

Received October 11, 2021, accepted November 4, 2021, date of publication November 11, 2021, date of current version November 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3127581

# Scalogram as a Representation of Emotional Speech

PAWEŁ POWROZNIK<sup>1</sup>, PIOTR WOJCICKI<sup>1</sup>, AND SŁAWOMIR W. PRZYLUCKI<sup>1</sup>

Department of Computer Science, Lublin University of Technology, 20-618 Lublin, Poland

Corresponding author: Piotr Wojcicki (p.wojcicki@pollub.pl)

**ABSTRACT** It is very hard to implement the emotion recognition system based on spoken text. Computer applications have a huge problem with understanding non-literal meaning of statements as well as irony or a situational joke. The article describes how to represent emotional speech in the form of scalograms which are the result of speech signal processing by Discrete Wavelet Transform (DTW). The method of processing scalograms in order to extract input data for natural language processing algorithms in order to recognise the emotional state is also presented. The following emotional states were considered during the research: joy, anger, boredom, sadness, fear and neutral state. The developed method has been tested on databases containing recordings of emotional speech in the following languages: Polish, English, German and Danish. Depending on the language and classifier used, obtained results ranged from over 62% to over 94%. The use of fuzzy classifiers greatly improves the time and efficiency of classification.

**INDEX TERMS** Discrete wavelet transforms, emotion recognition, fuzzy neural networks, speech analysis.

## I. INTRODUCTION

Affective computer science is a field of research that is closely related to the processing, interpretation and recognition of emotions and other affective phenomena. Nowadays, this area plays an increasingly important role in supporting technology. Research carried out all over the world shows that computers are no longer completely indifferent logical machines. They may be able to interpret user's needs or feelings and provide feedback in a much easier way for non-technical users to understand.

Speech is the most popular form of interpersonal communication. However, it is also the most difficult signal for understanding and correctly interpreting not only by machines but also by other people. Research carried out by Swiss scientists shows that under certain conditions a statistical person is able to correctly recognise the emotional state of an unknown person only in 60% of cases [1]. There are IT solutions that try to deal with non-verbal speech signal processing, but the results achieved are far from desirable. This is due to the characteristics of the speech signal, which, in addition to its semantic value, also transmits other information, e.g. related to the mood or emotional state of the speaker. Moreover, extra-linguistic values vary according to the age of

the speaker, the gender or the country of origin. Non-verbal communication, such as gestures, also affects the meaning of spoken text, which is understandable and interpretable for the participants of the conversation, but which is an additional difficulty in processing the speech signal. The same sentence spoken in a different context can have a completely different meaning or emotional character. For example, the word "OKAY" is used to express consent, assurance, but also admiration, disbelief or lack of interest. Therefore, just understanding the text is not enough to understand the whole sentence correctly. This makes it necessary for computer systems to recognise emotions correctly.

One of the biggest problems in recognising emotional speech is the multitude of emotional states with only a subtle difference. There is only a thin line between emotions such as fear and terror, rage and anger, or ecstasy and joy. Developing a system that correctly recognises the most complex emotional problem is not trivial. Also existing models of emotional speech focus primarily on extracting from the speech signal parameters commonly used in sound processing. Of course, some models also include parameters such as laryngeal tones and various types of coefficients. However, attempts at a graphic representation of emotional speech along with the way it is processed are very rare.

The article presents a graphic way of representing emotional speech in the form of scalograms, and provides a

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas<sup>1</sup>.

practical way of processing it, allowing for unambiguous classification of the speaker's emotional state. The model includes such parameters as the level of signal decomposition or the values of parameters simplifying scalograms. The influence of fuzzy logic assumptions on the quality of classification and the time of preparing the emotional speech model are also presented.

In research based on speech recordings, the response time of classifier had a great practical importance. A common approach was to use deep learning methods during image data analysing which produced very good results. However, due to the time needed to process the data, the authors decided to use fuzzy classifiers, which generated a satisfactory response in a significantly shorter time.

## II. RESEARCH ISSUE

The main research problem was to find and describe an emotional speech representation which would be easy to obtain and further process. It was therefore necessary to determine and extract those parameters on which emotions have the greatest impact. Another problem was to find an effective tool to correctly classify the emotional state of the speaker. There are several obstacles that complicate the desired solutions. Moreover, there was a research suspicion that due to the lack of strict unambiguity in the identification of emotions, the use of fuzzy classifiers may give better results.

First, from the psychological and social point of view emotions are real complex and subjective issues. Each person interprets emotions a bit differently. Moreover, there are many difficulties in defining emotions [2].

Secondly, assigning emotions to the recorded audio signal is also a difficult task. Its complexity increases when the emotional state of people speaking a different language is identified. The studies described in [3] show that even with geographical and cultural proximity, the results of emotional recognition are approximately 68%. In the case of geographical distance, they are 10-15% worse. The third obstacle is related to the complexity and cost of collecting the database. It seems that, of the millions of recordings available in the Internet, the separation of several hundred should not be a problem. However, publicly available amateur recordings are of relatively low quality, information services are usually devoid of emotion, and in films and radio programs, emotions are usually played back.

As has been shown, the recognition of emotional speech is a difficult task. However, there are practical applications used in everyday life. Mood recognition systems can be used in car systems to warn drivers. This helps to avoid accidents caused by the mental state of the driver [4]. Analysing conversations from a Call-Centre and e-tutorials [5], as well as storytelling [6] and medical applications [7] are further uses of the systems for recognising emotions [8]. In aeroplanes, emotional state identification systems are used to classify stressed speech [9].

Recognising emotional states on the basis of voice recordings only is an issue that is quite often discussed in the

literature. However, the available solutions do not give satisfactory results in the case of the Polish language, the processing of which was the main research topic. The existing methods allow to achieve positive classifications at the level of 86% [10]. This is a high value, however, in the era of dynamic development of artificial intelligence and human-computer systems it is unsatisfactory. The proposed method using the wavelet transform in combination with a number of classifiers has resulted in much better results reaching the level of 94%. In addition, the use of scalograms, along with a slight modification of the method of obtaining them (changing the value of the cut-off threshold – discussed more detailed in Chapter III), allows it to be used also for other languages. The difference in the threshold value, according to the authors, results from the presence of specific sounds, characteristic of each language, and not occurring in others. Another research task was to check how large the test set should be for the classification to bring satisfactory results. This is an important issue in the context of future research works due to the small number of recordings for some languages and the need to determine the cut-off threshold value. The conducted research has shown that relatively small training sets are able to correctly teach some (especially fuzzy) classifiers.

