ANALYSIS OF VOCAL PATTERNS AS A DIAGNOSTIC TOOL IN PATIENTS WITH GENETIC SYNDROMES

Lorenzo Frassineti[1,2], Alice Zucconi[3], Federico Calà[4], Elisabetta Sforza[3], Roberta Onesimo[3], Chiara Leoni[3], Mario Rigante[3], Claudia Manfredi[1*] and Giuseppe Zampino [3,5*]
[1] Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[2] Department of Medical Biotechnologies, Università degli Studi di Siena, Siena, Italy
[3] Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma Italy
[4] School of Engineering, Università degli Studi di Firenze, Firenze, Italy
[5] Università Cattolica del Sacro Cuore, Roma, Italy
*Claudia Manfredi and Giuseppe Zampino jointly coordinated this work
lorenzo.frassineti@student.unisi.it, ali.zucconi@gmail.com, federico.cala@stud.unifi.it, giuseppe.zampino@unicatt.it, claudia.manfredi@unifi.it

Abstract: **Acoustical analysis is widely used in the diagnosis of speech disorders related to several pathologies and helps in defining the severity of their clinical pictures. Recently it was proved that some genetic syndromes may have a specific language phenotype. In this work we apply acoustical analysis to the discrimination between four genetic syndromes: Down, Noonan, Costello and Smith-Magenis. The analysis is performed with Praat and BioVoice tools. Several estimated acoustical features are applied as input to machine-learning models. Though preliminary, the results are encouraging: the acoustical analysis of the sustained vowel /a/ give an average accuracy > 50% with both tools. Our findings confirm that for some syndromes a specific "vocal phenotype" exists that might support the clinician in highlighting syndrome's characteristics not yet studied.**

Keywords: **Language Phenotype, BioVoice, Praat, Genetic Syndrome, Costello Syndrome, Noonan Syndrome, Smith Magenis Syndrome, Down Syndrome.**

I. INTRODUCTION

Genetic syndromes have been extensively studied for a better definition of their clinical manifestation, natural history and etiopathogenetic mechanisms. Nevertheless, some relevant but still unexplored aspects of these multisystemic conditions are not yet fully exploited, one of them being the characterization of vocal production. Genetic factors play a pivotal role not only in the determination of distinct phenotypes and neurobehavioral profiles, but also in establishing voice patterns with recognizable sound characteristics. Therefore, perceptual and acoustical analysis of voice could be helpful for the evaluation of specific voice characteristics as a non-invasive approach to the assessment of genetic syndromes [1]. More than 240 genetic syndromes have distinctive abnormalities of voice quality, significant enough to be considered as diagnostic indicators [2]. For some genetic syndromes the existence of a specific language phenotype obtained by acoustical analysis was already discussed in the literature. For example, young subjects affected by Down Syndrome may have differences concerning tremor, biomechanical behaviour and vibration of the vocal folds as compared to normative subjects [3]. For the Smith-Magenis Syndrome, acoustical and biomechanical analysis was recently performed to detect possible differences between pathological subjects and control groups [4]. Also, for the Cornelia de Lange Syndrome, anomalies in speech such as high levels of speech impairment were found [5]. For the Noonan Syndrome some preliminary evaluation was made with acoustical and biomechanical analysis to explore different aspects of the syndrome [6]. These findings might contribute to the differential diagnosis between Noonan Syndrome and some RASopathies [7] that share several aspects with them, such as the Costello Syndrome [8]. Indeed the Costello Syndrome may have specific acoustical characteristics due to the craniofacial anomalies often related to this syndrome that could alter the process of phonation and articulation [9]. Finally, acoustical analysis could be helpful for an early intervention in patients with speech impairments, to improve their communication skills and reduce speech deficits [10]. Based on the above mentioned evidences, some genetic abnormalities of a recognizable phenotype are expected to determine a specific vocal phenotype. Therefore, vocal characterization could represent a useful tool in the diagnostic process and in defining the severity of some clinical pictures [4].

To this aim, machine-learning methods and supervised classifiers are applied here to acoustical parameters estimated with two analysis tools: Praat and BioVoice [13, 14]. Being based on non-invasive and easily administered tests, this approach could be helpful for obtaining additional features useful for diagnosis and for the automatic classification of

different syndromes. The paper is organized as follows: in Section II the dataset and machine-learning experiment are described. In Section III the main results obtained are presented. Section IV is devoted to the discussion of results, limits and possible future developments. Conclusions are reported in Section V.

