Article

AI-Based Procedure for the Assessment of Voice Characteristics in Genetic Syndromes

Federico Calà , Lorenzo Frassineti , Elisabetta Sforza , Roberta Onesimo , Lucia D'Alatri , Claudia Manfredi [*] , Antonio Lanata , Giuseppe Zampino

Article

AI-Based Procedure for the Assessment of Voice Characteristics in Genetic Syndromes

Federico Calà [1], Lorenzo Frassineti [1,2], Elisabetta Sforza [3], Roberta Onesimo [4], Lucia D'Alatri [5], Claudia Manfredi [1,*], Antonio Lanata [1] and Giuseppe Zampino [3,4,6]

[1] Department of Information Engineering, University of Florence, Florence, Italy
[2] Department of Information Engineering, Università degli Studi di Pisa, Pisa, Italy
[3] Department of Life Sciences and Public Health, Faculty of Medicine and Surgery, Catholic University of Sacred Heart, Rome, Italy
[4] Centre for Rare Diseases and Transition, Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy
[5] Unit for Ear, Nose and Throat Medicine, Department of Neuroscience, Sensory Organs and Chest, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy
[6] European Reference Network for Rare Malformation Syndromes, Intellectual and Other Neurodevelopmental Disorders – ERN ITHACA
[*] Correspondence: clauida.manfredi@unifi.it

Abstract: Perceptual and statistical evidence has underlined voice characteristics of individuals affected by genetic syndromes different from that of normophonic subjects. In this paper we propose a procedure for the systematic collection of such pathological voices and the development of AI-based automatic tools to support differential diagnosis. Guidelines are provided concerning most suitable recording devices, vocal tasks and acoustical parameters, in order to simplify, speed up and make the whole procedure homogenous and replicable. The proposed procedure was applied to a set of 56 subjects, affected by Costello syndrome (CS), Down syndrome (DS), Noonan syndrome (NS) and Smith-Magenis syndrome (SMS). The whole database has been divided into three groups, respectively called: paediatric subjects (PS, individuals < 12 years of age), female adults (FA) and male adults (MA) subjects. In line with literature results, the Kruskal-Wallis test and post-hoc analysis with Dunn-Bonferroni test highlighted several significant differences in acoustical features not only between healthy subjects and patients, but also across syndromes within PS, FA and MA groups. Machine learning provided for the PS group a k-nearest neighbour classifier with 75% accuracy, for the FA group a support vector machine (SVM) model with 84% accuracy and for the MA group a SVM model with 97% accuracy. These preliminary results suggest that the proposed procedure, based on acoustical analysis and AI, might be helpful for an effective non-invasive automatic characterization of genetic syndromes.

Keywords: genetic syndrome; acoustical analysis; artificial intelligence; classification

---

1. Introduction

Voice represents the result of complex configuration, arrangement and coordination of the elements that make up the phonatory apparatus, the respiratory system and the nervous central system: abnormal neurological and anatomical characteristics, that often concern genetic syndromes, could alter voice production. Over the years, several papers investigated voice pathology detection due to benign formations (e.g., nodules, polyps), neuromuscular disorders (e.g., vocal cords paralysis) [1,2] as well as to neurodegenerative diseases such as Parkinson and Alzheimer [3–5]. Vocal tract, larynx and vocal folds anomalies can be identified by analyzing important acoustical parameters that are perceptually rated by experienced physicians and objectively assessed by dedicated software. In this latter context, some of the most important parameters are [6]:

- The fundamental frequency (F0) that describes the frequency of vibration of the vocal folds.

- The first formant (F1), related to front-half oral cavity constriction: the greater the cavity, the lower F1. Furthermore, F1 is raised by pharyngeal tract constriction.
- The second formant (F2), linked to tongue movements: it is lowered by back-tongue constriction and increased by front-tongue constriction.
- The third formant (F3) that depends on lips rounding: the more this configuration is accentuated the lower is F3.
- F0 and formants F1-F3 are respectively inversely proportional to the size and thickness of the vocal folds and to the vocal tract length.

In the last two decades, acoustical analysis started to be applied to patients affected by genetic syndromes, such as Costello (OMIM #218040, CS), Down (OMIM #190685, DS), Noonan (OMIM #163950, NS) and Smith-Magenis syndromes (OMIM #182290, SMS), leading to interesting results that underline markedly irregular voices. Non-invasive semeiotics of these diseases is generally carried out by studying somatic traits. In order to obtain a more detailed phenotype, a promising approach consists in objective acoustical analysis for the identification of a set of parameters associated with individual pathological conditions. Perceptually, low tonality and intensity of voice, as well as hoarseness, are typical characteristics of CS adult individuals [7]. Norman [8] has detected in CS children a severe deficit of articulation and language by means of the GFTA-3 test in comparison to age-matched healthy subjects (HS). However, the rarity of this syndrome has made particularly difficult to outline a precise acoustical profile and indeed no objective acoustical analysis was carried out on these patients. In Down syndrome, Moura et al. [9] detected statistically significant differences in comparison to HS concerning F0 for sustained vowels / a /, / e /, / i / and / ɔ/ as well as for formants F1-F2 and HNR measures in Portuguese-speaking children. In adults, Bunton and Leddy [10] highlighted articulation difficulties that reduce vowel space and intelligibility. Other studies show higher values for F0 in DS adults with respect to normophonic individuals, while the perceptual assessment in DS subjects, and in children in comparison to healthy, age-matched subjects [11,12]. Turkyilmaz et al. [13] have analyzed with the MVDP software [14] the sustained vowel /a/ of 11 children affected by Noonan syndrome: no significant differences were detected with respect to a control group except for the soft phonation index (SPI). Pierpont et al. [15] highlighted that 20% of NS children and adolescents present articulatory deficits using the GTFA-2 test. In a single-case report, Wilson and Dyson [16] noticed vowel neutralization and nasalization in a female child. In Garayzabal-Heinze et al. [17] study, SMS adults showed higher F0 values for the vowel /a/ as compared to normophonic ones, but no significant differences were found in voice perturbation measures. In another work [18], the same authors performed a similar experiment with SMS children and analyzed formants F1-F2 and cepstral peak prominence (CPP) [19], an important parameter for assessing dysphonia. Only CPP showed a significant difference between patients and control group. These studies suggest that acoustical analysis can provide physicians and speech therapists useful information. However, they have focused on a single pathology and identified acoustical parameters that show statistically significant differences as compared only to the normophonic case. Thus, it is important to further extend this research by analyzing and comparing vocal phenotypes across syndromes, in order to find significant alterations that could support and speed differential diagnosis. In this work, we propose a procedure for the standardization of the recording and analysis of the voice of patients affected by several genetic syndromes. Specifically, this procedure focuses on the most suitable devices for voice recordings and on which vocal tasks should be uttered. To ensure repeatability, we also propose signal pre-processing steps and feature extraction. Furthermore, this work describes how to carry out statistical analysis, the development of machine learning models and the performance assessment. The validation, feasibility and robustness of this procedure were tested applying the proposed approach to 72 patients recruited at the Fondazione Policlinico Universitario A. Gemelli (FPUG) in Rome, Italy.

2. Materials and Methods

Figure 1 displays and summarizes the proposed procedure. Details are given in the following subsections: Recordings, Vocal tasks, Preprocessing of audio samples, Acoustical analysis, Machine Learning and Statistical Analysis. Procedure assessment is discussed in the subsection named Procedure validation with real data.
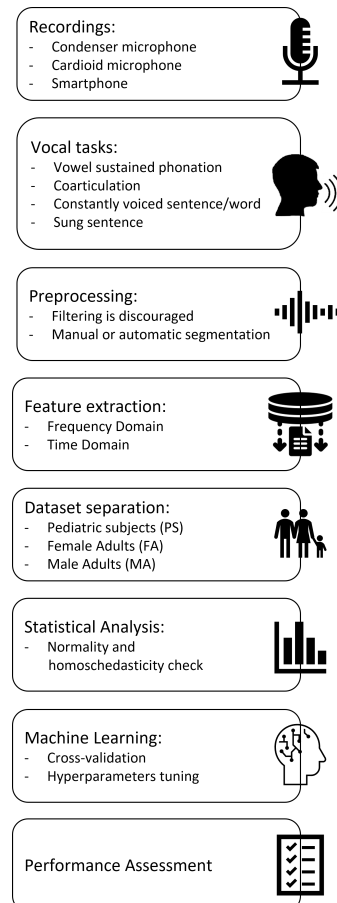

Figure 1. Procedure pipeline.

2.1. Recordings

In the guidelines provided by the Committee on Phoniatrics of the European Laryngological Society (ELS) [20], the importance of high-quality recordings is highlighted, both as far as perceptual and acoustical analysis are concerned. However, no further specification is given and this may cause difficulty in selecting an adequate recording system. Indeed, the choice of the most suitable device and its characteristics strongly depends on the application. Nevertheless, some general rules can be identified for the most appropriate microphone for different purposes [21]:

1. Flat frequency response.
2. Noise level at least 15dB lower than the sound level of the softest phonation.
3. Dynamic range upper limit higher than the sound level of the loudest phonation.
4. Distance between microphone and source for which the maximally flat frequency response occurs.