## A. RELATED WORKS

Constructing effective data structures describing objects of a subjective nature is a difficult issue. The most common emotional speech models include a set of parameters extracted directly from the speech signal, such as the Fundamental Frequency (F0), Formant Frequencies [11], Mel Frequency Cepstral Coefficients (MFCC) [12]–[14], Linear Predictive Coding (LPC) [15], [16]. Emotional speech models based on data obtained from statistical speech signal parameters constitute another group [17]. The next proposed models are supported by parameters related to speech energy and power. There are also publications in which the emotional speech model is presented in graphical form, e.g. spectrograms [18]–[20]. The combined spectral and prosody features are also considered for the task of emotion recognition [21].

Since an emotion can be expressed in a gentle or intense way, it is worth trying to distinguish it. In paper [22] the intensity in expressing emotions was examined. Frequency and amplitude were used as coefficients of feature extraction. This is directly because anger increases the speed of speech and sadness slows it down. Conversely, the intensity is higher if the speech is happy or angry. The research described Glottal Pulse Frequency (GPF) method with a model prediction accuracy of approximately 83%. A comparative analysis of GPF and MFCC methods with modifications was carried out. It has been shown that for specific emotions that greatly affect the frequency of speech (i.e. anger), the MFCC method works better, while the remaining emotions are better detected by the GPF method proposed in the study. This is due to the fact

that GPF is based on voice signal analysis, not only frequency domain.

Multi-model recognition of emotions (audio and video), due to the complementary of information is proposed as an extended approach to the detecting emotions challenge [23]. Emotions are analysed in two stages, recognition audio channel as the first and as the second video frame (depending on whether the audio channel contained relevant information). Multi-Layer Perceptron Back Propagation network is used for both classification – voice and image respectively with classification rate for isolated audio signal around 90%. The combination of audio and video classification increases the accuracy to about 95%, with much more complexity in data processing. A similar approach to multi-model analysis is presented in [24].

A combination of several emotions along with analysis is shown in [25]. A comprehensive approach is very similar to the human perception of emotions, because usually emotions do not only occur individually, and sometimes are components of complex emotional states. The study shows isolating the distinguishing features of the speech signal and analyses anxiety as a combination of three basic emotions, namely anger, fear and sadness. The overall classification accuracy of the neural network is around 70.3%

In paper [26], the problem of recognising emotions of whispered speech, which naturally differs from normal speech, is discussed. A solution extended by the autoencoder-based feature transfer learning is used. This approach allows the use of a known solution for purposes other than the original one, with the recognition accuracy at least at the same level as for normal speech.

Emotional speech classifiers are commonly used tools which are well known from speech signal processing problems as well as identification of other types of data. They include: Gaussian Models (GMM) [27], Hidden Markov Models (HMM) [28], Support Vector Machine (SVM) [29], the k-NN algorithm, or broadly understood artificial neural networks [30], [31]. There are also known publications in which several classifiers were involved in the process of identifying emotions. Different classification methods were juxtaposed in order to compare several alternative approaches for final voting [10].

The paper presents an innovative method of signal processing, not presented in the literature, allowing for an effective classification of the emotional state of the speaker. The obtained results are about 10% better than the currently available methods for the studied languages [32], [33]. In addition, it has been shown that the type of classifier used does not have a radical impact on the effectiveness of the classification, but is guaranteed by the method of processing the speech signal. Nevertheless, classifiers with the fuzzy part allow to obtain better results (on average by 10%) Moreover, it has been shown that the appropriate selection of classifiers, even with small training sets (about 120-150 elements), enables an effective classification of emotional states.

## B. DATABASES

Processing an emotionally charged speech signal requires an appropriate recording database. Publicly available databases can be divided into three categories:

- databases containing played back recordings,
- databases containing induced recordings,
- databases containing natural speech recordings.

It is obvious that the most valuable group of recordings are samples from the third group. However, due to the difficulties in obtaining recordings, such databases are almost inaccessible. The first two groups form the basis of research used in many studies [2], [3], [12]. For the purposes of this article, four different databases containing speech recordings in four languages were used.

The first database (Database A) was prepared and made available by the employees of the Technical University of Łódź [34]. It contains 240 sentences in Polish in six emotional states, i.e.: fear, joy, anger, boredom, sadness and neutral state. The sentences were uttered by eight professional actors of both sexes.

The second database (Database B) was prepared at the Centre for Strategic Technology Research [35]. The corpus contains 700 statements in English in five emotional states: anger, joy, sadness, fear and neutrality. The database was prepared by 30 actors. The assignment of a particular emotional state to a recording has been verified by 30 people not involved in acting.

An often used resource in studies related to the processing of emotional speech is the Berlin Emotional Database (BES) [36]. It contains recordings in six emotional states: joy, anger, anxiety, fear, boredom and indignation. The recordings were prepared by ten people professionally involved in acting and checked by 20 people, who, after listening to the recording once, had to assign an emotional state to it. Utterances with a recognition rate better than 80% were chosen for further analysis. Finally, out of over 800 recordings, 300 remain. This database (Database C) was also used in the research carried out by the authors of this article.

The fourth database was prepared in the Centre for Personal Communication at Aalborg University (Database D) [37]. The corpus consists of recordings in Danish. In the case of the Danish Emotional Speech Database (DES), registration was made of: 2 single words, 9 sentences and 2 fragments of fluent speech. The above sentences were spoken by four actors in four emotional versions: sadness, surprise, joy, anger. The database also contains 18 recordings spoken in a neutral voice. The correctness of assigning emotions to samples was verified by 20 people aged 18-58.

## C. WAVELET TRANSFORM

A Continuous Wavelet Transform (CWT) was developed by Jean Morlet and Alex Grossman for one-dimensional, time dependent  $t$ , signals  $x(t) \in L^2(\mathbb{R})$  is expressed

in the following form [38]:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\Psi^* \frac{t-b}{a} dt, \quad (1)$$

where  $a, (a > 0)$  – scale parameter,  $b$  – offset parameter,  $\Psi$  – mother wavelet,  $*$  – indicates Conjugate complex function.

Despite the many advantages of the continuous wavelet transform, issues related to signal processing more often employ its discreet form. One of the reasons for this state is that computer systems process discrete signal instead of its continuous form, so the use of Discrete Wavelet Transform (DWT) seems to be a natural approach. A discrete wavelet transform is defined as follows [38]:

$$DWT(m, n) = \frac{1}{\sqrt{a^m}} \sum_n s(k)\Psi(a^{-m}n - bk), \quad (2)$$

where  $a, (a > 0)$  – scale parameter,  $s(k)$  – input signal,  $\Psi$  – mother wavelet,  $b$  – offset parameter.

