

PUBLISHED BY

# INTECH

open science | open minds

World's largest Science,  
Technology & Medicine  
Open Access book publisher



**3,150+**  
OPEN ACCESS BOOKS



**104,000+**  
INTERNATIONAL  
AUTHORS AND EDITORS



**108+ MILLION**  
DOWNLOADS



**BOOKS**  
DELIVERED TO  
151 COUNTRIES

AUTHORS AMONG

**TOP 1%**  
MOST CITED SCIENTIST



**12.2%**  
AUTHORS AND EDITORS  
FROM TOP 500 UNIVERSITIES



Selection of our books indexed in the  
Book Citation Index in Web of Science™  
Core Collection (BKCI)

**WEB OF SCIENCE™**

Chapter from the book *Advances in Speech-language Pathology*

Downloaded from: <http://www.intechopen.com/books/advances-in-speech-language-pathology>

Interested in publishing with InTechOpen?  
Contact us at [book.department@intechopen.com](mailto:book.department@intechopen.com)

---

# The Advanced Voice Function Assessment Databases (AVFAD): Tools for Voice Clinicians and Speech Research

---

Luis M.T. Jesus, Inês Belo, Jessica Machado and  
Andreia Hall

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.69643>

---

## Abstract

A new open access resource called Advanced Voice Function Assessment Databases (AVFAD) was developed, based on a sample of 709 individuals (346 clinically diagnosed with vocal pathology and 363 with no vocal alterations) recruited in Portugal. All clinical conditions were registered according to the Classification Manual of Voice Disorders-I. Participants were audio-recorded, producing the following vocal tasks: Sustaining vowels /a, i, u/; reading of six CAPE-V sentences; reading a phonetically balanced text; spontaneous speech. The AVFAD are comprised of 8648 uncompressed audio files and an additional database file with 19 Praat Voice Report parameter values and 16 clinical data entries per participant. An annotated segment of the vowel /a/ for each participant was analysed automatically with a Praat script. Radial graphs were generated considering that all variables had an approximately normal distribution, and using previously calculated average and standard deviation values for all parameters. The normal and pathological  $f_0$  mean, Jitter ppq5, Shimmer apq11 and Harmonics-to-Noise-Ratio characteristics were compared. An additional analysis of the relation between the acoustic parameters and gender, age group, smoking habits, body mass index and voice usage, was considered. The AVFAD will allow future cooperative work and testing of non-invasive methods for voice pathology diagnosis.

**Keywords:** voice, voice disorders, database, assessment, multi-dimensional acoustic voice analysis, Praat, classification manual of voice disorders-I, Portuguese

---

## 1. Introduction

The multidimensionality of voice requires the use of several types of evaluation and measures to allow the correct characterization of vocal quality [2]. The instrumental evaluation of voice

---

[29] is considered as one of the most important elements for a correct vocal diagnosis and must precede intervention. It should include perceptive, acoustic, physiological, aerodynamic evaluation, and an auto-assessment of vocal quality.

Acoustic voice analysis [20, 21] is an effective and noninvasive tool that can be used to confirm an initial diagnosis and provide an objective determination of the impairment [38]. It is also an important tool to an early detection and treatment of laryngeal tumors that can reduce both morbidity and mortality.

The collection of voice databases for testing and comparing the analysis methods is regarded as an important research area. However, despite the variety of models and methods developed by signal processing engineers, voice clinicians still express their disappointment with regard to the performance of the existing approaches for assessing voice quality.

Reference acoustic databases allow the standardization of acoustic analysis, benchmarking and comparing the performance of different voice analysis techniques. They also allow to differentiate normal voice and pathological voice, to evaluate and monitor it clinically, and to diminish the subjectivity that underlies the acoustic-perceptual analysis [22] by establishing a correlation between quantitative data. Results can be interpreted reliably, as long as they are collected by the same equipment, and the same data collection methods and recording techniques are used [43]. However, the reliability of acoustic analysis of the voice signal is still hindered by the “scarcity of sufficiently comprehensive databases” [38, p. 4].

A database of normal and pathological voices is a reference for the identification of clinically relevant perturbations in voice quality and data collection and analysis suitability developed for specific applications [26, p. 131].

The most widely used clinical graphical and numerical representation of normal and pathological voices [47] is the Multidimensional Voice Program (MDVP), and even when the acoustic analysis of voice is performed with freeware [30], reference values from the MDVP can be found in the manuals. However, these values should be used with great caution because they are based only in 15 normal voices [47, p. 227] and “may not be appropriate for various age-sex subpopulations. At this time, the MDVP normative values should be regarded as preliminary and not as commonly recognized criteria by which abnormality is established. However, the concept of an integral database is important” [26, p. 135].

There have been various scientific studies along the past 40 years that compare, acoustically, normal and disordered voices [8, 10, 16, 27, 31, 37]. For Portuguese, there have been some voice research on vocal quality [3, 6, 33, 34, 41, 42, 49], distinction between pathological voice and normal voice through acoustic analysis [5, 12, 13, 21, 36] and the prevalence of laryngeal disorders [7, 39]. However, there are no known open access databases that allow the comparison of voice studies.

The University of Aveiro in Portugal collected, annotated and analyzed the Advanced Voice Function Assessment Databases (AVFAD), an open access resource that facilitates vocal evaluation, representing the first normative database for EP. Databases collected by clinicians enable the interpretation of automatically extracted descriptors of the speech signal and lead to the development of models for the interaction of these descriptors.

One of the purposes of this book chapter is to compare, acoustically, participants with normal voice to participants with voice pathology, regarding the parameters fundamental frequency ( $f_0$ ) mean (Hz), Jitter ppq5 (%), Shimmer apq11 (%) and Harmonics-to-Noise Ratio (HNR) in dB. An additional analysis of the relation between these two groups and participants' demographic data was considered, including gender, age group, smoking habits, body mass index (BMI), and voice usage. Generally, the main goal was to study EP speakers' acoustical characteristics and verify if it is possible to differentiate voice disorders through an acoustical analysis of voice.

The normative voice data presented in this book chapter is important to typify voice pathologies and when evaluating treatment success. For instance, it has long been known [32] that the voices of speakers with organic disorders of the larynx have higher Jitter and Shimmer and a lower HNR relative to the voices of normal speakers [8].

