Applied Medical Informatics

Original Research Vol. 40, No. 1-2 /2018, pp: 15-23

An Entropy-Based Computer Model for the Measurement of Phonetic Similarity: Dyslalia Screening in Early School-Age Children

Emilian Erman MAHMUT[1,*], Michele DELLA VENTURA[2], and Vasile STOICU-TIVADAR[3]

[1] Faculty of Automation and Computers, Politehnica University Timişoara, 2 Vasile Pârvan Blvd., 300223, Timişoara, Romania.
[2] Department of Technology, Music Academy "Studio Musica", 25 Andrea Gritti Str., 31100 Treviso, Italy.
[3] Faculty of Automation and Computers, Politehnica University Timişoara, 2 Vasile Pârvan Blvd., 300223, Timişoara, Romania.
E-mails: emilian.mahmut@aut.upt.ro; dellaventura.michele@tin.it; vasile.stoicu-tivadar@aut.upt.ro

* Author to whom correspondence should be addressed; Tel.: +4-0720-651-951

Received: December 11, 2017 /Accepted: June 22, 2018 / Published online: June 30, 2018

Abstract

This paper presents a computer model for the assessment of the similarity between two sound patterns, to identify phoneme mispronunciations circumscribed by dyslalic disorders in early school-age children (6-10 year olds). From a linguistic standpoint, it is the phonetic tier that is mainly engaged in dyslalia. Unlike other speech disorders, which involve meaning-coding and decoding mechanisms (semantics), dyslalia lends itself more easily to mathematical analysis in the screening stage. The method is based on the analysis of the sound waves and on the quantification of the information carried by every single sound pattern, by calculating its entropy. It is an empirical methodology that provides results that may be analyzed. An experimental study was conducted according to the model and method presented on a sample of 30 subjects. The results are assessed and conclusions are issued. The representation using an isometric diagram accommodates a better interpretation of the results.

Keywords: Entropy; Phonetic similarity; Dyslalia; Soundwave; Markov process

Introduction

The most frequent speech disorder among early school-age children is dyslalia, a pronunciation disorder caused by organic or functional problems with the peripheral organs of speech. The mispronunciations circumscribed by dyslalia consist of distorting, substituting, omitting and/or inverting speech sounds in articulation. Vowels /a/, /e/, /u/ and consonants /b/, /d/, /t/, /m/, /n/ are usually less affected and difficulties in their pronunciation are corrected more easily. Affricates /ʦ/, /ʧ/, /ʤ/, most likely due to their complex articulation (two-phase articulation starting as occlusion and ending as constriction), occur much later in children's pronunciation and are often affected. Sibilant consonants /ʃ/ - /ʒ/ (Romanian ş - j) as well as hissing consonants /s/- /z/ are in the same situation. The vibrant consonant /r/ also occurs at a later time, after the sibilant and hissing consonants, and it is often mispronounced, being omitted (rhotacism) or substituted (pararhotacism). Following a survey conducted on primary and middle school children in Romania, Verza reported that 13% of investigated children are affected by language disorders, which is considered by the author an incidence indicator for such disorders in Romanian language speakers [1]. The targeted age range of the study (6-10 year-olds) excludes cases of physiological dyslalia (developmental delays or insufficient development of the phono-articulatory apparatus), which do not usually persist beyond the age of 5. The existing applications used in speech therapy focus rather on stimulating and strengthening or instilling correct language segments, in relation to the specific language disorder being approached, and less on assessing the symptoms (screening).