For example, a cardioid microphone performs well in noisy environments for perturbation metrics, but on the other hand, its short distance to the sound source may distort spectral evaluation and also give unreliable sound pressure level measurements, as pointed out by Svec and Granqvist [21]. However, it can be easily miniaturized and mounted on the subject with a clip, so it is particularly

indicated for children or patients with behavioural issues to avoid any form of distraction, which could negatively affect phonation and task completion. This recording device has been successfully used to study vowel utterance of patients with Smith-Magenis syndrome [18] and Williams syndrome [22]. To reduce ambient noise, headset with incorporated microphones has been proposed as well [23], however this choice strictly depended on the experimental design (i.e., Down syndrome prosody evaluation through serious games) and its use with other syndromes should be considered with caution due to its possible discomfort. Another possibility is represented by condenser microphones: for instance, Bunton & Leddy [10] in their experiment have used a SHURE SM81 microphone to record Down patients' voices. Smartphones represent a promising alternative for voice recordings: several studies have reported their efficiency, portability, cost-effectiveness also with pathological voices [24–26], although they must be used with caution as far as noisy environments, distance and inclination with respect to the radiation source are concerned. In Manfredi et al. [27], two smartphones selected at the extremes of commercial price range performed similarly in acquisition, suggesting that almost any smartphone-integrated microphone could be used to reliably record audio signals for acoustical analysis purposes. Cavalcanti et al. [28] performed a similar analysis, finding that smartphones seem not to alter most acoustical properties as compared to professional microphones, with the exception of harmonic-to-noise ratio (HNR) and cepstral peak prominence (CPP). These results partially agree with the study of Glover and Duhamel [29], where noise measurements were significantly different when comparing audio samples from smartphones and from digital voice recorders. However, such differences could be caused by incorrect positioning of recording devices and the small number of participants. As Cavalcanti et al. [28] state, the highly dynamic smartphone industry, the lack of standardization and the fact that companies usually do not disclose microphones characteristics, put a limit to the effective usability of smartphones. Nevertheless, their ubiquity and ease-of-use offer a relevant opportunity to monitor the voice quality for longitudinal evaluation and also during daily activities. Sound-treated booths are advisable, especially when using smartphones, however they can make the acquisition process slower and more complex, undermining smartphone's usage advantages. Taking into account the pathological subjects under study, our procedure suggests the use of a smartphone as it allows collecting a large number of recordings quite easily and quickly. Possibly, only one model of smartphone should be used for all acquisitions to ensure experimental repeatability and uniformity of the recordings. This device should be kept at 15cm from the mouth and with a 45° inclination, to reduce lateral distortions [30]. Background noise should be <50dB and subjects have to speak with conversational tone and intensity.

2.2. Vocal tasks

The evaluation of voice condition is typically based on two types of utterances: sustained vowels and running speech. In this work we refer to [6] for vowel symbols. Several papers have highlighted that certain pathologies can be more easily described and identified when examining specific tasks. For instance, Hidalgo-De la Guia et al. [22] required the utterance of vowel /a/ to favour a less forced phonation in patients with neuromotor deficits due to stable tongue and jaw position. In Frassineti et al. [31], it has been proven that adding /I/ and /u/ phonations improved pathology detection. Vowels /e/ and /o/ can be used as well, e.g., Suppa et al. [32] performed acoustical analysis of a sustained /e/ to detect Parkinson in elderly patients, however these utterances are more sensitive to dialects and therefore results may be less reliable. Instead, /a/, /I/ and /u/, usually referred as cardinal or corner vowels, are characterized by a well-defined vocal tract configuration and remain stable during articulation, which makes them substantially independent from dialectal and even linguistic diversity [33]. Running speech could provide further information to describe voice quality, as some of its aspects are highlighted in a voiced context or after a glottal closure. According to the European Laryngology Society (ELS), it is important that running speech, in the form of a single sentence or a short, standardized passage, should be characterized by constant voicing and not containing fricatives. This reduces possible biases due to articulation noise on noise level metrics computation and better

highlights the habitual fundamental frequency during speaking. As suggested by Godino-Llorente et al. [34], a coarticulation task should be included as well to evaluate the influence of the preceding and succeeding acoustical unit on the current unit under analysis. The Società Italiana di Fonologia e Laringologia (SIFEL) considers singing voice to evaluate a different aspect of speech production [20]: it does not entirely reflect daily conversational performance but, since voice is used at a higher functional level, it can provide an interesting insight into vocal properties, even in non-professional singers or actors. Seok et al. [35] demonstrated that, adding such task to their voice evaluation protocol, better allowed the assessment of vocal properties before and after thyroid surgery, helping to monitor postoperative voice changes and improved the assessment of subjective voice discomfort. Of course, factors like age, scarce cooperation, language deficits and cognitive disorder pose a challenge to the feasibility of running speech and singing tasks in patients affected by genetic syndromes. Despite such difficulties, our procedure consists in three repetitions of the following items (Italian language):

- List of numbers from 1 to 10.
- Word /aiuole/ (IPA transcription «aˈjwɔle»).
- Vowels /a/, /e/, /I/, /o/, /u/, sustained for at least 3s.
- Sentence "io amo le aiuole della mamma" (IPA transcription: «ˈio ˈamo ˈle aˈjwɔle ˈdelːa ˈmamːa», English translation: "I love mother's flowerbeds").
- Sung sentence "Fra Martino campanaro, dormi tu" (Italian version of the first sentence of the very well-known European traditional song Frère Jacques).

Three repetitions are required to account for biological variability and obtain, usually by averaging, more reliable parameters [36]. However, this is not always possible depending on the severity of the pathology.

2.3. Preprocessing of audio samples

To enhance feature extraction, speech and voice signals may undergo several preprocessing techniques. One of the most used techniques consists in inverse filtering: indeed, in voice pathology detection, the vibratory dynamics of the vocal folds can be analyzed removing the influence of vocal tract resonances [37,38]. However, when dealing with genetic syndromes, vocal properties alterations may not be uniquely associated to vocal folds biomechanical factors, but also to several morphological anomalies that affect the vocal tract itself. Examples are given by laryngomalacia, redundant nasal tissue, hypopharyngeal veil collapse for Costello syndrome [7,39], arytenoids cartilages enlargement and pharyngeal constriction for Down syndrome [23,40], ogival palate and tongue malformations for Noonan syndrome [41], velopharyngeal insufficiency for Smith-Magenis syndrome [42]. Therefore, filtering to reduce background or convolutional noise is discouraged in our procedure, as it could remove important information about irregularities or turbulences connected with pathologies themselves [34]. Thus, more appropriate tools could consist in voice/unvoiced and/or silence detectors that might better allow finding those segments of the recording where vocal folds vibration really occurs [3,43]. This approach can be applied to sustained vowels as well. In our procedure we suggest selecting only the central part of these signals to obtain more reliable acoustical measures, as they correspond to the "steady state" part of the recording. Such selection has been already performed in the literature [28]. For the procedure validation, manual segmentation was carried out using the Audacity software [44].

2.4. Acoustical analysis

Characterization of voice signals aims at extracting features to describe the properties of the pathological groups under examination. Gomez-Garcia et al. [34] pointed out that identifying the presence of voice impairment represents a difficult task since certain phenomena typically associated with vocal disorders (i.e., aperiodicity) can be inherent to physiological phonation processes. Thus, it is important to implement techniques to obtain a large number of different features in order to maximize

the probability of finding a range of metrics capable of separating normophonic from pathological subjects and/or differentiate pathologies. In recent years, several methodologies have been proposed and implemented to improve feature extraction; however the interpretability of the results must be taken into consideration, especially when exploratory analyses are performed. The open-source BioVoice tool performs objective acoustical analysis [45] both in the time and in the frequency domain. In the time domain, the number, length, and percentage of voiced and unvoiced segments (V/UV) are detected. In the frequency domain, fundamental frequency (F0, in Hz), formant frequencies (F1-F3, in Hz), noise level (Normalized Noise Energy – NNE, in dB), and jitter (in %) are estimated. NNE ranges from 0dB downwards, thus the higher the noise, the closer its value to 0 dB. For F0 and for each formant, the mean, median, standard deviation, maximum, and minimum values are calculated. Moreover, the Power Spectral Density (PSD) is computed in the frequency range of each category (newborn, child, adult female, adult male and singer), and it is normalized with respect to its maximum value. This allows comparison among different PSDs, ranging 0dB downward. Excel tables and pictures are automatically saved in devoted folders, one for each recording. The BioVoice drop-down menu allows choosing gender, age, and type of emission without requiring any manual setting to be made by the user. This greatly simplifies its usage also by non-expert users, automatically adjusting the proper frequency ranges for F0 and formants estimation. Overall, BioVoice computes 37 parameters, listed in Table 1 along with their meaning. The effective set of features to take into account depends on the context: when analyzing phonation, metrics associated with voiced and pause units can be ignored unless voice production for certain genetic syndromes is affected by factors such as spasmodic muscle contraction or frequent voice breakings.