The AVFAD are distributed freely using a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, through the Advanced Communication and Swallowing Assessment (ACSA) platform at <http://acsa.web.ua.pt/>.

## 2. Method

The work reported in this book chapter is part of a larger ongoing project of the University of Aveiro in Portugal, which aims to build and validate a comprehensive set of resources for voice clinicians, including a standardized voice case history form [11, 24], a voice evaluation protocol [1, 25] and a reference voice database (AVFAD).

The sample used in this study includes 709 individuals, 346 of whom clinically diagnosed with vocal pathology and 363 with no vocal alterations, matched for gender and chronological age. Healthy controls were recruited at hospitals, from the University of Aveiro (UA) staff and students, and institutions with UA protocols. The recruitment process took place in the otorhinolaryngology departments of three hospitals that have a long-standing cooperation with the UA. Local clinicians discussed with the research team their medical diagnosis, and sociodemographic and anthropometric information were collected.

All clinical conditions were classified with the wording and numeric coding system proposed by Verdolini et al. [46]. The Classification Manual of Voice Disorders-I (CMVD-I) "lists most conditions that may negatively affect a patient's ability to produce voice, based on current understanding" [46, p. 2]. CMVD-I's Dimension 1 uses nine categories to classify these conditions [46, p. 4]. The 346 participants clinically diagnosed with vocal pathology are distributed as uniformly as possible distributed through these categories. Participants were recruited through a convenience sampling method, fulfilling a set of predefined inclusion criteria: aged 18 or older; Portuguese nationality; and EP as mother tongue.

Verdolini et al.'s [46] classification was derived from notes on diagnosis collected from the local hospitals voice clinicians, which were carefully analyzed by two independent speech and language therapists (SLTs), which reached a consensus after clarifying some participant's diagnosis with original clinical team.

Chronological age and gender matching participants with vocal pathology and healthy participants was implemented in 5-year clusters [15], that is, each participant with vocal pathology with a certain age and gender was matched to a control with the same gender and age within a 5-year range. For example, participants with vocal pathology aged 18–22 years were matched to controls within the same age range.

Informed consent was collected from all participants prior to any data collection, authorizing the use of recordings for the present study and also for other studies and by other researchers in the area of voice. The following participants' clinical and demographic data were then registered: smoking habits, age group, BMI, gender, and voice usage.

All participants were sitting in a comfortable chair, so they were as static as possible during the recording. A microphone was held on a tripod placed at a distance of 30 cm [44] from the participant's mouth (on-axis to the lips). Acoustic signals resulting from different voice assessment tasks were recorded via a Behringer ECM 8000 omnidirectional electret microphone connected to an audio interface (Presonus AudioBox USB; AudioBox Driver Version 1.57.0.5385; 16 bits and 48,000 Hz sampling frequency), using Praat version 5.3.56 [4].

Participants were recorded, producing the following vocal tasks: production of sustained vowels /a, i, u/—three repetitions each; reading of the six Portuguese CAPE-V sentences [19]—three repetitions each; reading a phonetically balanced text [23] and spontaneous speech.

Raw recordings were segmented into eleven.wav files (one for each speech sample), using Audacity 2.0.5 and Praat 5.4.04. The /a/ vowel repetition considered to be closer to the speaker's natural voice and produced with a comfortable pitch and volume was selected for analysis, and an interval corresponding to one hundred consecutive cycles, two hundred milliseconds after phonation began, was annotated and analyzed automatically (through a script written specifically for this purpose) with Praat version 5.4.08.

Firstly, for all files and for a 75-ms section of Praat's editor window, the incorrect identification of the periods by the program (e.g., situations in which period-doubling or period-halving occurred) was monitored for each participant. Those whose samples did not allow the correct identification of the periods or did not present a segment in which it was possible to identify, in the same sequence, 100 cycles of oscillation of the vocal folds were dropped out of the final version of the database. It was possible to identify all the participants that had to be dropped out based on these criteria, since the following parameters were also extracted automatically (also available within Praat's Voice Report): number of pulses; number of periods; mean period; standard deviation of period; fraction of locally unvoiced frames; number of voice breaks; and degree of voice breaks. These parameters were not included in the final database.

The parameterization of the Praat scripting language function used to extract the data were `voiceReport $ = Voice report... analysisStart analysisEnd 75 500 1.3 1.6 0.03 0.45`, where each of the parameters evoked had the following correspondence with the designations used in the system Of Praat menus (View & Edit → Pitch → Pitch settings... and Advanced Pitch settings...): time range (s): analysisStart-analysisEnd (beginning and end of the / a / segment that was noted previously); pitch range (Hz): 75–500; maximum period factor: 1.3; maximum amplitude factor: 1.6; silence threshold: 0.03; and voicing threshold: 0.45.

The following Praat Voice Report function default parameters were extracted with the script and stored in the databases: f0 median (Hz); f0 mean (Hz); f0 std (Hz) ; f0 min (Hz); f0 max (Hz); Jitter local (%); Jitter local\_abs (s); Jitter rap (%); Jitter ppq5 (%); Jitter ddp (%); Shimmer local (%); Shimmer local dB (dB); Shimmer apq3 (%); Shimmer apq5 (%); Shimmer apq11 (%); Shimmer dda (%); Autocorrelation mean; NHR mean; and HNR mean (dB).

The normal and pathological acoustic characteristics were compared using IBM SPSS Statistics 22, in order to explore differences between the parameters through the Mann-Whitney U test. In this book chapter, the following parameters are analyzed: f0 mean, Jitter ppq5, Shimmer apq11, and HNR.

An additional analysis of the relation between the acoustic parameters and gender, age group, smoking habits, BMI and voice usage was considered using the Kruskal-Wallis test. The participants ages were grouped for the purpose of this analysis and with the additional objective of analyzing voice changes across the life span [17], according to the following classification [45, p. 3]: young adulthood (18–45 years of age); middle adulthood (46–65 years of age); and older adulthood (older than 65). The BMI values were grouped into three categories according to WHOS's [48, p. 9] criteria: underweight (less than 18.5); normal range (18.5–24.99); overweight (25.00–29.99); and obese (greater or equal than 30.00).