Fourier transform is often used in speech signal processing, but in the current research it was decided to use discrete wavelet transform because DWT provides accurate and undistorted time information, which is a significant improvement in signal processing. The discrete wavelet transformer has been used to solve many problems directly related to audio signal analysis [39]–[41]

Using a wavelet transform, signal  $S$  can be represented in scale  $J$  by wavelet coefficients  $D_{J,K}$  and scaling coefficient  $A_{J,K}$  [42]

$$S = \sum_k A_{J,K} \phi_{J,K} + \sum_{j=1}^J \sum_k D_{J,K} \psi_{J,K} \quad (3)$$

where  $\phi_{J,K}$  and  $\psi_{J,K}$  are scaling function and wavelet, respectively, in scale  $J$  and shifted  $k$  times.

The scaling coefficient  $A_{J,K}$  and wavelet coefficients  $D_{J,K}$  can be determined as follows:

$$A_{J,K} = \int \phi_{J,K}(t)d(t), \quad (4)$$

$$D_{J,K} = \int \psi_{J,K}(t)d(t). \quad (5)$$

The wavelet function  $\psi_{J,K}(t)$  is described as the high pass filter with coefficients  $h_1(k)$ :

$$\psi_{J,K}(t) = \sum h_1(k)\phi_{J-1,K}(t). \quad (6)$$

The scaling function  $\psi$  is described as a low pass filter with coefficients  $h_0(k)$ :

$$\phi_{J,K}(t) = \sum h_0(k)\phi_{J-1,K}(t). \quad (7)$$

There is no absolute way to choose a particular wavelet. The choice depends on type the of signal to be analysed and that choice is very often based on the experience of researchers. There are several commonly used wavelet families like: Haar, Biorthogonal, Coiflets, Daubechies Morlet,

Mexican Hat or Meyer. At the beginning of this study, based on heuristic selection methods, reverse biorthogonal wavelet (rbio1.3) and Haar wavelet were used to extract the speech signal features. In the initial phase of the research, slightly better results were obtained with the use of the rbio1.3 wavelet, therefore the described studies consider only the use of this method.

#### D. SCALOGRAM CREATION

Scalogram is one of the ways of representing the results of the wavelet transform in the time-frequency space. It shows the energy density obtained by CTW. When applying the CWT to finite discrete time series, a choice for the discretization must be made. With an appropriate sampling rate, CWT can be interpreted as projecting the signal onto successive versions of the wavelet  $\psi$  shifted by  $t$  and scaled by  $a$ . The MATLAB environment used in the conducted research offers a set of tools enabling this operation to be performed. As a result, the scalogram  $S$  of  $f$  can be expressed in the following form:

$$S(s) = ||Wf(a, b)|| = \left( \int_{-\infty}^{+\infty} |Wf(a, b)|^2 da \right)^{\frac{1}{2}}. \quad (8)$$

Equation (8) represents the energy of  $Wf$  at a scale  $s$ . Thus, the scalogram allows the detection of the most representative frequencies of a signal, that is the frequencies that contribute the most to the total energy of the signal. An example of scalogram is shown in Figure 4 and the method of its processing is described by Algorithm 1.

#### E. DATASETS FUZZYFICATION

*Definition 1:* A non-empty fuzzy set  $f_s$  can be understood as an ordered pair  $(f_s, \eta_{f_s})$ , where  $\eta_{f_s}$  is a membership function  $\eta_{f_s} : f_s \mapsto [0, 1]$ , that allows to perform the fuzzyfication operation.  $\eta_{f_s}$  assigns to each element  $x$  in  $f_s$  a degree of membership,  $0 \leq \sigma(x) \leq 1$  [43].

*Definition 2:* A fuzzy relation on  $f_s$  is a fuzzy subset of  $f_s \times f_s$ . A fuzzy relation  $\eta_{f_s}$  on  $f_s$  is a fuzzy relation on the fuzzy subset  $\sigma$ , if  $\eta_{f_s}(x, y) \leq \sigma(x) \wedge \sigma(y)$  for all  $x, y$  from  $f_s$  and  $\wedge$  stands for minimum. A fuzzy relation  $\eta_{f_s}$  on  $f_s$  is said to be symmetric if  $\eta_{f_s}(x, y) = \eta_{f_s}(y, x)$  for all  $x, y \in f_s$  [43].

*Definition 3:* A fuzzy graph is a pair  $G : (\sigma, \eta_{f_s})$  where  $\sigma$  is a fuzzy subset of  $f_s$ ,  $\eta_{f_s}$  is a symmetric fuzzy relation on  $\sigma$  [43].

*Definition 4:*  $(\sigma', \eta'_{f_s})$  is a fuzzy sub graph if  $\sigma' \subseteq \sigma$  and  $\eta'_{f_s} \subseteq \eta_{f_s}$  [43].

The usage of fuzzy logic in classifiers allows modelling “uncertain” phenomena [44]. Therefore, the fuzzy classifiers is able to imitate the way people perceive the environment. Applying fuzzy logic rules, sharp boundaries between the analysed sets are blurred. Fuzzyfication and defuzzyfication procedures allow to transform sets from one state to another. The input data fuzzyfication process was carried out in the Matlab environment using the



**Algorithm 1** Scalogram Processing and Feature Vector Creation**Input:** Speech signal in form of sound file**Parameter:** Mother wavelet, decomposition level, threshold value, number of sub-areas**Output:** Feature vector

- 1: Normalised all signal samples to the range of [-1, 1].
- 2: Based on parameters create the scalogram.
- 3: Change scalogram into grey scale image.
- 4: Based on threshold value change grey scale image into binary image.
- 5: Divide modified scalogram into sub-areas.
- 6: Sum the values in each sub-area.
- 7: Add values to feature vector.
- 8: Add speaker's sex information to feature vector.

following membership functions:

$$\eta_{Rf_s}(x) = \begin{cases} 0, & (x > d) \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 1, & x < c \end{cases} \quad (9)$$

$$\eta_{f_s}(x) = \begin{cases} 0, & (x < a) \vee (x > d) \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x < c \\ \frac{d-x}{d-c}, & c \leq x \leq d \end{cases} \quad (10)$$

$$\eta_{Lf_s}(x) = \begin{cases} 0, & (x < a) \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases} \quad (11)$$

where  $a, b, c, d$  are trapezoidal functions parameters and  $a < b < c < d$ . In case of equation (5)  $a = b = -\infty$ , (7)  $c = d = \infty$ .

**III. EMOTIONAL SPEECH MODEL**

The speech signal processing was based on the wavelet transform described in previous section, which was the basis for the development of emotional speech representation. All studies were divided into two stages: the stage related to the creation of a scalogram and the extraction of features, and the stage of classification. The first of the stages was presented on Algorithm 1.