Table 1. Acoustical parameters estimated by BioVoice. ($^+$) denotes the parameters used in this study.

Feature	Description
F0 mean [Hz]$^+$	Mean fundamental frequency
F0 median [Hz]$^+$	Median fundamental frequency
F0 std [Hz]$^+$	Standard deviation of fundamental frequency
F0 min [Hz]$^+$	Minimum fundamental frequency
T0 (F0 min) [s]$^+$	Time instant at which the minimum of F0 occurs
F0 max [Hz]$^+$	Maximum fundamental frequency
T0 (F0 max) [s]$^+$	Time instant at which the maximum of F0 occurs
Jitter [%]$^+$	Frequency variation of F0
NNE [dB]$^+$	Normalized Noise Energy
F1 mean [Hz]$^+$	Mean value of the first formant
F1 median [Hz]$^+$	Median value of the first formant
F1 std [Hz]$^+$	Standard deviation of the first formant
F1 min [Hz]$^+$	Minimum value of the first formant
F1 max [Hz]$^+$	Maximum value of the first formant
F2 mean [Hz]$^+$	Mean value of the second formant
F2 median [Hz]$^+$	Median value of the second formant
F2 std [Hz]$^+$	Standard deviation of the second formant
F2 min [Hz]$^+$	Minimum value of the second formant
F2 max [Hz]$^+$	Maximum value of the second formant
F3 mean [Hz]$^+$	Mean value of the third formant
F3 median [Hz]$^+$	Median value of the third formant
F3 std [Hz]$^+$	Standard deviation of the third formant
F3 min [Hz]$^+$	Minimum value of the third formant
F3 max [Hz]$^+$	Maximum value of the third formant
Signal duration [s]	Total audio file duration
% voiced	Percentage of voiced parts inside the whole signal
Voiced duration [s]	Total duration of voiced parts
Number Units	Number of voiced parts
Duration mean [s]	Mean duration of voiced parts
Duration std [s]	Standard deviation of duration of voiced parts
Duration min [s]	Minimum duration of voiced parts
Duration max [s]	Maximum duration of voiced parts
Number pauses	Total number of pauses in the audio file
Pause duration mean [s]	Mean duration of pauses
Pause duration std [s]	Standard deviation of duration of pauses
Pause duration min [s]	Minimum duration of pauses
Pause duration max [s]	Maximum duration of pauses

To carry out a more detailed voice analysis, it is possible to include articulatory parameters as well. They are related to the so-called vowel triangle: Figure 2 displays the American English vowel triangle [6], as well as the Italian vowel triangle, referred to adult males. The only slight difference concerns the vowel /I/ (Italian) with the F2 mean value about 300Hz lower than for the American English /i/. This difference will be taken into account in this paper.
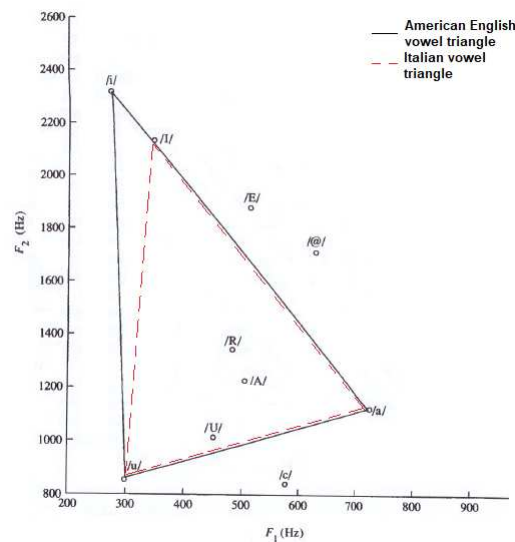

Figure 2. American English and Italian vowel triangles.

The vowel space area VSA (Equation 1) measures the vowel triangle area thus quantifying the articulatory ability [46].

$$VSA = \frac{\mid (F1I * (F2a - F2u) + F1a * (F2u - F2I) + F1u * (F2I - F2a) \mid}{2} \tag{1}$$

The formant centralization ratio (FCR) represents a normalization procedure to obtain an acoustical parameter that maximize dysarthria detection and minimize intervariability [47]. It is expressed as:

$$FCR = \frac{F2u + F2a + F1I + F1u}{F2I + F1a} \tag{2}$$

Formant ratios, proposed by Shapir et al., are other important parameters to evaluate tongue movements and therefore articulatory capabilities [48].

$$\frac{F1a}{F1I} \tag{3}$$

$$\frac{F1a}{F1u} \tag{4}$$

$$\frac{F1I}{F1u} \tag{5}$$

In particular, Equations 3 and 4 are sensitive to vertical tongue movements, Equation 5 to horizontal movements.

2.5. Dataset separation

Previous studies have highlighted that analyzing male and female voice altogether leads to less reliable results [31] mainly because of the different size and shape of the phonatory apparatus. Therefore, in this work the database was split into three groups: paediatric subjects (i.e., individuals < 12 years of age), female adults and male adults, respectively denoted by the acronyms PS, FA, MA.

2.6. Machine learning

The automatic assessment of voice quality based on artificial intelligence (AI) represents a well-established strategy that typically relies on supervised learning techniques such as k-nearest neighbours (KNN), support vector machines (SVM) and random forest (RF) [34,49]. Statistical analyses of voice properties of genetic syndromes have already highlighted significant differences in phonation and articulation. However, the proposed procedure focuses on prediction rather than inference; therefore it is important to develop a tool capable of generalizing the underlying pattern of training data in order to identify new observations. Machine learning (ML) is suitable for this task and several studies have demonstrated that such methods can separate data even if they do not show statistically significant alterations [50,51]. Moreover, ML requires few assumptions about the data generating systems, as opposed to statistical analysis where possible violations of the assumptions can lead to unreliable results. Our procedure suggests the use of ML techniques for the development of classifiers based on objective acoustical features to distinguish four pathological classes (Costello, Down, Noonan and Smith-Magenis syndromes) and normophonic subjects. Three different models are developed for the PS, FA and MA groups. Specifically, in this work KNN, SVM and RF classifiers are implemented. K-fold cross validation was performed, with k=5. To find the best model hyperparameters that maximize global accuracy, Bayesian optimization was carried out, with 30 iterations for KNN and 60 iterations for SVM and RF [49]. In our study:

1. For the KNN classifier: the number of neighbours k was evaluated between 2 and 27. The considered distance metrics were: "cityblock", "Chebyshev", "correlation", "cosine", "Euclidean", "hamming", "jaccard", "mahalanobis", "minkowski", "seuclidean", "spearman". The distance weight was chosen between "equal", "inverse", "squared inverse".
2. For the SVM classifier: coding was selected between "one vs. one" or "one vs. all". Box constraint and kernel scale were evaluated between 10-3 and 103. The kernel function was set as Gaussian.
3. For Random Forest: the fitcensemble.m function was used and the aggregation method was set as 'Bag'. The minimum number of leaves was selected between 2 and 27, the maximum number of splits between 2 and 27, the split criterion between "deviance", "gdi", "twoing", the number of variables to sample between 1 and 55.

All ML experiments were carried out in MATLAB® 2020b (The MathWorks, Inc., Natick, MS, USA). A code was developed to compute, for each class: recall, specificity, precision, F1-score, accuracy and AUC. Global accuracy was determined as well. In conclusion, for each cardinal vowel the first 24 parameters in Table 1, denoted by (+), plus 5 articulatory parameters (Equations 1-5) were considered, for a total of 77 features. We remark that the details provided here in points 1-3 allow the experimental repeatability, but they do not represent the unique procedure for future works. Indeed, with a larger dataset or when considering further syndromes, other models (including deep learning techniques) and hyperparameter tuning strategies would be tested.

2.7. Statistical analysis

In addition to machine learning methods, a statistical analysis was performed in order to understand whether the BioVoice acoustical parameters allow detecting significant differences among syndromes and if these results are in line with the literature. To find the more appropriate statistical test a Shapiro-Wilk test and a Lèvene test were applied to check normality and homoscedasticity, respectively. The SPSS tool (IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp) was used. Based on the outcome, a parametric one-way ANOVA or a non-parametric Kruskal-Wallis test was considered to carry a multivariate analysis of the acoustical features. The α-level of significance was set equal to 0.05. Post-hoc analysis considered t-tests with Tukey correction or Dunn-Bonferroni tests.

