Acoustic Description of Hearing-Impaired Children's Voice Quality

Abderrazak Rougab
Departement of Electronic,
Amar Thelidji University of Laghouat
UATL Laghouat Algeria

Mhania Guerti
Departement of Electronic Polytechnic
School of Algeria ENPA BP 182
El-harach 16200 Algeria

Abstract---**The voice quality is generally determined both by the physiology of the speaker and by their social environment . Concerning the hearing-impaired speakers, their voice quality can also be influenced by the type of 'handicap' or the type of cochlear implants they can have. Although there are lots of sources of variation, the hearing-impaired voice quality is still identifiable auditorily. Our study is one of the first to analyze the voice quality of Arabic hearing-impaired children. In order to measure the voice quality variations, we needed to find a tool that would analyze the signal on a long-term basis [1], thus the Long Term Average Spectrum was selected.**
In order to proceed at the analysis, a script was written with the Matlab software [2]. Two methods were used in order to compare the LTAS. The Principal Component Analysis allowed us to check whether the deaf speakers would form a separate group from the control speakers. The results showed that the hearing-impaired speakers formed a well separated group from the normally hearing speakers according to both axes.

Keywords—Hearing-Impaired; Speech Spectrum; LTAS; PCA

I. INTRODUCTION

Most of the time, the voice of a deaf speaker is easily identifiable by the human ear. However, the physical and acoustic characterization of the deaf speaker's voice quality is quite difficult to establish. The main acoustic characteristics of the hearing-impaired (HI) speech are generally described as a longer utterance duration, a higher F0 peak and more important changes in the FO contours than the normally hearing (NH) speakers [3], [4]. Concerning the spectral domain, less information is known about the frequency distribution of the energy. [5] Showed that for some of their HI subjects, they obtained spectra that were different in their harmonic structure and/or the spectral slopes declined to a greater rate the NH adolescents. However, they didn't find any characteristics that would allow identifying the HI group.

II. SCIENTIFIC BACKGROUND

A. General background

Considering the lack of information and technique of analysis in the pathological domain, our study was inspired mostly by non-pathological studies. Among those studies, [6] used the LTAS to measure the variation of voice quality between Castilian and Catalan dialects. This study was successful in showing that the LTAS was a good tool to measure voice quality variation. However, the LTAS can be influenced by many different variables. For instance, some intra-speaker variation was observed in [7]. Considering those problems, we were particularly attentive on the method to calculate the LTAS.

B. Evolution of the methods

The spectral analysis aims at developing our knowledge of the signal according to the frequency domain. One of the major tools is the measurement of the power spectral density (PSD). It is defined as the Fourier transform of the self-correlation function of the signal, shown in the following equations.

$$S(f) = \lim_{N \to \infty} E\left[\frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j2\pi f n} \right|^2 \right] \quad (1)$$

The mathematical expectation (*E*) means that an average of several realizations is calculated.

The method of Schuster ([8]) is the first to propose the periodogram, which is an estimation of PSD. Then, a modified version of Schuster's periodogram integrates a weighting window. The first version of an average of spectrums was proposed by [9]. He started to divide the signal in several segments and then used Schuster's method for each segment. Finally the last method, which is going to be used here, is the method of Welch ([10]). He proposed to modify Bartlett's version of the periodogram introducing an overlapping of the segments and a weighting window. The following figure represents a simplified version of the

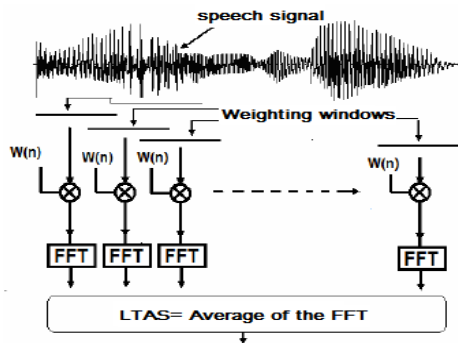process that is used to calculate the LTAS [11] according to the method of Welch.

Figure 1: Description of the calculation of the LTAS.