**A. INPUT PARAMETERS**

All algorithm parameters, except the mother wavelet, were determined in an experimental manner. The first problem was the choice of the level of signal decomposition used during wavelet processing operations. To determine this, the k-nearest neighbours (k-NN) algorithm was used. It works by assigning the analysed sample to the group of their neighbours whose occurrence is most numerous in the vicinity of  $k$  of its nearest neighbours [45]. In the case when several

competing groups are equidistant from the classified sample, the assignment to one of the classes is arbitrary. The quickest form of the k-NN rule is the 1-NN rule, which assigns an unknown test sample to the class of its nearest neighbour. The Statlog project [46], in which several classifiers were compared on more than twenty large databases, showed that for 75% of tests the best value of the  $k$  parameter was 1. This rule was used to define the input parameters of the Algorithm 1. The scalogram processing scheme was also presented in Figure 1. The first two stages: noise reduction and values normalisation were carried out in Matlab environment. In some databases (i.e. Database A) the noise reduction was omitted due to quality of recordings.

The level of signal decomposition by a DWT had a direct impact on the appearance of the scalogram, and thus on the way of representing the emotional speech signal. The higher the degree of decomposition, the more bands appear on the scalogram. Moreover, the increase of signal decomposition level influence on the time necessary to perform calculations. The obtained results were presented in Figure 2.

In all databases, single emotions were assigned to individual recordings. The efficiency of classification is understood as the correct identification of these emotions. In other words, if the emotion label assigned to the recording matches the label specified by the classifier, then the identification is said to be correct.

It can be observed that up to the seventh level of signal decomposition the classification efficiency increases, then up to the eleventh level it remains at almost the same level, and from the twelfth level it slightly decreases. Due to the time needed to process a single recording, decomposition level 7 was selected for further investigation.

Fuzzy logic is used in quite a different situation. Signal decomposition at the fourth level gives satisfactory classification results. Moreover, further decomposition of signals only slightly improves the effectiveness of the identification of emotion in speech, at considerable computational effort. It should also be noted that the fuzzification resulted in a significant improvement in classification efficiency for lower levels of signal decomposition. As shown later in the paper, for the analysed databases, as an example for the fourth level of decomposition, improvement is from 34-37% to 64-67%. In further research it was decided to use the fourth level of emotional speech signal decomposition.

Another parameter that had to be determined was the threshold value used when converting the scalogram into a binary image. The transformation consisted in changing the pixel value above the threshold to 1 and below to 0. Based on previous experiments [19], all pixel values between 50 and 200 were examined, while pixel values below 50 and above 200 were omitted. It has been observed that for each of the databases tested, and thus for different languages, the threshold level is not the same. The best efficiency of classification was achieved when the threshold was 101 for database A, 123 for database B, 111 for database C and 117 for database D.

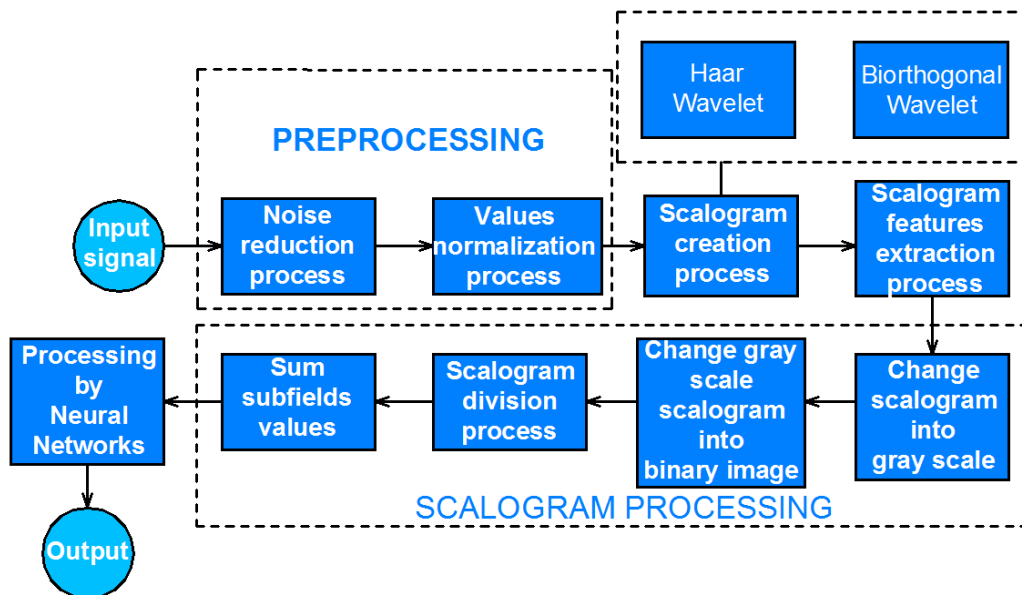


FIGURE 1. Speech signal processing scheme.

The first step in processing the scalogram was to convert it to a grey-scale image and then to a binary form. A similar method of determining the threshold was used here as in the case of scalograms. Values ranging from 50 to 200 were tested. The best results were obtained for the threshold of 100. Thus, all values below 100 were changed to 0, and those above to 1. The next step was to divide the scalogram into smaller areas. The division was made in such a way that the dividing lines coincided with the level of the wavelets used. In the case of using the 7th order wavelets, the scalogram was divided into: 70, 105, 140, 175, 210, 245, 270, 315, 350 and 385 areas, respectively.

The conversion of the image to binary form resulted in a decrease in the significance of information irrelevant to the problem of identifying emotions and sharpened those areas of the scalogram where the importance is greater. The difference in the threshold value for different databases may be due to the presence of characteristic sounds for each of the languages tested.

The degree of signal decomposition had a direct impact on the number of sub-areas of the scalogram, because it determined the number of bands, and thus the number of horizontal dividing lines. The vertical lines were determined in an experimental manner, determining the total number of sub-areas of the scalogram. Therefore the whole image consists of rectangles containing fragments of a binary scalogram. The impact of scalogram division for classification efficiency was presented in Figure 3.

It should be noted that, in the case of non-fuzzy data sets, the division of the scalogram into less than 175 sub-areas resulted in a much lower classification efficiency than when more input data was generated. The division into 210 sub-areas and more did not cause a significant increase in the effectiveness of classification; therefore, for further research

the division of the scalogram into 210 fields was used. Also the division of the scalogram into 210 sub-areas is important for the average processing time of a single recording (from the moment the file is loaded, through the generation of input parameters to the neural network, to the stage of classification). An example of scalogram division is presented in Figure 4.

If the fuzzy input data is considered, it can be observed that the presentation of the emotional speech signal in the form of a 160-subarea scalogram gave satisfactory results for the identification of emotions in speech.

Due to different ranges of speech frequency, differentiating the voices of men and women, the last information added to the parameters representing emotional speech was the speaker's sex.

### B. EMOTIONS REASONING

Recent years have brought a dynamic development of work related to artificial neural networks. Such structures as convulsive neural networks [47] and GMDH neural networks [48], [49] are commonly used. However, artificial neural networks (ANN) or Kohonen networks are still widely used. During the tests, artificial neural networks of different structures were used as classifiers. The structure of the network was closely related to the division of the scalogram.