Radial graphs were generated in Excel 2013, considering that all variables had an approximately normal distribution, and using previously calculated average and standard deviation values for all parameters. After the standardization of each variable, a grey circular area was drawn for each gender, corresponding, in each direction, to the average range of two standard deviations (that is, about 95% of normal distribution) of the healthy population. Applying that same standardization to each individual, a polygon in the radial graph was drawn, which allows the visualization of variables that are out of the expected range. The goal of radial graphs “is not only to determine if changes occur in the magnitude of certain parameters, but also to determine if there are configurational adjustments in a multi-dimensional profile” [26, p. 131] of voice.

Ethical approval was obtained from all authorities required by Portuguese bylaws for clinical research: national data protection committee; independent ethics committees.

### 3. Results

Data were collected during more than 150 sessions, over a period of three years (2012–2015). The AVFAD are comprised of 8648 data files (709 participants × (11.wav files + 1 annotated Praat binary file) + 140 background noise.wav files) and an additional Excel 2013 database file with 19 Praat Voice Report parameter values and 14 clinical data entries per patient, including: File ID; Visit date; Visit place; Age; Sex; Weight; Height; Surgery (Without laryngeal surgery; With laryngeal surgery); SLT Intervention (Without intervention; Under intervention; Postintervention ); Smoking (Nonsmoker; Former smoker; Smoker); Singing (Nonsinger; Regular use of singing voice); Job; Diagnosis (CMVD-I Dimension 1 numeric system); Diagnosis (CMVD-I Dimension 1 word system) and Notes.

The AVFAD include 709 participants, from 18 to 93 years old, of whom 346 (49%) had a medical diagnosis of vocal pathology and 363 (51%) did not present any vocal pathology; 499 (70%) were females, and 210 (30%) were males, which are typical male/female ratios in Portuguese hospitals where the present study was conducted. Within the group diagnosed with vocal pathology, there are 26 different diagnoses based on Verdolini et al. [46] classification, including 249 (72%) female and 97 (28%) male participants. The control group was composed of 250 (69%) females and 113 (31%) males.

The acoustic parameters  $f_0$ , Jitter ppq5, Shimmer aqp11, and HNR were compared between the participants without vocal pathology and the group of participants with a diagnosis of vocal pathology. The analysis considered gender, since it is generally accepted that the difference in anatomic structures affect the parameter  $f_0$ . **Table 1** shows the results for males.

The results show that for all of the assessed parameters, there were statistically significant differences ( $p < \alpha$ ) between the two groups. Normal participants presented lower  $f_0$ , Jitter ppq5 and Shimmer aqp11 values, and higher value HNR values, as expected.

**Table 2** shows the results for females.

Results showed that there are statistically significant differences ( $p < \alpha$ ) between the groups in three parameters: Jitter ppq5, Shimmer aqp11, and HNR. The fundamental frequency was unaffected by pathology in females.

Generally, the results showed that in both genders, there was a difference between normal participants' voices and pathological voices in most parameters that should be further considered and analyzed. For that purpose, multiple comparisons between normal participants and each of the six groups of pathology (nodules; polyp(s); cyst; Reinke's Edema; Reflux; Unilateral Vocal Fold Paralysis—UVFP) with the largest dimension were performed. The Bonferroni correction to control the chance of overall false-positive results leads to  $\alpha = 0.05/6 = 0.0083$ .

**Table 3** presents the results of the comparison between normal and each type of pathological voices for male participants. Note that only the Reflux group has  $n \geq 20$ .

The results show that the differences presented before, when pathological voices were analyzed as a single group, are just noticeable for two diagnosis, in the same two parameters. Males with Reinke's Edema or Reflux showed a statistically significant decrease in Shimmer aqp11 and HNR, when compared to normal voices.

	Normal	Vocal pathology	U test	<i>p</i> -value
$f_0$ mean (Hz)	120.68 ± 22.30	138.28 ± 40.09	4082.0	0.001*
Jitter ppq5 (%)	0.247 ± 0.190	0.354 ± 0.289	4132.0	0.002*
Shimmer aqp11 (%)	4.403 ± 2.652	8.297 ± 3.392	2638.5	<0.001*
HNR (dB)	16.315 ± 3.267	13.168 ± 4.105	2854.5	<0.001*

Nonparametric Mann-Whitney U test; \*statistical significant differences for  $\alpha = 0.05$ .

**Table 1.** Descriptive (Mean ± Std. dev.) and inferential statistics for the male gender.

	Normal	Vocal pathology	U test	p-value
f0 mean (Hz)	193.45 ± 28.47	198.80 ± 42.73	29883.5	0.441
Jitter ppq5 (%)	0.214 ± 0.126	0.447 ± 0.484	15658.5	<0.001*
Shimmer apq11 (%)	5.174 ± 2.696	9.816 ± 4.884	9792.5	<0.001*
HNR (dB)	17.335 ± 3.958	11.774 ± 3.422	8876.5	<0.001*

Nonparametric Mann-Whitney U test; \*statistical significant differences for  $\alpha = 0.05$ .

**Table 2.** Descriptive (Mean ± Std. dev.) and inferential statistics for the female gender.

**Table 4** presents the results of the comparison between normal and each type of pathological voices for female participants. Note that all groups have  $n \geq 20$ .

The results of the multiple comparisons show several statistically significant differences between normal and all the pathological groups. The parameters Jitter ppq5, Shimmer aqp11, and HNR are affected by pathology, that is, all the pathology groups presented statistically significant differences from the normal group. As far as the parameter f0 is concerned, results are not consistent across pathologies. The participants diagnosed with Nodules and Reinke’s Edema presented statistically significant differences in comparison with the normal group, when all the other groups did not. In other words, based in this sample of female voices, Polyp(s), Cyst, Reflux, and UVFP seem to cause alterations in Jitter ppq5, Shimmer apq11, and HNR but not in f0, and Nodules and Reinke’s Edema cause alterations in all the parameters.

**Figures 1–4** present additional information and provide a visual representation that allows the comparison of the previous acoustic parameters between the normal participants, and the participants diagnosed with the six most prevalent pathologies. Note that the boxes in **Figures 1–4** represent the 25–75th percentile range, black lines in the boxes represent medians, and whiskers correspond to the furthest observation within the  $\pm 1.5$  interquartile range; outliers are represented as circles and extremes ( $>3$  interquartile range from the box) as asterisks.

