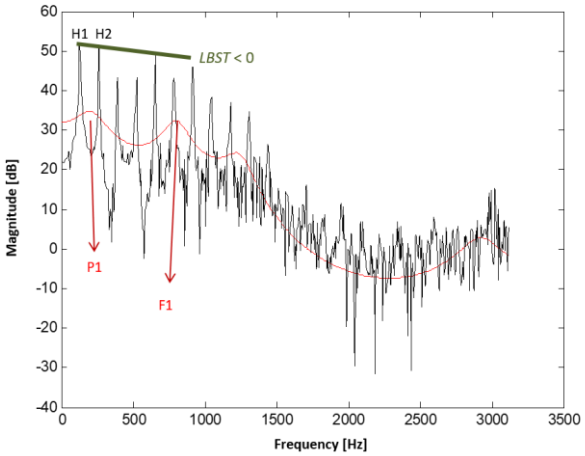
For a sustained vowel /a/ produced by healthy subjects, the harmonic with the highest energy correspond to the first formant (F1), matching to the first peak of the spectral envelope (P1), as can be seen in Fig. 1a). The frequency of the first formant, for the vowel /a/, is typically between 550 and 950 Hz. A first peak in the spectral envelope may also occur to model the first harmonics but with lower energy than that of the first formant, and with a bandwidth that tends to increase. Due to the vibration of the vocal folds, modelled by two poles at the origin, a spectral tilt occurs that diminishes the energy of the higher frequencies.

However, for the corpora extracted from the database of the Bioengineering Group of the Engineering School of São Paulo University (DBSP) first used in [11], for all the 31 unhealthy subjects, the first two harmonics have a higher energy comparing to the higher-order harmonics, creating a first peak (P1) with an energy higher than the first formant (F1), as can be seen in Fig. 1b). These two harmonics, whose amplitudes are designated in the literature by H1 and H2, are object of several studies [12], [13]. These studies reveal that the difference between these two amplitudes is related with breathy voices, since the second harmonic has less energy than the first harmonic, which is not the case in healthy voices. This analysis also reveals that there is significant information at low frequencies that allows discrimination among pathological voices and healthy voices.

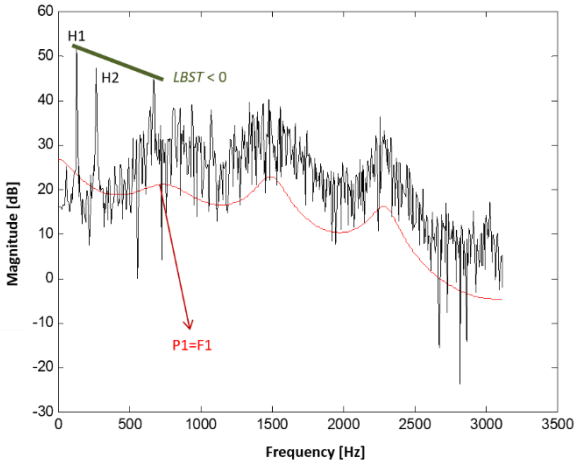
With the evolution of the disease, vocal folds tend to increase aperiodicity and the signal becomes noisier, especially at higher frequencies. The spectral tilt decreases and the energy in the high frequencies increases, which impairs the modelling of the first two harmonics, as show in Fig. 1c). In conclusion, this method tends to fail in more advanced disease states.



a) Healthy subject, P1=F1>550 Hz, Low noise, LBST > 0



b) Unhealthy subjects, P1<550 Hz, Low noise, LBST < 0



c) Unhealthy subject, P1=F1>550 Hz, High noise, LBST < 0

Fig. 1. Spectrum and spectral envelope.

III. DATABASE

The database used in this work is the DBSP database and is composed by 15 healthy subjects (9 males, 6 females), 16 subjects (2 males, 14 females) with Reinke's edema and 15 subjects (2 males, 13 females) with vocal fold nodules, with age distribution ranging from 21 to 48 years hold. These pathologies are both present in the vocal folds. A brief description of these pathologies is presented below. Further details about the corpus and the addressed pathologies can be found in [11].

Each subject produced a sustained vowel /a/ for 5 seconds, recorded with a 22050 Hz sampling rate, with comfortable pitch and loudness level

A. Nodules

The vocal fold nodules are benign lesions that can occur in both vocal folds, at anterior 1/3 and posterior 2/3 of the vocal folds, where friction between folds is most aggressive. This pathology is common in people who use voice intensively, as politicians, teachers, singers and children who often cry. The lumps are small sized, typically the size of a pinhead, and can appear in groups over a given area.