II. Material and Methods

Data were collected at the Università Cattolica del Sacro Cuore, (Roma), Faculty of Medicine and Surgery. Machine-learning methods are applied to several acoustical parameters estimated from the vocal emissions of a set of 72 subjects (36 male and 36 female, age range 4-33 years, mean 14±7 years), affected by 5 different genetic syndromes. Specifically, the dataset consists of: 22 subjects with Down syndrome (DS); 17 with Noonan syndrome (NS); 19 with Costello Syndrome (CS); 10 with Smith-Magenis syndrome (SMS) and 4 with Cornelia de Lange syndrome (CdLS). However, the CdLS syndrome was excluded from the analysis due to the small number of subjects in this class. The vocal samples come from a previous study based on the SIFEL protocol [11], [12]. After a training phase of the subject, the recorded audio files consist of the vowel /a/ sustained for at least 4 seconds. Recordings were obtained using a portable DAT (Digital Audio Tape) in a controlled environment (environmental noise < 40dB), with the microphone set at 15 centimetres from the subject's lips and with an angle of 45°. The sampling rate was 44100 Hz. Moreover, in the same sessions, the Italian word /aiuole/ (flower beds) as well as the vowels /i/, /u/ /o/ and /e/ were recorded. However, in this work we did not perform the acoustical analysis of these data with BioVoice, because some of them were corrupted or no more available. Only the acoustical analysis previously performed by Praat [11, 13] was available. The quasi-stationary central part of each sustained vowel (about 3s of duration) was manually extracted by an expert, disregarding onset and offset [11].

For the acoustical analysis and classification we considered here both the previously collected dataset of parameters estimated with Praat and new estimates obtained with the BioVoice tool [14, 15]. Only the sustained vowel /a/ was considered. With Praat, the following 34 acoustical parameters were taken into account: mean, standard error, coefficient of variation, maximum and minimum of the fundamental frequency F0; Jitter (local, absolute, Relative Average Perturbation, DDP and PPQ5, where PPQ is Period Perturbation Quotient); Shimmer (%, dB, APQ3, APQ5, APQ11, DDA, where APQ is the Amplitude Perturbation Quotient); mean Noise to Harmonic Ratio (NHR); mean Harmonic to Noise Ratio (HNR); the first four formants (F1, F2, F3 and F4); four clinical features: gender, age, weight and body mass index.

With BioVoice we extracted 24 acoustical features. Analysis is performed distinguishing between infants (<14 years) and adults [14] and in the case of adults between male and female. The 24 acoustical parameters from BioVoice are: maximum, minimum, mean, median and standard deviation for F0 and formants F1, F2 and F3; $T0_{min}$ and $T0_{max}$ for F0; jitter; Normalized Noise Energy (NNE). As before, the four clinical features: gender, age, weight and body mass index (BMI) were also included. In a first step, we compared the acoustical parameters in common between BioVoice and Praat. Then, we used those parameters considering separately each syndrome subgroup. All features except gender (0=male, 1=female) were normalized to zero mean and unit variance and the corresponding feature matrix was applied as input to the following supervised classifiers: k-nearest neighbours (KNN), support vector machine (SVM) and ensemble methods (we considered RUSBoost, AdaBoost and Random Forest). These methods are implemented under MATLAB 2020b computing environment [16]. K-fold cross validation (k=5) and Bayesian Optimization were applied for the selection of the hyper-parameters of the models. The optimization was performed considering the highest global Accuracy as validation metric (i.e. the average Accuracy between the four classes). To improve the classifier's performance the ReliefF algorithm [16] was used as feature selection method. During the model selection process we also varied the number of input features for the classifiers. All the experiments were repeated 5 times, to take into account possible variations of the performance due to the random selection of the subjects during cross-validation. We did not find significant differences in the performances (<5% Accuracy). Finally, we performed the same experiment on the Praat dataset, considering also features from the vowels /a/, /i/ and /u/. In this case the features given by the formant ratios between vowels were added (e.g., $F1_{[a]}/F1_{[u]}$) [13]. As said before, this analysis could not be performed with BioVoice due to missing data.

III. Results

Table 1 shows the comparison between Praat and BioVoice concerning the vowel /a/. We used a two-sample t-test with level of significance α=0.05. We checked the hypothesis of normality by Shapiro-Wilk Test (level of significance α=0.05). Table 2 shows the True Positive Rate (TPR) and the False Negative Rate (FNR) for the four genetic syndromes.

With BioVoice the 10 features obtained for the best model were: $T0_{maxF0}$ /a/, gender, age, median F3 /a/, BMI, min F1 /a/, $T0_{minF0}$ /a/, min F0, jitter and weight. The best model for BioVoice was a KNN with a

Global Accuracy of 53.1%. Instead with Praat the best model was made of 15 features: gender, mean F1 /a/, age, mean F2 /a/, BMI, max F0 /a/, min F0 /a/, weight, mean F0 /a/. median F0 /a/, Shimmer /a/ APQ11, Shimmer /a/ APQ5, Shimmer local /a/, mean F4 /a/, Shimmer /a/ DDA. The best model with Praat was a KNN with 52.9% of Global Accuracy.