The set of the most popular classification methods such as Multilayer Perceptron Classifier (MPC), Naive Bayes Network Classifier (NBNC), Decision Trees Classifier (DTC), Probabilistic Neural Network Classifier (PNNC), Random Forest Classifier (RFC), Fuzzy Multilayer Perceptron Classifier (FMPC), Fuzzy Rule Classifier (FRC) and Fuzzy Decision Trees Classifier (FDTC) were considered during research. All classifications were conducted separately for each language.

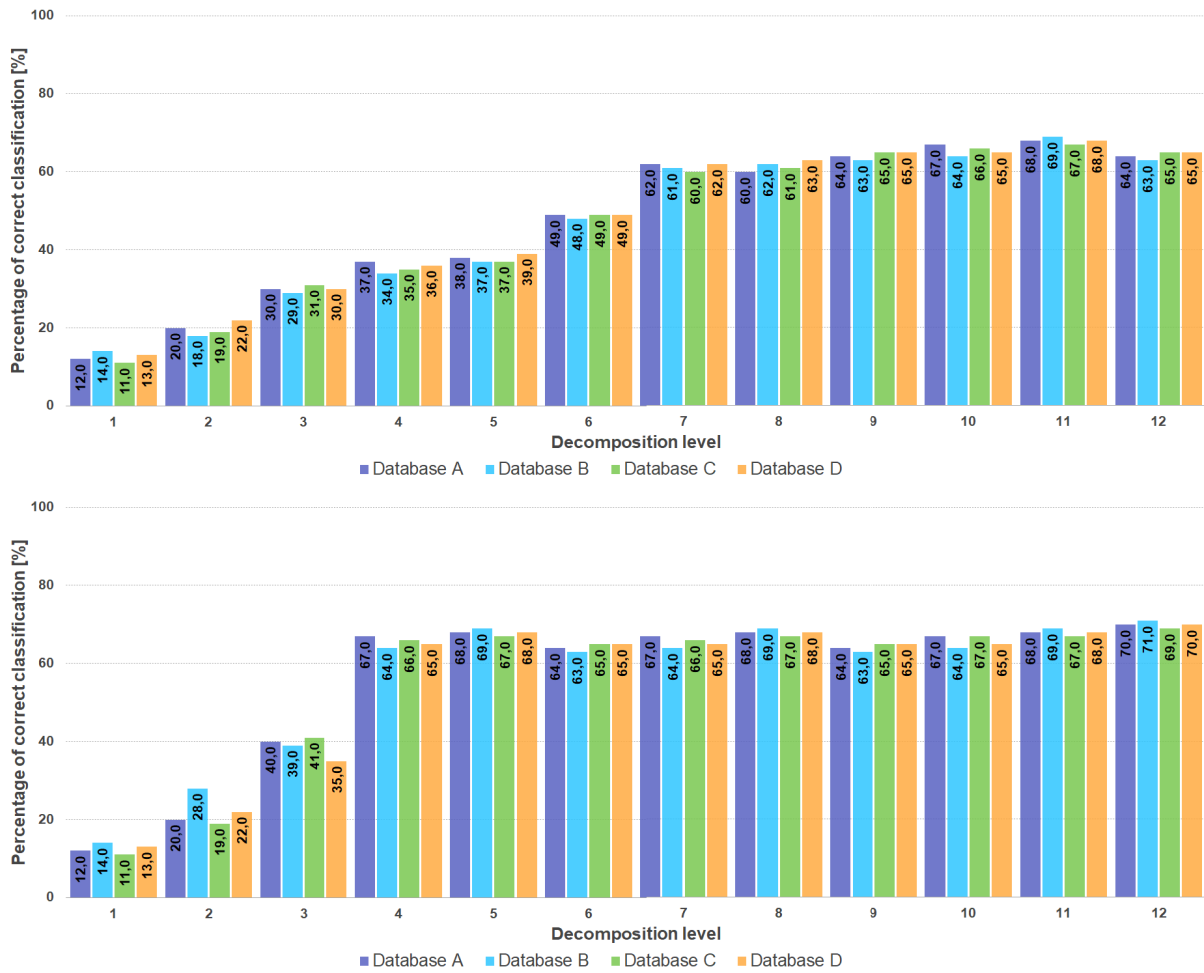


FIGURE 2. Results obtaining by k-NN classifier depending on the signal decomposition level for data sets without fuzzification (top) and with fuzzification (bottom).

1) MULTILAYER PERCEPTRON CLASSIFIER

A multilayer perceptron is a class of feedforward artificial neural network (ANN). This type of network usually consists of one input layer, several hidden layers and one output layer. Hidden layers usually consist of McCulloch-Pitts neurons [50]. Determining the right number of hidden layers and the number of neutrons in individual layers is a difficult issue. The number of neurons in the hidden layer was determined experimentally on the basis of the growth up technique. In all studies, hyperbolic tangent was used as a function to activate neurons. The network was taught using a backpropagation algorithm with adaptive learning rate and momentum.