2.8. Procedure validation

Fifty-six patients were recruited at Fondazione Policlinico Universitario Gemelli (FPUG), Rome, Italy. Genetic syndromes involved in this study are: Down syndrome (13 subjects), Costello syndrome (10 subjects), Noonan syndrome (17 subjects) and Smith-Magenis syndrome (16 subjects). Data from sixteen healthy subjects were also collected to make up the control group. Inclusion criteria were: absence of voice pathologies, airways acute or chronic inflammation (such as rhinosinusitis or asthma). A Huawei 10 Mate was used for the recordings. All items of the SIFEL protocol described in [52] were recorded, but in this exploratory analysis only those concerning the three sustained vowels /a/, /I/ and /u/ are considered. Data were treated anonymously, and informed consent was obtained by each participant or their parents/tutor in case of minors. Table 2 shows the mean, standard deviation (in brackets) and number of recordings (in square brackets) of the considered genetic syndromes and control subjects in each group.

Table 2. Mean age with standard deviation (in brackets) and number of recordings (in square brackets) in each group.

	PS	FA	MA
CS	9.9 (2.0) [9]	16.4 (4.3) [15]	29.5 (2.1) [6]
DS	7.2 (3.6) [18]	21.2 (11.7) [12]	18.3 (2.2) [9]
NS	10.7 (2.3) [15]	22.4 (7.7) [18]	23.7 (8.4) [18]
SMS	8.0 (2.0) [24]	17.5 (1.3) [15]	16.3 (1.5) [9]
HS	8.9 (3.1) [21]	18.3 (6.8) [9]	21.3 (6.4) [18]

3. Results

None of the groups gave a positive result for the normality test, therefore only the Kruskal-Wallis test and post-hoc analysis with Dunn-Bonferroni test were performed. Tables 3–5 report both H-statistic and p-value only for the acoustical parameters that showed significant differences for the PS, FA and MA groups, respectively. An acoustical parameter able to discriminate between the normophonic class and one (or more) genetic syndrome is marked with (*), whereas for a separation across two (or more) pathological classes a (#) is used. The details about multiple comparisons are reported in Appendix A, B and C for the PS, FA and MA groups, respectively.

Table 3. Statistically significant differences in the acoustical parameters for the PS group. (*) denotes difference between one (or more) pathological classes and control subjects. (#) denotes one (or more) difference across genetic syndromes.

Parameter	Kruskal-Wallis H-statistic	p-value
F0 std /a/#	11.58	0.021
T0 (F0 min) /a/*	19.68	<0.001
T0 (F0 max) /a/*	23.40	<0.001
NNE /a/	11.14	0.025
F1 median /a/*#	20.02	<0.001
F1 min /a/*#	21.56	<0.001
F1 max /a/*#	16.50	0.002
F2 mean /a/*#	20.29	<0.001
F2 std /a/#	13.27	0.01
F2 min /a/*	13.84	0.008
F2 max /a/*#	29.77	<0.001
F3 mean /a/*	10.80	0.029
F3 std /a/#	22.01	<0.001
F3 min /a/*	10.09	0.039
F3 max /a/*	15.69	0.003

Table 3. *Cont.*

Parameter	Kruskal-Wallis H-statistic	p-value
T0 (F0 min) /I/*	19.58	<0.001
T0 (F0 max) /I/	10.75	0.03
F2 mean /I/*#	20.62	<0.001
F2 max /I/#	17.44	0.002
F1 mean /u/#	10.93	0.027
F1 std /u/#	10.44	0.034
F1 min /u/#	15.70	0.003
F2 std /u/#	14.29	0.006
F2 max /u/#	10.93	0.027
F3 std /u/#	12.80	0.012
F3 max /u/#	10.50	0.033
F1a/F1i*#	18.14	0.001
F1a/F1u*	18.07	0.002
F2i/F2u#	11.94	0.018
VSA*#	17.53	0.002
FCR*#	26.98	<0.001

Table 4. Statistically significant differences in the acoustical parameters for the PS group. (*) denotes difference between one (or more) pathological classes and control subjects. (#) denotes one (or more) difference across genetic syndromes.

Parameter	Kruskal-Wallis H-statistic	p-value
F0 mean /a/*#	18.70	<0.001
F0 min /a/#	14.76	0.005
F0 max /a/*#	17.37	0.002
NNE /a/	11.50	0.022
F1 mean /a/	14.07	0.007
F1 std /a/*#	18.53	<0.001
F1 min /a/*#	18.14	0.001
F2 mean /a/*	19.01	<0.001
F2 std /a/*#	16.00	0.003
F2 min /a/	10.20	0.04
F2 max /a/*#	24.78	<0.001
F0 mean /I/*#	18.70	<0.001
F0 std /I/#	11.07	0.026
F0 min /I/*#	13.05	0.011
F0 max /I/*#	19.55	<0.001
Jitter /I/#	21.09	<0.001
NNE /I/#	10.41	0.034
F1 std /I/#	15.94	0.003
F1 min /I/#	13.07	0.011
F2 mean /I/#	14.13	0.007
F2 std /I/#	12.57	0.014
F2 min /I/#	15.65	0.004
F3 mean /I/#	17.60	0.001
F3 min /I/#	14.07	0.007
F3 max /I/#	15.14	0.004
F0 mean /u/#	17.24	0.002
F0 std /u/#	12.73	0.013
F0 min /u/*#	19.72	<0.001
F0 max /u/#	11.87	0.018
Jitter /u/*#	11.77	0.019
F1 mean /u/#	17.38	0.002
F1 min /u/#	17.77	0.001
F2 mean /u/*	13.89	0.008
F2 std /u/#	14.65	0.005
F2 min /u/*#	13.38	0.01
F2 max /u/*#	17.54	0.002
F3 std /u/	10.50	0.033

Table 5. Statistically significant difference in the acoustical parameters for the PS group. (*) denotes difference between one (or more) pathological classes and control subjects. (#) denotes one (or more) difference across genetic syndromes.

Parameter	Kruskal-Wallis H-statistic	p-value
F0 mean /a/*	22.61	<0.001
F0 min /a/*	21.28	<0.001
F0 max /a/*	22.29	<0.001
F1 std /a/*#	23.17	<0.001
F1 min /a/*	14.70	0.005
F2 mean /a/*#	20.67	<0.001
F2 min /a/*#	29.86	<0.001
F2 max /a/*#	15.38	0.004
F3 mean /a/*#	19.49	<0.001
F3 min /a/*#	18.36	0.001
F3 max /a/*#	18.19	0.001
F0 mean /I/*	18.31	0.001
F0 max /I/*	21.74	<0.001
NNE /I/*	24.75	<0.001
F2 mean /I/*	15.58	0.004
F2 std /I/*	13.60	0.009
F2 min /I/*	16.94	0.002
F2 max /I/*	11.81	0.019
F0 mean /u/*#	25.06	<0.001
T0(F0 min) /u/*#	15.99	0.003
F0 max /u/*#	24.86	<0.001
NNE /u/*	16.51	0.002
F1 mean /u/*	11.67	0.02
F1 std /u/*	17.51	0.002
F1 min /u/#	14.64	0.006
F1 max /u/*	12.66	0.013
F2 mean /u/*	16.32	0.003
F2 std /u/	9.46	0.05
F2 min /u/	10.08	0.039
F2 max /u/*	27.40	<0.001
F3 mean /u/	11.58	0.021
F3 std /u/*	12.71	0.013
F3 min /u/*	18.99	<0.001
F1a/F1u*	10.27	0.036
F2i/F2u*	23.07	<0.001
VSA*	22.82	<0.001
FCR*	19.33	<0.001

For the PS group, Table 6 shows the performance of a KNN model with k = 2, distance metric = standardized Euclidean, weight = equal.

Table 6. Performance of a KNN classifier for the PS group.

Parameter	SMS	NS	CS	DS	HS
Precision	0.79	0.69	0.75	0.60	0.86
Recall	0.65	1	1	0.43	0.80
Specificity	0.97	0.91	0.96	0.96	0.85
F1-score	0.71	0.82	0.86	0.50	0.83
AUC	0.83	0.99	0.97	0.77	0.95
Validation Accuracy			75%		

Table 7 displays the performance of a SVM model with box constraint = 62 and kernel scale = 13.

Table 7. Performance of a SVM model for the FA group.

Parameter	SMS	NS	CS	DS	HS
Precision	0.85	0.86	1	1	1
Recall	0.92	1	0.86	0.5	1
Specificity	0.92	0.92	1	1	1
F1-score	0.88	0.92	0.92	0.67	1
AUC	0.97	0.98	0.94	0.91	1
Validation Accuracy			89%		

Table 8 reports the performance of a SVM model with box constraint = 462 and kernel scale = 21.

Table 8. Performance of SVM model for the MA group.

Parameter	SMS	NS	CS	DS	HS
Precision	1	0.92	1	1	1
Recall	0.83	1	1	1	1
Specificity	1	0.95	1	1	1
F1-score	0.91	0.96	1	1	1
AUC	1	1	1	1	1
Validation accuracy			97%		

Figures 3–5 show the vocalic triangle for the PS, MA and FA groups, respectively. The solid line refers to the healthy subjects considered in our study, whereas:

- The dotted line refers to SMS patients.
- The dashed line with circle markers refers to NS patients.
- The simple dashed line refers to CS patients.
- The dash-dotted line refers to DS patients.