A study published in AECE (Advances in Electrical and Computer Engineering) in 2011 presents a method for the identification of the mispronunciation of consonant /r/ in Romanian using Kohonen networks [2]. Another interesting study by a group of researchers of the Stefan cel Mare University of Suceava and the Alexandru Ioan Cuza University of Iasi aimed at creating and implementing an intelligent CBT (Cognitive Behavioral Therapy) system for the therapy of dyslalic disorders as a complementary, personalized patient-oriented speech therapy method. *Logomon*, the software application developed on this study, includes a complex examination, by collecting data related to the child, a general therapy framework (mobility development, airflow control, hearing development) and a specific therapy framework (achieving and consolidating correct sound pronunciation) [3]. In the European Union, a series of research projects have been initiated within the interface between speech technology and its practical applications in speech therapy [4-6]. The model proposed by this study aims to present a speech analysis methodology, based on the evaluation of the information that each message carries. Our study had two objectives. The first objective was to identify a distinct ranges of values of the information content, by determining a minimum value and a maximum value, deducible from the information values of the single changes of direction of the sound wave. The second objective was to check, for every single peak of the two sound patterns being compared, what range they identify within: the value of the first peak of the standard pattern shall be compared with the value of the first peak of the sample pattern, and so on. If the information value of the two peaks falls within the same range, they share an equality/similarity. This procedure is due to the fact that, in accordance with the Information Theory, described below, the information value depends on the alphabet of the message that determines the transition probabilities between the system states (*conditional probability*): the probability, in our case, for a peak of the standard pattern to resolve to a peak of the sample pattern.

Material and Method

Background

The analysis based on Information Theory considers the audio message as a linear process endowed with a syntax formulated not by preset rules but on the probability of occurrence of each element of the audio message in relation to the element preceding it [7,8]. The speech "units of meaning" coincide with the minimum elements of an audio message considering the definition of a "message" as a chain of discontinuous speech "units of meaning" [7]. Any element of a chain built in this fashion requires a prevision in relation to the element that will follow it [9,10]. In a communication happening using a given alphabet of symbols, the information is associated with every single transmitted symbol [7]. Information, therefore, may be defined as *the reduction of the uncertainty that might have been, a priori, present on the transmitted symbol.* The ampler the range of messages that the source may transmit (and the larger the uncertainty of the receiver about the possible message), the larger the quantity of transmitted information is – and, along with it, its own measure: *the entropy* [8]. In the Information Theory, the *entropy* measures the quantity of uncertainty or information existing in a random signal. If every message has the probability p_i of being transmitted, the entropy is obtained as the sum of all the set of functions $p_i \log_2 p_i$, each of them being related to a message, i.e.:

$$H(X) = E[I(x_i)] = \sum_{i=1}^{n} I(x_i) \times P(x_i) = \sum_{i=1}^{n} P(x_i) \times log_2 \frac{1}{P(x_i)} \tag{1}$$

Description of the Sample Population Selected for Audio Recordings

A speech therapy worksheet was given to 30 subjects aged 6 (from infant classes within the Banatean National College of Timişoara) affected by dyslalic disorders (mainly rhotacism) from

November 2016 to May 2017. The children were asked to pronounce a series of Romanian words containing target phonemes in the initial, medial and final position and audio recordings were made. The experiment focused on the pronunciation of consonant /r/ given its high level of acoustic intensity oscillations that enables clear discrimination of the amplitude values [11].

Audio Recording Equipment and Sound Wave Analysis Software

An Olympus LSP1 Linear PCM Recorder (PCM 44,1 kHz/16 bit) was used for the observational experiment. The sampling rate of 44.1 kHz is considered a *de facto* standard of the audio recording equipment. The human ear is sensitive to air vibrations with frequencies within the 20 Hz - 20 kHz range, therefore the sampling rate must be higher than 40 kHz (Nyquist-Shannon sampling theorem) [12]. The amplitude diagrams above were obtained using the *Audacity 1.3 Beta* (Unicode) open-source application [13].

Phoneme Analysis

To compare two audio messages (sound patterns), respectively the standard pronunciation of a certain phoneme (hereinafter referred to as standard pattern) and the pronunciation of the same phoneme by the child (hereinafter referred to as the sample pattern), the algorithm must calculate the information content of each of them (therefore, the entropy) and then display the results in an isometric diagram. Figure 1 below illustrates the data flow diagram of the algorithm.

The first stage of the analysis process consists in checking the duration of each sound pattern to make them homogeneous for the analysis. If the duration of the sample file is shorter or longer than the duration of the standard pattern, the former is automatically readapted: in this stage where the duration of a digital file is changed, its form does not change. Both patterns are then segmented [14] to calculate the entropy and, therefore, the information they contain. In this stage, the algorithm identifies the peak values of each sound wave, based on the direction changes (from ascending to descending and vice versa). To calculate the entropy, we must refer to a specific alphabet: the alphabet is language-specific [15] and, as it may be immediately inferred from the formula (based on the probability of certain symbols rather than other symbols to be transmitted) it proves to be associated to language. In case of an audio file, the peak values of the sound wave were considered to be symbols of the alphabet [14,15]. The peak value is (in digital audio), the point where the sound wave assumes its maximum amplitude value, before changing its direction (from ascending to descending and vice versa) (see Figure 2). The amplitude may assume a value ranging from 0 (zero) to 1 or from 0 (zero) to -1.

