DE GRUYTER OPEN

1862
RIGA TECHNICAL UNIVERSITY

Electrical, Control and Communication Engineering

ISSN 2255-9159 (online)
ISSN 2255-9140 (print)
July 2016, vol. 10, pp. 35–41
doi: 10.1515/ecce-2016-0005
https://www.degruyter.com/view/j/ecce

Multi-Stage Recognition of Speech Emotion Using Sequential Forward Feature Selection

Tatjana Liogienė (*Doctoral student, Vilnius University Institute of Mathematics and Informatics*),
Gintautas Tamulevičius (*Associate Professor, Vilnius Gediminas Technical University*)

Abstract **– The intensive research of speech emotion recognition introduced a huge collection of speech emotion features. Large feature sets complicate the speech emotion recognition task. Among various feature selection and transformation techniques for one-stage classification, multiple classifier systems were proposed. The main idea of multiple classifiers is to arrange the emotion classification process in stages. Besides parallel and serial cases, the hierarchical arrangement of multi-stage classification is most widely used for speech emotion recognition. In this paper, we present a sequential-forward-feature-selection-based multi-stage classification scheme. The Sequential Forward Selection (SFS) and Sequential Floating Forward Selection (SFFS) techniques were employed for every stage of the multi-stage classification scheme. Experimental testing of the proposed scheme was performed using the German and Lithuanian emotional speech datasets. Sequential-feature-selection-based multi-stage classification outperformed the single-stage scheme by 12–42 % for different emotion sets. The multi-stage scheme has shown higher robustness to the growth of emotion set. The decrease in recognition rate with the increase in emotion set for multi-stage scheme was lower by 10–20 % in comparison with the single-stage case. Differences in SFS and SFFS employment for feature selection were negligible.**

Keywords **– Classification algorithms; Emotion recognition; Human voice.**

I. Introduction

Human–computer interaction has become an everyday routine nowadays. The knowledge of computer science, sociology psychology, data visualization, and other fields are integrated to improve this interaction. Implementing speech communication into interaction process brings some challenges. Besides obvious tasks like noise removal problem, speech recognition or speaker identification, the question of speaker's emotional state arises. The emotional state of the speaker affects his speech inevitably, thus making aforementioned tasks more complicated.

In some cases, the emotional information is vital in the interaction process. A well-timed and accurate identification of particular emotional states of the client would help to optimize the work process of customer service centers and call-centers by redirecting the calling person to agents with appropriate qualification [1]. Speech emotion identification could also be integrated into personal assistance systems helping us to drive car, staying at home, hospitals or shopping.

Speech emotion recognition is a common classification task with three major steps: the feature set formation, the training process of the classifier, and the process of decision-making about an unknown emotional pattern. In spite of numerous studies and research results, there are still unsolved issues on speech emotion recognition. The language-independent or language-specific emotion features, feature selection and feature sets, classification scheme, language effect on emotion recognition and other questions need to be answered to achieve a robust and reliable speech emotion recognition.

II. Speech Emotion Recognition

A. *Feature Sets*

The intensive study and research of speech emotion recognition problem introduced a huge collection of potential speech emotion features. They include various prosodic (estimated from pitch and formant frequencies, vocal intensity and energy, amount of pauses, speech duration and rate) and spectral (based on linear prediction model, mel-frequency spectrum) features [2]. Additional features like jitter and shimmer in voice, Zipf characteristics of speech rhythmics and prosody [3], glottal closure parameters (namely, strength and sharpness of closure), glottal flow features [4] are proposed to obtain additional discriminating power of feature sets. Besides, various derivative statistics of features are also included in feature sets (like average, median, standard deviation, dispersion, minimum and maximum values, quantiles and others) [2].

Such a variety of features establishes the sets of a few thousands of features [5]. Large feature sets complicate the speech emotion recognition task as the number of analyzed speech patterns becomes lower than the feature order. In this case, the classification results become unreliable and meaningless. Therefore, the order of the feature sets need to be reduced [6].