The nodules do not allow complete closure of the vocal folds, resulting on a noisy speech signal. To compensate this effect the patient tends to increase tension in the muscle, increasing the collision forces of the vocal folds. This disease also causes irregularities in the fundamental frequency (jitter) and variations in amplitude (shimmer).

Symptoms include dysphonia, with voice timbre changing through a variable hoarseness. In extreme cases it can reach aphonia preventing altogether the voicing.

B. Reinke's Edema

The Reinkes's edema affects both vocal folds and is characterized by their swelling, caused by accumulation of fluid dispersed in the Reinke space. As the fluid accumulates, this space increases so that the vocal folds also increase in thickness and protrude into the interior of the larynx. In consequence, the voice becomes rougher due to the swelling that causes changes in the elasticity of the vocal folds. In extreme cases the swelling may even hinder the passage of air. In reaction to this situation, the patient increases the vocal effort, resulting in excessive opening of the glottis, causing an asymmetrical and irregular vibration of the vocal folds. The main symptom is the reduction on pitch frequency, making female subjects easier to detect since they typically have higher pitch.

IV. PROPOSED METHOD

When analysing the sustained vowel /a/ spectrum for healthy subjects, Fig 1 a), and the spectrum for unhealthy subjects, Fig. 1 b) and c), it can be seen in all the presence of the first formant around 660 Hz, represented by a peak in the spectral envelope.

For healthy subjects, in the first two harmonics the spectral envelope does not reveal a peak or can exhibit a peak smaller than the formant peak, Fig 1 a). For initial stages of the diseases the peak in the first two harmonics, Fig 1 b) is bigger than the formant peak. In advanced states of the diseases, the peak in the first two harmonics disappears due to the increase of the noise and can be confused with healthy subjects.

However, if the spectrum of the signal is considered instead of the spectral envelope in the case of healthy voices, a spectral positive tilt occurs between the first two harmonics

and the first formant, Fig 1 a). For the unhealthy voices this spectral tilt is always negative, Fig. b) and c), regardless the stage of the disease.

In the proposed method the signal spectrum was analysed in two bands. The first band contains the first two harmonics and the second band is characterized by the interval between the third and the tenth harmonics. The interval of the second band is enough to estimate the first formant in subjects with the lowest fundamental frequency. For example, it guarantees that a subject with first harmonic of 80 Hz, and first formant of 720 Hz, the formant is estimated in the 9th harmonic. For each band on the signal spectrum the maximum spectrum energy of each band is computed.

The spectrum of the signal and the maximum energies of the spectrum bands is computed with an algorithm like the applied in [15], but the noise component is ignored and only the harmonics are analysed. Briefly, the algorithm consists of estimating the fundamental frequency by the autocorrection method, and then the signal spectrum is estimated in a window of 60 fundamental periods. This analysis allows for a long-term spectrum with higher resolution. For each subject 1 second of speech is selected in a stable zone of the signal and its amplitude is normalized. The signal spectrum is estimated with 5 ms steps. The first band maximum energy (FBME) and the second band maximum energy (SBME) corresponding to the first formant are plotted in Fig.2.

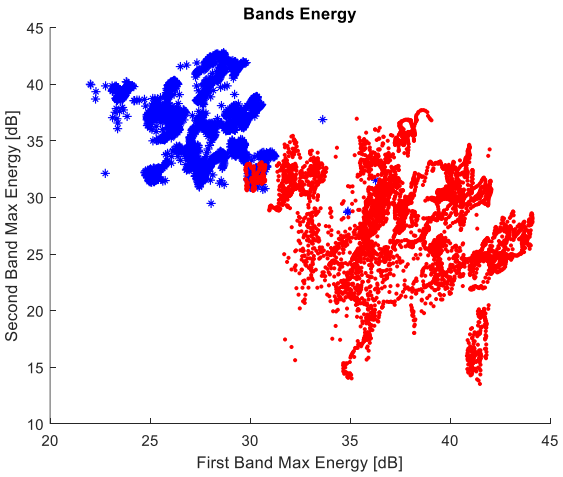


Fig. 2. Max Band Energy. Blue/star-healthy subjects, Red/point-unhealthy subjects.

Healthy subjects have the SBME higher than the FBME. Contrary situation is found for unhealthy subjects. Note that with the difference of these two parameters a considerable discrimination between healthy and unhealthy subjects is achieved, as can be seen in Fig. 3.

