Assessing Movement of Articulatory Organs in Patients with Parkinson’s Disease

K. Novotný[1], J. Mekyska[2]

[1]Master Degree Programme (1), FEEC BUT

[2]Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 12, 61600 Brno, Czech Republic

E-mail: 212566@vut.cz, mekyska@vut.cz

Abstract—Hypokinetic dysarthria is a motor speech disorder often present during Parkinson’s disease. It affects the speech system, including articulatory abilities. There are several speech parameters describing this domain, so it is suggested to deal with their mutual comparison. This work aims to design and describe an algorithm for calculating the parameters of articulation, adapted for the Czech language, and then compare their discriminative power. The acoustic analysis of speech included in it is done via the Praat program and basic machine learning algorithms such as Expectation-Maximization, K-means and linear regression are used for the subsequent data processing. The Mann-Whitney U test, descriptive statistics and Random Forest machine learning model using cross-validation and balanced accuracy is used for evaluation. The results are scripts for automatic assessment of vowel space area, for calculating articulation parameters and for their evaluation. The outputs of the analysis of speech recording database prove that differences in articulation can indeed be observed between normal and dysarthric speech. Based on the mutual comparison of results, it is therefore proposed in the work which parameters are being appropriate for further dealing with this issue.

Keywords—Hypokinetic dysarthria, acoustic analysis, speech signal processing, speech parametrization, formant frequencies, articulation, machine learning

1. INTRODUCTION

Motor speech disorders (MSD) are manifested in areas such as respiration, phonation, articulation or prosody. According to [1] 9.3 % of all MSD is a disorder called hypokinetic dysarthria (HD), which occurs in up to 90 % of patients with Parkinson’s disease (PD).

HD is often associated with an articulation disorder (speech is wiped, faded, etc.), which can be quantified using various methods based on formant frequencies [2, 3, 4, 5, 6, 7]. However, these methods are often language dependent and their generalization has not been tested yet.

The aim of this work is to adapt the existing methods for the Czech language and design new algorithms for quantification of articulation based on formant frequencies, and subsequent testing of the ability of these parameters to differentiate healthy controls (HC) and patients with PD.

2. MATERIALS AND METHODS

Database – The input data for this work were audio recordings from the PARCZ database, which contains 100 patients with PD (40 women/60 men; mean age 67.36 ± 8.35) and 52 HC (26 women/26 men; mean age 63.69 ± 9.09). It was recorded at the 1st Department of Neurology, St. Anne’s University Hospital Brno within the project of the Ministry of Health of the Czech Republic no. NT13499.

Formants – The main regions of energy concentration in the sound spectrum of an acoustic signal are called formants. Thanks to the formants, it is possible to distinguish individual vowels – in the Czech language: [a], [ɛ], [i], [o], [u]. The order of speech formants is determined from those at lower frequencies to higher ones. The most mobile is the second formant – F2. This formant corresponds to the oral cavity and is strongly influenced mainly by the position of the tongue and the size of the jaw opening. The second most mobile is the formant of the throat cavity – F1. It is also influenced by tongue, although not to such an extent. This time it is the position of the root of the tongue. The movement of the tongue therefore changes the volume ratio of these two cavities. By plotting the pairs F1 – F2 in the graph, where the frequency F2 is on the *x*-axis and the frequency F1 on the *y*-axis, it is possible to obtain a space with their higher concentration called the Hellwag triangle. An important hypothesis of this work is that when the mobility of articulatory organs deteriorates, there is a centralization in the distribution of formant pairs.

Proposed algorithm – To get a better description of articulatory disorders, it is more appropriate to determine the formants from continuous speech. Therefore, in 2013, an algorithm was developed to automatically calculate the Hellwag triangle [8]. Its advantage is automation, as well as the ability to process a speech signal of any length (assuming that it contains all the necessary vowels) and also considering all the vowels contained in the signal (not only those at the vertices of the triangle according to the classical method).

The steps of its equivalent after optimization for Czech speakers proceed as follows. First, F1 and F2 are calculated frame by frame from the speech signal in places with voiced speech (see Figure 1). Subsequently, outliers are removed based on the Gaussian Mixture Model (GMM). Parameters for GMM calculation are obtained using Expectation-Maximization algorithm (EM). Then, those pairs that do not reach the set limit derived from the average likelihood of all observations are discarded (see Figure 2). From the remaining pairs of formants, the cluster centers belonging to the individual vowels are determined. To do this, the k-means algorithm is used, based on a predetermined number of resulting clusters (*k*), which corresponds to the total number of different vowels. After finding the centers, you can draw their convex hull and get its vertices (see Figure 3).