To overcome this problem, feature selection and transformation techniques are proposed for the reduction of feature sets. The principle of feature selection is to keep only dominant features by excluding the insignificant ones. Sequential Forward Selection (SFS) [3], Sequential Floating Forward Selection (SFFS) [7], Sequential Backward Selection, Promising First Selection, genetic algorithms [6] and the Maximum Relevance–Minimum Redundancy approach [8] are well known feature selection techniques. The features are selected using the criterions of the individual classification power of separate features [3], the cross-correlation of features [9], Fischer rates, and feature information gain.

Feature transformation techniques represent feature sets into lower order space. Standard techniques like Principal Components Analysis, Linear Discriminant Analysis, Multidimensional Scaling, Lipschitz spacing method, Fisher

Discriminant Analysis, neural networks, decision trees are used for this transformation. Despite the vigorous mathematical basis of transformation techniques, the selection ones are preferred in speech emotion classification task.

B. Classification of Speech Emotion

It is a common practice to identify speech emotion by performing a one-step classification. Large feature sets with a high degree of variability and overlap burdens the classification process. Recently, various multiple classifier systems were proposed as an alternative to one-stage classification. The main idea of multiple classifiers is to arrange the emotion classification process in stages, thus obtaining a multi-stage classification. Various implementations of multi-stage classification were proposed for speech emotion identification task.

The parallel ensemble of classifiers was designed by arranging the classifiers into an ensemble one by one [7]. Each classifier employed a different feature set formed using the Sequential Floating Forward Selection algorithm. This classification scheme (based on Bayesian classifiers) gave a 92.6 % overall recognition rate for six emotions.

The class-specific multiple classifier scheme was proposed for a seven-emotion case (anger, boredom, disgust, fear, happiness, sadness, and neutral) and implemented in parallel manner too [10]. Each classifier was dedicated for a single emotion and used a particular feature set. The classifiers and the features were selected depending to their performance on specific emotion. Fusion technique was applied for final decision making by combining the results of the class-specific classifiers. This classification scheme gave higher classification accuracy (80.6 % on average) in comparison with single-stage classification case.

A serial organization of classifiers for the identification of six emotions (happiness, boredom, neutral state, sadness, anger, and anxiety) was proposed in [7]. Again, a few single-emotion dedicated Bayesian classifiers were arranged in a cascade scheme. Each classifier operated using a particular feature set obtained by using the modified version of SFFS algorithm. Each emotion is identified by emotion-specific classifier in a separate stage of the cascade scheme. The serial organization of classifiers gave an average recognition rate of 96.5 %.

Another group of multi-stage classification schemes employs the hierarchical organization of classification process. In this case, the classifiers are combined in hierarchical manner according to some predefined structural assumptions.

The hierarchical structure of subsystems for the analysis of emotions in pairs was proposed in [4]. Six emotions (anger, joy, sadness, fear, boredom, and neutral state) were grouped into 15 different pairs. Fifteen separate subsystems were trained to distinguish between two emotions in pair using a particular feature set. For example, recognition of joy required five sub-systems analyzing the following pairs: joy/anger, joy/neutral, joy/sadness, joy/fear, and joy/boredom. The final result of classification was obtained by using majority voting over the results of sub-systems. The overall emotion recognition accuracy was 85.2 % in gender-dependent experiment and 80.1 % in gender-independent case by using 112 features in total.

A three-level classification scheme was proposed for the identification of sadness, anger, surprise, fear, happiness, and disgust [11]. Five classifiers were used for pairwise emotion classification. Emotional classes were constituted as having the highest Fisher rate for features values (Fig. 1). The average rate of emotion recognition for each level was 86.5 %, 68.5 %, and 50 %, respectively. The total number of analyzed features was 288.

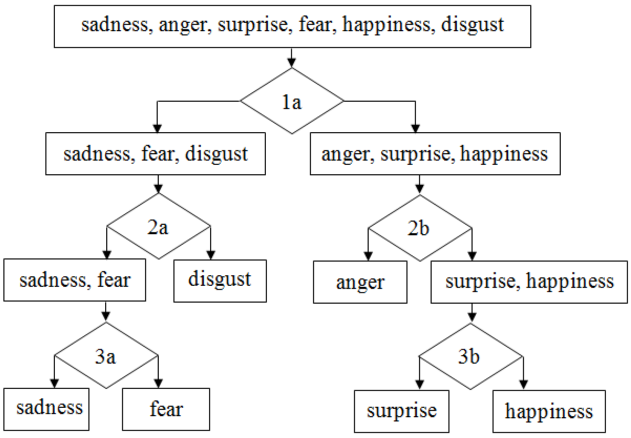

Fig. 1. Structure of the three-level model [11].

