ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2009, Vol. 38, No. 4

IMPLEMENTATION OF HIERARCHICAL PHONEME CLASSIFICATION APPROACH ON LTDIGITS CORPORA

Kęstutis Driaunys, Vytautas Rudžionis, Pranas Žvinys

Vilnius University, Kaunas Faculty, Department of Informatics
Muitines St. 8, LT-44280, Kaunas, Lithuania
e-mail: kestutis.driaunys@vukhf.lt, vytautas.rudzionis@vukhf.lt, pranas.zvinys@vukhf.lt

Abstract. Better discrimination of phonemic units still remains one of the most important problems in automatic speech recognition. Direct phoneme recognition in speaker independent automatic speech recognition systems is unable to provide good enough recognition results. There is made an assumption that better results could be achieved through the recognition of phoneme groups using group characteristic features: voiced/unvoiced, vowel/consonant, etc. This paper describes and statistically motivates features and rules for the detection of phoneme groups using phonetically labeled data. Algorithms for recognition of stop and fricative consonants are presented. Experimental research confirmed the advantages of the hierarchical classification of phonemes. Combination of knowledge and rules for detection of acoustic events with the classical statistical classification methods produced an overall 3% improvement of phoneme recognition accuracy and a 52-55% reduction of time taken by classification.

Keywords: LTDIGITS, phoneme classification, stop consonants recognition, fricative consonants recognition.

1. Introduction

Speech technologies are rapidly developing field of information technologies and more and more often find their way to the practical applications. Also it is well known fact that human auditory system possesses significantly better capabilities to recognize and to discriminate natural speech than the best automatic speech recognition algorithms today. There is common agreement in the speech technologists' community that it is necessary to apply other approaches in order to achieve better overall speech recognition accuracy. Several approaches and methods to achieve higher speech recognition accuracy were proposed: to use more efficient adaptation to the environment methods, to reduce impact of background noise with novel speech enhancement algorithms, to look for the better suited features for speech signal description, to look for the ways to exploit supra-segmental information and to use better syntax models. Anyway, it is clear that acoustical-phonetical information of speech signal isn't exploited enough in modern speech recognition algorithms. We believe that more efficient modeling of acoustical processes and exploitation of methods with better discriminational abilities could lead to overall increase in the performance of speech recognition systems.

Our previous experiments using Lithuanian speech corpora LTDIGITS [4,5] showed that direct phoneme classification using MFCC based features did not provide good recognition results and it could be reasonable to perform classification to the phoneme groups prior the phoneme discrimination [6]. This approach is based on the assumption that phonetic features have enough information to capture ant exploit structural properties (characteristic only to some groups of phonemes) of speech signal. Experiments with a subset of fricative consonants [7] and stop consonants [9] confirmed feasibility of this approach so it could be reasonable to extend the investigation to the other groups of phonemes.

This paper presents attempts to implement speaker-independent hierarchical phoneme recognition method which combines differential acoustical features of phoneme groups with MFCC type coefficients and integrating knowledge about acoustic events with classic statistical classification methods. Our algorithms were tested using LTDIGITS corpora recordings.

2. Theoretical framework

Automatic speech recognition most frequently is based on pattern recognition methods. Their main principle consists in preparation of templates of relevant speech units and comparing them with the recognized vector during the recognition stage. The simplest pattern-based algorithm of phoneme classification is the direct phoneme classification, when part

of the speech signal is compared with each template value and, at the stage of decision making, the fragment is attributed the symbol having exhibited the highest correspondence results.

It has to be noticed that such classification requires large amounts of time, since an unknown features vector has to be compared against all the available templates. Another flaw of this process is that the error weights between phoneme groups and within a group differ. Errors within a group can be corrected in the later stages of recognition. Errors, leading to false differentiation between phoneme groups, have a bigger negative influence on the final recognition reliability. Such drawbacks can be at least partially eliminated by the hierarchical structure of phoneme classification: recognizing class of the phoneme first and then recognizing phoneme itself.

The phonetic theory treats the set of Lithuanian phonemes as a phonetic tree-like hierarchy, where phonemes correspond to "leaves" and, depending on their acoustic and articulational features, are combined into specific groups (vowels, consonants).