III.   METHOD

A. Corpus

8 hearing impaired and 10 control speakers. Reading three times the same text LTAS was calculated on the three texts. Try to eliminate the variation of reading between the texts.

Recording with Praat software, the microphone was placed at the same distance from the mouth thanks to the system attached to headphones.

The specialized school for hearing-impaired children and the school of the control group are both in Laghouat city, meaning that the deaf speakers are from the Laghouat region.

The sampling rate is of 22 KHz. Most studies used a sampling rate of 16 KHz (minimal limit).Easier now to record at a higher rate.

B.   Analysis Procedure

B.1. Calculation tool

As it was mentioned in section 3, the method of Welch with an overlapping of 50% was selected for this study. The algorithm used to calculate the LTAS are shown the following equation (2).

$$\hat{S}_{Welch}(f) = \frac{1}{KLU} \sum_{i=0}^{k-1} \left| \sum_{n=0}^{L-1} \omega(n)x(n+iD)e^{-j2\pi fn} \right| \quad (2)$$

$$U = \frac{1}{N} \sum_{n=0}^{N-1} |\omega(n)|^2$$

The length of the window is of 512 points of measurement, which means that its duration is of 32 milliseconds.

Once the method for the calculation of the LTAS was chosen, a program was developed with Matlab. Before the analysis of the LTAS is executed, the programme allows the user to choose between different types of windows and also different durations. Then, the programme asks for the number of files to be treated.

Once the LTAS has been calculated, the programme provides a list of variables that can be used in order to sum up the different parameters or for the representation of the results.

B.2 Preliminary studies

Before analysing our corpus, two preliminary studies were done, one in order to test the type of weighting window and the other to test different durations.

As it was mentioned in section 3.2.1, several types of weighting windows can be selected in the script. The first preliminary study was, then, to observe the effect of the type of windows on the LTAS. A comparison was made between twelve different windows for one speaker.
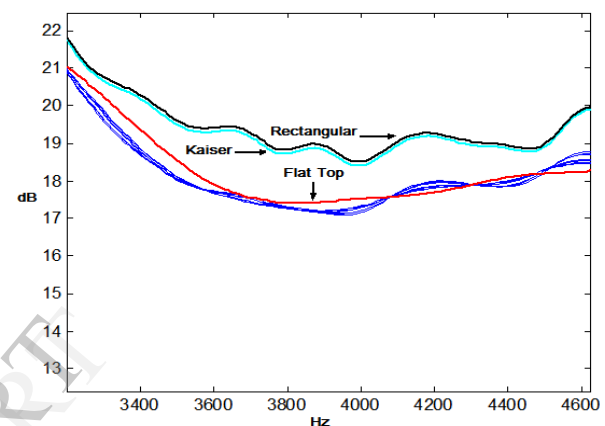

Figure 2: Comparison between the sixteen types of windows

The results showed that three windows were particularly different from the others, the flat top, the Kaiser and the rectangular ones. Thus we decided not to use them for our study. Among the others, we decided to use the Hamming window because most of the studies on the subject used this one.

The other preliminary study consisted in comparing three different durations of windows. In order to obtain the durations in milliseconds, the number of samples has to be divided by the sampling frequency. For example, if the number of samples if 256 and the sampling frequency is of 22050 Hz, the duration of the window is of 11 ms.
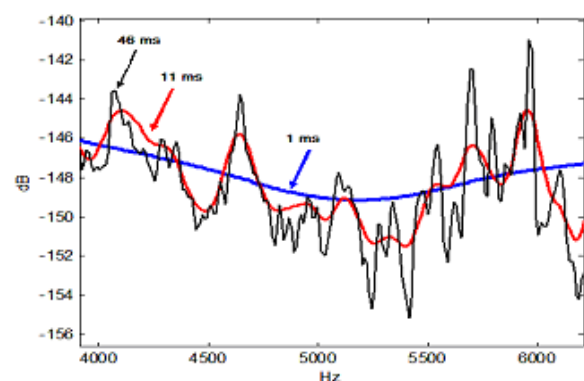

Figure 3: Comparison of three durations of window

For our study, we finally, chose a duration of 46 ms (1024 samples) in order to have the most precise spectrum and to remain within a stationary signal.