A two-stage hierarchical classification scheme for two emotions (anger and neutral) was based on gender information [12]. During the first stage, all utterances were classified into three emotional groups: male (or neutral), female (or anger), and unknown group. During the next stage, the utterances of the unknown group were classified into two subgroups: anger or male, and neutral state or female. Different feature sets were used in each classification stage. The total order of features in this scheme was 56. The obtained average emotion recognition rate was 80.7 %.

Gender information was also employed in Enhanced co-training algorithm [13]. Two feature sets and two classifiers were applied for the classification of six emotions (female and male utterances were analyzed separately). Equally labeled utterances (by both classifiers) were assigned to temporal collection for further examination. The final decision was obtained by combining the results of both classifiers. The obtained average emotion recognition accuracy was 75.9 % for female speakers and 80.9 % for male speakers.

A psychologically-inspired binary cascade classification scheme employs dimensional descriptions of the emotions: valence, activation, and stance [14]. During the first stage, seven emotions are classified into two groups: non-negative valence (happiness and neutral) and negative valence (anger, boredom, disgust, anxiety, and sadness). During the second stage, the non-negative valence group is classified into happiness and neutral, and the negative valence group is classified into two groups: negative activation (boredom and sadness), and positive activation (anger, disgust, and anxiety). The negative activation group during the next stage is separated into boredom and sadness, and positive activation is classified into the lower stance (disgust) and higher stance (anger and anxiety) groups. Lastly, the anger and the anxiety are separated. The highest obtained emotion recognition accuracy in this scheme was 97 % using the 75th order feature set.

Another multi-stage hierarchical classification scheme is driven by a dimensional emotion model [3]. Firstly, all six emotions are classified by arousal dimension into three groups: active, median, and passive emotions. Secondly, each of these emotion groups is divided into two emotions using different classifiers. As a result, active emotion group is divided into anger and joy, and median arousal emotion group is divided into fear and neutral state. Passive emotion group is divided into sadness and boredom. Gender-based classification was also applied in this scheme. The average recognition accuracy of 76.4 % was obtained using 68 different features in this scheme.

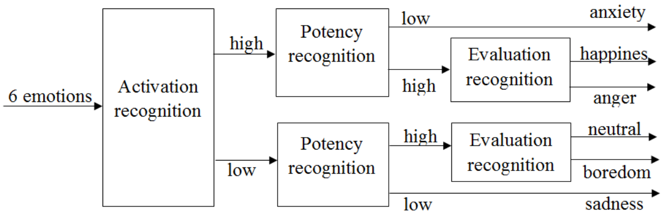

Fig. 2. Two-step hierarchical classification of two emotions [14].

The alternative hierarchical classification approach was introduced by applying the three-dimensional (activation, potency, and evaluation) emotion model [7]. Six emotions (anger, happiness, anxiety, neutral state, boredom, and sadness) were classified in three steps by using multiple Bayesian classifiers (Fig. 2). Firstly, all emotions are classified into the high activation and low activation classes. Each of these classes are divided into the low and high potency groups. Separate emotions are labeled in the third level. The average recognition rate of 88.8 % was stated for this scheme.

Concluding this section, we can notice that the hierarchical arrangement of multi-stage classification is the most widely applied for speech emotion classification. Hierarchical arrangement facilitates the integration of emotional models, additional information, and the formation of emotion (or emotion group) specific feature subsets into classification scheme. This property guarantees the superiority of multi-stage speech emotion classification over the single-stage classification case.

III. Multi-Stage Classification Using SFS and SFFS Techniques

A. *Multi-Stage Classification Scheme*

A multi-stage classification scheme was employed for speech emotion classification task in this study [15].

The main idea of multi-stage scheme is to classify all emotional speech utterances in several stages using different feature sets (subsets) for every stage. During the first stage, all emotional speech utterances are separated into emotional classes that are determined by the first-level feature subset. During the second stage, each of these emotional speech classes are divided into lower level classes or separate emotions using different second-level feature subsets. There can be any number of classification stages L with different feature subsets for each class. The number of classes in every stage is also unlimited.