The Italian reference triangle of Figure 2, represented in Figure 3–5 by a solid line with diamond markers, is added in order to compare the general results (healthy adult male subjects) to those related to our cohort of healthy subjects that differ for age and gender.
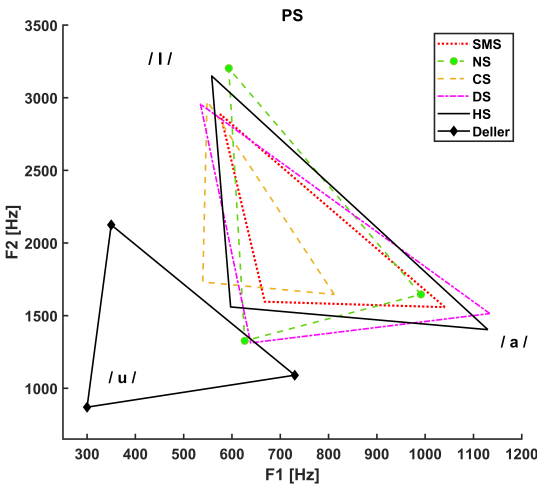

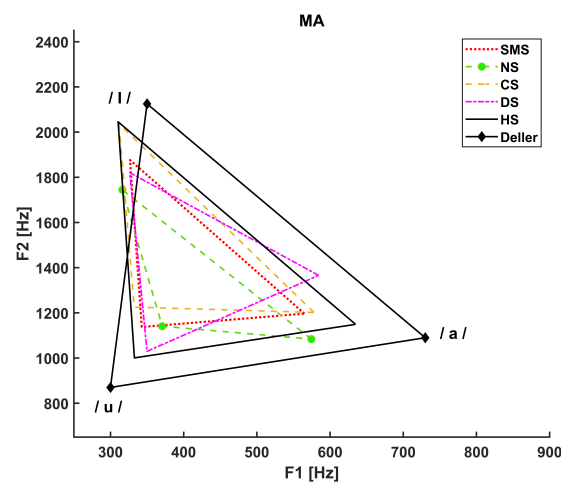**Figure 3.** Vocalic triangle for the PS group.

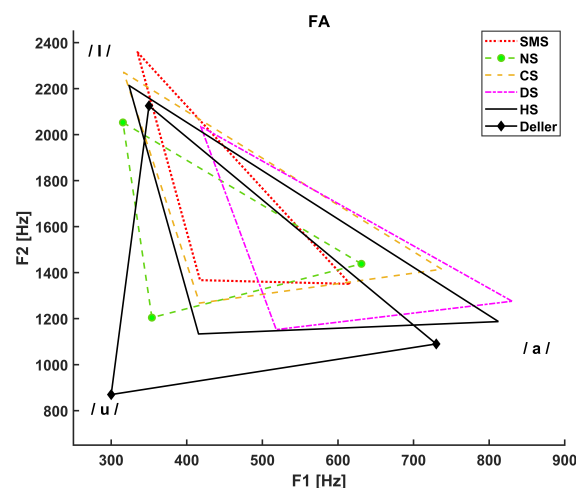**Figure 4.** Vocalic triangle for the MA group.

Figure 5. Vocalic triangle for the FA group.

4. Discussion

This paper proposes an innovative and detailed procedure for the assessment of voice characteristics of patients affected by genetic diseases. It was developed following the general guidelines provided by otolaryngological societies and associations and by reviewing literature articles on voice analysis and automatic voice quality assessment. This is a first attempt to standardize the acquisition, analysis and classification processes of voice samples of subjects affected by genetic syndromes. Acoustical analysis represents a promising non-invasive approach in this clinical field, therefore this paper aims at establishing first ground rules for uniform and comparable results. For the recordings, quiet rooms and a Huawei Mate 10 Lite (RNE-L21) smartphone were used. Though rigorous, the proposed procedure is easily adaptable to other pathologies. Moreover, this procedure might be also applicable to languages different from Italian, taking into account specific vocal tasks. Being an exploratory analysis, we validated it with statistical analysis and machine learning techniques and reported the outcome. Age range and gender were taken into account: this allowed obtaining more reliable acoustical parameters. Voice properties were compared not only between healthy and pathological subjects but also among genetic syndromes. Specifically, the considered syndromes were: Costello (CS), Down (DS), Noonan (NS) and Smith-Magenis (SMS). The results will be discussed following this order.

CS paediatric subjects did not show any statistically significant difference in acoustical parameter with the exception of F0 std /a/, which could reflect the lower ability to sustain vowel emission with respect to (w.r.t.) SMS patients due to generalized hypotonia or neck muscle spasticity [53]. Articulation deficits were highlighted by the vowel triangle (shrunk and left-shifted diagram in Figure 3 that may depend on detectable deformations of the vocal tract such as ogival palate, macroglossia, hypopharyngeal velum laxity and supraglottic stenosis [53]: these signs as well as pharynx structure malformations might cause difficulties in tongue movements. Indeed, statistical analysis detected significant differences in formant ratios (related to tongue motor ranges) and articulatory measures, e.g., F1a/F1i w.r.t. DS (p-value=0.006) and HS (p-value=0.014), FCR w.r.t. NS (p-value=0.021), DS (p-value=0.015) and HS (p-value<0.001). In the FA group of the Costello syndrome, the fundamental frequency of vowel /a/ presents significant differences w.r.t. NS (p-value=0.004) and of /u/ w.r.t. NS (p-value=0.006) and DS (p-value=0.018). Vocal instability and noise metrics computed for /I/ showed significant differences as well: jitter w.r.t DS (p-value=0.018) and NNE w.r.t. NS (p-value=0.026). This latter finding is in agreement with perceptual evaluation of the CS voice, which is defined as hoarse [7]. Hypotonia constraints of lips and tongue movements, especially in reaching their limit positions, and pharyngeal space reduction due to macroglossia, could be the reason for significant differences in F2 mean /a/ and F2 max /a/ w.r.t. the control group (p-value=0.031 and p-value=0.005, respectively).

Male subjects diagnosed with CS did not present significant differences for F0-related parameters, whereas statistical analysis showed differences concerning articulation, specifically w.r.t. NS (p-value=0.044) for F2 min /a/ and w.r.t. HS (p-value=0.024) for F2 mean /u/, that is also supported by the vowel triangle in Figure 4. This could be related to structural alterations of the posterior fossa, that can cause dysarthria [54], macroglossia or generalized hypotonia. This last medical evidence is related as well to a significant difference in F3 min /u/ w.r.t. HS (p-value=0.023).

In the DS PS group, differently from the results by Moura et al. [9] and Zampini et al. [12], F0 of vowels and jitter did not significantly differ from the HS group. Such discrepancy could be related to different spoken languages (Brazilian Portuguese in [9]), recorded utterances (speech fragments in [12]), numerousness ([9] applied acoustical analysis on a group of patients ten times larger than the one of this study), and the software used for acoustical analysis (PRAAT [55] in both cases). In the review by Kent [40] it was also stated that voice impairments with neurologic origin cause large variability in results, especially when evaluating F0 and its perturbations. As far as formant analysis is concerned, multiple comparisons showed statistical differences w.r.t. CS in F1 mean /a/ (p=0.002) and VSA (p=0.015), and w.r.t. NS in F2 max /I/ (p=0.021). These could be related to larger tongue dimensions which affect its movements and therefore modify vocal tract resonances. Statistical analysis on DS female subjects highlighted significant differences in F0 mean /I/ w.r.t. HS (p-value=0.003) and in F0 mean /u/ w.r.t. CS (p-value=0.018). Multiple comparisons showed significant differences between DS and CS (p-value=0.018) and NS (p-value<0.001) for jitter /I/ and between DS and CS (p-value=0.037) and HS (p-value=0.038) for jitter /u/. Articulation problems, which are still present in adults, determine significant differences in F1 mean /u/ and F1 min /a/ w.r.t. NS (p-value=0.001 and p-value=0.028) and F2 max /u/ w.r.t. SMS (p-value=0.003). In the MA DS group, post-hoc analysis detected significant statistical differences for FCR w.r.t. HS (p-value=0.007), for F2 mean /a/ w.r.t. NS (p-value<0.001) and HS (p-value=0.015), for F2 mean /I/ with HS (p-value=0.004), and for F3 mean /a/ w.r.t. NS (p-value=0.008) and HS (p-value=0.004). Neurologic abnormalities, located in the low temporal regions of the motor cortex, could be the reasons for these results.