Taking into account the above-mentioned motives, a simplified classification method of Lithuanian phonemes is proposed. The method combines the group features with features within a group, i.e., at the initial stage, a phoneme is attributed to one of four phoneme groups (vowels, semivowels, plosives, fricatives). After that, precise phoneme recognition is carried out, comparing the phoneme's features with the template values of the relevant phoneme group (as opposed to comparing the features with the available template values of the whole phoneme set).

The overview of the literature in the speech recognition field showed that an increasing attention is being paid recently to speech recognition research, using the knowledge and rules for the detection of acoustical-articulational events. These methods exploit the information, contained at the phonetic acoustic level, in a more precise manner. However, when trying to use only features and rules of acoustic events for recognition, the amount of errors increases significantly due to the number of inserted, omitted and replaced phonemes. Moreover, in order to use the acoustic knowledge of the phonetic units, one must possess a unified and complete theoretical model of acoustic articulational features. But the development of such a model for Lithuanian consonants is only in the initial stage.

In order to exploit the advantages of both the statistical classification and acoustic events detection methods, they were combined into a hierarchical phoneme classification method which is presented in Figure 1. So the process of phoneme recognition was subdivided into two steps:

1. The phoneme is attributed to the one of the 3 main phoneme groups (plosive consonant, fricative consonant or a sonant).
2. Recognition within the phoneme group is performed.

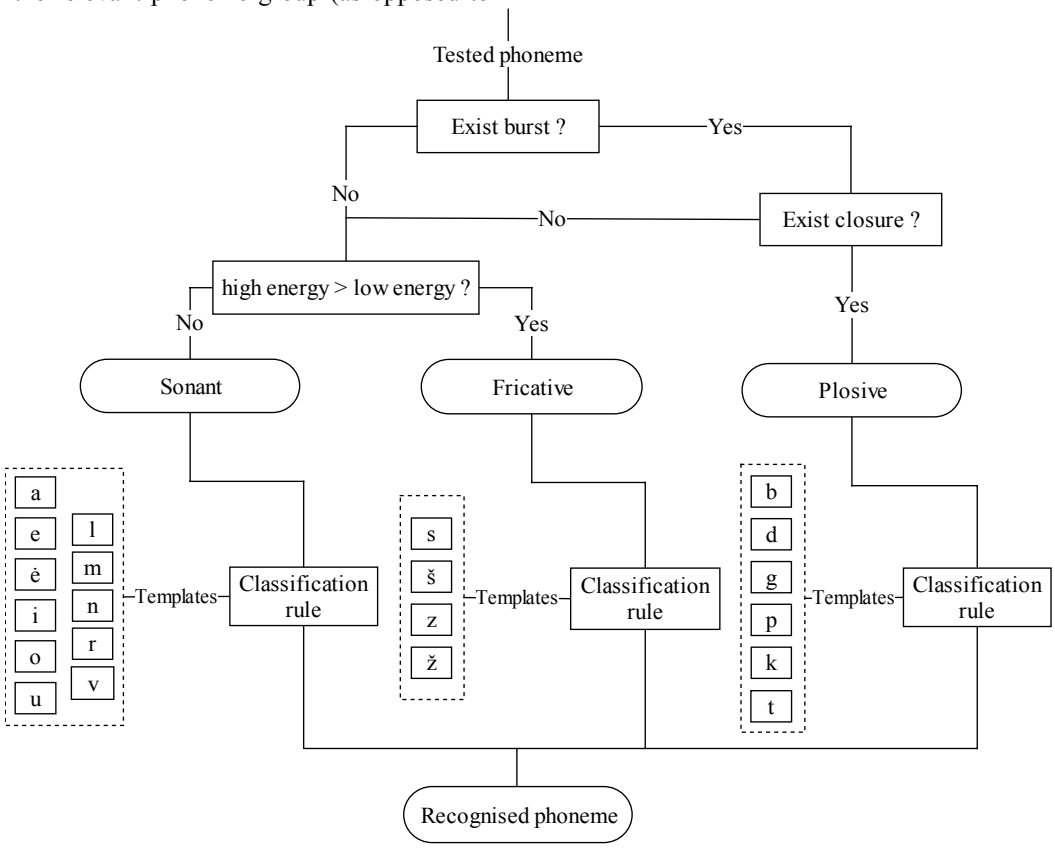


Figure 1. Scenario of hierarchical phoneme classification

In the first step, we need to find the phoneme group the analyzed speech fragment belongs to. It has to be noted that, at the present time, there have been no studies analyzing and describing differential features of Lithuanian phonemes. Therefore, one can use the postulates of the classical phonetic theory that all the plosives consist of two main events: occlusion and plosion. A phoneme is considered fricative when the energy of its high frequencies is higher than that of low frequencies, while the sonants contain much more energy in the lower frequencies than in the higher frequencies. These propositions are implemented in the algorithms for recognition of plosives and fricatives. Experimental research is to be used in order to check the efficiency of the mentioned algorithms and to establish the logical rules, which differentiate the sets of sonants, fricatives and plosives in the most appropriate way.