The Low Band Spectral Tilt (LBST) is given by the quotient of the difference between the difference of maximum amplitude of the two bands and the difference between the corresponding frequencies, as in Eq. (1).

$$LBST = \frac{SBME - FBME}{SBME_{freq} - FMBE_{freq}} \left[\frac{dB}{Hz}\right] \qquad (1)$$

The introduction of the LBST parameter allows a more complete characterization, as presented in Fig. 4. Detailed analysis of the results and discussion will be presented in the next chapter.

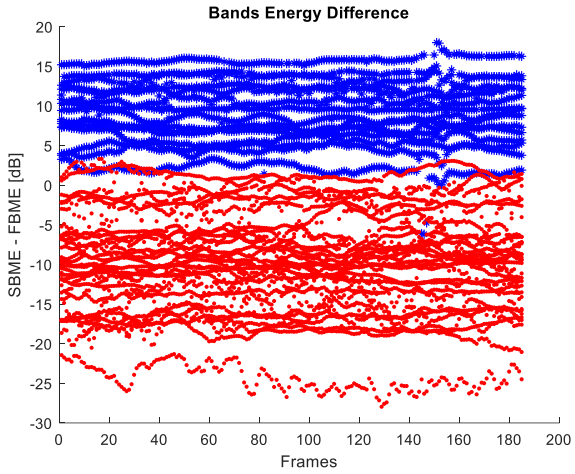


Fig. 3. Diference between maximum energies in the two bands. Blue/star-healthy subjects, Red/point-unhealthy subjects.

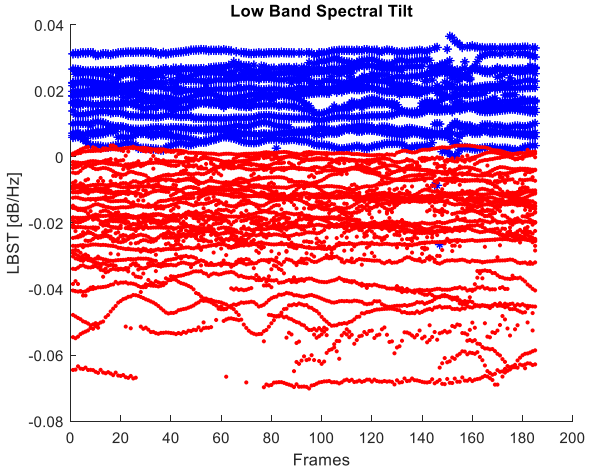


Fig. 4. LBST Blue/star-healthy subjects, Red/point-unhealthy subjects.

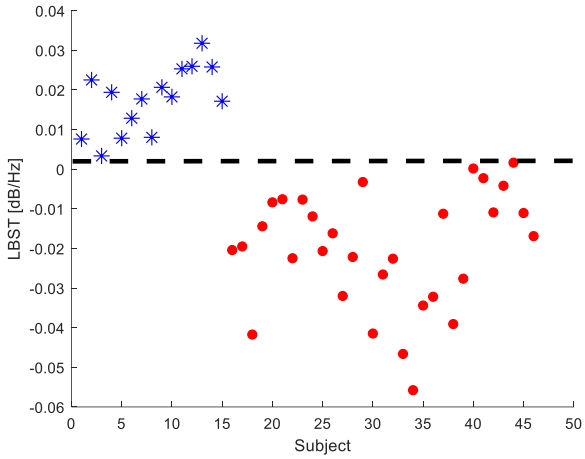
V. Results and Discussion

From the data presented in Fig. 4, the average values are computed for each subject and plotted in Fig. 5 a). As can be seen the data are separable, that is, there is complete discrimination between healthy and unhealthy subjects.

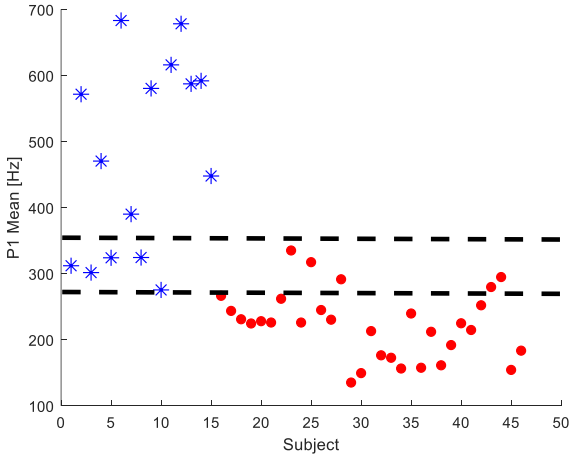
For healthy subjects the LBST mean are always positive. For unhealthy subjects, 29 out of 31 subjects present a negative LBST mean. Although two of the unhealthy subjects have positive LBST mean, these values are lower than the minimum obtained for the healthy subjects, thus ensuring complete discretion between healthy and unhealthy subjects, as the classes are separable.