For every single sound pattern a table, which represents its alphabet, will be filled in (see Table 1 below). For an easy operation, a decision was made to approximate the peak values read by the algorithm to reduce the number of symbols. If, for instance, a peak value is 0.67 it is approximated to 0.7.

Determining the alphabet thus identified is not enough to calculate the entropy of a sound pattern: it is necessary to consider the manner in which the peak values succeed one another within the sound pattern. The Markov process (or Markov's stochastic process) is used: we chose to deduce the transition probability that determines the passage from a state of the system to the next uniquely from the immediately preceding state [16,17]. Therefore, the transition matrix is created by the transition probabilities between the states of the system (conditional probability) [16-19]. In our case, the matrix represents the probabilities for a peak value to resolve to another peak value and is presented in Table 2.

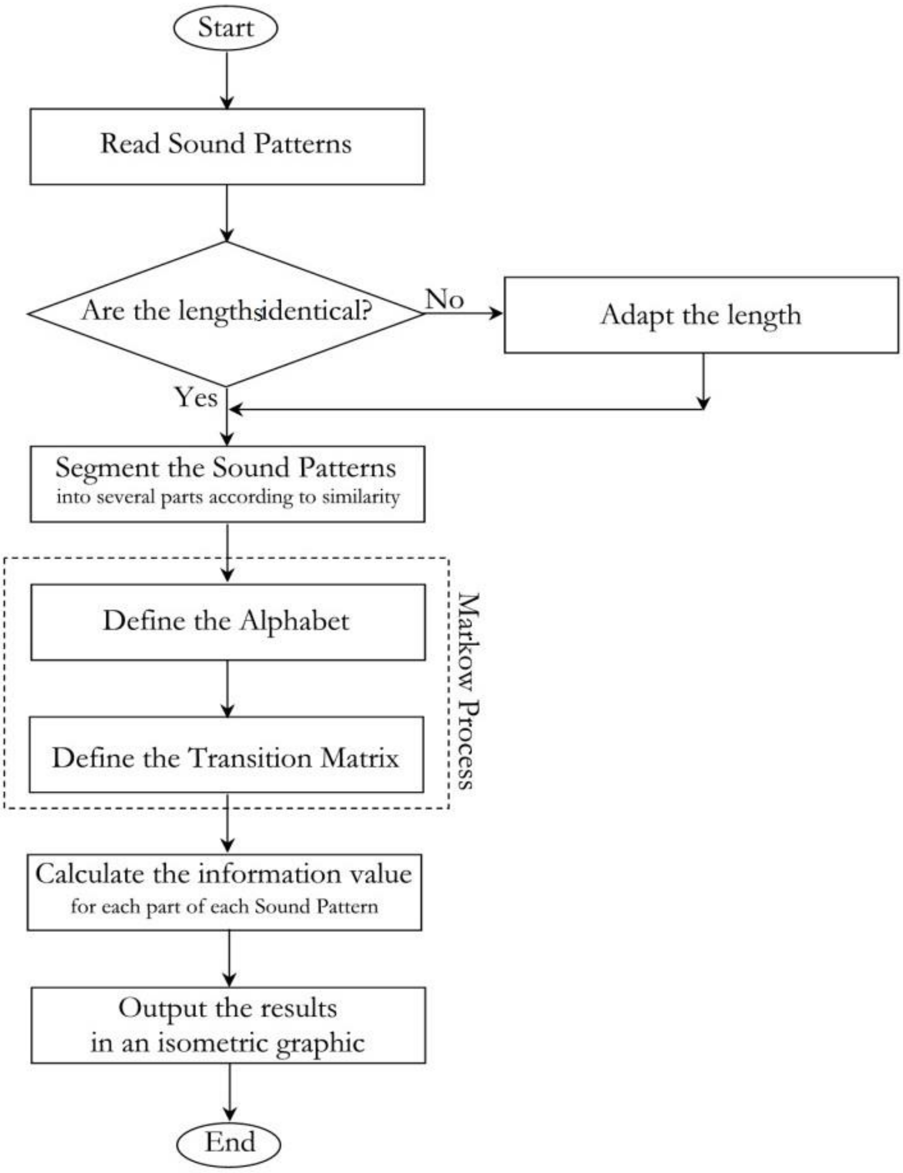

Start
Read Sound Patterns
Are the lengthsidentical?
No
Adapt the length
Yes
Segment the Sound Patterns
into several parts according to similarity
Define the Alphabet
Define the Transition Matrix
Markow Process
Calculate the information value
for each part of each Sound Pattern
Output the results
in an isometric graphic
End

Figure 1. Data flow diagram

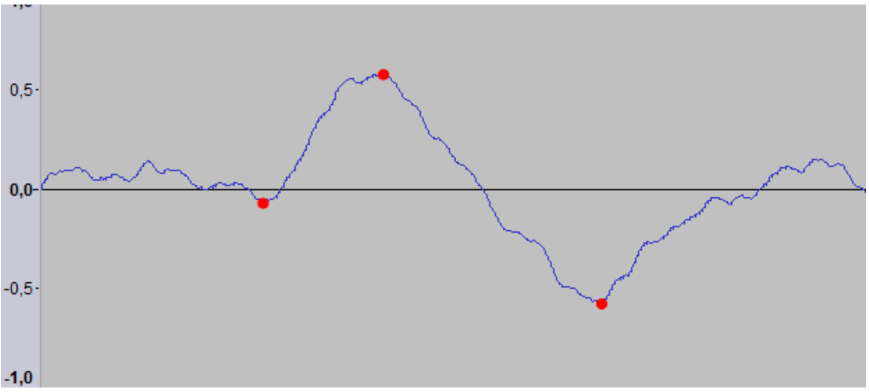
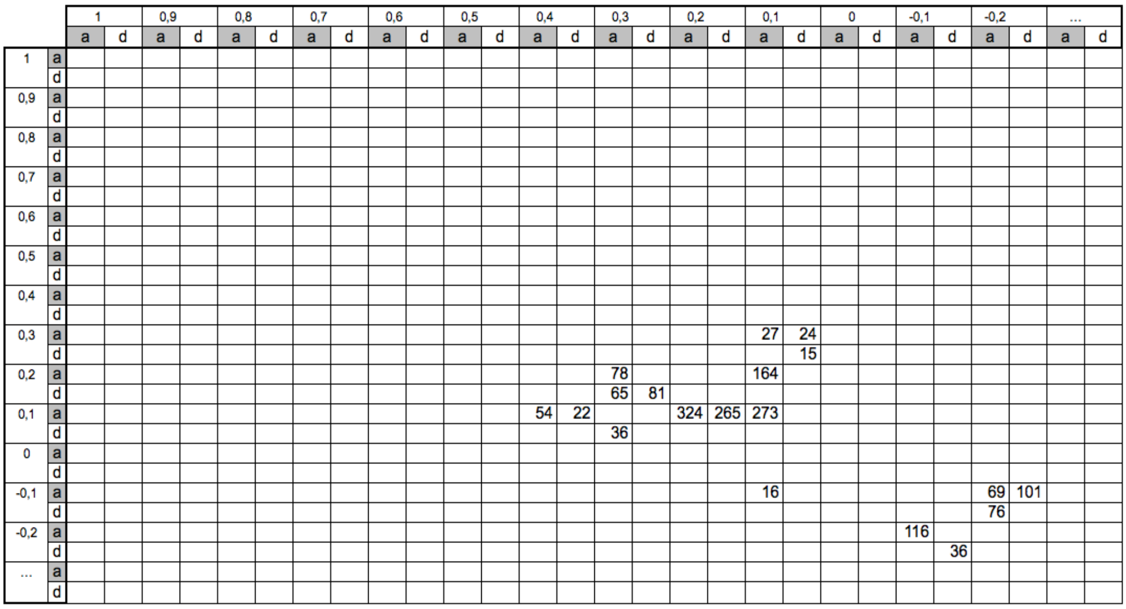