In the second step, using the algorithm for calculation of the templates of a phoneme, the feature vectors are calculated for the stationary part and the left and right contexts of the phoneme under consideration. Depending on the group to which the phoneme corresponds, a single vector (of the stationary part or one of the context parts) representing its group the best can be prepared for further recognition. The feature vector, produced for the phoneme under consideration, will be compared with the templates of each phoneme, constituting the group, in order to find the closest one.

2.1. Recognition of stop consonants

Stop consonant group in Lithuanian speech as well as in many others consists of six phonemes (/b/, /d/, /g/, /p/, /k/, /t/). All stop consonants are formed from two main acoustic events: closures and bursts [10]. Looking at the acoustic properties of closures, we could distinguish voiced and aspirated stops while looking at the articulation place we could distinguish labials (/p/, /b/), velars (/k/, /g/) and alveolars (/t/, /d/) [1]. But our goal isn‘t to discriminate particular phoneme: we need to make a decision if analyzed part of speech signal belongs to the any of stop consonants.

One common feature of all stop consonants is the fact that closed wall is formed initially in the vocal tract. This wall is called closure. Closure is such an obstacle in vocal tract for air stream when different parts of vocal tracts are squeezed together to close it. So air stream can‘t go out through the lips (acoustically closure results in short pause in speech signal). Finally air stream opens vocal tract rapidly when air pressure level increases enough like during burst (acoustically burst is similar to the white noise and this means that signal energy should increase in all frequency ranges). These acoustic events could be seen in Figure 2 which shows oscilogram and spectrogram of the Lithuanian word *sustoti*. Regions marked with white indicate burst of stop consonants. As could be seen, the absence of energy is characteristic for closure (region prior to the marked burst) while during burst energy spreads over the whole spectrum. Taking into account all these physical properties of stop consonants, we proposed and developed an algorithm for the detection of these phonemes. It means that we need to detect inter-phoneme pauses and bursts (rapid sound energy changes through the whole spectrum) as reliable as possible.

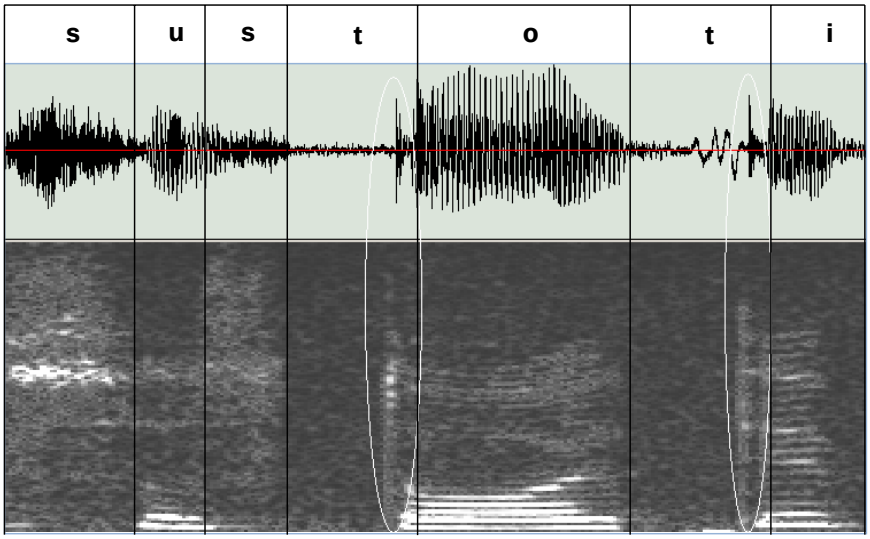


Figure 2. Waveform and spectrogram of the word “sustoti”

Further we will describe the proposed algorithm for recognition of stop consonants. We will make an assumption that acoustic event of burst should be observed as a local maximum of energy level variations in different frequency ranges (variables *Spro1* and *Proc*) and could be used for the detection of stop consonant.