The features used after the selection process are listed in descending order according to their relevance.

Table 1 – Vowel /a/ - Comparison between BioVoice and Praat on the 4 syndromes. Statistically significant differences are highlighted in bold.

	Syndrome (p-value)			
Feature	**DS**	**NS**	**CS**	**SMS**
Median F0 /a/	0.91	0.74	0.99	0.77
Mean F0 /a/	0.80	0.80	0.95	0.66
Min F0 /a/	**0.01**	0.05	**p<0.01**	0.13
Max F0 /a/	**p<0.01**	0.44	**0.02**	0,16
Mean F1 /a/	0.55	0.43	0.92	0.56
Mean F2 /a/	**p<0.01**	**p<0.01**	**0.03**	0.11
Mean F3 /a/	**p<0.01**	0.12	0.23	**p<0.01**

Table 2 – Vowel /a/ - Comparison between BioVoice and Praat - Results of k-fold cross validation.

Genetic	**BioVoice**		**Praat**	
Syndrome	**TPR**	**FNR**	**TPR**	**FNR**
DS	61.9%	38.1%	63.6%	36.4%
NS	26.7%	73.3%	17.6%	82.4%
CS	68.4%	31.6%	73.7%	26.3%
SMS	55.6%	44.4%	40.0%	60.0%

Table 3 shows the results obtained for the four genetic syndromes considering all the available Praat features for vowels /a/, /u/ and /i/.

Table 3 - Vowels /a/, /i/ and /u/ - KNN's Multiclass confusion matrix with Praat parameters. Main diagonal: TPR for each class. Other values: FNR for a single class.

	Predicted Class			
True Class	**DS**	**NS**	**CS**	**SMS**
DS	68.2%	13.6%	18.2%	0%
NS	17.6%	64.7%	17.6%	0%
CS	31.6%	5.3%	63.2%	0%
SMS	20.0%	10.0%	10.0%	60.0%

The best model was a KNN with Global accuracy 64.7%. In this case, the following 15 features were selected: mean F1 /a/, age, gender, formant ratio $F1_{[a]}/F1_{[u]}$, max F0 /a/, mean F2 /a/, Shimmer APQ11 /a/, mean F0 /a/, median F0 /a/, min F0 /a/, Shimmer /a/ (dB), BMI, Shimmer APQ5 /a/, weight, Shimmer /a/ (local).

IV. Discussion

This work presents preliminary results concerning the discrimination among some genetic syndromes: Down Syndrome, Noonan Syndrome, Costello Syndrome and Smith-Magenis Syndrome. The analysis was performed with acoustical parameters estimated on the sustained vowel /a/ with BioVoice and Praat and applying machine-learning models. The aim of this work was the definition of a proper language phenotype able to distinguish the genetic syndromes considered. The results shown in Table 2 and 3 confirm a possible relationship between genetic syndromes and their specific acoustical characteristics. The results obtained with BioVoice and Praat are comparable. Statistical analysis highlights some differences between the two tools as far as the estimation of formants F2 and F3 for some syndromes is concerned (Table 1, p-values <0.05). This might be related to different techniques for formants estimation implemented in the two tools, as discussed in [14]. Moreover, differences between BioVoice and Praat exist concerning F0 max and min. This could be due to different ranges for F0 estimation defined by the two software tools. We remark that with BioVoice the selection of the frequency range for adults (male or female), infants and newborns is automatically made by BioVoice, while Praat requires some skill of the user to manually set the best frequency range. However, the results shown in Table 2 are preliminary, suggesting that the analysis of the vowel /a/ alone might not be enough for defining a vocal phenotype (TPRs<50%). This is confirmed in Table 3, where the acoustical analysis of vowels /i/ and /u/ performed with Praat was added for all the syndromes, giving Accuracy>50%. In particular, the formant ratio $F1_{[a]}/F1_{[u]}$ was classified as one of the most relevant features by the ReliefF algorithm. This result suggests that a multi-vowel analysis might add more information than a single vowel analysis and should be preferred for the characterization of these genetic syndromes. Our results also confirm evidences previously found for some genetic syndromes. Indeed, for DS, NS and SMS acoustical analysis was already proved useful to find differences between pathological and control groups [3, 4, 6]. Table 3 also shows that SMS has the lowest false negative rate (0%), confirming that acoustical analysis can provide characteristics strictly related to the pathology [4]. Our results suggest that acoustical analysis could be useful also for CS. Indeed, as shown in Table 3, the false negative rates between CS and NS were 5.3% and

17.6% respectively, thus acoustical analysis might be useful to discriminate between these two syndromes.