Based on these results, it is possible to conclude that males with a diagnosis of Reinke’s Edema or Reflux presented a higher Shimmer aqp11 and lower HNR than the males without any vocal pathology. For females, there was no consistency in the behavior of f0, because it was only affected by the diagnosis of Nodules or Reinke’s Edema. All the other parameters (Jitter ppq5, Shimmer apq11 and HNR) showed a consistent behavior: the group without vocal pathology showed, in comparison with all the other groups, lower values of Jitter ppq5 and Shimmer aqp11 and higher values of HNR.

The influence of age, BMI, smoking habits, and voice usage of participants on the acoustic characteristics of voice was also investigated. **Table 5** shows the results for age, grouped by gender and diagnosis.

The results showed some significant differences between the groups (young adulthood; middle adulthood and older adulthood) for the variable age. These differences are noticed in f0, independently of the gender or diagnosis, also in Shimmer apq11 in all participants except males with normal voice and in HNR in the group of females with normal voice.



	Nodules		Polyp(s)		Cyst		Reinke's Edema		Reflux		UVFP	
	n = 2	n = 7	n = 3	n = 8	n = 29	n = 2						
	U	p-value	U	p-value	U	p-value	U	p-value	U	p-value	U	p-value
f0 (Hz)	30.0	0.076	352.0	0.626	153.0	0.774	301.0	0.115	1375.0	0.182	19.0	0.044
Jitter ppq5 (%)	90.0	0.623	272.0	0.167	46.0	0.032	310.5	0.140	1176.0	0.019	16.0	0.038
Shimmer apq11 (%)	73.0	0.392	202.0	0.30	89.0	0.161	198.0	0.008*	606.0	<0.001*	6.0	0.022
HNR (dB)	106.0	0.881	238.0	0.078	37.0	0.021	192.0	0.007*	812.0	<0.001*	11.0	0.013

Nonparametric Mann-Whitney U test; \*statistical significant differences for  $\alpha = 0.0083$  (Bonferroni correction).

Table 3. Multiple comparisons between normal and pathological voices for the male gender.

	Nodules		Polyp(s)		Cyst		Reinke's Edema		Reflux		UVFP	
	n = 26	n = 20	n = 22	n = 71	n = 24							
	U	p-value	U	p-value	U	p-value	U	p-value	U	p-value	U	p-value
f0 (Hz)	2211.0	0.007*	2297.0	0.546	1948.0	0.100	1121.5	<0.001*	8117.0	0.272	2134.0	0.020
Jitter ppq5 (%)	1744.0	<0.001*	984.0	<0.001*	1165.5	<0.001*	1158.5	<0.001*	5506.0	<0.001*	1016.0	<0.001*
Shimmer apq11 (%)	1197.0	<0.001*	685.0	<0.001*	778.5	<0.001*	1270.0	<0.001*	2716.0	<0.001*	739.5	<0.001*
HNR (dB)	1264.0	<0.001*	547.0	<0.001*	688.0	<0.001*	836.0	<0.001*	2366.0	<0.001*	1056.0	<0.001*

Nonparametric Mann-Whitney U test; \*statistical significant differences for  $\alpha = 0.0083$  (Bonferroni correction).

Table 4. Multiple comparisons between normal and pathological voices for the female gender.

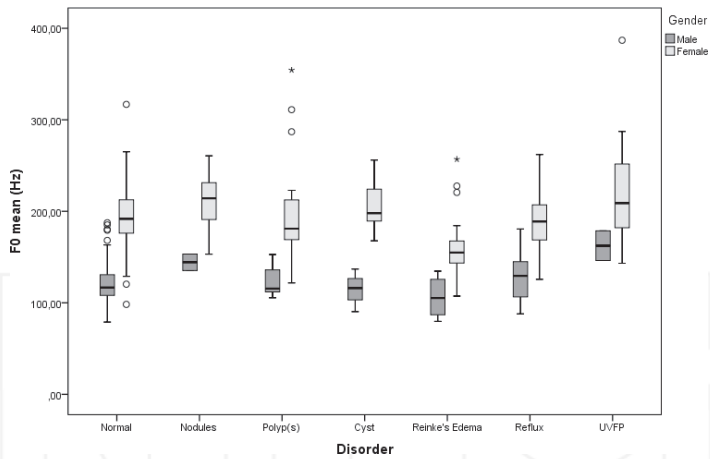


Figure 1. Fundamental frequency by voice disorder for both genders.

Table 6 shows the influence of smoking habits on the acoustic parameters, by gender and diagnosis.

The results show significant differences between groups (nonsmoker; former smoker; smoker) in f0 for females and males with normal voice and Shimmer apq11 for males with normal voice.

Table 7 shows the influence of BMI on the acoustic parameters by gender and diagnosis.

The results show that BMI only had an influence on females with normal voice. All the others did not show statistical differences between groups (underweight; normal range; overweight; obese).

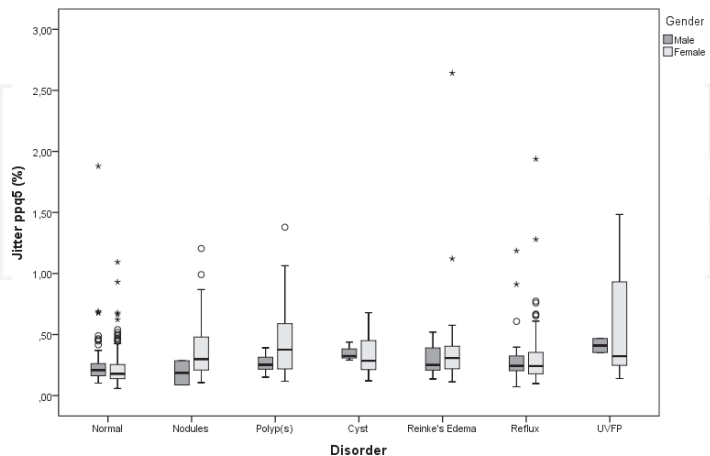


Figure 2. Jitter ppq5 by voice disorder for both genders.

