Automatic Recognition of Improperly Pronounced Initial 'r' Consonant in Romanian

Valentin VELICAN, Rodica STRUNGARU, Ovidiu GRIGORE
Politehnica University of Bucharest, 720229, Romania
ovidiu.grigore@upb.ro

***Abstract*—Correctly assessing the degree of mispronunciation and deciding upon the necessary treatment are fundamental activities for all speech disorder specialists. Obviously, the experience and the availability of the specialists are essentials in order to assure an efficient therapy for the speech impaired. To overcome this deficiency a more objective approach would include the existence of a tool that independent of the specialist's abilities could be used to establish the diagnostics. A complete automated system based on speech processing algorithms capable of performing the recognition task is therefore thoroughly justified and can be viewed as a goal that will bring many benefits to the field of speech pronunciation correction. This paper presents further results of the authors' work on developing speech processing algorithms able to identify mispronunciations in Romanian language, more exactly we propose the use of the Walsh-Hadamard Transform (WHT) as feature selection tool in the case of identifying rhotacism. The results are encouraging with a best recognition rate of 92.55%.**

***Index Terms*—speech processing, pronunciation evaluation.**

I. Introduction

Automatic speech recognition (ASR) is a widely developed field with many applications spanning from telephony, military, robotics to video-games or transcriptions. Yet the most common shortage of the existing algorithms used in ASR is their inability to perform well in cases when the actual speech to be recognized is affected by noise or is badly pronounced. Of interest for this paper is the latter case when the speaker has an inability to correctly reproduce the spoken word and is required to pass different stages of speech therapy.

There are different known attempts of research teams that surveyed the performance of existing techniques when applied to impaired speech, designed algorithms capable of evaluating the level of pronunciation or even have actually built a system capable of automatic pronunciation evaluation [1-5], but there is still much work to be done in this domain for the simple fact that pronunciation rules, and as a consequence the common mistakes and/or accepted mispronunciations, differ from language to language. For Romanian, to our knowledge and up to this moment, only two research teams worked on such projects [6-10] plus the authors of this paper [11-15].

The importance of adjusting ASR systems in order to identify speech problems or if necessary designing new such techniques is proven by the potential high demand of such tools in the field of speech therapy. Mispronunciations are most frequently encountered at children of elementary school age and are identified through screening activities in schools or when the parents decide to visit a specialized cabinet in search for advice. [16] ASR systems capable of identifying impaired speech can be very useful in such cases as they could relief the specialists of some of the workload of therapy and at the same time could be a very interesting at home-practicing means for the child.

This paper presents further results of our research team in the case of identifying rhotacism (mispronunciation of /r/ consonant) as initial phoeneme in spoken words.

II. Dyslalia, a Brief Overview

Statistically, the most extent category of language disorders is represented by pronunciation disorders or dyslalia. Dyslalia is the language disorder characterized by mispronunciations of sounds or group of sounds [17]. This is further classified in *simple* dyslalia - only isolated sounds are affected or a fundamental category of sounds is affected – *complex* dyslalia – when mispronunciations are extended, severely affecting the pronunciation – and *total* dyslalia – when the pronunciation is unintelligible. [18]

In Romanian, one of the most common examples of speech difficulties is the *rhotacsim* - mispronunciation of /r/ consonant. Rhotacism is a case of *simple dyslalia*, but for specialist is of great importance as rhotacsim in some cases can be a clue of more severe disorders. [19] A similar simple dyslalia example and also very common for Romanian is the so called *para-rhotacsim*. In para-rhotacism the affected pronunciation is characterized by the fact that a /r/ phoneme is being replaced by another sound.

More examples of speech difficulties may be reminded (mispronunciations of /s/ or /z/ - *sigmatism*, mispronunciations of /c/, /g/, etc.) yet before moving on it is important to emphasize that the informational pattern that will characterize every such affected sound will be dependent on the speech disorder and on the position of the affected sound, potentially making the recognition algorithm very complex - build around a numerous amount of dedicated recognition tools (for every instance of speech difficulty).

Therefore, in our work we decided to investigate at first only the case of rhotacism with initial /r/, maybe the most common example of dyslalia in Romanian.