The idea of multi-stage classification was formulated considering these presumptions on speech emotion classification task:

- significant overlap of features in single-stage classification of all emotions is the main reason of classification errors. This can be solved by reducing the number of emotions analyzed at the same time (during one stage);
- increase in the average classification rate not necessarily implies the classification rate increase for each emotion individually. The reason for this is the combined feature set for all emotions. To avoid this, each emotion (or emotional group) should be characterized by its own feature set (subset);
- all analyzed emotions can be divided into some groups depending on selected feature set or another objective criterion. The derived groups can be divided again using some other feature set and so on until we obtain separate emotions.

Two different feature selection criterions were proposed in our previous works for the multi-stage scheme: maximal efficiency, and minimal cross-correlation. In the first case, feature subsets were formed by selecting the first most efficient features for every classification stage [15]. The second criterion determines the selection of linearly independent features, thus maximizing the discriminating power of the entire feature set [16]. The employment of fundamental frequency-based feature sets revealed the superiority of the proposed multi-stage scheme against the direct (single-stage) classification technique using the full set of 234 features. The average results of the classification of four emotions (anger, joy, sadness, and neutral state) were 59.6 % and 56.3 %, respectively, using maximal efficiency and minimal cross-correlation criterions for multi-stage classification scheme. Single-stage classification using aforementioned full feature set gave a 30.5 % classification accuracy.

B. *Sequential Feature Selection*

In this study, we employed the Sequential Forward Selection (SFS) and Sequential Floating Forward Selection (SFFS) techniques for the selection of feature subsets during every classification stage. SFS and SFFS are greedy search algorithms and are supposed as sub-optimal feature selection techniques as not all possible feature combinations are analyzed during selection process.

SFS is one of the simplest and fastest feature selection techniques. It composes the feature set by adding new features one by one. However, SFS technique does not provide the possibility of removing the included feature which can lose its positive effect on the entire subset with the increase of the subset size. SFFS technique can be considered as the extension of the SFS technique as it contains feature removal step. Consequently, SFFS can result in smaller feature sets in comparison with the SFS technique.

The following steps are performed in SFS-based feature selection:

- initialization of the empty feature subset F_0:

$$F_0 = \{\varnothing\}; \tag{1}$$

- the subset F_i is extended with the feature f_m making the new subset F_{i+1} more effective:

$$F_{i+1} = \{F_i + f_m\}; \tag{2}$$

$$f_m = \arg\max[E(F_{i+1})] > E(F_i), \quad m = 1, 2, ..., M. \tag{3}$$

This step (we call it incremental) is repeated while the efficiency of a new feature subset F_{i+1} increases or while $(i + m) < M$, where M – the number of analyzed speech emotion features.

In general, the selection of P features will require $(P + 1) \times M$ feature set evaluations, which for large M values can become a computationally quite intensive task.

SFFS algorithm contains one additional step (a decremental one). This step is intended for the removal of the feature f_n making the obtained subset F_{i+1} more effective:

$$F_{i-1} = \{F_i - f_n\}; \tag{4}$$

$$f_n = \arg\max[E(F_{i-1})] > E(F_i), \quad n = 1, 2, ..., P, \tag{5}$$

where P is the size of the analyzed feature subset F_i.

In this case, the computational load of the feature selection process grows only slightly. Every decremental step requires P evaluations, which is minor in comparison with the load of incremental step.

The idea of sequential feature selection employment is as follows. During the first classification stage, all emotional speech utterances are separated into an initially predefined number of classes. The feature subsets for this classification are formed applying SFS (or SFFS) technique for the entire collection of the extracted speech emotion features. This process is repeated for every classification stage and every emotional class until the last stage classification is performed and the set of identified emotions is obtained.

IV. Experimental Study

Sequential-forward-selection-based multi-stage classification scheme was experimentally tested in the speech emotion recognition task. Two different databases were used for this study: Berlin emotional speech database [17], and Lithuanian spoken language emotion database [18]. Both databases contain an acted emotional speech recorded by actors: professional (German database), and non-professional (Lithuanian case).

The experiments carried out four cases of the speech emotion recognition task: two emotions (joy, anger), three emotions (joy, anger, neutral), four emotions (joy, anger, neutral, sadness), and five emotions (joy, anger, neutral, sadness, fear). In total, 60 German patterns and 300 Lithuanian utterances per emotion were selected, thus giving the sets of 300 German and 1500 Lithuanian utterances.