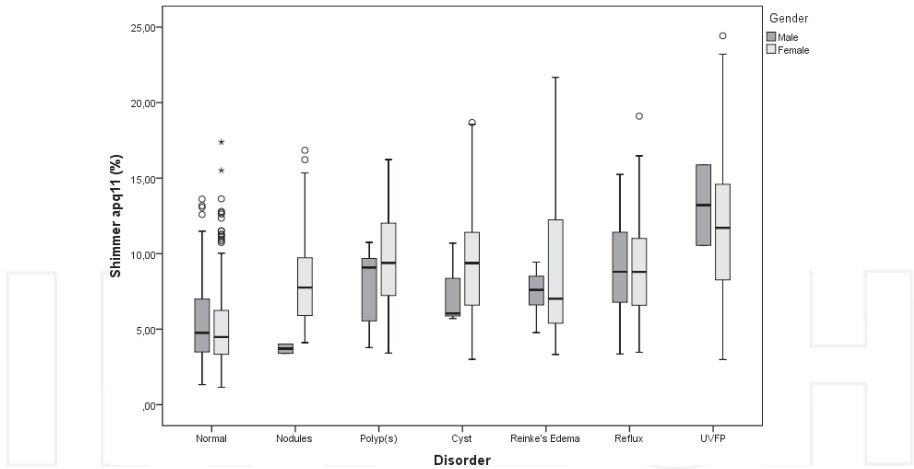


Figure 3. Shimmer apq11 by voice disorder for both genders.

Table 8 shows the influence of voice usage on the acoustic parameters, by gender and diagnosis.

For the variable voice usage, the results showed that there are no differences between groups (singers and nonsingers), except for the parameter Jitter ppq5 in males with vocal pathology.

Figure 5 shows examples of radial graphs for two participants, randomly chosen and included in AVFAD database: the male participant FAX and the female participant MLY. Through the analysis of the graphics, it is possible to verify that only the NHR and autocorrelation mean parameters of FAX are out of the normal range (grey circular area), but a much broader range of parameters for MLY are not within the reference interval.

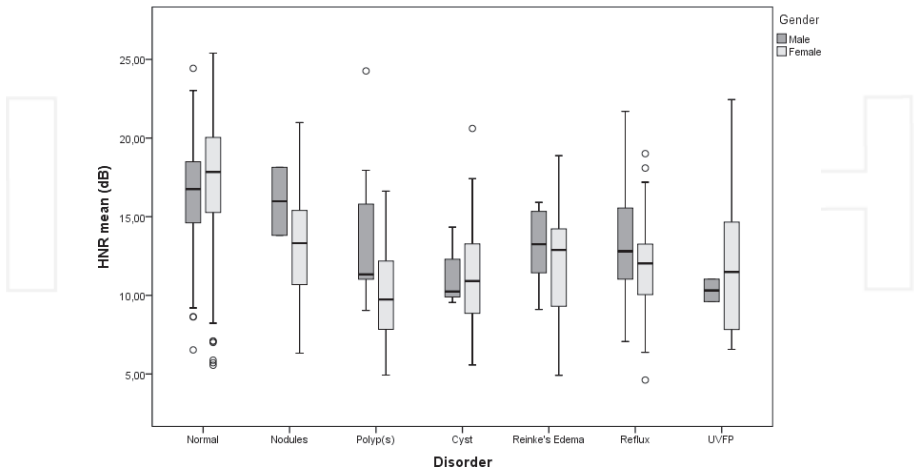


Figure 4. HNR by voice disorder for both genders.

	♂ Male				♀ Female			
	Normal		Vocal pathology		Normal		Vocal pathology	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
f0 mean (Hz)	27.817	0.000*	8.229	0.016*	10.277	0.006*	11.050	0.004*
Jitter ppq5 (%)	1.976	0.372	3.765	0.152	6.630	0.036	2.300	0.317
Shimmer apq11 (%)	9.842	0.007*	2.940	0.230	27.521	0.000*	9.945	0.007*
HNR (dB)	0.313	0.855	2.572	0.276	11.990	0.002*	5.942	0.051

Nonparametric Kruskal-Wallis Test; \*statistical significant differences for  $\alpha = 0.05$ .

**Table 5.** Inferential statistics, grouped by gender and diagnosis, for the variable age.

	♂ Male				♀ Female			
	Normal		Vocal pathology		Normal		Vocal pathology	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
f0 mean (Hz)	8.371	0.015*	5.839	0.054	9.590	0.008*	32.882	<0.001*
Jitter ppq5 (%)	1.240	0.538	1.587	0.452	1.499	0.473	5.688	0.058
Shimmer apq11 (%)	9.097	0.011*	0.528	0.768	0.352	0.839	2.885	0.236
HNR (dB)	3.912	0.141	0.455	0.797	0.609	0.737	0.352	0.838

Nonparametric Kruskal-Wallis Test; \*statistical significant differences for  $\alpha = 0.05$ .

**Table 6.** Inferential statistics, grouped by gender and diagnosis, for the variable smoking habits.

	♂ Male				♀ Female			
	Normal		Vocal pathology		Normal		Vocal pathology	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
f0 mean (Hz)	6.123	0.106	5.272	0.153	9.658	0.022*	3.775	0.287
Jitter ppq5 (%)	0.725	0.867	4.968	0.174	2.028	0.567	1.515	0.679
Shimmer apq11 (%)	0.520	0.914	0.505	0.918	1.574	0.665	2.425	0.487
HNR (dB)	0.032	0.999	2.502	0.475	3.017	0.389	0.501	0.919

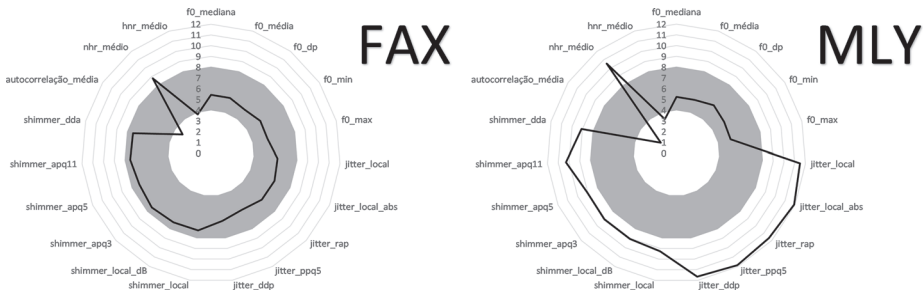
Nonparametric Kruskal-Wallis Test; \*statistical significant for  $\alpha = 0.05$ .