III. Algorithm Overview and Speech Database

As with most pattern recognition applications, the algorithm presented in this paper relies on the same general principles: A test database comprising of unlabeled (correct / incorrect) recordings of speech is being fed as an input to a flow of tasks. These processing rules, for every recording in

Digital Object Identifier 10.4316/AECE.2012.03012

1582-7445 © 2012 AECE

the test database gather and then select relevant informational features (from within the speech sample) that will help in the classification / decision stage. At this latter step an additional reference database is used - all the recordings in the reference database are labeled and serve as models in the process of deciding to which class belongs every recording in the test database. (Fig.1)

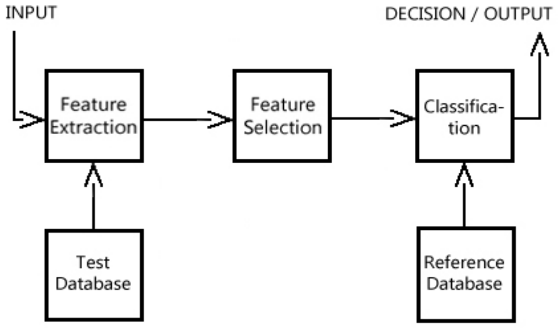

Figure 1. Processing Flow of the Recognition Algorithm. An input phoneme passes the feature extraction and feature selection stages before it is fed to the classifier. The output is a decision regarding the class in which the current input belongs: correct / incorrect (pronunciation).

All the recordings used in this work were gathered with help from personnel in Centrul Logopedic nr. 8. Overall 100 incorrect and 44 correct occurrences of /r/ as initial phoneme in words were used. These phoneme were selected from voice recordings of 10 children pronouncing correctly and 30 children with rhotacism / pararhotacism. The segmentation of the desired sounds was performed manually, concentrating on correctly separating the /r/ of the subsequent vowel. It is also worth mentioning that all the children involved were of elementary school age: 6 - 8 years old thus making the not only the timbre but also the pitch of the sound to resemble more from child to child.

IV. Feature Extraction and Selection

Voice is considered to be a stationary signal only on small time intervals of about 15ms - 30ms. It is therefore important to analyze the sounds in the database (which are considerably longer than the time-frames required) on overlapped windows and for every such recording to concentrate the relevant information into one single object that uniquely identifies it.

In speech processing, voiced sounds are considered to be the result of a convolution between a fixed frequenc sequence of pulses (which represent the vibration of the vocal cords) and a filter representing the vocal tract. Consequently, unvoiced sounds differ only by the fact that the source is represented by noise thus modeling the airflow which will pass through the vocal tract. This abstracted model of speech production is known as the *source-filter model* and its implementation represents one of the mathematical foundations on which speech processing techniques are build. One such common technique used in many applications is the *real cepstrum*.

Cepstrum is widely used in voice representation for applications ranging from speech / speaker recognition to voice synthesis. The important characteristic of cepstrum that makes it so extensively used is the ability to separate the two non-linearly combined components that define the voice signal: the source and the filter (speech is thus the result of a convolution).

$$s(n) = e(n) * h(n) \tag{1}$$

This is even more important when one takes into consideration that *h(n)*, the transfer function of the filter is highly linked to the spoken phoneme and is essential in determining the shape, the concept, whereas the source, i.e. *e(n)*, is the support, the carrier, relatively to the occurred sound. [20]