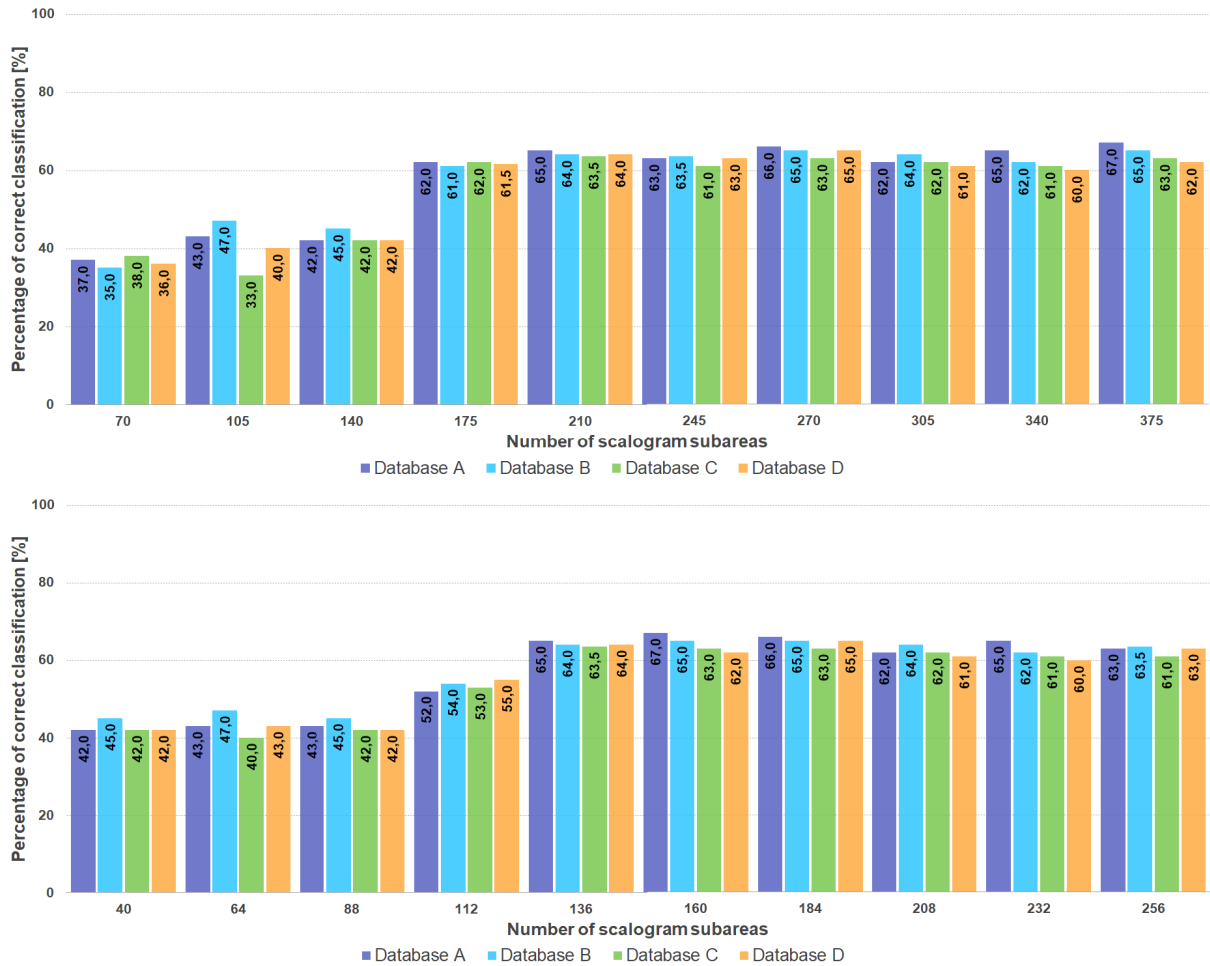
2) NAIVE BAYES NETWORK CLASSIFIER

NBC is a family of classifiers, where it was assumed that the individual predictors are independent of each other. In many practical aspects is unfortunately not true. The great advantage of this classic tool, apart from its simplicity, is its high stability and resistance to failure. This classifier uses the maximum a posteriori decision rule. The classification is valid as long as the correct class is more likely than others.

Methods using Bayesian classifiers are used in many empirical studies [51].

3) DECISION TREES

The decision tree is a graph-tree, which consists of a root, nodes, edges and leaves. The leaves are nodes from which no edges come out. The root of the tree is created by the selected attribute, while the individual branches represent the values of this attribute. Due to this construction, new objects that did not participate in the tree creation process can be classified. Decision trees are characterised by a hierarchical structure. In the next steps, the set of objects is divided by answering questions about the values of selected features or their linear combinations. The final decision depends on the answer to all questions. In the tree construction algorithms, one of the key elements is the selection of the order of the features according to which, at individual stages, the set of objects will be divided. The decision tree technique complements classical methods. An example would be discriminant analysis. Decision hierarchy is a feature that distinguishes decision trees from other methods [52].



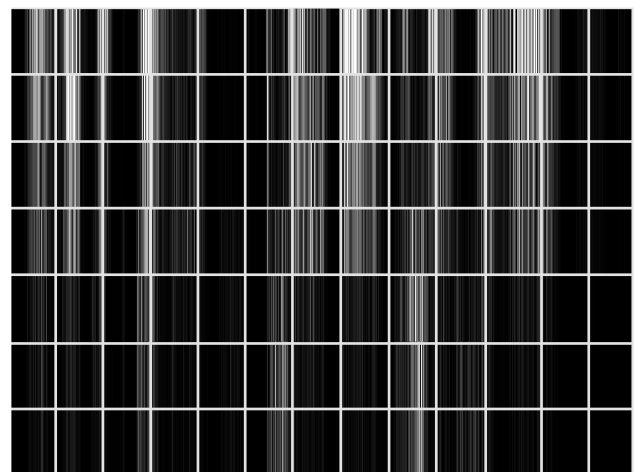
**FIGURE 3.** Results obtaining by k-NN classifier depending on the scalogram division for data sets without fuzzification (top) and with fuzzification (bottom).

#### 4) PROBABILISTIC NEURAL NETWORK CLASSIFIER

These are neural networks in which outputs are treated as the probabilities of individual possible solutions. PNN are radial networks usually with a number of neurons in a hidden layer equal to the number of learning cases. The basic feature of probabilistic networks is normalisation of output signal values such that their sum (on all network outputs) is 1. Then it can be assumed that the values at individual network outputs represent the probabilities of the categories (recognition) assigned to these outputs. A characteristic feature of this classifier is replacing the sigmoid activation function often used in neural networks with an exponential function, namely a probabilistic neural network that can compute nonlinear decision boundaries [53].

#### 5) RANDOM FOREST CLASSIFIER

Random forest is a team method of machine learning for classification, regression and other tasks, which consists in constructing many decision trees during learning and generating a class, which is the dominant of classes (classification) or the predicted average (regression) of individual trees. For



**FIGURE 4.** An example of scalogram division. The horizontal and vertical lines are auxiliary and physically absent in the scalogram. Marked to better visualise the division.

a given observation each tree returns a multiple probability of classification. Probabilities (decisions) from trees in the forest are treated as votes, and the result that receives the





FIGURE 5. Efficiency of the multilayer perceptron classifier.



FIGURE 6. Efficiency of the naive Bayes network classifier.

most votes is returned as the result. Votes can be weighted or independent of the voting method chosen by the user. Random decision forests improve the tendency of decision trees to over-adapt to the training set [54].

6) FUZZY MULTILAYER PERCEPTRON CLASSIFIER

A modified form of the Multilayer Perceptron Model is its fuzzy form. The structure of the network as well as the way it is trained are identical to the MLP model. The main difference is the existence of the membership function, whose task is the fuzzification of input data [55]. Due to the fuzzy membership function, this model is used to work on incomplete or imprecise data [56].

7) FUZZY RULE CLASSIFIER

In recent years, fuzzy classifier methods have been commonly used. In the literature there are a lot of studies that have worked on this type of classification [57], [58]. The rule for fuzzy classifiers is as follows: “Any classifier that

uses fuzzy sets or fuzzy logic during training or operation can be called fuzzy classifier” [59]. The choice of FRC is dictated by the fact that its algorithm generates rules based on numerical data, which are fuzzy intervals in spaces of higher dimensions. These hyper rectangles are defined by trapezoidal fuzzy membership functions for each dimension.