Comparing the results obtained only with the analysis of P1, Fig.5 b), it is verified that the data are not separable, being the success rate around 93% with a decision value of 300 Hz. In this case the misclassification will be one healthy and two unhealthy subjects. The presented method based on the tilt between bands presents an improvement in the classification rate, having also a simple interpretation that can be used in

visual analysis of the spectrum or as more one parameter in high complexity automatic classifiers.



a) Mean LBST.



b) First Formant

Fig. 5. Parameter comparation. Blue/star-healthy subjects, Red/point-unhealthy subjects.

This method also guarantees the identification of pathological voices for signals with reduced bandwidth. In fact, the maximum necessary bandwidth is just the vowel /a/ first formant frequency, typically below 1000 Hz.

VI. Conclusions

This work presents a new parameter for discrimination of pathological voices based on the low band spectrum analysis. The signal spectrum is divided into two bands. The first band contains the first and second harmonics. The second band contains the first formant. The LBST is computed from the maximum energy value of each band. With this parameter it is possible to discriminate all the subjects of the DBSP database.

In the future LBST will be tested in other databases, as de MEEI database. Judging by the results obtained by a previous analysis, promising results are expected. Combined with other parameters will certainly enhance the discriminating ability of future classifiers.

References

[1] S. Iwata, "Periodicities of pitch perturbations in normal and pathological larynges," *Laryngoscope*, vol. 82, pp. 87–96, 1972.

[2] J. P. Teixeira and A. Gonçalves, "Algorithm for Jitter and Shimmer Measurement in Pathologic Voices," *Procedia Comput. Sci.*, vol. 100, no. March, pp. 271–279, 2016.

[3] K. Shama, A. Krishna, and N. U. Cholayya, "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology," *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007.

[4] Z. Ali, M. S. Hossain, G. Muhammad, and A. K. Sangaiah, "An intelligent healthcare system for detection and classification to discriminate vocal fold disorders," *Futur. Gener. Comput. Syst.*, vol. 85, pp. 19–28, 2018.

[5] R. Fraile, N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and C. Fredouille, "Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex," *Folia Phoniatr. Logop.*, vol. 61, no. 3, pp. 146–152, 2009.

[6] J. D. Arias-Londoño, J. I. Godino-Llorente, M. Markaki, and Y. Stylianou, "On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices.," *Logoped. Phoniatr. Vocol.*, vol. 36, no. 2, pp. 60–69, 2011.

[7] V. Majidnezhad, "A novel hybrid of genetic algorithm and ANN for developing a high efficient method for vocal fold pathology diagnosis," *EURASIP J. Audio, Speech, Music Process.*, vol. 2015, no. 1, p. 3, 2015.

[8] J. W. Lee, H. G. Kang, J. Y. Choi, and Y. I. Son, "An investigation of vocal tract characteristics for acoustic discrimination of pathological voices," *Biomed Res. Int.*, vol. 2013, 2013.

[9] H. T. Cordeiro, J. M. Fonseca, and C. M. Ribeiro, "LPC Spectrum First Peak Analysis for Voice Pathology Detection," *Procedia Technol.*, vol. 9, pp. 1104–1111, 2013.

[10] M. N. Vieira, F. R. McInnes, and M. a Jack, "On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures.," *J. Acoust. Soc. Am.*, vol. 111, no. 2, pp. 1045–1055, 2002.

[11] P. R. Scalassara, M. E. Dajer, C. D. Maciel, R. C. Guido, and J. C. Pereira, "Relative entropy measures applied to healthy and pathological voice characterization," *Appl. Math. Comput.*, vol. 207, no. 1, pp. 95–108, 2009.

[12] R. Wayland and A. Jongman, "Acoustic correlates of breathy and clear vowels: The case of Khmer," *J. Phon.*, vol. 31, no. 2, pp. 181–201, 2003.

[13] B. R. Gerratt, J. Kreiman, and M. Garellek, "Comparing Measures of Voice Quality From Sustained Phonation and Continuous Speech," *Am. J. Speech-Language Pathol.*, vol. 59(5), no. October, pp. 994–1001, 2016.

[14] H. Cordeiro, J. Fonseca, and C. Meneses, "Spectral envelope and periodic component in classification trees for pathological voice diagnostic," *Conf. Proc. .. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, 2014.

[15] A. Stráník and R. Čmejla, "an Analysis of Iterative Algorithm for Estimation of Harmonics-To-Noise Ratio in Speech," pp. 0–6.