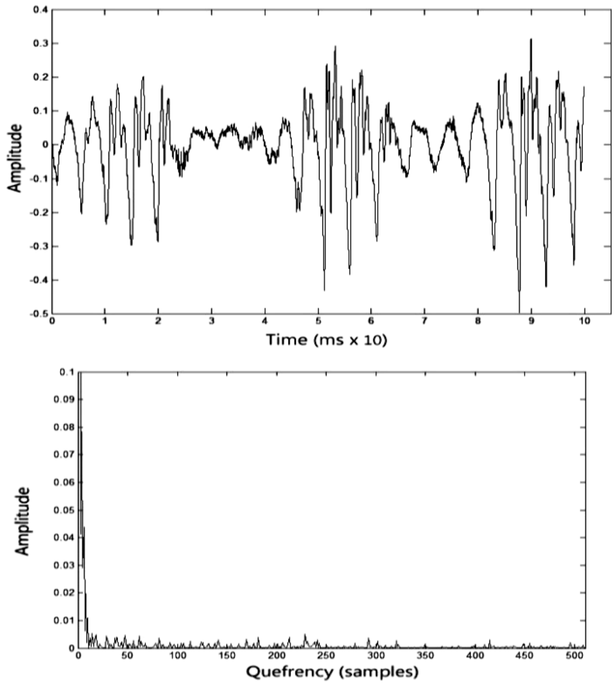

Figure 2. Real Cepstrum of the /r/ phoneme in 'rac' (Romanian for 'crab') pronounced correctly by an adult male. The Cepstrum corresponds to the first time-frame of 23ms (index 0 - 2) of the above signal. Before processing, the extracted signal was multiplied with a Hamming analyzing window.

The brief algorithm for the representation of the real cepstrum of a time-frame of speech signal is the following:

- take the *Fourier Transform* of the speech signal
- take the *absolute value* of the above results
- take the *log* of the above results
- take the *Inverse FT* of the above results

In short:

$$c_s(n) = IFT\{\log|FT(s(n))|\} \tag{2}$$

The key feature in the above equation is the logarithmic function which, according to its properties plus the properties of the Fourier transform and Eq.(1), will generate:

$$\log(e(n) \cdot h(n)) = \log(e(n)) + \log(h(n)) \tag{3}$$

Thus, the two components *e(n)* and *h(n)* become additively separable and, moreover, this separation is unaffected by the inverse Discrete Fourier Transform which projects the information back in the original time domain. It is important to mention though, that the more common name of *quefrency* is used for the resulting x-axis as a similar anagram to *cepstrum*, and thus marking a difference from the unprocessed time domain signal. The extracted information has the following meaning when related to the source-filter model of speech:

- contributions to the cepstrum due to excitation *e(n)* will occur at integer multiples of the fundamental period (this is mostly visible when the analyzed sound is a voiced sound, the excitation being in this case a sequence of pulses of fixed frequency) and will be placed in the high quefrency region.

- contributions to the cepstrum due to the vocal tract *h(n)* will be concentrated in the low quefrency region and will decay rapidly with *n*.

An example for an /r/ phoneme in 'rac' (crab) is provided in Fig.2.

For the current study, the cepstrum was computed of length 512, on 20ms time frames overlapped by half and windowed with Hamming windows. Then only coefficients from 5 to 170 (out of 256 unique coefficients) were chosen for the next step of feature selection. The process of choosing the mentioned coefficients was based on gradually reducing the length of the cepstrum (from both ends) and measuring the results of the classification algorithm. Considering that /r/ is a phoneme that presents some vocalized excitation [21] we therefore believed that not only the low quefrency region is relevant in the process of classification but higher quefrencies are useful as well.

For every such observation (cepstrum of time frame) in an analyzing phoneme we had to reduce the feature vector length from 166 to a smaller value, more acceptable from running time considerations, that is to be used in the classification stage. This process, known as feature selection is often implemented in data classification problems because it has the theoretical role of both reducing the dimensions of the feature space and growing the distances between similar classes thus making the whole classification process more precise. The classic approach implies the use of a linear or non-linear transform that maps the input feature space into a lower dimension output space with the relevant information for the classification process concentrated on fewer coefficients of the feature vectors. One of the most popular techniques, also used in this work, implies the computation of the discrete cosine transform - Eq.(4). As it is well known the spectrum of the DCT gathers most of the energy (and thus the relevant information) on its lower end coefficients. Ignoring the higher ranked coefficients presents the chance of building feature vectors of smaller length thus accelerating the speed at which the classification algorithm performs.

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}(n+\frac{1}{2})\cdot k\right],\ k = 0 \ldots N\text{ - }1 \quad (4)$$

The novelty of this work is represented by the fact that it takes in consideration the Walsh-Hadamard Transform for the process of feature selection. The WHT is a unitary transform used very often in voice transmission applications (CDMA standard) but to a lesser extent in speech recognition and in the case of pronunciation evaluation our team did not discover any references.

The forward and inverse WHT for an input *x* of length *n* are defined as:

$$\begin{aligned} \overline{X} &= H\cdot x \\ x &= H\cdot\overline{X} \end{aligned} \quad (5)$$

where *H* is a Hadamard matrix of order *n*. The *n* x *n* Hadamard matrix is build out of Walsh functions, every row in the matrix representing a Walsh function of length *n* and every two such rows being orthogonal.