8) FUZZY DECISION TREES CLASSIFIER

Linking the assumptions of fuzzy logic and previously discussed decision trees allows the introduction of a more generalised creation known as fuzzy decision trees. They are commonly used wherever it is not possible to strictly separate the set of data analysed, in particular those related to language uncertainty.

9) CONDUCTED CLASSIFICATION TESTS

For each of the proposed classifiers, a series of tests was carried out consisting in randomly dividing the available data into a training set and a test set. When dividing the considered

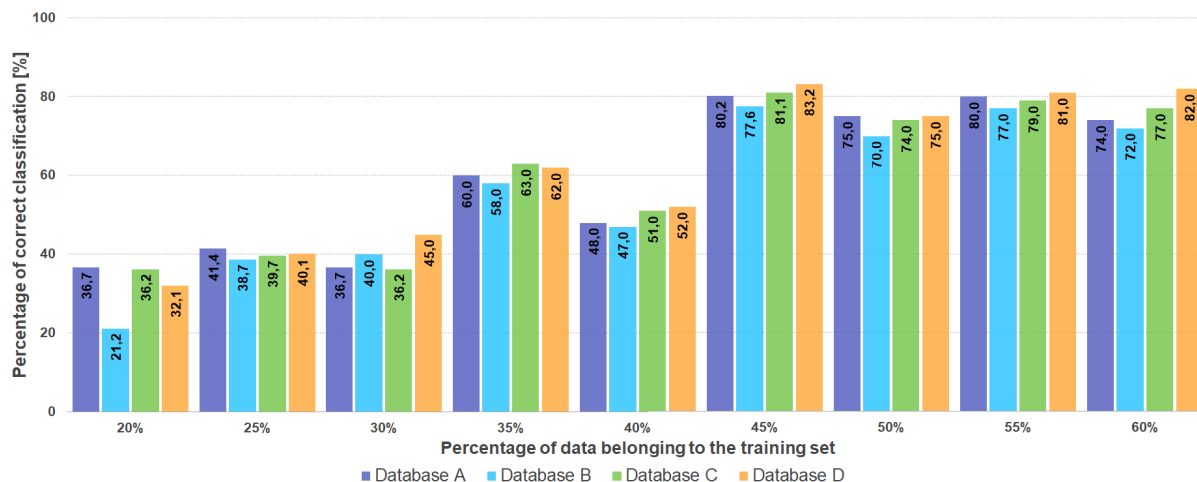


FIGURE 7. Efficiency of the decision trees classifier.

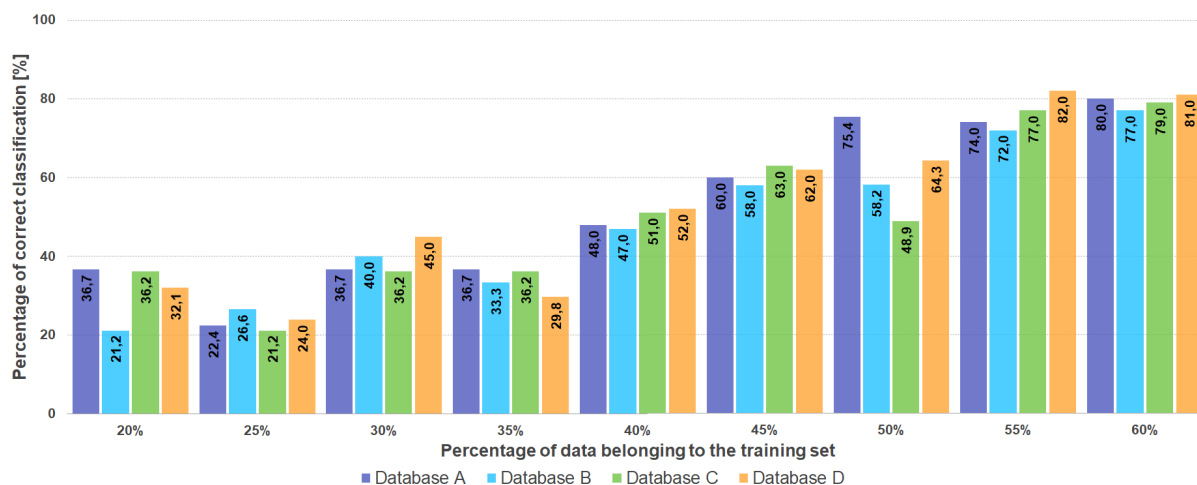


FIGURE 8. Efficiency of the probabilistic neural network classifier.

sets, different proportions were used between the sets, starting from 20% to 60% share of teaching collections. Each time the elements of the training and test set were added by randomisation. For example, first a basic training set covering 20% of the data set is selected at random. The rest of the data are taken as a test set. In the next approach, randomly selected elements are added, increasing the previously selected training set. Such a maneuver does not allow for the selection of 'better' data, because only part of the training set changes. Three independent replicate trials were performed for each division and the results obtained were collected.

The Figures 5–12 show the average classification results for all emotions in the databases. The results obtained in three consecutive tests were also averaged.

For Multilayer Perceptron Classifier (Fig. 5) the selection of elements for the training set has a large impact on the level of classification. The greatest variability of correct classifications for individual databases was noticed when the training set constituted 45% of the data. It should also be emphasised

that a significant jump in the correctness of classification in a situation where the training set constituted not less than 55% of all the data in each of the databases. Above this value, the percentage of correct classifications remains at a similar and high level.

In the case of a Naive Bayes Network (Fig. 6) a very good result of proper classification is noted. Even when the training set amounts to 35% of the elements, then around 90% correct classifications are noted. Unfortunately, in the case of different (random) selection of elements for the teaching set, the number of teaching set elements increases and the correct classification is not stable. Moreover there are also large differences in the correctness of classification between individual databases.