Figure 2. Peak values of the sound wave

Table 1. Example of alphabet

Peak value	Number
1.0	
0.9	
0.8	
0.7	
0.6	
0.5	
0.4	
0.3	
0.2	
0.1	
0.0	
-0.1	
-0.2	
-0.3	
-0.4	
-0.5	
-0.6	
-0.7	
-0.8	
-0.9	
-1.0	

Table 2. Transition matrix is drawn on the sound segment of Table 1



		1		0,9		0,8		0,7		0,6		0,5		0,4		0,3		0,2		0,1		0		-0,1		-0,2		...	
		a	d	a	d	a	d	a	d	a	d	a	d	a	d	a	d	a	d	a	d	a	d	a	d	a	d	a	d
1	a																												
	d																												
0,9	a																												
	d																												
0,8	a																												
	d																												
0,7	a																												
	d																												
0,6	a																												
	d																												
0,5	a																												
	d																												
0,4	a																												
	d																												
0,3	a																			27	24								
	d																				15								
0,2	a															78				164									
	d															65	81												
0,1	a													54	22			324	265	273									
	d															36													
0	a																												
	d																												
-0,1	a																			16						69	101		
	d																									76			
-0,2	a																							116					
	d																								36				
...	a																												
	d																												

The model of analysis described in this article was verified by elaborating an algorithm, the structure of which takes into consideration every single aspect described above. The results of the analysis are illustrated using isometric diagrams that allow an immediate visualization for interpretative purposes. The algorithm does not provide any limitation concerning the dimensions of the table representing the alphabet and the matrix of transitions that will be automatically dimensioned in every single analysis by the characteristics of the analyzed sound segment. This confers generality to the algorithm and specificity of every single analysis (Strength). The algorithm performs the analysis stages considering the two (standard and sample) sound patterns as if they were one single sound pattern: the two sound waves are viewed one after the other (Figure 3) in order to determine the alphabet, the transition table and then to quantify the information value of every sound by calculating the entropy.

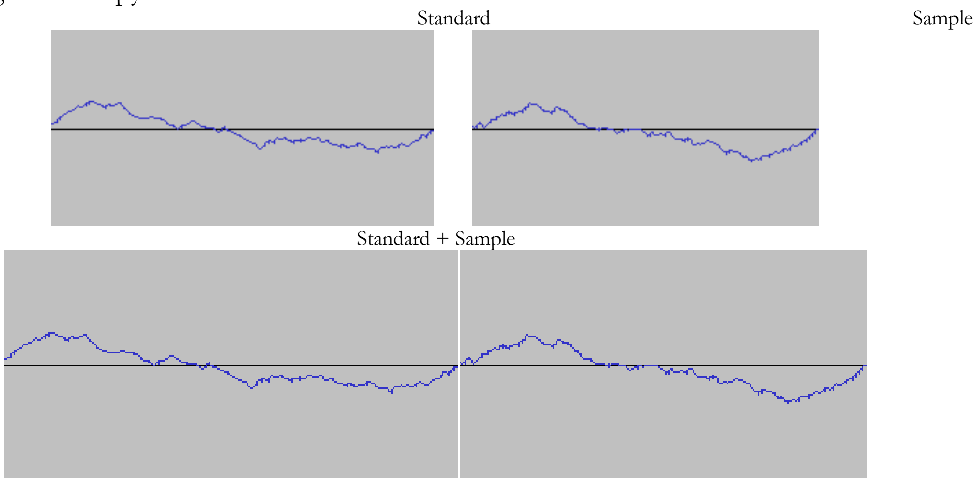


Figure 3. Representation of the sound patterns

Results and Discussion

The results of the comparison between the two sound segments are shown in Figure 4 to explain how the data is interpreted.

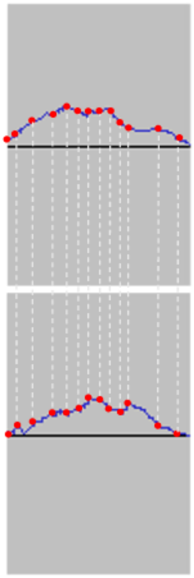


Figure 4. Segmentation and identification of the peaks (the initial part of the standard and sample patterns)

Once the various tables (alphabet and transitions) are defined, the information value of every peak is calculated, and then the information intervals are identified, each of them representing a certain range of values. The obtained results are presented in Table 3.

