Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200182

Optimization of Entropy-Based Automated Dyslalia Screening Algorithm

Emilian Erman MAHMUT[a,1], Dorin BERIAN [a], Michele DELLA VENTURA[b] and Vasile STOICU TIVADAR[a]
[a] *Politehnica University Timisoara, Romania*
[b] *Department of Music Technology, Music Academy "Studio Musica", Treviso, Italy*

Abstract. This paper presents the current state of progress of a project aimed at achieving an automated information entropy-based discrimination of phoneme mispronunciations in utterances of early school-age children. The introductory part briefly describes the dyslalia symptomology and the incidence of dyslalic disorders. This section also reviews the current challenges posed by the main research objective in other similar projects sharing the same objective and summarizes the current results thereof. The Material and Method section presents the conditions, the technology and the feature-extraction technique used in the experiment. The same section also describes the computation of the information entropy values of each analyzed speech sample. The highest match rate of 93.33% was achieved in the classification of words containing the phoneme /r/ in the initial position. A synthesis of the achieved results is provided in the Results section based on which conclusions are drawn and exposed in the Discussion and Conclusions section.

Keywords. Dyslalia, Information Entropy, Polynomial Trendline

1. Background

Dyslalia is the most frequent Speech Sound Disorder (SSD) among early-school aged children affecting mainly the phonetic tier of communication and it consists in the mispronunciation (distortion, omission, inversion or substitution) of speech sounds (phonemes). Its prevalence indicator for Romanian children is 13% [1] and its incidence is on a global ascending trend. Unless diagnosed and treated in due time, dyslalic disorders may have a dramatic impact on children's school performance and even lead to behavioral disorders. From a neuroscience perspective dyslalia has been linked to right-brain dominance and left-handedness and roughly a quarter of dyslalic subjects are affected by dyslexia-dysgraphia [2].

Phonological assimilation, which is both progressive and regressive within an utterance, poses a major challenge to the attempts to devise mathematical models for phoneme classification. It follows that the phonetic context of a given target consonant is of utmost importance for rigorous analysis of its articulation.

There are several projects aiming to provide a computerized solution for the discrimination of misarticulated phonemes. Logomon [3] is a complex speech therapy

1 Emilian-Erman MAHMUT, Politehnica University Timisoara, P-ta Victoriei no. 2, Timisoara, Romania; E-mail: emilian.mahmut@aut.upt.ro.

software for Romanian equipped with a fuzzy-logic based expert system dedicated to pronunciation assessment. No specific pronunciation assessment results are provided by the authors, but human intervention (phoneme scores assigned by the Speech Therapists) is massively required to achieve such objective. Intelligent System for Impaired Speech Evaluation [4] focuses on the pronunciation of consonant /r/ in Romanian in an attempt to generate real-time numerical information on mispronunciations by analyzing Mel-Cepstral coefficients extracted from phonemes and by computing, for each item of the unclassified classes (test data), the (Euclidian) distance to each item in the known (learning) data classes (manually selected correct or incorrect pronunciations). The related distances are ordered and the final decision is made by using a K-Nearest Neighbor (KNN) algorithm. This study reports a 78% match rate for a K of 11. Self-Organizing Maps for Identifying Impaired Speech [5] describes an automated method for the discrimination of heavily impaired pronunciations of /r/ in Romanian based on Kohonen neural networks. The algorithm uses the parameters extracted from the alternating component of the signal envelope as feature vectors in a neural network-based classification stage. The study reports an 82.5% match rate with the best mispronunciation classification results within the 1000-1500 training ages range and concludes that mispronounced speech sounds have a tendency to organize on a single neuron while correct pronunciations spread on several map nodes. Vocaliza [6] is a computer-based speech therapy application for Spanish speaking children. Its phoneme mispronunciation module proposes a detection method based on feeding single phonemes (segmented from words uttered by children) to a phonetic decoder (automatic speech recognition algorithm). The speech sound discrimination scenario consists of 3 stages: identifying the phoneme-level confidence measure by using posterior probabilities resulting from a confusion network (CN) generated from the phone lattices (lattice compression algorithm), adapting the acoustic model by using Hidden Markov Models (HMM) trained on a Spanish language audio-record database containing 63193 words (Albayzin) and non-linear mapping of CN posteriors (confidence measures) to improve prediction as compared to the classification performed by non-expert human labelers. The authors conclude that the decision made by an expert (speech pathologist) will invariably have a subjective component. The study reports a match rate of 85.81% for the human labeling tasks. Human intervention is still required to a significant extent in order to reach efficient SSD assessment match rates enabling rigorous detection of phoneme mispronunciations.

The main objective of our research project is the development of a software solution for the fully automated entropy-based screening of dyslalic disorders in early school-age children. The screening system architecture is presented in [7].