The matrix can be defined recursively:

$$H_n = H_1 \otimes H_{n-1} = \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & H_{n-1} \end{bmatrix} \quad (6)$$

where

$$H_1 \overset{\Delta}{=} \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (7)$$

and $\otimes$ represents the *Kronecker product* of two matrices.

The output data of a WHT can be similarly interpreted as the output of the Fourier transform where the coefficients show how much of the input signal lay within a bandwidth of frequencies (frequency content of the signal). In the case of WHT the coefficients replace the concept of frequency with the concept of sequency due to the fact that the Walsh functions do not necessarily have fixed cycles in contrast with the sinusoidal Fourier waves. [22] Yet the resulting coefficients in a Walsh-Hadamard transform show how much of the specified sequence is contained in the input signal.

After computing the WHT or DCT of the observations, the mean was computed for every set of equivalent coefficients in the corresponding spectrum of observations of a phoneme. Thus one feature vector was obtained per phoneme of length equal to the number of chosen WHT / DCT coefficients. At this point the search for the relevant coefficients commenced: when the overall mean of incorrect feature vectors and the overall mean of the correct feature vectors are computed (in order to define the two classes feature prototypes) it can be observed that the region spanning from coefficient 5 to coefficient 16 clearly present different magnitudes for the two classes - Fig.3. In the figure, only coefficients 1 to 22 (out of 256) of the feature prototypes are presented. Clearly for the span 5 - 16 an important difference can be seen between correct pronunciations (continuous line) and incorrect pronunciations (dashed line) - which have a higher magnitude. Therefore reducing the feature vector only to these type of regions yields an increase in performance and precision of the classification algorithm.

Not to go any further it is important to mention firstly that only the case of the WHT is depicted in the figure, the problem for selecting the coefficients out of the DCT being similar and secondly that the choice of coefficients was subsequently optimized by observing both the results of the classification process and analyzing the distribution of the magnitude of the coefficients over the sequency indexes.

Even further, the resulting feature space was reduced to a two dimensional one by constructing feature vectors of length two out of the mean and standard deviation of the selected coefficients. The result is encouraging - Fig.4: the two classes, correct pronunciations and incorrect pronunciations, are well defined. The figure only presents the situation when coefficients 2 to 11 out of the WHT were chosen but similar results are obtained in both the case of the DCT and when other sub-optimal coefficient choices are performed.

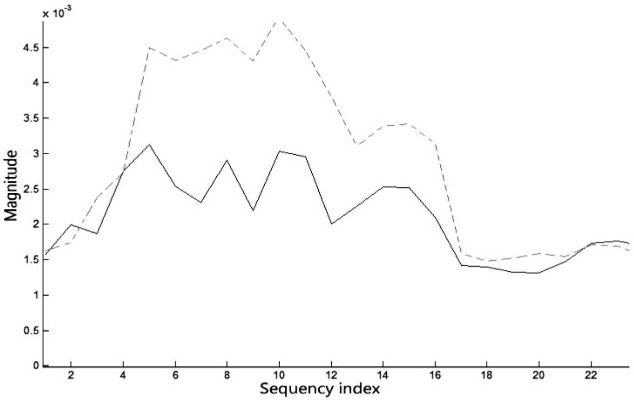

Figure 3. Correct and incorrect pronunciations prototypes represented in the intermediate feature space. The prototypes were obtained by performing the mean of all the feature vectors in the two classes. The correct pronunciations correspond to the continuous trace whereas the dashed line corresponds to incorrect pronunciations.

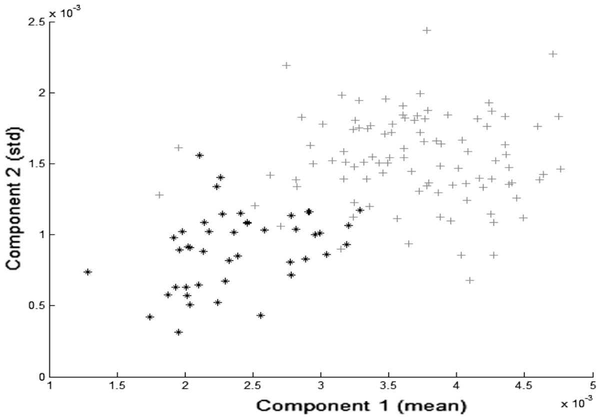

Figure 4. Correct and incorrect pronunciations represented in the final feature space obtained by performing the mean and standard deviation of the selected coefficients out of the FWHT. The correct pronunciations correspond to the black points whereas the light gray corresponds to incorrect pronunciations.