A non-parametric classifier was selected for the test considering a moderate size of data sets. We have used the K-Nearest Neighbor classifier (with $K = 3$). In order to get a more reliable evaluation of the performances, 3-fold cross-validation was applied. Again, the number of folds was limited by the size of datasets.

A total of 6552 different speech emotion features were extracted for the emotion recognition experiment using OpenEAR toolkit [19]. The features included zero crossing rate, energy, fundamental frequency, mel-scale features, spectral band features, probabilities of voicing in speech, and various their derivatives (like smoothed envelope values, first and second order differential features, feature value statistics, parameters of value distribution).

We made assumption about low-pitch and high-pitch emotion classes during the first classification stage. Pre-experimental analysis of pitch values for different emotions has shown that neutral state and sadness should be labeled as low-pitch emotions and the high-pitch group should consist of anger, joy, and fear.

Sequential feature selection procedures were applied for all classification stages. This ought to have ensured the employment of the most effective feature set in every stage.

For comparison purposes, we analyzed three different cases of the single-stage scheme and two cases of the multi-stage classification scheme:

- *S-ALL* – single-stage classification using the entire set of 6552 features. This case will be the basic level for the comparison of results;
- *S-SFS* – single-stage classification scheme using the feature set obtained by SFS technique;
- *S-SFFS* – single-stage classification using the feature set obtained by SFFS technique;
- *MS-SFS* – our proposed multi-stage classification scheme using the SFS-based feature set;
- *MS-SFFS* – multi-stage classification using the SFFS-based feature set.

The averaged recognition results of emotions for the German and Lithuanian datasets are given in Tables I and II, respectively.

Recognition results in the cases of *S-SFFS* and *MS-SFFS* were almost identical to the results of *S-SFS* and *MS-SFS*, respectively; therefore, they are not given in the Tables. There were only three different results between the SFS and SFFS procedures: two different results in the German case, and one difference in the Lithuanian case. As these differences varied form 0.1 % to 0.6 %, they can be considered as negligible.

We can see that recognition rates for an individual emotion depend heavily on the size of the analyzed emotion set. For example, the identification of anger in a single-stage scheme decreased on average by 25.9 % (for the cases of *S-ALL* and *S-SFS*) with the increase in the emotion set. In the case of multi-stage classification, this decrement was 19.7 %. For the emotion of joy, these values were 18.9 % and 9.8 %, respectively. Thus, multi-stage classification scheme using selection procedures is more robust to the increase in analyzed emotions than the single-step scheme.

TABLE I

RECOGNITION RESULTS OF GERMAN SPEECH EMOTIONS

Scheme	Number of emotions	Anger	Joy	Neutral	Sadness	Fear	Average
		Recognition accuracy, %					
S-ALL	2	53.3	58.3	—	—	—	55.8
	3	53.3	28.3	50.0	—	—	43.9
	4	53.3	28.3	48.3	58.3	—	47.1
	5	50.0	20.0	38.3	48.3	36.7	38.7
S-SFS	2	88.3	76.7	—	—	—	82.5
	3	91.7	75.0	95.0	—	—	87.2
	4	90.0	73.3	86.7	85.0	—	83.8
	5	86.7	63.3	80.0	81.7	46.7	71.7
MS-SFS	2	88.3	76.7	—	—	—	82.5
	3	86.7	75.0	96.7	—	—	86.1
	4	88.3	73.3	98.3	98.3	—	89.6
	5	88.3	80.0	83.3	96.7	60.0	81.7

TABLE II

RECOGNITION RESULTS OF LITHUANIAN SPEECH EMOTIONS

Scheme	Number of emotions	Anger	Joy	Neutral	Sadness	Fear	Average
		Recognition accuracy, %					
S-ALL	2	38.3	78.0	—	—	—	58.2
	3	16.0	59.0	67.7	—	—	47.6
	4	13.7	55.7	38.0	29.3	—	34.2
	5	8.7	48.3	23.7	24.7	29.7	27.0
S-SFS	2	81.3	87.3	—	—	—	84.3
	3	49.7	69.7	90.7	—	—	70.0
	4	50.7	77.0	72.7	54.7	—	63.8
	5	7.3	77.3	68.3	50.3	52.0	51.1
MS-SFS	2	81.3	87.3	—	—	—	84.3
	3	78.0	80.7	89.0	—	—	82.6
	4	68.0	84.7	85.3	68.7	—	76.7
	5	42.0	71.0	82.7	58.7	62.3	63.3