2. Material and Method

An observational experiment was conducted on a target population of 30 children aged 5-7 from the Banatean National College in Timisoara (Romania), focusing on the pronunciation of the vibrant lateral consonant /r/ in initial, median and final position within the Romanian words: “RAFT” (shelf), “PARĂ” (pear) and “FAR” (headlight). The audio samples were recorded in the speech therapy practice using an Olympus LSP1 Linear PCM Recorder (PCM 44.1 kHz/16 bit). Each pronunciation was assessed by the speech language pathologist (SLP) so as to be compared with the results of the

screening application. The application was built in C# (Microsoft.NET) using Accord.NET audio processing libraries to display the polynomial trendline based on the positive amplitude values of the audio signal (Figure 1).

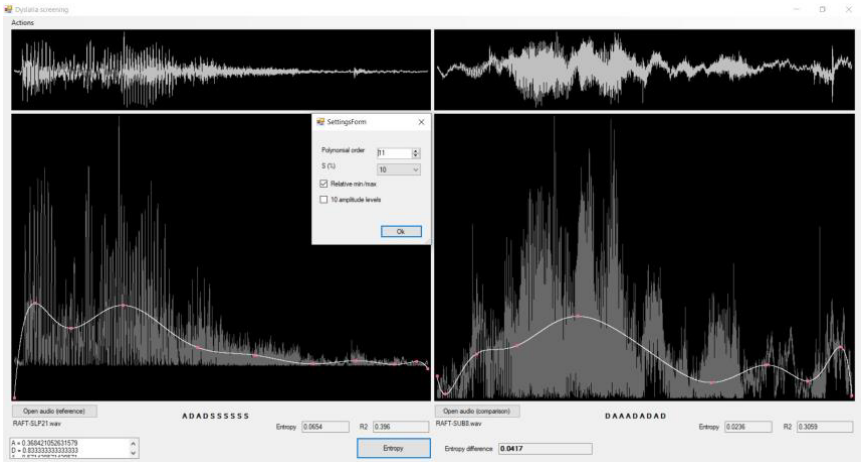

Figure 1. Dyslalia Screening Application

The algorithm extracts the polynomial trendline of a variable order (from 9 to 12) from the audio signal and describes its evolution by assigning 3 possible values to its peaks and troughs: A (ascending), D (descending) or S (stable, i.e. curve variation from previous value remains within a predefined range, in percentages of the max. amplitude value). A transition matrix shown in Figure 2 (Markov process) [8] is then filled in with such values (letters) for the reference signal (SLP's pronunciation whose corresponding polynomial trendline has the highest R-squared value) and the test signal (subject's pronunciation of the same word) taken together, e.g.: SLP's pronunciation and Subject's pronunciation. The probability of each alphabet letter is given by the quotient of the number of occurrences of the specific letter $n+1$ given another previous letter n (numerator) and the total number of occurrences of the specific letter in the 2 segments combined (denominator) [9]. The corresponding information entropy values for each segment are then calculated based on the probabilities provided by the transition matrix and compared. To have a drastic assessment of the test signal, the similarity threshold was set to a maximum value of 0.00099, therefore the test-signal entropy values belonging to the [0-0.00099] range were deemed to match the reference-signal entropy value, whereas test-signal entropy values of higher magnitude orders (i.e. in the [0.001 – 1] range) were labeled as mismatch (false positive or false negative cases).

	A	D	S
A	2	5	0
D	4	0	1
S	0	1	5

Alphabet	Total
A	7
D	6
S	6
	19

Figure 2. Transition matrix and alphabet

3. Results

Calculations were made based on the algorithm described above for 1440 values in all possible configurations (combinations of polynomial orders and S-letter ranges).

Table 1. Highest match rates for initial /r/

Initial /r/	Poly 9-S 15	Poly 10-S 2	Poly 10-S 5	Poly 11-S 10	Poly 12-S 15	SLP Opinion
Subject 1	-0.0099	-0.0614	-0.0125	0.0163	0.0049	Pararhotacism (/iaf-/)
Subject 2	0.0043	-0.0194	-0.0194	0.0102	0.0019	Rhotacism (/Raft/)
Subject 3	0.0057	-0.0617	-0.0703	-0.0218	-0.0137	Polymorphic (/rarf/)
…						
Subject 30	0.0019	-0.1449	-0.018	0.0341	0.0086	Rhotacism (/Ra:ft/)
Match Rate	86.66%	90%	90%	**93.33%**	86.66%	

The resulting match rate was compared to the SLP's opinion on each analyzed utterance. The highest match rates achieved were: 93.3% for initial /r/, 80% for median /r/ and 83.3% for final /r/ as shown in Figures 3 and 4.