Butterworth second order filters were used to divide speech signal into frequency bands. The total number of bandpass filters in the filterbank is denoted

by *Z*; filtered speech signal at the output of each filter *z* is denoted by yf_z.

For each frame and for each filter *z*, the energy Ef_z is calculated using the following formula:

$$Ef_z = \log\left(\frac{1}{n}\sum_{i=1}^{n} yf_{z,i}\right),$$

where *n* is the number of samples per frame.

Then energy change rates and their maximums for each frame *k* are derived using formulas:

$$\Delta Ef_{z;k} = Ef_{z;k} - Ef_{z;k-1},$$

where *k* is the index of frame

$$Del = \arg\max_{z} \Delta Ef_{z,k},$$

where *z* = 1, …, *Z*.

From *Del* we obtain two threshold levels *Burst* and *Proc*. *Burst* is obtained by evaluating mode while *Proc* is obtained by evaluating the reliability of energy change rate maximums in all frequency bands and finding in which frequency areas the energy change rate is fastest.

The ratio between the average of filter energy change rate («delta») and the maximum of change rate is used to evaluate the quality of burst threshold:

$$echrate_z = \frac{1}{Burst}\sum_{i=1}^{Burst} \Delta Ef_{z;i},$$

where *z* = 1, …, *Z*,

$$SVM_z = echrate_z \Big/ \Delta Ef_{z;Burst}.$$

The average of the ratio above over all filters is used to determine threshold levels for pause or silence.

The ratio (SVM) between average of each filter output energy change rate and the maximum of energy change rate is obtained. Later the average of these ratios is calculated. If this ratio is smaller, equal or only slightly bigger than 1, this part of signal is marked as burst.

Closure search is performed at the third stage of recognition of stop consonant acoustic events. Here we check if energy level of analyzed frame (*E*) is lower than the threshold level of the pause and we find the number of lower energy frames within phoneme.

The proposed algorithm detects and identifies the above mentioned acoustic events within each phoneme. During decision making, the final solution if observed phoneme belongs to the stop consonants class is done according to the information from those acoustic events.

2.2. Recognition of fricative consonants

We need to perform classification of fricative consonants and vowels in the hierarchical phoneme classification algorithm and to find when consonant belongs to the class of plosives.

A speech signal is called sonant if its energy in low frequencies is significantly bigger than energy in higher frequencies. It is a specific property of fricative consonants that in higher frequencies they have more energy than sonants. Implementing this knowledge, the algorithm for recognition of fricatives was proposed. The following notations will be used below:

yzd – sample representing phoneme filtered with low frequency filter;

yad – sample representing phoneme filtered with high frequency filter;

FS – ratio of fricativity,

For the *i*-th frame, the ratio of fricativity is obtained using this formula:

$$FS = \log\left(\frac{1}{n}\sum_{i=1}^{n} yad_i\right) \Big/ \log\left(\frac{1}{n}\sum_{i=1}^{n} yzd_i\right).$$

At first, filtering of speech signal using two filters with low and high frequency pass-bands was performed. The ratio of fricativity (*FS*) was obtained by division of energy level in the high frequencies to the energy level in the low frequencies. This ratio was calculated for each analysis frame. The *FS* value is compared with the threshold level of fricativity and, if threshold is exceeded, frame is marked as fricative. During decision making, the following rule could be used – if more than a half of phoneme frames were marked as fricative, then the whole phoneme could be marked as a fricative. This rule is empirical and optimal thresholds should be obtained experimentally.

3. Experimental investigation

The goal of experimental research was to implement an algorithm of hierarchical phoneme classification, integrating knowledge and rules of detection of acoustic events with classical statistical classification methods, to evaluate the results of phoneme recognition offered by this algorithm, and to compare them with the results of previous research. Hierarchical phoneme classification was implemented according to the scenario provided in Figure 1. Thus, the process of a phoneme's recognition is divided into two stages. In the first stage, with the help of phoneme group recognition rules, the phoneme under consideration is attributed to one of the main phoneme groups (plosive consonant, fricative consonant or sonant sound). In the second stage, phoneme classification has to be performed, comparing it only to the template values of the respective phoneme group. Different phoneme groups are classified using different parts of phoneme as well as different lengths of those parts. According to results obtained in our previous experiments [8], sonants and fricative sounds are classified using phoneme templates formed from 6 frames of the phoneme's left context, while the templates of plosive consonants are constructed from 3 frames of the

phoneme's right context. It is expected that implementation of characteristics of phoneme groups will increase the efficiency of phoneme recognition.