For NS paediatric subjects, generalized low muscular tone, which tends to make difficult lateralization and protrusion of lips and tongue and limits jaw opening as well, might explain statistical differences in F1 min /a/ w.r.t. HS (p=0.024), in F2 mean /a/ w.r.t. HS (p=0.001), in F2 mean /I/ w.r.t. SMS (p=0.001) and F2i/F2u w.r.t. CS (p=0.049). Indeed Lee et al. [56] has demonstrated, with ultrasonographic measures, that F1 and F2 are strongly correlated to the oral cavity anterior length and the tongue posterior superficial length. Moreover, T0(F0 min) /a/ and T0(F0 max) /a/, show significant statistical differences w.r.t. HS (respectively p=<0.001 and p=0.002), that could be related to patients'

difficulty in maintaining a stable and regular vocal folds vibration during phonation. Statistical analysis of FA diagnosed with NS has highlighted differences in F0 mean /a/ and F0 mean /I/ w.r.t. HS (p-value=0.005 and p-value=0.028, respectively) and in F0 mean /u/ with CS (p-value=0.006): these alterations might depend on shorter height and shorter neck w.r.t. control subjects, which are common phenotypical feature for this syndrome. Moreover, jitter /I/ showed a significant difference w.r.t. CS (p-value=0.018) and SMS (p-value=0.025). As shown in Figure 5, NS FA vowel triangle is characterized by a small area, but VSA did not show any statistical significance. Nevertheless, formant coordinates have shown significant differences in F2 mean /a/ w.r.t. HS (p-value=0.001), F1 mean /u/ w.r.t. DS (p-value=0.001) and in F2 mean /I/ w.r.t. SMS (p-value=0.024): such alterations can be associated with difficulties in lips protrusion and lateralization [41]. Regarding the NS MA group, multiple comparisons showed statistical differences for F0 mean /a/ w.r.t. HS (p-value=0.028) and for F0 mean /u/ w.r.t. SMS (p-value=0.014): besides short stature, possible causes of this alteration could be the presence of an anterior glottis web [57] or a tendency to incur in vocal fold paralysis. These conditions can be associated as well to NNE values closer to 0, especially for /I/ and /u/ w.r.t. HS (p-value<0.001 and p-value=0.001, respectively). Figure 4 shows vowel area reduction and centralization: both VSA and FCR detected significant differences w.r.t. HS (p-value<0.001 and p-value=0.001, respectively). F2 also showed significant differences: F2 mean /a/ w.r.t. DS (p-value<0.001), F2 mean /I/ w.r.t. HS (p-value=0.004), and F2 mean /u/ w.r.t. HS (p-value=0.028). Such alterations might depend on structural properties, such as choanal atresia, supraglottic stenosis and soft palate laxity, as well as neurologic problems.

In PS SMS subjects, articulation measures and formants showed significant statistical differences for: F1 mean /u/ w.r.t. CS (p=0.034), F2 mean /a/ w.r.t. HS (p=0.037), F2 mean /I/ w.r.t. HS (p=0.015) and CS (p=0.001), F1a/F1u with HS (p=0.004) and FCR with HS (p=0.001). In Garayzabal-Heinze [18], nor F1 or F2 were able to discriminate SMS individuals from the control group: this difference could depend on the use of different acoustical analysis software tools. First formant alterations may be linked to velopharyngeal insufficiency: this incomplete closure typical of SMS patients causes a constant leak of airflow through nasal cavities, consequently altering resonant frequency along the vocal tract [42]. F0-related features did not show any significant difference for the FA group diagnosed with SMS. They were found for F2 mean /a/ w.r.t. HS (p-value=0.026), F2 mean /I/ w.r.t. NS (p-value=0.024) and F3 median /I/ w.r.t. NS (p-value=0.001) and CS (p-value=0.039). Hypotonia and structural lip malformations [17], in addition to frontal lobe calcification and cortical atrophy, could be the reasons for these anomalies. For the MA SMS group, post-hoc analysis highlighted significant differences in F0 mean /a/, F0 mean /I/ and F0 mean /u/ w.r.t. HS (p-value<0.001 for the three cardinal vowels). Alterations of F0 can be associated with greater vocal efforts during phonation due to vocal cords stiffness. Orofacial dysfunctions, worsened by hypotonia, soft palate clefts and posterior fossa anomalies might be responsible for articulation disabilities and therefore related to significant differences that were identified for F1 min /a/ and F2 max /u/ w.r.t. HS (p-value=0.050 and p-value=0.009, respectively).

Figures 3–5 show that age and gender strongly influence F1 and F2: with respect to the reference adult males (solid line with diamond markers), the PS group, and to a less extent the FA group, presents higher formant values due to the shorter and smaller size of the vocal folds and vocal tract. These results underline the importance to carry out the acoustical analysis taking into account also age and gender. Moreover, as shown in Figure 4, a difference exists also between the healthy adult male subjects considered in this study (simple solid line) and the reference adult males (solid line with diamonds) possibly because of our limited sample size.

In the KNN classifier of the PS group, NS and CS classes showed the highest AUCs (97% and 99%, respectively), whereas DS subjects present lower performance, in particular an AUC of 77% and recall of 43%, which may mean that their acoustical properties are not specific and therefore DS subjects are misclassified as other syndromes, possibly as SMS since this class shows a low recall value as well (65%). The SVM model of the FA group was able to detect healthy subjects with 100%

precision; moreover, no pathological subject was misclassified as HS, leading to 100% F1-score. This is an important result: as supported by statistical analysis, voice quality between female normophonic subjects and genetic syndrome patients is different. Moreover, SMS and NS class presented high precision values as well (85% and 86%), whereas DS present a low specificity of 50% meaning that their vocal properties are probably similar to the other considered diseases. In the male cohort, the overall best classification results were obtained: all the observation belonging to CS, DS and HS classes were correctly identified. High performance characterized SMS and NS classes as well. In this latter case, in the future it will be important to understand whether the same performance will be achieved by reducing number of parameters.

Summing up, these preliminary results are promising for defining a phonatory profile for genetic diseases. However, we remark that this outcome was obtained with a limited dataset and therefore more voice samples need to be collected. By applying the proposed procedure to a larger dataset, it will be possible to carry out reliable comparisons in order to validate and possibly find new acoustical features that could reliably describe genetic syndromes. Indeed, with a higher amount of data, new models could be developed to understand whether the same differences in the acoustical parameters between syndromes found in this work will be confirmed and if any improvements in classification results are feasible. Indeed, the low numerousness of subjects analyzed in this first study did not allow investigating feature selection or feature engineering techniques to obtain better classifiers. Such methods will be implemented once a larger database will be available.

5. Conclusions

In the present work, acoustical voice analysis of patients affected by genetic syndromes was performed according to a new procedure that can be easily applied both in clinical and domestic environments, as it does not require any special equipment. These guidelines allowed obtaining reliable acoustical parameters and assessing voice properties not only comparing healthy and pathological subjects, but also looking for acoustical differences among the four genetic syndromes considered here, i.e., Costello, Down, Noonan and Smith-Magenis syndromes, and with respect to healthy subjects. Acoustical parameters represent an important phenotypical aspect that can be measured non-invasively and, in addition to somatic traits analyzed by dysmorphologists, can be of help in addressing further medical examinations for diagnosis when genetic screenings are not available or when the syndrome's genome is still under evaluation. We analyzed these four syndromes taking into account the results in the literature that demonstrated the presence of neurological and structural problems associated with organs involved in voice production that alter phonation and articulation. The present paper aims at developing an easy, robust and efficient procedure for the analysis and classification of vocal traits specific of a number of genetic syndromes that can be used to set up and organize a large database. This will help extending existing results, comparing voice production between pathological and healthy subjects, to better highlight differences and finding the best parameters for each syndrome. Through this procedure, we also aim at performing more detailed statistical analyses and implementing new artificial intelligence approaches. A larger dataset will allow carrying on further studies to identify which morphological anomalies are linked to altered voice properties and verify the existence of possible vocal phenotype variability within single syndromes.

Appendix A

Figure A1 displays the results of post-hoc analysis for PS group. Black boxes highlight a significant difference between groups.

Parameter	DS vs. CS	DS vs. NS	DS vs. SMS	DS vs. HS	CS vs. NS	CS vs. SMS	CS vs. HS	NS vs. SMS	NS vs. HS	SMS vs. HS
F0 std /a/						■				
T0 (F0 min) /a/				■					■	
T0 (F0 max) /a/				■			■			
F1 median /a/	■					■	■			
F1 min /a/	■								■	
F2 mean /a/									■	■
F2 std /a/		■								
F2 min /a/	■						■			■
F2 max /a/	■								■	
F3 mean /a/										
F3 std /a/						■				
F3 max /a/							■		■	
T0 (F0 min) /I/				■					■	
F2 mean /I/								■		■
F2 max /I/		■						■		
F1 mean /u/						■				
F1 min /u/	■									
F2 std /u/	■			■						
F3 std /u/		■						■		
F3 max /u/										
F1a/F1i	■						■			
F1a/F1u							■			■
F2i/F2u				■				■		
VSA		■					■			
FCR	■			■			■			■

Figure A1. Black boxes refer to significant difference between syndromes and w.r.t. healthy subjects. DS = Down Syndrome; CS = Costello Syndrome; NS: Noonan Syndrome; SMS = Smith-Magenis Syndrome; HS = Healthy Subjects; VSA = Vowel Space Area; FCR = Formant Centralization Ratio.