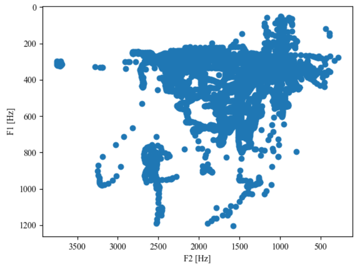

Figure 1: Extracted F1 – F2

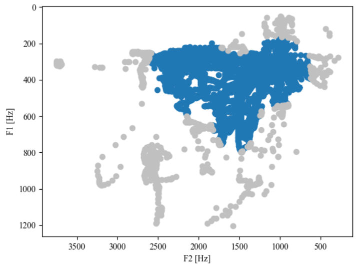

Figure 2: Filtered F1 – F2

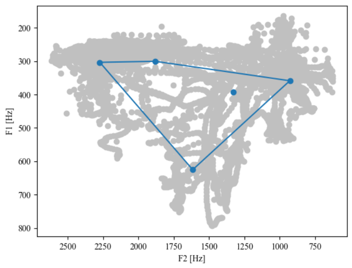

Figure 3: Convex Hull

Subsequently, from these centers (F1 – F2 pairs), those representing the vowels [i], [a] and [u] are selected. Thanks to these three vertices of the imaginary triangle, it is possible to calculate the basic parameters characterizing the mobility of the articulatory organs – Vowel Space Area (VSA), Logarithmic Vowel Space Area (LnVSA) and Vowel Articulation Index (VAI). The Vowel Space Hull Area (VSHA) parameter is calculated from all vertices of the convex hull. For the Articulatory-Acoustic Vowel Space (AAVS) parameter, a matrix of all values after filtration is used, and for our own Density Percentage (DP) parameter (representing the graphic parameter – Vowel Space Density – VSD) and the relative standard deviations of the first and second formant it is the matrix of all values, including those filtered out. A detailed explanation of the properties and calculation procedure for each parameter is described in work [9].

The calculated parameters for all speakers are processed collectively in the next step. First, the effect of age and gender is removed from all results using linear regression. Subsequently, linear regression is used again, but only for speakers with PD. Thanks to it, the effect of Levodopa Equivalent Dose (LED) medication is removed.

Statistical analysis – Descriptive statistics of individual parameters are determined for each group separately. Then all values of one parameter from both groups of speakers undergo a non-parametric Mann-Whitney U test (M-W test). It assesses the statistical significance of the PD effect on a given parameter. The value $\alpha = 0.05$ was chosen as the level of significance.

Machine Learning – To determine the common ability of these parameters to differentiate HC and patients with PD, a Random Forest (RF) machine learning model with fine-tuned (thanks to the grid search method) optimal hyperparameters was used. The comparison of the results was performed after three times repeated 10-fold cross-validation based on the comparison metric – balanced accuracy (BA).

3. RESULTS

Eight parameters were calculated for each speaker in the database and these parameters were consequently evaluated. According to the M-W test, three parameters were identified as statistically significant – VSHA ($p_{\mathrm{VSHA}} = 0.0378$, HC > PN), AAVS ($p_{\mathrm{AAVS}} = 0.0038$, HC > PN) a DP ($p_{\mathrm{DP}} = 0.0138$, HC > PN).

In the best trained RF model, an average BA of 70.2 % was achieved with a sensitivity (SEN) of 89.0 % and a specificity (SPE) of 51.3 %. Complete results for both the descriptive statistics and machine learning with tables and graphs can be found in work [9].

4. DISCUSSION

According to the results of statistical analysis, VSHA and AAVS appear to be the parameters with the highest discrimination power. The fact that the VSA and LnVSA parameters did not appear as good as the VSHA may be evidence of the significance of the size of the entire vowel space area, including the “cropped” areas when plotting the Hellwag vocal triangle. Thus, it is not only the position of the formants in its vertices that is decisive, but the overall distribution of all F1 – F2 pairs. Similar results are reported by the study from 2019 [5]. It achieves the highest resolution for the second mentioned AAVS parameter too. This study also evaluates positively the distinction between PD and HC by observing differences in the size of areas in VSD. It can be said that in this work it represents our DP parameter whose discriminative power was also strong enough. As in studies [6, 7], VSA and LnVSA do not show promising results. The VAI parameter is on the bottom limit of statistical significance. The results confirm the theory that its values are indeed close to one and lower for speakers with PD than for HC. However, this difference is not very significant. A lower VAI in the vowel space area indicates a higher centralization of the F1 – F2 formant pairs. In speech it can be caused by the stiffness of the tongue or the articulatory organs in general. The tested relative deviations of the formants result in a better resolution for the relative standard deviation of the second formant. This one is related to the front-backness of the vowel and is therefore mainly influenced by the mobility of the tongue. Thus, this outcome could support the hypothesis of its stiffness in patients with PD.