3.1. Recognition of stop consonants

At the initial stage, acoustic properties of plosive consonants were investigated and the algorithm for recognition of plosive consonants was realized. This algorithm makes decision if analyzed phoneme belongs to the class of plosive consonants and evaluates recognition accuracy on Lithuanian plosives from the LTDIGITS corpora. Phoneme boundaries were found and set manually by expert labelers.

Experiments for stop consonants recognition were performed using the stop consonant recognition method. It is based on the assumption that the acoustic event of burst should be observed as a local maximum of energy change rate in all frequency ranges. A system using 14 band pass filters was realized to check this assumption. Each filter had pass band of 500Hz and the whole system covered frequency range between 500 and 7500 Hz.

Signal at the output of each filter has been divided into the analysis frames of 10 ms length with overlaps of 5 ms. For each frame, energies and energy change rates ("deltas") were obtained. We looked for the maximum of each filter output energy change («delta») within each phoneme when searching for burst. The place of the burst was detected and stored in the memory. Then mode of energy "deltas" from filter outputs and "deltas" reliability was obtained within each phoneme. If reliability is equal to 1, it means that in all 14 pass bands maximums of energy change rate are observed simultaneously. If reliability is equal to 0.5, it means that places of maximums coincide in 7 out of 14 pass bands, etc. So we realized the rule that burst is detected if the local maximum of energy change rate coincides in more than a half filters in the filter bank.

Additionally, we calculated the ratio of energy change rate maximum for each filter with the average change rate in frames prior to frame with maximum change rate (SVM). This is done to locate burst place more precisely. If this ratio is bigger than one, then it means that change rate maximum has been fixed in relatively noisy environment and this isn't characteristic for stop consonants since burst should exist prior to closure. The bigger SVM ratio is the clearer burst is seen.

It should be noted that the above described rules for burst localization treat all sharp energy change variations as bursts which are characteristic not only for stop consonants but also for the transitions from fricative sound to the voiced sound. The search for closure was carried out to eliminate errors of this type.

The low energy level is characteristic for closure. Also inspecting spectrograms we could see that some low frequency oscillations called pitch are also characteristic for closures of voiced stops. In some cases pitch is characteristic also for closures of unvoiced stops. This occurs when vowel of longer duration is pronounced before the stop consonant as in the words *se**p**tyni, a**t**gal*. This effect is called assimilation in vocal cord activities in the theory of phonetics [10].

The low frequency filtering (frequency range up to 500Hz) has been carried out trying to eliminate oscillations generated by the context and to reject pitch.

At the next step, the closure threshold for each utterance in the LTDIGITS corpora has been determined according to the energy level between words in this utterance. We are looking if there are parts of signals within phoneme whose threshold is lower than threshold for closure.

Statistical evaluation of acoustic events shows that it was impossible to find features which could unambiguously separate subsets of stop consonants and non-stop consonants. So a series of experimental evaluations were performed trying to achieve the optimum performance. They included manipulations with the above described features (testing various fricativity thresholds and testing the impact of rules weights).