Algorithmically for a given voice recording - *crt_rec*, the overall processing flow is presented below. The output, *crt_feat_vect*, is the current feature vector associated with the recording.

```
MAX_WIN   = Get_No_of_Overlapped_Win(crt _rec);
sel_feat        = 0;

for i = 1 to MAX_WIN
        win        = Extract_ Win (crt_rec, i);
        ceps       = Real_Ceps (window);
        obs        = Sel_Ceps_Coeff(ceps);
        t_coeff   = Sel_Coeff ( FWHT_or_DCT (obs) );
        sel_feat  = sel_feat + t_coeff;
end

c1 = Mean(sel_feat  ./ MAX_WIN);
c2 =     Std (sel_feat  ./ MAX_WIN) ;
crt_feat_vect   = Cat ( c1, c2 );
```

V. Classification

The classification algorithm used in this paper is the well-known, easy to implement, *k-th Nearest Neighbor - kNN*. [23], [24] The classic approach was chosen for implementation even though more efficient versions are known [25-27]. The *kNN* was chosen for both implementation convenience reasons and for the good performance it provides. Not least important, using the same classification algorithm as in most of our past work related to pronunciation evaluation is essential for correctly comparing the different feature extraction methods developed.

As a brief reminder the *kNN*, for every instance in the *database of unknown items*, computes the Euclidean distances to all the elements in a set of feature vectors known as the *reference database* (a database which is already labeled into classes). Then, the results are sorted in ascending order, the first *k* closest neighbors deciding that the current unknown item belongs to the class in which most of them also reside. The correct and incorrect decisions are updated at every iteration in order to construct the correct classification rate and / or the error rate. It is thus obvious that not only the overall database should be divided into two subsets in order to run the algorithm but also the choice of the *k* parameter greatly influences the results. More on the experimental conditions are presented below.

VI. Experimental Setup

Overall the experiment was performed in order to investigate the best coefficient choices for both the WHT and the DCT at the feature selection stage level. For every such selection of coefficients tests were conducted using a *kNN* classification algorithm for *k* ranging from 3 to 17 in odd steps (3, 5, 7, etc.) in order to investigate the variation of the correct classification rate depending on *k*. The database of 144 recordings was divided in the two needed subsets of data: one for testing purposes and the other, already labeled, for comparison. The latter was built out of 35 randomly chosen pronunciations (10 correct and 25 incorrect) whereas the remaining 109 was the test database of unknown pronunciations. The algorithm was performed 500 times for each *k* in order to compensate the size of the available data, at each iteration the databases being reshuffled.

The implementation was based on a multi-threaded setup, in order to accelerate the running time, with 8 simultaneous threads running on separate cores on a machine with 4 physical and 8 logical cores available.

VII. Results

Figures for these experiments are presented in Appendix A at the end of the paper. We shall emphasize those that are relevant for the conclusions in the investigation for determining which are the best choices of coefficients in the feature selection process.

Overall these results were encouraging and prove that a good feature selection method for the evaluation of rhotacism in the case of the initial /r/ consonant is the above presented algorithm. Out of the two feature selection methods used, better performance was obtained with the Walsh Hadamard Transform compared to the Discrete Cosine Transform. The best correct classification rate of 87.63%, in the case of the DCT, was obtained when the selected region spanned from the 2nd to the 6th coefficient and the *k* parameter of the classifier was equal to 13. At the same time the best classification rate when the WHT was used as feature selection tool was of 92.55%, when *k* = 3 and the selection region spanned from the 3rd to the 12th coefficient. All the classification tests conducted with the mentioned selected regions are represented in Fig.5. It can be observed the variation of the result versus the *k* parameter.