Once the type and duration of the weighting window were determined, we were able to proceed at the calculation of the LTAS.

IV. RESULTS AND DISCUSSION

A. Spectral Analysis

The following figures represent the results of the LTAS analysis according to the gender of the speakers.
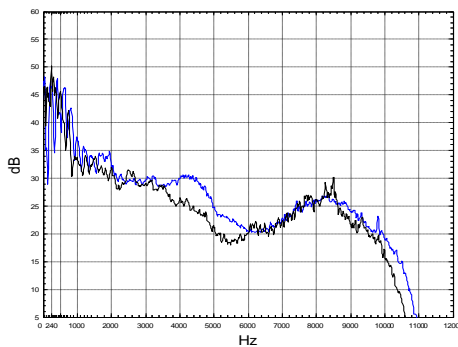

Figure 4: Comparison between the average spectra of the HI speakers and the NH speakers (male speakers).

As far as the male speakers are concerned, figure 4 shows that there is an important separation between the two spectrums within the frequency range of 3500 and 6000 Hz. Concerning the results for the female speakers (fig.5), the separation of the two spectrums is located around 5000 Hz onwards.
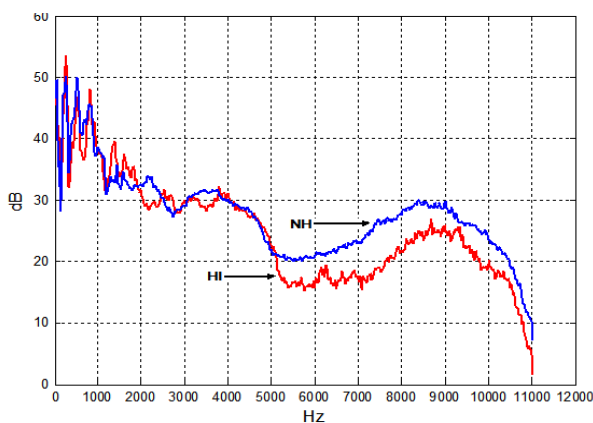

Figure 5: Comparison between the average spectra of the HI speakers and the NH speakers (female speakers).

Finally, the comparison between the average of the HI and NH spectrums (fig.6) shows that the

difference between the two is visual small though constant from 3000 Hz onwards.
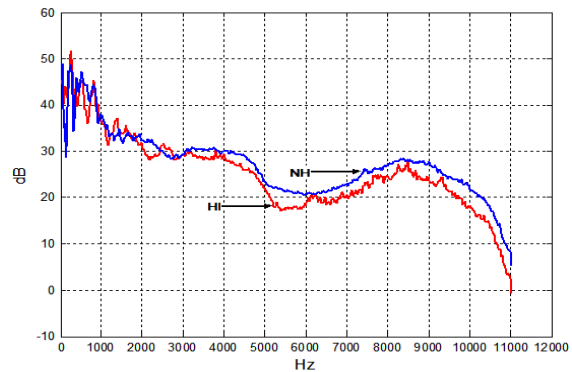

Figure 6: Comparison between the average spectra of the HI speakers and the NH speakers.

The visual observation of the results is necessary but not sufficient to determine whether the LTAS allows the differentiation between the HI and the NH speakers. In order to do so, the multidimensional scaling analysis seems an appropriate tool.

B. Principal Component Analysis

The Principal Component Analysis (PCA) is going to be used in this study in order to compare the results according to a different type of representation. The PCA is a mathematical technique that reduces the number of dimensions of a complex system of correlation. It represents the majority of the total variance of the data. The first principal component corresponds to the maximum variance that can be obtained out of the analysis, the second one being the maximum of the remaining variance etc. Figure n°6 shows that the percentage of representation of the variance is around 70% for the first two components, meaning that a great majority of the variance between the LTAS are represented with those two principal components.
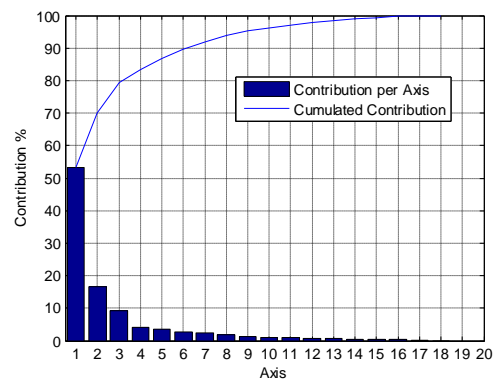

Figure 6: Percentage of absorbed variance for the twenty main axes.