Our results are preliminary and further study is required to confirm them. First, the number of subjects was poor, thus more cases must be recruited especially for SMS and CdLS. Moreover, we did not perform a comparison between pathological subjects and control cases. This will be done in future work, also taking into account previous studies that already presented such differences for some genetic syndromes [3,4,6]. Considering the promising results obtained, further studies will be made to investigate if some of the acoustical features could be specific of a single genetic syndrome. The acoustical analysis of vowels /i/ and /u/ made with the Praat dataset was found useful, therefore we are planning to perform the same analysis with BioVoice on the same recordings, when available, and/or new ones. Another limit of the work presented here is the wide age range of the subjects, also due to the low number of cases in some syndromes (e.g. CdLS or SMS). If other subjects will be available, a more detailed analysis at different age ranges will be made. If successful, acoustical analysis may be included in the process of differential diagnosis as a completely non-invasive approach to detect specific acoustical characteristics related to speech or phonation impairment for several genetic syndromes, along with e.g. the analysis of facial characteristics and expressions [17].

V. Conclusions

The work presented here is a first step towards the analysis and disentangle of the complex mosaics behind the detection of "voice" phenotypes related to some genetic syndromes. Preliminary results suggest that acoustical parameters and supervised classifiers might provide additional information about genetic syndromes through the characterization of voice. Future work will be devoted to the definition of a protocol for data recording and will concern a larger number of subjects and syndromes, as well as different supervised classifiers and feature selection approaches.

Acknowledgements

Partially funded by the Italian Ministry of Foreign Affairs and Scientific Cooperation (MAECI) MX18MO14.

References

[1] Villafuerte-Gonzalez, R., et al., Acoustic analysis of voice in children with cleft palate and velopharyngeal insufficiency. Int J Pediatr. Otorhinolaryngol, 2015. 79(7): 1073-6.

[2] Hamosh, A., et al., Online Mendelian Inheritance in Man (OMIM), a knowledge base of human genes and genetic disorders. Nucleic acids research, 2005. 33(Database issue): D514-D517.

[3] Hidalgo-De la Guía, et al. (2021). Specificities of phonation biomechanics in Down syndrome children. Biomedical Signal Processing and Control, 63, 102219.

[4] Garayzábal-Heinze, E, et al. (2020). Voice characteristics in smith–magenis syndrome: an acoustic study of laryngeal biomechanics. Languages, 5(3), 31.

[5] Moore, M. V. (1970). Speech, hearing, and language in de Lange syndrome. Journal of Speech and Hearing Disorders, 35(1), 66-69.

[6] Lazzaro, G., Zampino G., et al. (2020). Defining language disorders in children and adolescents with Noonan Syndrome. Molecular genetics & genomic medicine, 8(4), e1069.

[7] Myers, A., et al. (2014). Perinatal features of the RASopathies: Noonan syndrome, cardiofaciocutaneous syndrome and Costello syndrome. American journal of medical genetics Part A, 164(11), 2814-2821.

[8] Zampino, G., et al. (1993). Costello syndrome: further clinical delineation, natural history, genetic definition, and nosology. American journal of medical genetics, 47(2), 176-183.

[9] Mori, M., et al. (1996). Elastic fiber degeneration in Costello syndrome. American journal of medical genetics, 61(4), 304-309.

[10] Moura, C. P., et al. (2008). Voice parameters in children with Down syndrome. Journal of Voice, 22(1), 34-42.

[11] Zucconi A., (2018). Analisi della voce dei bambini con sindromi genetiche: verso l'identificazione di un "fonotipo". [Master Thesis] Università Cattolica del Sacro Cuore, Faculty of Medicine and Surgery.

[12] Ricci Maccarini A, et al. Relazione Ufficiale del XXXVI Congresso Nazionale SIFEL. Acta Phon Lat 2002.

[13] Paul Boersma & David Weenink (2018): Praat: doing phonetics by computer [Computer program].Version 6.0.37, retrieved 28 August 2021 from http://www.praat.org/

[14] Morelli, M. S., Orlandi, S., & Manfredi, C. (2021). BioVoice: A multipurpose tool for voice analysis. Biomedical Signal Processing and Control, 64, 102302.

[15] Manfredi, C., et al. (2015). Automatic assessment of acoustic parameters of the singing voice: application to professional western operatic and jazz singers. Journal of Voice, 29(4), 517-e1.

[16] MATLAB and Statistics and Machine Learning Toolbox Release 2020b. The MathWorks, Inc., Natick, Massachusetts, United States.

[17] Bandini, A., ... & Manfredi, C. (2016). Markerless analysis of articulatory movements in patients with Parkinson's disease. Journal of Voice, 30(6), 766-e1.