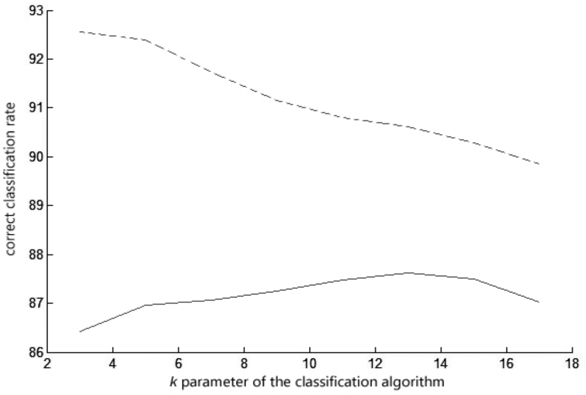

Figure 5. Correct classification rate versus *k* parameter in the case of the best two choices of coefficients. The dashed line corresponds to the WHT as feature selection tool and the choice [3 .. 12] whereas the continuous line corresponds to the DCT and the choice [2 .. 6] .

For WHT reasonably close (to the overall best classification rate) classification rates of around 90% were obtained for a feature vector much smaller of only four elements (coefficients 7 to 10) meaning that most of the relevant decision region lies within that area. This can be explained by the fact that in this region the incorrect pronunciations are characterized by cepstral coefficients that are better correlated to the corresponding Walsh sequences in the Hadamard matrix. In contrast, the correct pronunciations of the initial /r/ have cepstral coefficients less correlated to the sequences but still with magnitudes close enough to each other in order to define a common model that separates them from their incorrect counterparts. The fewer selected coefficients also means an increased computational speed. The only downfall in this case was the fact that *k* also increased, though slightly, (90% was obtained for *k* = 5 or higher) but when compared to the DCT case, *k* still has a lower value meaning that WHT is the better choice for feature selection from computational efficiency point of view, too.

VIII. Conclusion

This work used the Walsh-Hadamard Transform as a novel approach for feature selection in identifying mispronunciations. For comparison reasons the Discrete Cosine Transform was also used in our experiments. We investigated the case of initial /r/ consonant in Romanian language and used selected coefficients out of the real cepstrum as the feature vectors.

The overall results were encouraging with a best correct classification rate of more than 92% in the case of the WHT better than the 87% obtained with the DCT. At the same time it was observed that high classification rates (close to 90%) were also obtained for relatively few selected coefficients meaning that our approach can also be computational efficient.

As an overall conclusion we consider our results as a good foundation for potential future applications like automated software that could be highly useful in assisting fields that involve speech development or therapy.

References

[1] C. Vaquero, O. Saz, W. Rodriguez, E. Lleida, “Human Language Technologies for Speech Therapy in Spanish Language”, Available at: http://dihana.cps.unizar.es/~alborada/docu/2008cvaquero2.pdf

[2] G. Potamianos, C. Neti, “Automatic Speechreading of Impaired Speech”, in *Proceedings of the Audio-Visual Speech Processing Workshop*, Scheelsminde, Denmark, 2001. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.2169&rep=rep1&type=pdf

[3] T. Starner, J. Weaver, A. Pentland, “Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, no. 12, december 1998, pp. 1371 - 1375. Available at: http://luthuli.cs.uiuc.edu/~daf/courses/Signals%20AI/Papers/HMMs/00735811.pdf

[4] N. Moustroufas, V.Digalakis, “Automatic Pronunciation Evaluation of Foreign Speakers Using Unknown Text”. Available at: http://www.telecom.tuc.gr/~vas/papers/csl-pronunciation-evaluation.pdf

[5] V. Young, A. Mihailidis, “Difficulties in Automatic Speech Recognition of Dysarthric Speakers and Implications for Speech-Based Applications Used by the Elderly: A Literature Review”, in *Assistive Technology: The Official Journal of RESNA*, 2010, Vol.22, Issue 2, pp. 99-112. Available at: http://www.resna.org/dotAsset/19865.pdf.