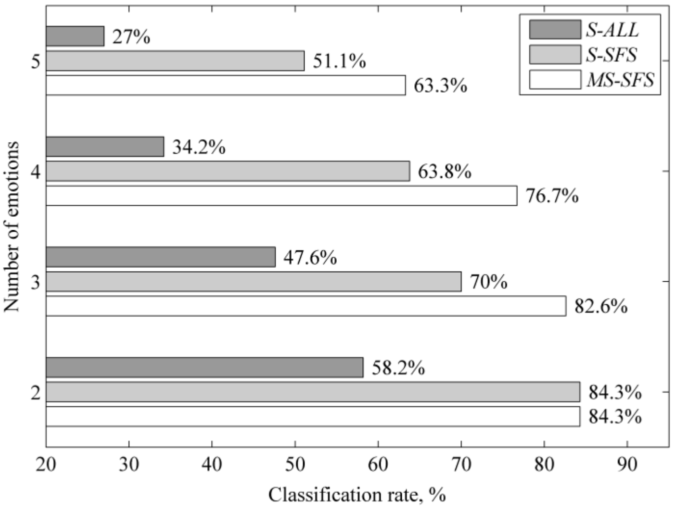

Fig. 3. Average recognition rates for Lithuanian dataset.

Fig. 3 shows the average emotion recognition rates and their dependence on the number of analyzed emotion for Lithuanian dataset (the German case is very similar therefore it is not given).

As we could expect, the expansion of the emotion set up to five emotions reduced the recognition accuracy. In the case of the entire feature set, the decrease was substantial: 31 % (17 % for German data-set). Similar values were obtained in the case of the SFS-based feature set: 33 % (and almost 11 % for German dataset).

The employment of multi-stage classification gave a more robust emotion recognition – the decrease was 21 % (and almost 1 % for German case). Thus, the decrease in classification rate caused by the increase in emotion set was 10 % to 20 % smaller for the multi-stage scheme.

In general, the multi-stage classification outperformed the single-stage scheme by 12 % to 42 %. The advantage of the multi-stage classification scheme is obvious.

In the publications surveyed, the average emotion recognition rates varied from 80 % up to 97 %. From this point of view, the recognition rates of 82 % to 90 % (for German dataset) obtained in our research seem quite competitive.

The differences in the results between the German and Lithuanian datasets presumably can be explained by the different acting level of speakers. Emotions in German were conveyed more precisely and expressively than in Lithuanian, which capacitated for a higher classification rate of German speech emotions. The morphological nature of the Lithuanian language could also make impact on the final results.

Fig. 4 represents the dependence of the size of the selected feature set on the number of analyzed emotions (again, the SFFS cases of the single-stage and multi-stage scheme are not given because of almost identical results to the SFS case).

Firstly, both SFS and SFFS procedures gave almost identical results (there were two cases when SFFS returned one-feature smaller sets). Considering the higher computational load of the SFFS procedure, the usage of this procedure is arguable. Slightly smaller feature sets is not a sufficient argument for a higher load during the feature selection step (the training phase).

Secondly, the Lithuanian speech emotion case is characterized by higher-order feature sets (both for single- and multi-stage schemes). This could be explained by the five times larger dataset. A larger number of speech utterances brings higher variability and overlap of emotion features. This means lower recognition rate for the same feature size or larger feature sets to achieve a higher classification rate. The specificity of Lithuanian could also condition the final size of the feature set.

Thirdly, we can notice that the multi-stage scheme requires for higher-order feature sets than the single-stage case. The difference varied from 1 (case of three emotions for German dataset) up to 14 (case of five emotions for Lithuanian dataset). This is an expected result of applying separate classifiers for different emotional groups.

Finally, the total order of the feature sets obtained in our research varied from the 4th up to the 30th (in the case of multi-stage classification). In comparison with the data in the literature (where the 56th–228th order feature sets were presented), these values are really low.