**Table 7.** Inferential statistics, grouped by gender and diagnosis, for the variable BMI.

	♂ Male				♀ Female			
	Normal		Vocal pathology		Normal		Vocal pathology	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
f0 mean (Hz)	0.656	0.418	0.025	0.874	1.503	0.220	1.762	0.184
Jitter ppq5 (%)	1.510	0.219	6.567	0.010*	0.225	0.635	1.532	0.216
Shimmer apq11 (%)	2.714	0.099	2.642	0.104	0.164	0.686	0.672	0.413
HNR (dB)	3.986	0.046	3.203	0.074	2.376	0.123	0.097	0.755

Nonparametric Kruskal-Wallis Test; \*statistical significant differences for  $\alpha = 0.05$ .

**Table 8.** Inferential statistics, by gender and diagnosis, for variable voice usage.



**Figure 5.** Radial graphs for patients FAX (male) and MLY (female).

### 4. Discussion

The number of patients with voice disorders has been increasing dramatically over the last decade, due mainly to unhealthy social habits and voice abuse. It has been reported that approximately 30% of the global population suffer from some kind of voice disorder during their lives. Previous studies have shown that impairment of vocal function can have a major impact on the quality of life, severely limiting communication at work and affecting all social aspects of daily life.

With an increasing concern to improve the assessment of voice for as many European languages as possible, we collected the first complete and representative EP pathological voice database. This database will be of huge significance for voice clinicians’ assessment and for testing and developing innovative, automated methods and devices for voice analysis. Besides the great importance of the database for this project, it will also be of great significance for future studies in the field of voice function assessment.

The development of voice is influenced by individual characteristics, but there is no consensus among authors about which specific characteristics affect vocal quality. Gender, however, has been identified as having a significant impact [28, 17], so the male and female samples of the study were analyzed individually.

The male  $f_0$  results showed a statistically significant difference between normal and pathological voices, but literature [40] indicates that both normal and pathological voices  $f_0$  values in our databases are within the normal range. Female  $f_0$  values were also within the normal range according to previous studies [16, 40], but normal and pathological voices in the AVFAD did not present statistically significant differences. For both genders, the results support previously published scientific evidence [16, 18, 35], suggesting that the  $f_0$  parameter does not allow to distinguish between a normal and a pathological voice.

According to literature [8], the voices of speakers with organic disorders of the larynx have higher Jitter and Shimmer and a lower HNR relative to the voices of normal speakers. Thereby, the results of this study (presented in **Tables 1** and **2**) are supported by literature [32], since for both genders, normal participants had statistically significant lower values of Jitter and Shimmer. According to [14], Jitter and Shimmer could be used as an acoustic parameters to differentially diagnose vocal pathology.

For the HNR parameter, as expected, normal participants had statistically significant higher values than pathological participants. The results of this study were below the threshold (30 dB) established by Deliyiski et al. [9], since the recording conditions (quiet room / audiology booth) used in their study were not reproducible in a hospital environment.

The results of multiple comparisons between normality and the six different vocal pathologies, for the male gender (as shown in **Table 3**), suggest statistically significant differences between the normal group and the Reinke's Edema group, and the normal group and the Reflux group, for Shimmer and HNR parameters. For the other groups, no statistically significant differences have been found. However, one should consider that the male pathological groups' sample size was too reduced to draw meaningful conclusions.

On the other hand, for the female gender, with much larger sample size, the results showed statistically significant differences between normality and the six pathologies, in relation to Shimmer, Jitter, and HNR parameters. When the Nodules and Reinke's Edema groups were compared to the normal group, significant differences in the  $f_0$  parameter were found as well. Accordingly, this study suggests that, in females, the studied vocal pathologies affect strongly the acoustic parameters Jitter, Shimmer, and HNR. The first two are increased, and HNR is decreased.

Concerning the four demographic characteristics in study (age group, smoking habits, BMI and voice usage), only the age group and smoking habits seem to interfere with voice quality. However, this was not the main goal of the present study and requires further investigation.

The radial graphs generated (with an Excel 2013 spreadsheet tool distributed with the AVFAD) from all the acoustic parameters automatically extracted from Praat were used in order to establish a threshold between normality and vocal pathology and are easy to interpret. It is believed that it might be a good tool helping voice clinicians to establish a diagnosis,

to communicate results with other health professionals, clients and their families. It allows a quick visualization of the acoustic speech signal parameters and to verify if they are within the normal range. They are also a functional way of monitoring progress in therapy.

## 5. Conclusions

The empirical work developed in this project produced new insights into voice disorder assessment and provides clinicians with a critical resource for voice assessment. The voice pathology database provides a new and valuable tool for voice clinicians and for speech research. This database will enable the interpretation of automatically extracted descriptors of the speech signal. The evaluation of some specific parameters such as  $f_0$ , Jitter, Shimmer and HNR has established differences of speech signal behaviors.

Acoustic databases aid professionals in the design of studies about the different pathologies, compare those pathologies with normal patterns and, furthermore, establish a threshold between normal and pathological voices. The AVFAD provide important tools for the study of voice, as they can objectively complement a differential diagnosis and help select an adequate intervention strategy.

In this study, we have analyzed the six most recurrent vocal pathologies (Reflux; Reinke's Edema; Nodules; Polyps; Cysts; UVFP) and defined standardized acoustic values for the  $f_0$  mean (Hz), Jitter ppq5 (%), Shimmer apq11 (%) and HNR (dB), correlating these with age, gender, BMI, voice usage and smoking habits.

The main conclusion drawn from this study was the existence of statistically significant differences between normal and pathological voice groups. Additionally, there are some evidences that suggest an association between age group, smoking habits, and the acoustic parameters, which should be further studied.

The radial graphs, drawn for each participant, allow a multidimensional acoustic analysis of patients and comparison with a normal (reference) range radial, giving clinician information on an individual's vocal quality, and allowing them to immediately detect changes in parameters.