Appendix B

Figure A2 displays the results of post-hoc analysis for FA group. Black boxes highlight a significant difference between groups.

Parameter	DS vs. CS	DS vs. NS	DS vs. SMS	DS vs. HS	CS vs. NS	CS vs. SMS	CS vs. HS	NS vs. SMS	NS vs. HS	SMS vs. HS
F0 mean /a/									■	
F0 min /a/					■					
F0 max /a/					■					
F1 std /a/		■								
F1 min /a/										
F2 mean /a/						■			■	
F2 std /a/		■							■	
F2 max /a/		■				■			■	
F0 mean /I/	■			■						
F0 std /I/							■			
F0 min /I/	■			■						
F0 max /I/	■									
Jitter /I/	■							■		
NNE /I/	■									
F1 std /I/		■			■					
F1 min /I/		■								
F2 mean /I/								■		
F2 std /I/								■		
F2 min /I/								■		
F3 mean /I/								■		
F3 min /I/								■		
F3 max /I/								■		
F0 mean /u/	■				■					
F0 std /u/						■				
F0 min /u/	■				■				■	
F0 max /u/	■									
Jitter /u/			■							
F1 mean /u/		■								
F1 min /u/		■								
F2 mean /u/										■
F2 std /u/		■								
F2 min /u/								■		
F2 max /u/			■					■		■

Figure A2. Black boxes refer to significant difference between syndromes and w.r.t. healthy subjects. DS = Down Syndrome; CS = Costello Syndrome; NS: Noonan Syndrome; SMS = Smith-Magenis Syndrome; HS = Healthy Subjects; VSA = Vowel Space Area; FCR = Formant Centralization Ratio.

Appendix C

Figure A3 displays the results of post-hoc analysis for MA group. Black boxes highlight a significant difference between groups.

Parameter	DS vs. CS	DS vs. NS	DS vs. SMS	DS vs. HS	CS vs. NS	CS vs. SMS	CS vs. HS	NS vs. SMS	NS vs. HS	SMS vs. HS
F0 mean /a/									■	
F0 min /a/									■	
F0 max /a/									■	■
F1 std /a/					■					■
F1 min /a/										■
F2 mean /a/		■		■					■	
F2 min /a/		■			■				■	
F2 max /a/		■		■						
F3 mean /a/		■		■						
F3 min /a/		■								
F3 max /a/		■								
F0 mean /I/										■
F0 max /I/										■
NNE /I/									■	
F2 mean /I/									■	
F2 std /I/									■	
F2 min /I/									■	
F2 max /I/									■	
F0 mean /u/								■		■
T0(F0 min) /u/			■					■		
F0 max /u/								■		
NNE /u/									■	
F1 mean /u/									■	
F1 std /u/				■						■
F1 min /u/								■		
F1 max /u/										
F2 mean /u/							■		■	
F2 max /u/									■	
F3 std /u/				■						
F3 min /u/								■	■	
F1a/F1u										
F2i/F2u										
VSA										
FCR				■						

Figure A3. Black boxes refer to significant difference between syndromes and w.r.t. healthy subjects. DS = Down Syndrome; CS = Costello Syndrome; NS: Noonan Syndrome; SMS = Smith-Magenis Syndrome; HS = Healthy Subjects; VSA = Vowel Space Area; FCR = Formant Centralization Ratio.

References

1. Muhammad, G.; Altuwaijri, G.; Alsulaiman, M.; Ali, Z.; Mesallam, T.A.; Farahat, M.; Malki, K.H.; Al-Nasheri, A. Automatic voice pathology detection and classification using vocal tract area irregularity. *Biocybernetics and Biomedical Engineering* **2016**, *36*, 309–317.

2. Al-Nasheri, A.; Muhammad, G.; Alsulaiman, M.; Ali, Z.; Malki, K.H.; Mesallam, T.A.; Ibrahim, M.F. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. *Ieee Access* **2017**, *6*, 6961–6974.

3. Bandini, A.; Giovannelli, F.; Orlandi, S.; Barbagallo, S.D.; Cincotta, M.; Vanni, P.; Chiaramonti, R.; Borgheresi, A.; Zaccara, G.; Manfredi, C. Automatic identification of dysprosody in idiopathic Parkinson's disease. *Biomedical Signal Processing and Control* **2015**, *17*, 47–54.

4. Mittal, V.; Sharma, R. Classification of Parkinson Disease Based on Analysis and Synthesis of Voice Signal. *International Journal of Healthcare Information Systems and Informatics (IJHISI)* **2021**, *16*, 1–22.

5. Naranjo, L.; Perez, C.J.; Martin, J.; Campos-Roca, Y. A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications. *Computer methods and programs in biomedicine* **2017**, *142*, 147–156.

6. Deller Jr, J.R. Discrete-time processing of speech signals. In *Discrete-time processing of speech signals*; 1993; pp. 908–908.

7. Gripp, K.W.; Morse, L.A.; Axelrad, M.; Chatfield, K.C.; Chidekel, A.; Dobyns, W.; Doyle, D.; Kerr, B.; Lin, A.E.; Schwartz, D.D.; others. Costello syndrome: Clinical phenotype, genotype, and management guidelines. *American Journal of Medical Genetics Part A* **2019**, *179*, 1725–1744.

8. Norman, B.I. An examination of communication skills of individuals with Cardio-facio-cutaneous syndrome and Costello syndrome compared to typically developing individuals. PhD thesis, California State University, Sacramento, 2020.

9. Moura, C.P.; Cunha, L.M.; Vilarinho, H.; Cunha, M.J.; Freitas, D.; Palha, M.; Pueschel, S.M.; Pais-Clemente, M. Voice parameters in children with Down syndrome. *Journal of Voice* **2008**, *22*, 34–42.

10. Bunton, K.; Leddy, M. An evaluation of articulatory working space area in vowel production of adults with Down syndrome. *Clinical linguistics & phonetics* **2011**, *25*, 321–334.

11. Albertini, G.; Bonassi, S.; Dall'Armi, V.; Giachetti, I.; Giaquinto, S.; Mignano, M. Spectral analysis of the voice in Down syndrome. *Research in developmental disabilities* **2010**, *31*, 995–1001.

12. Zampini, L.; Fasolo, M.; Spinelli, M.; Zanchi, P.; Suttora, C.; Salerni, N. Prosodic skills in children with Down syndrome and in typically developing children. *International Journal of Language & Communication Disorders* **2016**, *51*, 74–83.

13. Türkyilmaz, M.; Tokgöz Yılmaz, S.; Özcebe, E.; Yüksel, S.; Süslü, N.; Tekin, M. Voice characteristics of children with noonan syndrome Noonan sendromu olan çocuklarda ses özellikleri. *Turkiye Klinikleri Journal of Medical Sciences* **2014**, *34*.

14. Elemetrics, K. Operations manual, multi-dimensional voice program (MDVP) model 4305. *Issue A. August. Pine Brook NJ: Kay Elemetric Corporation* **1993**.

15. Pierpont, M.E.M.; Magoulas, P.L.; Adi, S.; Kavamura, M.I.; Neri, G.; Noonan, J.; Pierpont, E.I.; Reinker, K.; Roberts, A.E.; Shankar, S.; others. Cardio-facio-cutaneous syndrome: clinical features, diagnosis, and management guidelines. *Pediatrics* **2014**, *134*, e1149–e1162.

16. Wilson, M.; Dyson, A. Noonan syndrome: Speech and language characteristics. *Journal of Communication Disorders* **1982**, *15*, 347–352.

17. Hidalgo-De la Guía, I.; Garayzábal-Heinze, E.; Gómez-Vilda, P. Voice characteristics in smith–magenis syndrome: an acoustic study of laryngeal biomechanics. *Languages* **2020**, *5*, 31.

18. Hidalgo-De la Guía, I.; Garayzábal-Heinze, E.; Gómez-Vilda, P.; Martínez-Olalla, R.; Palacios-Alonso, D. Acoustic Analysis of Phonation in Children With Smith–Magenis Syndrome. *Frontiers in Human Neuroscience* **2021**, *15*, 661392.

19. Hillenbrand, J.; Houde, R.A. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research* **1996**, *39*, 311–321.

20. Dejonckere, P.H.; Bradley, P.; Clemente, P.; Cornut, G.; Crevier-Buchman, L.; Friedrich, G.; Van De Heyning, P.; Remacle, M.; Woisard, V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *European Archives of Oto-rhino-laryngology* **2001**, *258*, 77–82.