3.2. Recognition of fricative consonants

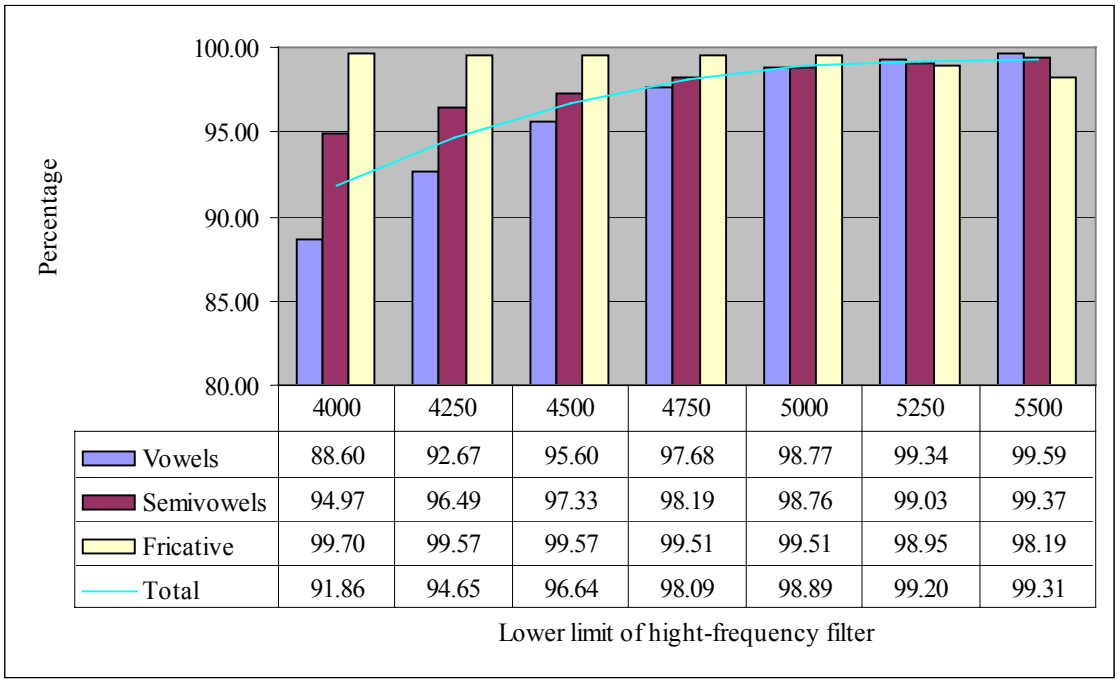
The aim of the experiment is to examine acoustic properties of fricative consonants and on that basis construct a recognition algorithm, which would provide the decision whether a phoneme under consideration is a fricative and to assess the recognition accuracy using the set of Lithuanian fricative consonants formed from the records of the LTDIGITS corpora.

A speech signal is considered sonant when its low frequencies energy is significantly higher than that of high frequencies, while fricatives carry more energy in high frequencies than sonant sounds. On the basis of this knowledge, a system of two band filters was implemented with such pass-band zones: low frequencies (50-2500 Hz) and high frequencies (5000-7000 Hz). The band limits of the low frequencies filter were chosen taking into account the results of experimental phonetic research stating that the first two formants of all the Lithuanian vowels are concentrated in this zone. The limits of the high frequency band were set on the basis of the best results obtained during the tests of the influence of the width of this band on the accuracy of fricatives recognition (summarized test results are provided in Figure 3).

The figure demonstrates the influence of the lower limit of the high-frequency filter on the accuracy of recognition – including phoneme groups in general and different classes of phoneme groups. As one can see, as the band of the high frequency filter gets narrower, the recognition accuracy increases for all the phoneme groups, and the highest value (98.89%) is achieved at the lower limit of 5000 Hz. Although further narrowing of the band leads to the increase of the recognition accuracy even more (up to 99.31%), this is caused by the improved results of vowel

recognition (this is significant in the overall accuracy since the number of vowels in the sample is high), while the recognition of fricative consonants begins to deteriorate (98.2%).

The next stage involves calculating the ratio between high frequencies and low frequencies energy and its comparison to the threshold fricative value. If the ratio is higher than or equal to the threshold of fricative value, a decision is taken that the interval belongs to a fricative sound; otherwise it is attributed to a sonant sound.



	4000	4250	4500	4750	5000	5250	5500
Vowels	88.60	92.67	95.60	97.68	98.77	99.34	99.59
Semivowels	94.97	96.49	97.33	98.19	98.76	99.03	99.37
Fricative	99.70	99.57	99.57	99.51	99.51	98.95	98.19
Total	91.86	94.65	96.64	98.09	98.89	99.20	99.31

Figure 3. Influence of the lower limit of high-frequency filter on the recognition accuracy of phoneme groups

Table 1. Results of phoneme recognition to the groups of plosive and non-plosive sounds (in percentages)

	Vowels	Semivowels	Plosive	Fricative
Number of phonemes	11614	4422	6472	3045
Recognised as plosive	2.9	17.2	**93.8**	1.9
Recognized as non-plosive	**97.1**	**82.8**	6.2	**98.1**

4. Results

The algorithm for the detection of acoustic events in consonants was tested using recordings from the LTDIGITS corpora. Utterances of 100 speakers (50 male, 50 female) were selected for the experiments. The overall number of phonemes in these experiments was more than 25000. The experiments demonstrated an overall accuracy of 93.95% in assigning phonemes to the groups of plosive and non-plosive sounds. More detailed results are provided in Table 1.