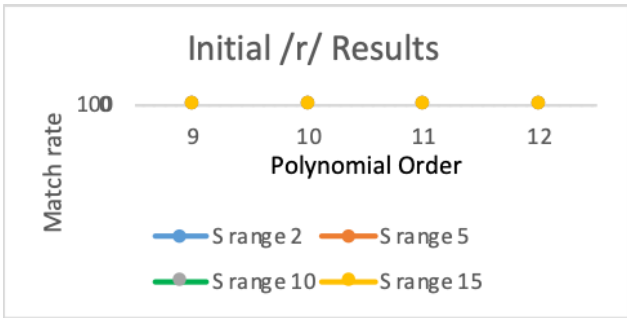

Figure 3. Initial /r/ results.

The highest match rate for the pronunciation containing the phoneme /r/ in the initial position (93.33%) was obtained using a polynomial order of 11 and an S-letter range of 10, while the best results for median /r/ (80%) were achieved using a polynomial order of 10 and an S-letter range of 5.

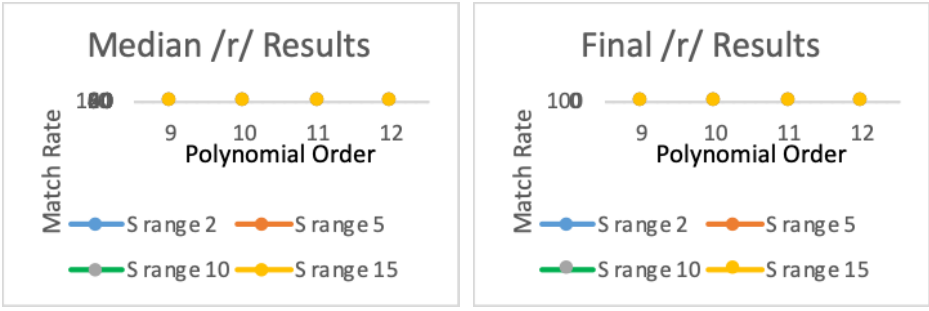

Figure 4. a) Median /r/ results **b)** Final /r/ results

For final /r/ the highest match rate (83.33%) was obtained by setting the polynomial order to 10 and the S-letter range to 10 or 15. It can be inferred from the synthetic diagrams (Figures 3 and 4) that an S-letter range of 2 produces the most stable results across the polynomial order range [9-12] in all three cases. The S-letter range of 15 generates the largest variation in the initial and final /r/ cases, while it remains fairly stable for the median /r/ case.

4. Discussion and Conclusions

The application computes and stores the R-squared value for each polynomial trendline to measure the goodness of fit of the model. Further analysis of such stored values is required to establish useful correlations. Higher polynomial orders (12 and above) generate poor match rates (33.33% to 70%) across the 3 cases, i.e. initial, median and final /r/.

The algorithm produced better results than previous studies and requires little human intervention.

Future work will focus on optimizing the audio signal quality (noise filtering) and segmentation by performing a normalized correlation of the pair of analyzed signals and using logatomes (monosyllabic pseudowords) with specific structures in terms of vowels (V) and consonants (C): VCV/CVC/CVV/VVC, where C stands for the target consonant. The inflection points of the polynomial trendline will also be considered for assignment of additional letters or sub-letters to be used in the transition matrix, by computation of the first and second derivatives.

Formant frequencies of the analyzed signals will also be explored to assess the possibility of devising alternative means to validate the segmentation described above.

References

[1] E. Verza, Tratat de logopedie, Editura Semne, Bucuresti, Romania 2000.
[2] J. Law, J. Boyle, F. Harris, A. Harkness and C. Nye, Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature, International Journal of Language and Communication Disorders 2 (2000).
[3] S.Gh. Pentiuc, O. A. Schipor, M. Danubianu, D.M. Schipor, Automatic Recognition of Dyslalia Affecting Pre-scholars, arXiv:1406.0495v1 [cs.CY] (2014).
[4] O. Grigore, C. Grigore, V. Velican, Intelligent System for Impaired Speech Evaluation, Recent
[5] Advances in Circuits, Systems and Signals, Book Series: International Conference on Circuits Systems Signals (2010).
[6] O. Grigore and V. Velican, Self-Organizing Maps For Identifying Impaired Speech, Advances in
[7] Electrical and Computer Engineering (AECE) 3 (2011).
[8] S.C. Yin, R. Rose, O. Saz and E. Lleida, A study of pronunciation verification in a Speech
[9] Therapy Application, IEEE International Conference on Acoustics, Speech and Signal Processing (2009).
[10] E.E. Mahmut, S. Nicola, V. Stoicu-Tivadar, A Computer-based Speech Sound Disorder Screening System Architecture, Proceedings of the 17th International Conference on Informatics, Management and Technology in Healthcare (ICIMTH) (2018).
[11] S. Meyn and R.L. Tweedie, Markov Chains and Stochastic Stability, Second Edition, Cambridge University Press, 2009.
[12] M. Della Ventura, The Influence of the Rhythm with the Pitch on Melodic Segmentation, Proceedings of the Second Euro-China Conference on Intelligent Data Analysis and Applications (Springer) (2015).