V. Conclusion

Sequential Forward Selection and Sequential Floating Forward Selection techniques were applied for multi-stage classification of speech emotions. Selection techniques were employed for every stage of the classification thus obtaining most effective feature subsets for every analyzed emotion group. Experimental results prove the advantage of the multi-stage classification of speech emotions.

We conclude our research study with the following statements:

- Sequential-forward-feature-selection-based multi-stage classification gives higher recognition than the single-stage scheme. Superiority of multi-stage scheme was 12 % to 42 % for different emotion sets.
- Multi-stage scheme has shown higher robustness to the growth of the set of analyzed emotions. The decrease in classification rate with the increase in emotion set for the multi-stage scheme was lower by 10 % to 20 % in comparison with the single-stage case.
- The SFS and SFFS techniques determined almost identical results in emotion classification. Differences in the obtained feature sets and classification rates were negligible.
- The order of feature sets was lower for the single-stage scheme. The difference between the single-stage and multi-stage schemes varied from 1 to 14 features. The highest obtained feature order was 30th.
- The dependence of feature order on the dataset size was observed. The order of feature set was approximately two times higher for the five times larger Lithuanian dataset.

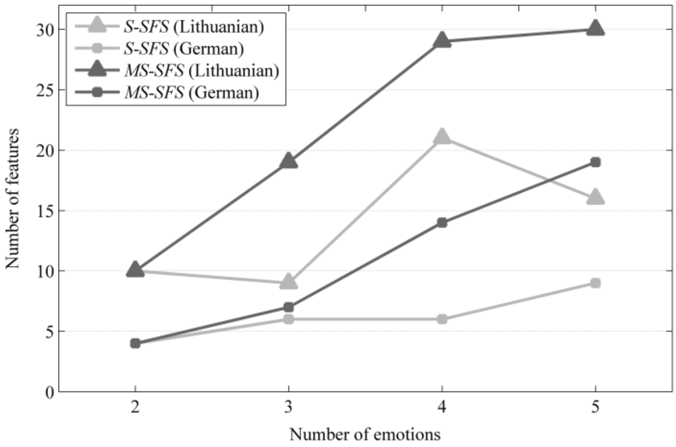

Fig. 4. Dependence of the feature set size on the number of analyzed emotions.

References

[1] S. Ramakrishnan and I. M. M. El Emary, ”Speech emotion recognition approaches in human computer interaction,” *Telecommun. Systems*, vol. 52, issue 3, pp. 1467–1478, Mar. 2013. https://doi.org/10.1007/s11235-011-9624-z

[2] S. G. Koolagudi and K. S. Rao, “Emotion recognition from speech: a review,” *Int. J. of Speech Technology*, vol. 15, issue 2, pp. 99–117, June 2012. https://doi.org/10.1007/s10772-011-9125-1

[3] Z. Xiao, E. Dellandrea, L. Chen and W. Dou, “Recognition of emotions in speech by a hierarchical approach,” in *2009 3rd Int. Conf. on Affective Computing and Intelligent Interaction and Workshops*, Amsterdam, 2009, pp. 1–8. https://doi.org/10.1109/acii.2009.5349587

[4] P. Giannoulis and G. Potamianos, “A hierarchical approach with feature selection for emotion recognition from speech,” in *Proc. of the Eighth Int. Conf. on Language Resources and Evaluation*, 2012, pp. 1203–1206.

[5] B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll and A. Wendemuth, “Acoustic Emotion Recognition: A Benchmark Comparison of Performances,” in *2009 IEEE Workshop on Automatic Speech Recognition & Understanding*, Merano, 2009, pp. 552–557. https://doi.org/10.1109/asru.2009.5372886

[6] A. Origlia, V. Galatà and B. Ludusan, “Automatic classification of emotions via global and local prosodic features on a multilingual emotional database,” in *Proc. of Speech Prosody*, 2010.

[7] M. Lugger, M.-E. Janoir and B. Yang, “Combining classifiers with diverse feature sets for robust speaker independent emotion recognition,” in *2009 17th European Signal Processing Conf.*, Glasgow, 2009, pp. 1225–1229.