These recognition results are comparable to the best results achieved in the studies for the English stop consonant recognition. Colotte and Laprie [3] obtained 86.8% stop consonant recognition accuracy using spectral variation function and the same system for detection of acoustic events. They used stop consonants from TIMIT corpora which were uttered by 215 different speakers. Abdelatty Ali with colleagues [1,2] realized acoustic knowledge-based system for phoneme group discrimination and classification. They achieved 90% recognition accuracy using 30 speakers' utterances from TIMIT corpora in their experiments. Accuracy increased by 2% comparing these results with our previous experiments using linear discriminant analysis and MFCC coefficients [11].

Looking at the results of phoneme groups recognition (Table 1) we see that relatively large number of errors was obtained on the semivowels. 42% of errors were obtained on phoneme *n*, 27% were obtained on phoneme *v* and 22% were obtained on semivowel *r*. Analyzing errors of semivowels according to the individual words in the LTDIGITS corpora it was observed that 56% or errors occur on the stressed transitions from semivowel to vowel (as in words ***nu**lis, dev**y**ni, **vi**enas*). This is because acoustic nature of signal is similar to the voiced stop consonant. Semivowels have characteristic low energy level in the high frequencies range while energy rapidly increases in the transition to the vowel. This is the reason why

stop consonants recognition algorithm fixes this effect as a plosive.

Looking at the errors of stop consonants recognition we observed that about 50% of errors were caused by the improper selection of closure level for voiced stop consonants. It is reasonable to evaluate this problem separately and to select threshold level individually for each utterance.

The described recognition rules for acoustic events of fricative consonants were tested using the same utterances from the LTDIGITS corpora. In the experiments, the accuracy of assignment of phonemes to the groups of fricative or sonant phonemes was 98.9%. More detailed results are provided in Table 2.

Table 2. Recognition accuracy of phoneme groups (in percentages)

	Vowels	Semivowels	Fricative
Number of phonemes	11614	4422	3045
Recognized as sonant	**98.8**	**98.8**	0.5
Recognized as fricative	1.2	1.2	**99.5**

The obtained results of fricatives recognition could be considered as similar to the results obtained in the recognition of English fricatives. Colotte and Laprie [3] used the same observation that energy in high frequencies of fricatives is higher than the energy in low frequencies. They achieved recognition accuracy of these sounds, equal to 91.2%, by performing their experiments with 171 phonemes pronounced by 25 different speakers in sentences. Abdelatty Ali and colleagues [1] implemented a system of phoneme group discrimination and classification, based on acoustic knowledge. During tests with records of TIMIT corpora, recorded by 30 speakers, a 93% accuracy of fricatives recognition was achieved [1].

After implementation of hierarchical phoneme classification algorithm, which integrates knowledge of acoustic events detection and rules with statistical classification methods, the overall recognition accuracy increased from 65.2% to 68.4%. In Table 3 we present results of hierarchical phoneme groups recognition which are comparable with the best results achieved using classification of left phoneme context obtained from 6 frames [8].

It could be observed that overall accuracy of recognition to the classes of vowels, semi-vowels, plosive and fricative consonants improved modestly from 85.8% to 86.1% using hierarchical phoneme recognition approach. Looking at the dynamics of each group's recognition accuracy (Table 3) it could be seen that the biggest improvement (more than 10 percent) was obtained for fricative consonants. At the same time, the recognition accuracy of plosives increased only slightly while the recognition accuracy of vowels remained the same. The recognition accuracy decreased by 7% for semi-vowels.