[6] S.-G. Pentiuc, I. Tobolcea, O. A. Schipor, M. Danubianu, M.D. Schipor, “ Translation of the Speech Therapy Programs in the Logomon Assisted Speech Therapy System”, in *Advances in Electrical and Computer Engineering*, Vol. 10, no.2, 2010. Available at: aece.ro/displaypdf.php?year=2010&number=2&article=8

[7] O. A. Schipor, S.-G. Pentiuc, M.D. Schipor, “Improving Computer Based Speech Therapy Using a Fuzzy Expert System”, in *Computing and Informatics*, Vol. 22, 2003. Available at: http://www.eed.usv.ro/~schipor/publications/10_1.pdf

[8] M. Danubianu, S.-G. Pentiuc, O.A. Schipor, M. Nestor, I. Ungureanu, “Distributed Intelligent System for Personalized Therapy of Speech Disorders”, in *The Third International Multi-Conference on Computing in the Global Information Technology*, 2008, Athens, Greece. Available at: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4591364

[9] Gladiola Andruseac, H. Costin, C. Rotariu, "eLearning Platform for Rehabilitation of the Romanian Patients with Neurological Diseases," in *Proc. of ICIW 2009, Fourth Int. Conf. on Internet and Web Applications and Services,* Venice, Italy, May 2009, pp.573-577, ISBN: 978-0-7695-3613-2

[10] Gladiola Andruseac, H. Costin, C. Rotariu, “Design of a Virtual Learning Environmnent for Romanian Patients with Dyslexia” in *Proc. of 1st International Conference on Computer Supported Education, CSEDU 2009*, 23-26 March 2009, Lisbon, Portugal, vol. II, pp. 301-304, ISBN: 978-989-8111-83-8

[11] O. Grigore, C. Grigore, V. Velican, “Intelligent System for Impaired Speech Evaluation”, in *Proceedings of the International Conference on Circuits, Systems, Signals,* 10/2010, pp. 365-368. Available at: http://www.wseas.us/e-library/conferences/2010/Malta/CSS/ CSS-60.pdf

[12] O. Grigore, C. Grigore, V. Velican, “Impaired Speech Evaluation using Mel-Cepstrum Analysis”, in *International Journal Of Circuits, Systems And Signal Processing*, pp. 70-77. Available at: http://www.naun.org/journals/circuitssystemssignal/19-537.pdf

[13] O. Grigore, V. Velican, “Pattern Recognition Based Method Used in Identifying Impaired Speech”, in *Proceedings of the 2nd International Conference on Applied Informatics and Computing Theory (AICT '11): Recent Researches in Applied Informatics*, 2011, Prague - Czech Republic, pp. 190-194, Available at: http://www.wseas.us/e-library/conferences/2011/Prague/AICT/AICT-31.pdf

[14] I.Gavat, O.Grigore, V.Velican, “ Impaired Speech Recognition. Case Study: Recognition of Initial 'r' Consonant in Rhotacsim Affected Pronunciations”, in *Proceedings of the 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 2011, Brasov, Romania, pp.1-6.

[15] O.Grigore, V.Velican, “Self-Organizing Maps for Identifying Impaired Speech”, in *Advances in Electrical and Computer Engineering*, 2011, Vol. 11, Issue 3, pp.41 - 48. Available at: http://www.aece.ro/abstractplus.php?year=2011&number=3&article=7

[16] D.V. Popovici, C. Buica-Belciu, V.Velican, “From ASR to CAST: Intelligent Systems for the Diagnosis and Therapy of Speech-Language Disorders”, in *Revista de Psihopedagogie*, 2010, Issue 2, pp.25-37.

[17] P. Popescu-Neveanu, *Psychology Dictionary*, ed: Albatros, Bucuresti, 1976.

[18] D.V. Popovici, C. Buica-Belciu, A. Iordan, “Phonetic Particularities of Dyslalic Children Pronunciations”, in *O Scoala Deschisa,* 2/2009, Ed.SS6SN, pp. 116-124.

[19] E.Verza, *Dyslalia and its Therapy*, E.D.P., Bucuresti, 1977

[20] I.Gavat, M.Zirra, O.Grigore et al, *Fundamentals of Speech Synthesis*, Ed. Printech, Bucuresti, 2000.