Figure 5 illustrates the diagram of the analysis of the sound patterns in Figure 4. The upper part of the diagram shows the various ranges of information of every single peak of the standard segment through colors, while the lower part of the diagram shows the various ranges of information of every single peak of the sample segment. The representative color of a certain band will have a darker or a lighter shade based on the higher or lower information value of the identifying range within the same band. In case two peaks have the same information value, the diagram displays a column with a single color of the related information range. If, instead, two values differ from each other, the color within the column changes by fading out, going from the color of the preceding value to the current value. The larger this color difference is, the lower the equality or the similarity between the two segments is.

Table 3. The information value of every peak

Standard pattern	Sample pattern
0.052	0.032
0.109	0.125
0.203	0.123
0.219	0.203
0.284	0.214
0.258	0.232
0.261	0.0261
0.253	0.248
0.252	0.234
0.184	0.221
0.168	0.240
0.165	0.162
0.121	0.341

Band 1	Band 2	Band 3
[0-0.1]	(0.1-0.2]	(0.2-0.3]

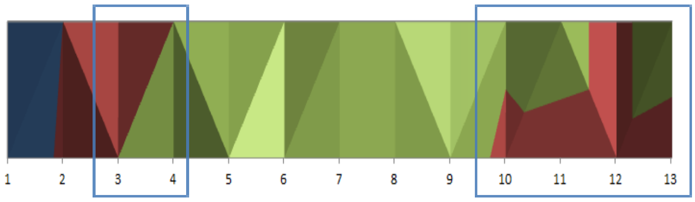


Figure 5. Identification of the dissimilarity between two segments

The two segments share a clear similarity. The color columns of every single peak belong to the same range even if they have different values (see the shades of green or of red), except for the highlighted points where the peaks of the sample pattern differ from the peaks of the standard pattern, even though moving gradually toward the range of the standard pattern. These points allow the discrimination of a pronunciation mismatch.

The algorithm was initially conceived to analyze the entropy of music scores [16]. This raised questions about its structural and functional compatibility in relation with the new type of input. Music and language share vibration as a way to transmit the message, they use a finite set of minimal elements (notes and phonemes) as *substance-energy carriers* [20] for the selection from the infinite continuum of *sema* and, as substantiated by neurophysiology research, depend on the same brain areas and mechanisms to process the message.

From what has been discussed above, it is easy to infer that for a more detailed and rigorous analysis it will be necessary to:

1. Take into consideration the largest possible number of peaks for every sound pattern;

2. Take into consideration a larger number of identifying levels of the sound-wave amplitude (see Table 1): for instance, 1, 0.95, 0.9, 0.85, …..;
3. Increase the number of decimals representing the entropy value: this allows for the management of a larger number of bands (see Table 3), thus obtaining a more accurate analysis about the differences of the pronunciation of a certain phoneme. This paper considered the possibility to compare two sound segments on a symbolic level, by measuring the information content of every sound. The approach proposed by the paper aimed at representing the relationships between the sound segments in an isometric diagram, built by the entropy concept. This allowed the identification of a *potential equality, similarity, and homology between the analyzed segments.*

The typical instrument which offers a synthetic description of the complex set of problems of language disorders and/or a matrix of the language-use level is the speech therapy evaluation form. Performing a traditional screening (using pen and paper and the numerous observation sheets, forms, and other speech therapy materials) on 25-30 children (a school class or a kindergarten group), within a limited time, may be overwhelming for the speech therapist. Moreover, there are significant risks concerning the objectivity, completeness, and accuracy of the information collected by the speech therapist. The shortage of specialists in the urban areas and their total absence in rural areas raise the issue of time-saving/efficiency in relation with this initial activity meant to detect potential patients. Furthermore, the global trend of dyslalic disorders is ascending, with 25-30% of dyslalic children being also affected by dyslexia-dysgraphia. Speech Pathology research also associates dyslalic disorders and rhythm disorders (stuttering) with left-handedness and right-brain dominance [1]. The initial detection stage is extremely important as much as it is unanimously accepted that early interventions have the highest rate of success in language-disorder treatment. An automated screening incorporated in a software application, drawing input from the audio recordings of the administered speech tests has clear advantages regarding accuracy of the collected data, mobility and time required to carry out the entire process, patient confidentiality and motivation, possibility to generate statistics.