One of the limitations of this study were the recording conditions. Although not optimal, they represent the real clinical setting where clinicians usually work and record their voice samples. For the male gender, some pathologies' sample size is limited and does not allow a reliable analysis of the results. We suggest that this sample may be increased in order to run a new statistical analysis and confirm the results.

In the future, considering the AVFAD database potentiality, the authors suggest their use in new voice studies, taking advantage of its complexity. For example, the other vocal tasks recorded could be analyzed. A detailed analysis of participants' clinical and demographic data is also recommended. These studies will represent a very important contribution to improve the knowledge about the acoustical characteristics of voice. Further work on performance improvement of assessment methods used by voice clinicians can be based on AVFAD samples and thus clinically applied.

## Acknowledgements

The authors would like to thank the Otorhinolaryngology teams from Hospital de Santo António, Hospital de São João and Hospital Pedro Hispano. We would also like to thank all the undergraduate and postgraduate students that contributed along the years to data collection, segmentation and annotation.

This research was partially funded by National Funds through FCT (Foundation for Science and Technology) in the context of the projects UID/MAT/04106/2013 and UID/CEC/00127/2013.

The AVFAD project was supported by the School of Health Sciences (ESSUA), University of Aveiro, Portugal.

This chapter was developed as part of the M.Sc. in Speech and Hearing Sciences at the University of Aveiro, Portugal: Belo I (2015). Valores de Referência de Parâmetros Acústicos para a Voz Normal no Português Europeu [Normal Voice Reference Values for Acoustic Parameters in European Portuguese], M.Sc. Thesis, University of Aveiro, Portugal; Machado J (2015). Características Acústicas de Patologias Vocais no Português Europeu [Acoustic Characteristics of Vocal Pathologies in European Portuguese], M.Sc. Thesis, University of Aveiro, Portugal.

The AVFAD are freely distributed through the ACSA <http://acsa.web.ua.pt/> platform.

## Author details

Luis M.T. Jesus\*, Inês Belo, Jessica Machado and Andreia Hall

\*Address all correspondence to: [lmtj@ua.pt](mailto:lmtj@ua.pt)

University of Aveiro, Aveiro, Portugal

## References

- [1] Alves M. Validade e Fiabilidade do Protocolo de Avaliação da Qualidade Vocal da Universidade de Aveiro [Validity and Reliability of University of Aveiro's Voice Evaluation Protocol] M.Sc. Thesis, Mestrado em Ciências da Fala e da Audição [M.Sc. in Speech and Hearing Sciences]. Aveiro: University of Aveiro; 2015
- [2] Barsties B, Bodt M. Assessment of voice quality: Current state-of-the-art. *Auris Nasus Larynx*. 2015;**42**:183-188
- [3] Behlau M. Consensus auditory-perceptual evaluation of voice (CAPE-V), ASHA 2003. *Revista da Sociedade Brasileira de Fonoaudiologia*. 2004;**9**(3):187-189



- [4] Boersma P. PRAAT, a system for doing phonetics by computer. *Glott International*. 2001;**5**:341-345
- [5] Brinca L, Batista A, Tavares A, Gonçalves I, Moreno M. Use of cepstral analyses for differentiating normal from dysphonic voices: A comparative study of connected speech versus sustained vowel in European Portuguese female speakers. *Journal of Voice*. 2014;**28**(3):282-286
- [6] Brinca L, Batista A, Tavares A, Pinto P, Araújo L. The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. *Journal of Voice*. 2015;**29**(6):776.e7-776.e14
- [7] Brinca L, Nogueira P, Tavares A, Batista A, Gonçalves I, Moreno M. The prevalence of laryngeal pathologies in an academic population. *Journal of Voice*. 2015;**29**(1):130.e1-130.e9
- [8] Cantarella G, Baracca G, Pignataro L, Forti S. Assessment of dysphonia due to benign vocal fold lesions by acoustic and aerodynamic indices: A multivariate analysis. *Logopedics Phoniatrics Vocology*. 2011;**36**(1):21-27
- [9] Deliyiski D, Shaw H, Evans M. Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*. 2005;**19**(15):15-28
- [10] Dollinger M, Kunduk M, Kaltenbacher M, Vondenhoff S, Ziethe A, Eysholdt U, Bohr C. Analysis of vocal fold function from acoustic data simultaneously recorded with high-speed endoscopy. *Journal of Voice*. 2012;**26**:726-733
- [11] Ferreira M, Jesus L, Couto P, Vilarinho H. University of Aveiro's standardised voice case history form. *Revista de Saúde Pública*. 2014;**48**(Special Number):297
- [12] Freitas S, Pestana P, Almeida V, Ferreira A. Audio-perceptual evaluation of Portuguese voice disorders—An inter- and intrajudge reliability study. *Journal of Voice*. 2014;**28**(2):210-215
- [13] Freitas S, Pestana P, Almeida V, Ferreira A. Integrating voice evaluation: Correlation between acoustic and audio-perceptual measures. *Journal of Voice*. 2015;**29**(3):390.e1-390.e7
- [14] Gelzinis A, Verikas A, Bacauskiene M. Automated speech analysis applied to laryngeal disease categorization. *Computer Methods and Programs in Biomedicine*. 2008;**91**(1):36-47
- [15] Gordis L. *Epidemiology*. 4th ed. Philadelphia: Saunders Elsevier; 2009
- [16] Guimarães I, Abberton E. Fundamental frequency in speakers of Portuguese for different voice samples. *Journal of Voice*. 2005;**19**:592-606
- [17] Hazan V. Speech communication across the life span. *Acoustics Today*. 2017;**13**(1):36-43
- [18] Hecker M, Kreul E. Description of speech of patients with cancer of the vocal folds. Part I measures of fundamental frequency. *Journal of Acoustical Society of America*. 1971;**49**:1275-1282
- [19] Jesus L, Barney A, Santos R, Caetano J, Jorge J, Couto P. Universidade de Aveiro's voice evaluation protocol. In: *Proceedings of Interspeech*. Brighton; 2009. pp. 971-974