[21] C. Paunescu et al., *An Introduction in Logophedics*, Bucuresti, 1976

[22] R. Tallia, P. Morello, G. Castellano, “The Walsh-Hadamard Transform: An Alternative Means of Obtaining Phase and Amplitude Maps”, in *The Journal of Nulcear Medicine*, 1984, Vol.25, Issue 5, pp.608-612. Available at: http://jnm.snmjournals.org/content/25/5/608.full.pdf

[23] X. Wu, V. Kumar, J.R. Quinlan et al. “Top 10 Algorithms in Data Mining” – survey paper, 2007. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=BF91E8B9BC8CEEB9881E2CBA4EF35C7B?doi=10.1.1.144.5575

[24] T.-.H. Cho, R. Conners, P. Araman, “A Comparison of Rule-Based, K-Nearest Neighbor, and Neural Net Classifiers for Automated Industrial Inspection”, in Proceedings of the IEEE/ACM International Conference on Developing and Managing Expert System Programs, 1991, Washington, USA, pp. 202 - 209. Available at: http://content.imamu.edu.sa/Scholars/it/net/9114.pdf

[25] L. Baoli, Y. Shiwen, L. Qin, “An Improved *k*-Nearest Neighbour Algorithm for Text Categorization”, Available: http://arxiv.org/ftp/cs/papers/0306/0306099.pdf

[26] W. Wang, S. Li, C. Wang, ICL at NTCIR-7: “An Improved KNN Algorithm for Text Categorization”, in *Proceedings of NTCIR-7 Workshop Meeting*, 2008, Tokyo, Japan, pp.385-388. Available at: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings7/pdf/NTCIR7/C3/PATMN/13-NTCIR7-PATMN-WangW.pdf

[27] R. Hassan, M. Hossain, J. Bailey, “Improving k-Nearest Neighbour Classification with Distance Functions Based on Receiver Operating Characteristics”, in *Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases*, 2008, Berlin, Germany, pp.489 - 504. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.146.321

APPENDIX A - TEST RESULT

Correct classification rates in percentage obtained for the usage of **DCT** as feature selection method.
The column on the left presents the chosen coefficients.

Chosen *k* parameter **Selected Coefficients**	***k* = 3**	***k* = 5**	***k* = 7**	***k* = 9**	***k* = 11**	***k* = 13**	***k* = 15**	***k* = 17**
[1 .. 5]	81.96	82.27	82.70	83.10	83.11	82.98	83.14	83.11
[2 .. 6]	86.43	86.96	87.06	87.25	87.47	**87.63**	87.50	87.03
[3 .. 7]	83.13	83.56	83.88	84.55	85.06	85.57	85.50	85.14
[2 .. 11]	84.98	86.01	85.81	83.05	85.73	85.43	85.14	84.32
[3 .. 12]	85.01	85.74	85.98	86.11	86.04	85.38	85.53	84.67
[4 .. 13]	83.3	84.03	84.38	84.24	84.35	84.22	83.70	82.82
[5 .. 14]	78.24	78.94	79.51	79.67	79.57	79.36	79.82	78.12
[6 .. 15]	71.61	72.82	73.55	73.63	72.89	72.26	71.50	70.77

Correct classification rates in percentage obtained for the usage of **WHT** as feature selection method.
The column on the left presents the chosen coefficients.

Chosen *k* parameter **Selected Coefficients**	***k* = 3**	***k* = 5**	***k* = 7**	***k* = 9**	***k* = 11**	***k* = 13**	***k* = 15**	***k* = 17**
[1 .. 10]	91.63	91.98	91.63	91.44	91.02	90.90	90.72	90.49
[2 .. 11]	92.43	92.50	91.78	91.41	91.10	90.93	90.92	90.64
[3 .. 12]	**92.55**	92.39	91.74	91.15	90.80	90.62	90.28	89.85
[4 .. 13]	91.55	91.68	91.14	91.06	91.01	90.93	90.83	90.61
[5 .. 14]	90.81	91.30	91.01	91.18	91.18	91.25	91.43	91.56
[6 .. 15]	89.39	89.86	89.94	90.29	90.38	90.20	89.68	89.93
[1 .. 5]	71.43	71.51	71.40	71.68	71.01	70.03	69.99	69.63
[3 .. 7]	80.72	81.67	82.11	82.34	82.11	81.93	81.91	80.66
[6 .. 10]	89.60	90.73	91.18	91.54	91.68	91.58	91.52	91.07
[7 .. 10]	88.42	89.22	89.65	89.96	90.08	90.06	90.06	89.86