Table 3. Phoneme group recognition results with and without hierarchical classification method

	Vowels	Semivowels	Plosive	Fricative
Number of phonemes	11614	4422	6472	3045
	Recognition without hierarchical classification			
As vowel	**86.4**	16.0	1.9	0.4
As semivowel	9.4	**73.6**	5.6	7.7
As plosive	4.1	10.4	**92.1**	4.5
As fricative	0.1	0	0.3	**87.3**
	Recognition with hierarchical classification			
As vowel	**86.4**	15.6	0.7	0.1
As semivowel	9.8	**66.2**	4.3	0.4
As plosive	2.9	17.2	**93.8**	1.9
As fricative	0.9	1.0	1.2	**97.6**

Summarizing achieved results, it must be noted that hierarchical recognition approach allowed us to increase the overall recognition accuracy by 3%. At the same time, recognition accuracy into phoneme groups improved only by 0.3%. The main reason of overall improvement was better recognition of fricative and plosive consonants. The set of these consonants was relatively small comparing with the set of vowels. Also, the overall result was affected by the decreased accuracy of semi-vowels recognition. This fact may be viewed as a regular since major attention has been paid to the consonants while properties of vowels and semi-vowels weren't analyzed thoroughly.

Evaluating computational efficiency of algorithms, we observed that hierarchical recognition reduced the processing time by 52–55% (tests were performed using two computers of different productivity).

5. Conclusions

1. Hierarchical phoneme recognition algorithm was implemented. This algorithm integrates knowledge and rules for acoustic events detection with the methods of statistical classification. The recognition accuracy was 68.4 % using the LTDIGITS corpora data.
2. Using the hierarchical classification approach, we obtained the overall improvement of recognition accuracy of about 3% comparing with the best results presented earlier.
3. Hierarchical classification approach is faster than direct classification approach (the computational time was reduced by 52–55 percent).
4. Knowledge of acoustic properties of fricative and plosive consonants was implemented in phoneme group recognition algorithm. This algorithm enabled to improve fricative consonant recognition accuracy by 10.3%. The plosive consonants recognition accuracy was improved by 1.7%.

References

[1] **A.M. Abdelatty Ali, J. Van Der Spiegel, P. Mueller, G. Haentjens, J. Berman.** An Acoustic-Phonetic Feature-based System for Automatic Phoneme Recognition in Continuous Speech. *IEEE International Symposium on Circuits and Systems* (*ISCAS*-99), *Vol*. III, 1999, 118-121.

[2] **A.M. Abdelatty Ali, J. Van Der Spiegel, P. Mueller.** Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants. *IEEE Transactions on Speech and Audio Processing*, *Vol*. 9, 2001, 833-741.

[3] **V. Colotte, Y**. **Laprie.** Automatic enhancement of speech intelligibility. *In IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP*'2000, *Istanbul, Turkey*, 2000, 1057-1060.

[4] **K. Driaunys, V. Rudžionis, P. Žvinys.** The classification of Lithuanian language phonemes through the application of Fisher linear discrimination function and mel frequency cepstral coefficients. *Information Sciences, ISSN* 1392-0561, *Vilnius, Vol*. 31, 2004, 213-218. (*in Lithuanian*).

[5] **K. Driaunys, V. Rudžionis, P. Žvinys.** Mel frequency cepstral coefficients analysis of Lithuanian phonemes. *Human language technologies: the Baltic perspective: the* 1*st Baltic conference, April* 21-22, *Riga*, 2004, 162-165.

[6] **K. Driaunys, V. Rudžionis, P. Žvinys.** The classification of LTDIGITS phonemes based on hierarchical phonetic structure. *In Proceedings of Information Technology* 2005, *Kaunas,* 2005, 283-288 (*in Lithuanian*).

[7] **K. Driaunys, V. Rudžionis, P. Žvinys.** Analysis of Vocal Phonemes and Fricative Consonant Discrimination based on Phonetic Acoustic Features. *Information Technology and Control*, *Vol*.34, *No*.3, 2005, 257-262.

[8] **K. Driaunys, V. Rudžionis, P. Žvinys.** Averaged Templates Calculation and Phoneme Classification. *Information Technology and Control, Vol*.36, *No*.1A, 2007, 139-144.

[9] **K. Driaunys, V. Rudžionis, P. Žvinys.** Analysis of recognition of acoustic events of stop consonants. *In Proceedings of the 14th International Conference on Information and Software Technologies, Kaunas, Lithunia*, 2008, 42-48.

[10] **A. Pakerys.** Phonetics of appellative Lithuanian language. *Vilnius, Mokslas,* 1986. (*in Lithuanian*).

[11] **V. Rudžionis, K. Driaunys, P. Žvinys.** Lithuanian Speech Recognition by Improved Phoneme Discrimination. *In Proceedings of the second Baltic conference on Human Language Technologies*, *Tallinn, Estonia*, 2005, 173-178.

Received October 2008.