- [20] Jesus L, Castilho S, Hall A. Is the relative fundamental frequency an acoustic correlate of laryngeal tension in Portuguese speakers? In: Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015). Glasgow; 2015
- [21] Jesus L, Martinez J, Hall A, Ferreira A. Acoustic correlates of compensatory adjustments to the glottic and supraglottic structures in patients with unilateral vocal fold paralysis. *BioMed Research International*. 2015;1-9, Article ID 704121
- [22] Jesus L, Tavares A, Hall A. Cross-cultural adaption of the GRBAS and CAPE-V scales for Portugal and a new training program for perceptual voice evaluation In: Fernandes F, editor. *Advances in Speech-Language Pathology*. Rijeka: InTech; 2017
- [23] Jesus L, Valente A, Hall A. Is the Portuguese version of the passage "The North Wind and the Sun" phonetically balanced? *Journal of the International Phonetic Association*. 2015;45(1):1-11
- [24] Jesus L, Vilarinho H, Ferreira M. Protocolo de Anamnese Vocal (Jovens e Adultos) da Universidade de Aveiro [University of Aveiro's Standard Voice Case History Form]. Aveiro: University of Aveiro; 2010. INPI Registration Number 463523 and IGAC Registration 27/05/2010. Available from Advanced Communication and Swallowing Assessment (ACSA) <http://acsa.web.ua.pt/>
- [25] Jesus L, Vilarinho H, Santos R. Protocolo de Avaliação da Qualidade Vocal da Universidade de Aveiro [University of Aveiro's Voice Evaluation Protocol]. Aveiro: University of Aveiro; 2010-2015. INPI Registration Number 463522 and IGAC Registration 27/05/2010 (Version 2 IGAC Registration Number 355/2012 and Version 2.3.6 IGAC Registration Number 3474/2015). Available from Advanced Communication and Swallowing Assessment (ACSA) <http://acsa.web.ua.pt/>
- [26] Kent R, Vorperian H, Duffy J. Reliability of the multi-dimensional voice program for the analysis of voice samples of subjects with dysarthria. *American Journal of Speech-Language Pathology*. 1999;8:129-136
- [27] Lazic M, Jovanovic N, Kulic M, Babac S, Jurisic V. Acoustic and perceptual characteristics of the voice in patients with vocal polyps after surgery and voice therapy. *Journal of Voice*. 2015;29(2):241-246
- [28] Linville S. The aging voice. In: Kent R, Ball M, editors. *Voice Quality Measurements*. San Diego: Singular; 2000. pp. 359-376
- [29] Ma E, Liu E. *Handbook of Voice Assessments*. San Diego: Plural; 2011
- [30] Madill C, McCabe P. Acoustic analysis using freeware: PRAAT. In: Ma E, Liu E, editors. *Handbook of Voice Assessments*. San Diego: Plural; 2011. pp. 231-251
- [31] Murry T. Speaking fundamental frequency characteristics associated with voice pathologies. *Journal of Speech and Hearing Disorders*. 1978;43:374-379
- [32] Murry T, Doherty E. Selected acoustic characteristics of pathologic and normal speakers. *Journal of Speech and Hearing Research*. 1980;23(2):361-369

- [33] Nemr K, Zenari M, Cordeiro G, Tsuji D, Ogawa A, Ubrig M, Menezes M. GRBAS and Cape-V scales: High reliability and consensus when applied at different times. *Journal of Voice*. 2012;**26**(6):812.e17-812.e22
- [34] Nemr K, Zenari M, Souza G, Hachiya A, Tsuji D. Correlation of the Dysphonia Severity Index (DSI), Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V), and gender in Brazilians with and without voice disorders. *Journal of Voice*. 2016;**30**(6):765.e7-765.e11
- [35] Peppard R, Bless D, Milenkovic P. Comparison of young adult singers and non-singers with vocal nodules. *Journal of Voice*. 1988;**2**:250-260
- [36] Pontes P, Vieira V, Gonçalves M, Pontes A. Características das vozes roucas, ásperas e normais: análise acústica espectrográfica comparativa. *Revista Brasileira de Otorrinolaringologia*. 2002;**68**(2):182-188
- [37] Pribuisiene R, Uloza V, Kupcinskas L, Jonaitis L. Perceptual and acoustic characteristics of voice changes in reflux laryngitis patients. *Journal of Voice*. 2006;**20**:128-136
- [38] Ravary L, Beet S. *Automatic Speech Signal Analysis for Clinical Diagnosis and Assessment of Speech Disorders*. New York: Springer; 2013
- [39] Rocha L, Behlau M, Souza L. Behavioral dysphonia and depression in elementary school teachers. *Journal of Voice*. 2015;**29**(6):712-717
- [40] Sebastian S, Babu S, Ballraj A, Oommen N. Acoustic measurements of geriatric voice. *Journal of Laryngology and Voice*. 2012;**2**(2):81-84
- [41] Silva R, Zenari M, Nemr N. Impacto de treinamento auditivo na avaliação perceptivo-auditiva da voz realizada por estudantes de Fonoaudiologia. *Revista da Sociedade Brasileira de Fonoaudiologia*. 2012;**24**(1):19-25
- [42] Silvestre I, Guimarães I, Teixeira A. Voice quality on women with vocal nodules diagnosis: Preliminary study. *Otorrinolaringologia E Cirurgia Cérvico-Facial (Revista Portuguesa)*. 2011;**49**(2):69-77
- [43] Stemple J, Roy N, Klaben B. *Clinical Voice Pathology: Theory and Management*. 5th ed. San Diego: Plural; 2014
- [44] Švec J, Granqvist S. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*. 2010;**19**(4):356-368
- [45] UN. *Provisional Guidelines on Standard International Age Classifications*. New York: United Nations; 1982
- [46] Verdolini K, Rosen C, Branski R. *Classification Manual for Voice Disorders-I*. Mahwah: Lawrence Erlbaum; 2006
- [47] Vogel A. Multidimensional analysis of voice: Computerized speech lab. In: Ma E, Liu E, editors. *Handbook of Voice Assessments*. San Diego: Plural; 2011. pp. 219-230

- [48] WHO. Obesity: Preventing and Managing the Global Epidemic. Geneva: World Health Organization (WHO); 2000
- [49] Yamasaki R, Madazio G, Leão S, Padovani M, Azevedo R, Behlau M. Auditory-perceptual evaluation of normal and dysphonic voices using the voice deviation scale. *Journal of Voice*. 2017;**31**(1):67-71

INTECH

INTECH