The following figure is the representation of the results of the PCA analysis for the eighteen speakers of the corpus.
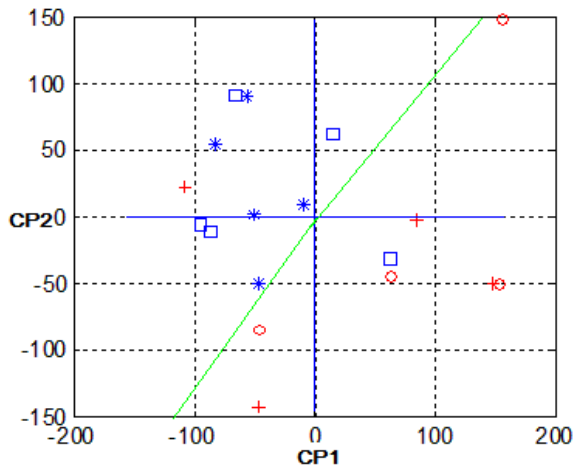
Figure 7: Representation of the results of the PCA according to the first two components

V. CONCLUSION

To conclude, this is, to our knowledge, the first study that provides a spectral description of Arabic hearing-impaired speakers.

The visual observation of the spectra showed that the energy distribution was less important for the HI speakers than for the NH speakers. Then, the comparison of the spectra thanks to the PCA analysis confirmed the existence of differences between the two groups. Indeed, we were able to differentiate the HI speakers from the NH speakers.

Our perspectives of research are to test this method on a larger corpus that would include HI speakers of different ages and from other regions of Algeria

VI. REFERENCES

[1] Pittam, J. The long term spectral measurement of voice quality as a social and personality marker: a review. In Tadjen Kris. Jan-March, 1-12. 1987.

[2] Rougab, A., Coadou, M. Long Term Average Spectrum: A Tool to Measure the Voice Quality of Deaf Speakers. SPiAP07 Speech Prosody in Atypical Groups. University of Reading, UK Monday 2nd April 2007.

[3] Laver, J. The Phonetic Description of voice Quality. Cambridge University Press: Cambridge 1980.

[4] Clement, C.J., Koopmans-van Beinum, F.J. and Pols. L.C.W. Acoustical characteristics of sound production of deaf and normally hearing infants. In. H.T Bunnel and W. Idsardi (Eds). Proceedings ICSLP96, Fourth Philadelphia Vol3,1549-1552 1996.

[5] Formbly,C. And Monsen, Randall B 1982 Long-term average speech spectra for normal and hearing-impaired adolescents. In: Journal of the Acoustical Society of America 71:1- pp. 196-

[6] Bruyninckx, M., Harmegnies, B., Llisterri,J. and Poch, D.. Language induced voice quality variability in bilinguals. Journal of Phonetics, 22:19-31. 1994

[7] Harmegnies, B. and Landercy. .Intra-speaker variability of the long term speech Spectrum. Speech Communication 7, 81-86. 1988

[8] Schuster A. On the investigation of hidden periodicities with application to a supposed 26 day period of metrological phenomena, Terrestrial Magnetism (now Journal for Geophysical Research), 3, 13-41 1898

[9] Bartlett,M. S" Smoothing periodograms from time series with continuous spectra. Nature, 161 686-687 1948

[10] Welch, P.D. The use of fast Fourier Transform for the estimation of power spectra IEEE Trans Audio Electroacoustics 15: 70-73 1967

[11] Marion Coadau. Abderrazak Rougab. Voice Quality and Variation in English, ICPHS XVI Saarbrücken., 6-10 August2007