21. Svec, J.G.; Granqvist, S. Guidelines for selecting microphones for human voice production research **2010**.

22. Hidalgo, I.; Vilda, P.G.; Garayzábal, E. Biomechanical Description of phonation in children affected by Williams syndrome. *Journal of Voice* **2018**, *32*, 515–e15.

23. Corrales-Astorgano, M.; Escudero-Mancebo, D.; González-Ferreras, C. Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome. *Speech Communication* **2018**, *99*, 90–100.

24. Flanagan, O.; Chan, A.; Roop, P.; Sundram, F. Using acoustic speech patterns from smartphones to investigate mood disorders: scoping review. *JMIR mHealth and uHealth* **2021**, *9*, e24352.

25. Yoon, H.; Gaw, N. A novel multi-task linear mixed model for smartphone-based telemonitoring. *Expert Systems with Applications* **2021**, *164*, 113809.

26. Amir, O.; Anker, S.D.; Gork, I.; Abraham, W.T.; Pinney, S.P.; Burkhoff, D.; Shallom, I.D.; Haviv, R.; Edelman, E.R.; Lotan, C. Feasibility of remote speech analysis in evaluation of dynamic fluid overload in heart failure patients undergoing haemodialysis treatment. *ESC Heart Failure* **2021**, *8*, 2467–2472.

27. Manfredi, C.; Lebacq, J.; Cantarella, G.; Schoentgen, J.; Orlandi, S.; Bandini, A.; DeJonckere, P.H. Smartphones offer new opportunities in clinical voice research. *Journal of voice* **2017**, *31*, 111–e1.

28. Cavalcanti, J.C.; Englert, M.; Oliveira Jr, M.; Constantini, A.C. Microphone and audio compression effects on acoustic voice analysis: A pilot study. *Journal of Voice* **2021**.

29. Glover, M.; Duhamel, M.F. Assessment of Two Audio-Recording Methods for Remote Collection of Vocal Biomarkers Indicative of Tobacco Smoking Harm. *Acoustics Australia* **2023**, *51*, 39–52.

30. Lebacq, J.; Schoentgen, J.; Cantarella, G.; Bruss, F.T.; Manfredi, C.; DeJonckere, P. Maximal ambient noise levels and type of voice material required for valid use of smartphones in clinical voice research. *Journal of voice* **2017**, *31*, 550–556.

31. Frassineti, L.; Zucconi, A.; Calà, F.; Sforza, E.; Onesimo, R.; Leoni, C.; Rigante, M.; Manfredi, C.; Zampino, G. Analysis of vocal patterns as a diagnostic tool in patients with genetic syndromes. *PROCEEDINGS E REPORT* **2021**, p. 83.

32. Suppa, A.; Costantini, G.; Asci, F.; Di Leo, P.; Al-Wardat, M.S.; Di Lazzaro, G.; Scalise, S.; Pisani, A.; Saggio, G. Voice in Parkinson's disease: a machine learning study. *Frontiers in Neurology* **2022**, *13*, 831428.

33. Lenoci, G.; Celata, C.; Ricci, I.; Chilosi, A.; Barone, V. Vowel variability and contrast in childhood apraxia of speech: Acoustics and articulation. *Clinical linguistics & phonetics* **2021**, *35*, 1011–1035.

34. Gómez-García, J.; Moro-Velázquez, L.; Arias-Londoño, J.D.; Godino-Llorente, J.I. On the design of automatic voice condition analysis systems. Part III: Review of acoustic modelling strategies. *Biomedical Signal Processing and Control* **2021**, *66*, 102049.

35. Seok, J.; Ryu, Y.M.; Jo, S.A.; Lee, C.Y.; Jung, Y.S.; Ryu, J.; Ryu, C.H. Singing voice range profile: New objective evaluation methods for voice change after thyroidectomy. *Clinical Otolaryngology* **2021**, *46*, 332–339.

36. Carrón, J.; Campos-Roca, Y.; Madruga, M.; Pérez, C.J. A mobile-assisted voice condition analysis system for Parkinson's disease: assessment of usability conditions. *Biomedical engineering online* **2021**, *20*, 1–24.

37. Kohler, M.; Vellasco, M.M.; Cataldo, E.; others. Analysis and classification of voice pathologies using glottal signal parameters. *Journal of Voice* **2016**, *30*, 549–556.

38. Gómez-Vilda, P.; Fernández-Baillo, R.; Nieto, A.; Díaz, F.; Fernández-Camacho, F.J.; Rodellar, V.; Álvarez, A.; Martínez, R. Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *Journal of Voice* **2007**, *21*, 450–476.

39. Gripp, K.W.; Lin, A.E. Costello syndrome: a Ras/mitogen activated protein kinase pathway syndrome (rasopathy) resulting from HRAS germline mutations. *Genetics in Medicine* **2012**, *14*, 285–292.

40. Kent, R.D.; Vorperian, H.K. Speech impairment in Down syndrome: A review **2013**.

41. Torres, G.X.; Santos, E.d.S.; César, C.P.H.A.R.; Irineu, R.d.A.; Dias, I.R.R.; Ramos, A.F. Clinical orofacial and myofunctional manifestations in an adolescent with Noonan Syndrome: a case report. *Revista CEFAC* **2020**, *22*.

42. Rinaldi, B.; Villa, R.; Sironi, A.; Garavelli, L.; Finelli, P.; Bedeschi, M.F. Smith-magenis syndrome—clinical review, biological background and related disorders. *Genes* **2022**, *13*, 335.

43. Vieira, M.N.; McInnes, F.R.; Jack, M.A. On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures. *The Journal of the Acoustical Society of America* **2002**, *111*, 1045–1055.

44. Schroder, C. *The book of Audacity: Record, edit, mix, and master with the free audio editor*; No Starch Press, 2011.

45. Morelli, M.S.; Manfredi, S.O.C. BioVoice: A multipurpose tool for voice analysis. Proceeding 11th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications, MAVEBA 2019. Firenze University Press, 2019, pp. 261–264.

46. Kent, R.D.; Kim, Y.J. Toward an acoustic typology of motor speech disorders. *Clinical linguistics & phonetics* **2003**, *17*, 427–445.

47. Sapir, S.; Ramig, L.O.; Spielman, J.L.; Fox, C. Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech **2010**.

48. Blog, C. Effects of Intensive Voice Treatment (LSVT) on Vowel Articulation in Dysarthric Individuals With Idiopathic Parkinson Disease: Acoustic and Perceptual Findings Shimon Sapir, Jennifer L. Spielman, Lorraine O. Ramig, Brad H. Story, and Cynthia Fox. *Journal of Speech, Language, and Hearing Research* **2018**, *50*, 899–912.

49. Harar, P.; Galaz, Z.; Alonso-Hernandez, J.B.; Mekyska, J.; Burget, R.; Smekal, Z. Towards robust voice pathology detection: Investigation of supervised deep learning, gradient boosting, and anomaly detection approaches across four databases. *Neural Computing and Applications* **2020**, *32*, 15747–15757.

50. Bur, A.M.; Shew, M.; New, J. Artificial intelligence for the otolaryngologist: a state of the art review. *Otolaryngology–Head and Neck Surgery* **2019**, *160*, 603–611.

51. Costantini, G.; Di Leo, P.; Asci, F.; Zarezadeh, Z.; Marsili, L.; Errico, V.; Suppa, A.; Saggio, G. Machine Learning based Voice Analysis in Spasmodic Dysphonia: An Investigation of Most Relevant Features from Specific Vocal Tasks. BIOSIGNALS, 2021, pp. 103–113.

52. Maccarini, L.R.; Lucchini, E. La valutazione soggettiva e oggettiva della disfonia. Il Protocollo SIFEL, Relazione Ufficiale al XXXVI Congresso Nazionale della Società Italiana di Foniatria e Logopedia. *Acta Phoniatrica Latina* **2002**, *24*, 13–42.

53. Choi, N.; Ko, J.M.; Shin, S.H.; Kim, E.K.; Kim, H.S.; Song, M.K.; Choi, C.W. Phenotypic and genetic characteristics of five Korean patients with Costello syndrome. *Cytogenetic and Genome Research* **2019**, *158*, 184–191.

54. De Smet, H.J.; Catsman-Berrevoets, C.; Aarsen, F.; Verhoeven, J.; Mariën, P.; Paquier, P.F. Auditory-perceptual speech analysis in children with cerebellar tumours: a long-term follow-up study. *european journal of paediatric neurology* **2012**, *16*, 434–442.

55. Boersma, P.; Van Heuven, V. Speak and unSpeak with PRAAT. *Glot International* **2001**, *5*, 341–347.

56. Lee, S.H.; Yu, J.F.; Hsieh, Y.H.; Lee, G.S. Relationships between formant frequencies of sustained vowels and tongue contours measured by ultrasonography. *American journal of speech-language pathology* **2015**, *24*, 739–749.

57. Yellon, R.F. Prevention and management of complications of airway surgery in children. *Pediatric Anesthesia* **2004**, *14*, 107–111.