Acknowledgments

Special thanks to Mrs. Mirela Viziteu, the speech therapist of the Banatean National College in Timişoara (Romania), for the heartfelt collaboration and openness demonstrated in support of this study.

References

1. Verza E. Capitolul VI. Tulburările de pronuntie de tip dislalic. In: Tratat de logopedie. Bucuresti: Editura Semne; 2000, pp. 124-147.
2. Grigore O, Velican V. Self-Organizing Maps For Identifying Impaired Speech, Politehnica University of Bucharest, 061071, Romania, Department of Applied Electronics and Telecommunications Faculty [Internet]. 2011 [accessed 2017 September 22]. Available from: http://www.aece.ro/abstractplus.php?year=2011&number=3&article=7
3. Pentiuc ŞG, Tobolcea I, Schipor O A, Danubianu M, Schipor MD. Translation of the Speech Therapy Programs in the Logomon Assisted Therapy System [Internet]. 2010 [accessed 2017 September 22]. Available from URL: http://www.aece.ro/abstractplus.php?year=2010&number=2&article=8
4. Ortho-logo-paedia Project coordinated by the Institute for Language and Speech Processing of Athens, Greece [Internet]. [accessed 2016 August 15]. Available from: http://www.xanthi.ilsp.gr/olp/
5. Isaeus Project coordinated by E.T.S.I. Telecomunicaciones Consortium of Madrid, Spain [Internet]. [accessed 2016 August 15]. Available from: http://cordis.europa.eu/project/rcn/35120_en.html

6. Speco Project coordinated by the Technical University of Budapest [Internet]. [accessed 2016 August 15]. Available from: http://alpha.tmit.bme.hu/speech/paperFON99.php
7. Weaver W, Shannon C. The mathematical theory of information. Urbana: Illinois Press; 1964.
8. Angeleri E. Information, meaning, and universalit. Turin: UTET; 2000.
9. Moles A. Teorie de l'information et Perception esthetique. Paris: Flammarion Editeur; 1958.
10. Lerdhal F, Jackendoff R. A Grammatical Parallel between Music and Language. New York: Plenum Press; 1982.
11. Mahmut EE, Della Ventura M. Prototip de aplicaţie pentru screeningul dislaliei la copiii de vârstă şcolară mică (6-10 ani). 11th Eastern and Central European Regional Conference For Augmentative And Alternative Communication Proc., Bucharest; 4-6 July 2017.
12. The Nyquist-Shannon Sampling Theorem [Internet]. [accessed 2017 September 22]. Available from: https://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem
13. Audacity 1.3 Beta (Unicode) [Internet]. [open-source, accessed 2016 May 21]. Available from: http://audacity.sourceforge.net
14. Della Ventura M. Analysis of algorithms' implementation for melodic operators in symbolical textual segmentation and connected evaluation of musical entropy. Proc. 1st Models and Methods in Applied Sciences, Drobeta Turnu Severin; 2011, pp. 66-73.
15. Monelle R. Linguistics and Semiotics in Music. Harwood Academic Publisher; 1992.
16. Della Ventura M. Rhythm analysis of the sonorous continuum and conjoint evaluation of the musical entropy. Proc. Latest Advances in Acoustics and Music; Iasi; 2012, pp. 16-21.
17. Nattiez JJ. Fondements d'une sémiologie de la musique. Paris: Union Générale d'Éditions; 1975.
18. Lemstrom K. Towards More Robust Geometric Content-Based Music Retrieval. Proceedings of the Conference of the International Society for Music Information Retrieval, 2010, pp. 577-582.
19. Urbano J. A Geometric Model Supported with Hybrid Sequence Alignment. Proceedings of the Annual Music Information Retrieval Evaluation Exchange [Internet]; 2013 [accessed 2017 September 22]. Available from: music-ir.org/mirex/abstracts/2013/JU1.pdf.
20. Golu M. Capitolul XI Limbajul. In: Bazele Psihologiei Generale. Ed. a 2-a Bucureşti: Editura Universitară; 2005, pp. 488-520.