[8] H. Peng, F. Long and C. Ding, “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 1226–1238, Aug. 2005. https://doi.org/10.1109/TPAMI.2005.159

[9] A. Mencattini, E. Martinelli, G. Costantini, M. Todisco, B. Basile, M. Bozzali and N. Di Corrado, “Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure,” *Knowledge-Based Systems*, vol. 63, pp. 68–81, June 2014. https://doi.org/10.1016/j.knosys.2014.03.019

[10] A. Milton and S. Tamil Selvi, “Class-specific multiple classifiers scheme to recognize emotions from speech signals,” *Comput. Speech and Language*, vol. 28, issue 3, pp. 727–742, May 2014. https://doi.org/10.1016/j.csl.2013.08.004

[11] L. Chen, X. Mao, Y. Xue and L. L. Cheng, “Speech emotion recognition: Features and classification models,” *Digital Signal Processing*, pp. 1154–1160, Dec. 2012. https://doi.org/10.1016/j.dsp.2012.05.007

[12] W.-J. Yoon and K.-S. Park, “Building robust emotion recognition system on heterogeneous speech databases,” in *2011 IEEE Int. Conf. on Consumer Electronics (ICCE)*, Las Vegas, NV, 2011, pp. 825–826. https://doi.org/10.1109/ICCE.2011.5722886

[13] J. Liu, C. Chen, J. Bu, M. You and J. Tao, "Speech Emotion Recognition using an Enhanced Co-Training Algorithm," in *2007 IEEE Int. Conf. on Multimedia and Expo*, Beijing, 2007, pp. 999–1002. https://doi.org/10.1109/ICME.2007.4284821

[14] M. Kotti and F. Paternò, "Speaker-independent emotion recognition exploiting a psychologically-inspired binary cascade classification schema," *Int. J. of Speech Technology*, vol. 15, issue 2, pp. 131–150, June 2012. https://doi.org/10.1007/s10772-012-9127-7

[15] G. Tamulevicius and T. Liogiene, "Low-order multi-level features for speech emotion recognition," *Baltic J. of Modern Computing*, vol. 3, no. 4, pp. 234–247, 2015.

[16] T. Liogiene and G. Tamulevicius, "Minimal cross-correlation criterion for speech emotion multi-level feature selection," in *Proc. of the Open Conf. of Electrical, Electronic and Information Sciences (eStream)*, Vilnius, 2015, pp. 1–4. https://doi.org/10.1109/estream.2015.7119492

[17] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier and B. Weiss, "A database of German emotional speech," in *Proc. of Interspeech*, Lissabon, 2005, pp. 1517–1520.

[18] J. Matuzas, T. Tišina, G. Drabavičius and L. Markevičiūtė, "Lithuanian Spoken Language Emotions Database," Baltic Institute of Advanced Language, 2015. [Online]. Available: http://datasets.bpti.lt/lithuanian-spoken-language-emotions-database/

[19] F. Eyben, M. Wollmer and B. Schuller, "OpenEAR – Introducing the munich open-source emotion and affect recognition toolkit," in *2009 3rd Int. Conf. on Affective Computing and Intelligent Interaction and Workshops*, Amsterdam, 2009, pp. 1–6. https://doi.org/10.1109/acii.2009.5349350

Tatjana Liogienė received the B.Sc. and M.Sc. degree in informatics from Lithuanian University of Educational Sciences in 2003 and 2005, respectively. Since 2005, she is a Lecturer at the University of Applied Sciences. At present, she is a Doctoral student at the Recognition Process Department of the Institute of Mathematics and Informatics of Vilnius University. Her fields of technical interest are speech signal processing and speech emotion recognition.

Address: Vilnius University, Institute of Mathematics and Informatics, Akademijos str. 4, Vilnius, LT-08663, Lithuania.

E-mail: tatjana.liogiene@mii.vu.lt

Gintautas Tamulevičius is an Associate Professor at the Department of Electronic Systems of Vilnius Gediminas Technical University. He received the Ph.D. degree in informatics engineering from Vilnius Gediminas Technical University in 2008.

His research interests include signal modeling, speech recognition, and speech emotion recognition.

Address: Vilnius Gediminas Technical University, Department of Electronic Systems, Naugarduko str. 4, Room 427, Vilnius, LT-03227, Lithuania.

E-mail: gintautas.tamulevicius@vgtu.lt