The large range between the minimum and maximum of the VSA, LnVSA and VSHA parameters is most likely due to inappropriate filtering of outliers. In general, this block appears as one of the biggest weaknesses in the data processing algorithm. The filtration parameters strongly influence the calculated values of the parameters for assessing the movement of articulatory organs, which consequently turn out to be very sensitive. We can propose a new way of filtering, based on VSD parameter, i.e. DP, which would allow only pairs of F1 – F2 formants from areas with a normalized density of internal distribution higher than a set limit. Imaginary smooth borders could provide a fairer sieve than GMM.

For trained machine learning models, the SEN was in the vast majority of cases higher than the SPE. This may be due to the fact that the database contains more speakers with PD than HC and thus offers better conditions for training the model to recognize them better. The BA result itself may not seem very delightful. However, according to [10] from 2021, only a fraction of people with PD suffer from articulation problems, so in light of this information, the results are not so bad after all.

Apart from the way of filtering the outliers, a small sample size of our dataset is also a limitation of this work. A higher number of speakers would provide both a more objective evaluation of individual parameters and better generalisation of the machine learning models.

5. CONCLUSION

This work had four main goals. The first goal was to design and implement new algorithm for determination of Hellwag vocal triangle from a speech recording. Despite its successful completion, a future work is planned – the limitations of the algorithm and the design of the solution were described above.

The second goal was to use the proposed algorithm to calculate the parameters assessing the movement of articulatory organs. Eight parameters were obtained here. A possible limitation of the simpler ones is the incorrect choice of formant pairs for the vertices of the Hellwag triangle.

The third goal, i.e. the mutual comparison and evaluation of the discrimination power of each individual parameter, was also achieved. Based on descriptive statistics and the M-W test, we can report that VSHA and AAVS have the highest discrimination power.

The last goal was to test the discrimination power of the combination of these parameters. Employing the RF algorithm, we were able to classify PD with BA = 70.2 %.

ACKNOWLEDGMENT

This study was supported by a grant from the Czech Ministry of Health no. NU20-04-00294.

REFERENCES

[1] J. R. Duffy, *Motor Speech Disorders E-Book: Substrates, Differential Diagnosis, and Management*. Mayo Clinic Colege of Medicine, Rochester, Minnesota: Elsevier Health Sciences, 3 ed., 2019.

[2] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech," *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 01, pp. 114–125, 2010.

[3] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Acoustic metrics of vowel articulation in parkinson's disease: Vowel space area (vsa) vs. vowel articulation index (vai)," *Models and Analysis of Vocal Emissions for Biomedical Applications - 7th International Workshop, MAVEBA 2011*, vol. 7, no. 01, pp. 173–175, 2011.

[4] B. Story and K. Bunton, "Vowel space density as an indicator of speech performance," *The Journal of the Acoustical Society of America*, vol. 141, no. 05, pp. EL458–EL464, 2017.

[5] J. Whitfield, "Exploration of metrics for quantifying formant space: Implications for clinical assessment of parkinson disease," *Perspectives of the ASHA Special Interest Groups*, vol. 4, no. 04, pp. 1–9, 2019.

[6] J. Whitfield and A. Goberman, "Articulatory-acoustic vowel space: Application to clear speech in individuals with parkinson disease," *Journal of Communication Disorders*, vol. 51, 09 2014.

[7] J. Whitfield and A. Goberman, "Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech," *International journal of speech-language pathology*, vol. 19, no. 04, pp. 184–194, 2017.

[8] S. Sandoval, V. Berisha, R. Utianski, J. Liss, and A. Spanias, "Automatic assessment of vowel space area," *The Journal of the Acoustical Society of America*, vol. 134, no. 11, pp. EL477–83, 2013.

[9] K. Novotný, "Hodnocení hybnosti mluvidel na základě akustické analýzy řeči," Master's thesis, Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, Brno, 2021.

[10] J. Rusz, T. Tykalova, M. Novotny, D. Zogala, K. Sonka, E. Ruzicka, and P. Dusek, "Defining speech subtypes in de novo parkinson disease," *Neurology*, vol. 97, no. 21, pp. e2124–e2135, 2021.