The effectiveness of classifiers based on Decision Trees (Fig. 7) depends to a large extent on the selection of elements for the training set. Unfortunately, no correlation can be noted between the number of elements in the training set and the classification efficiency. There are also large differences in

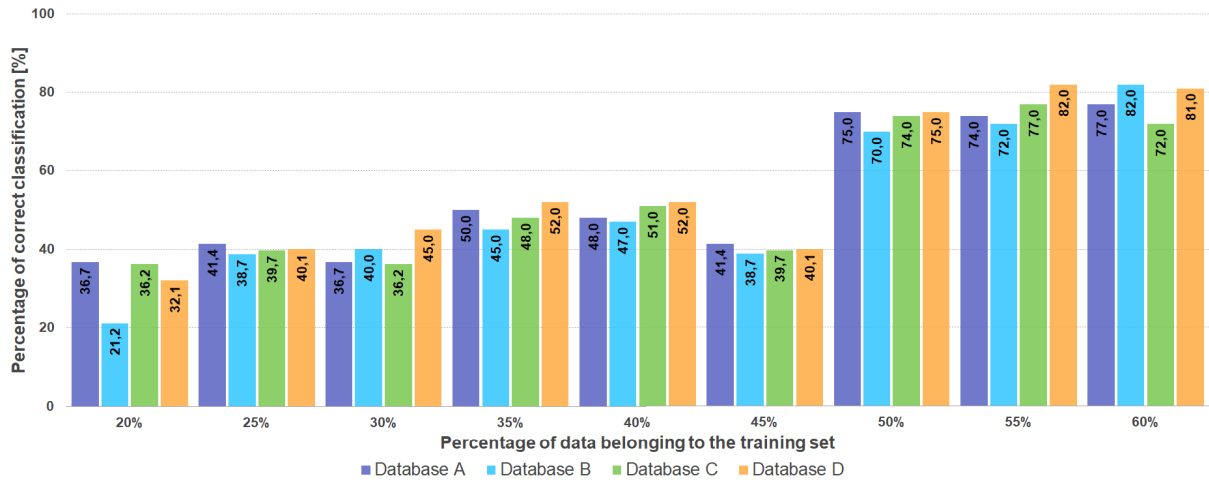


FIGURE 9. Efficiency of the random forest classifier.



FIGURE 10. Efficiency of the fuzzy multilayer perceptron classifier.

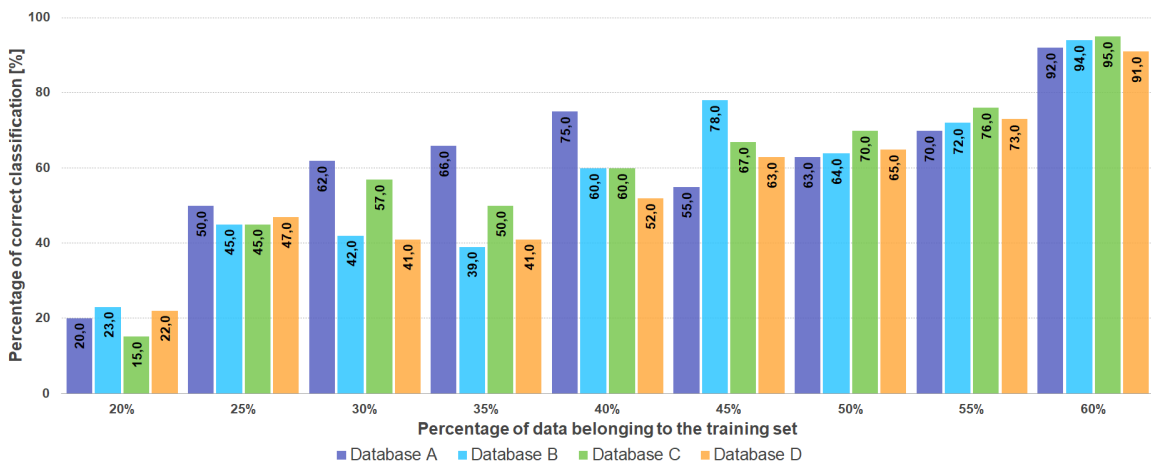


FIGURE 11. Efficiency of the fuzzy rule classifier.

classification in individual databases with the same size of the teaching set. However, for small training sets (20%-25%), the classification efficiency is better than when other structures are used.

The effectiveness of a classifier based on the Probabilistic Neural Network (Fig. 8) is very unstable. Despite the increase in the number of elements in the training set, there are large discrepancies in the degree of proper classification. There

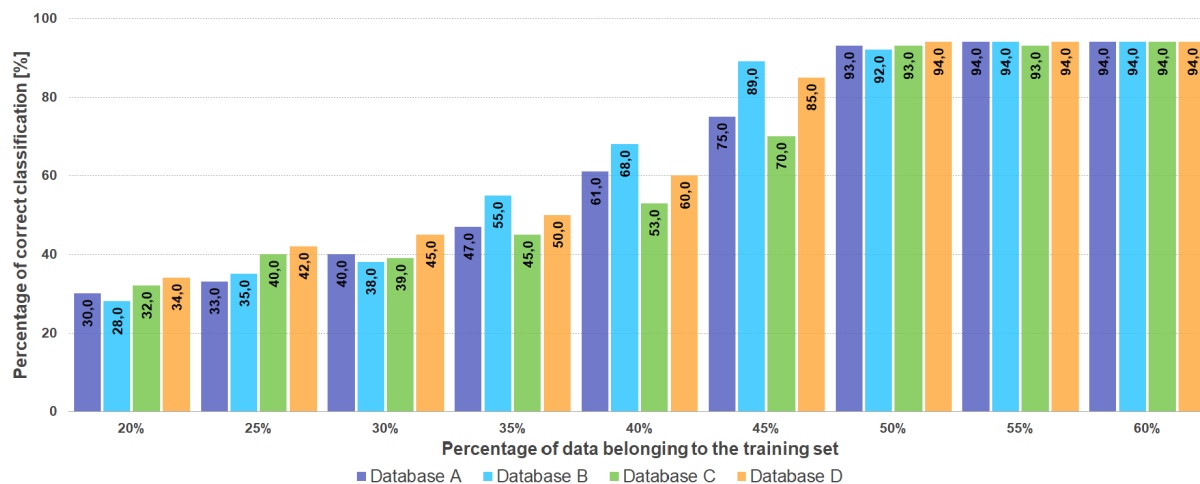


FIGURE 12. Efficiency of the fuzzy decisions trees classifier.

are also significant discrepancies in classification efficiency between individual databases, e.g. for 50% of the training set, the average classification efficiency for Database A is 75.4% and for Database C 48.9%.

As in the case of Decision Trees (Fig. 7), the use of the Random Forest Classifier (Fig. 9) was characterised by high instability of classification efficiency.

The use of the fuzzy multilayer perceptron classifier (Fig. 10) ensures excellent recognition results, which confirms the belief that the use of fuzzy numbers significantly increases the quality of classification. Satisfactory quality of classification is achieved very quickly, which remains stable despite the increasing number of learning sets.

In the case of Fuzzy Rule Classifier (Fig. 11), very different results are noted with equally numerous learning sets. Classification efficiency in this case is very unstable and an increase in the number of teaching sets does not always contribute to the improvement of the quality of the classification. Stability is only achieved with large learning collections.

As with FMLP (Fig. 10), the use of Fuzzy Decision Trees (Fig. 12) gives satisfactory results. After the correctness of the classification has stabilised, it is constantly maintained. A certain concern is the slight decrease in the correctness of classification in case of using small training sets; however, it is most likely related to the random assignment of learning elements. It is worth noting that the results of correct classifications significantly exceed those obtained with the use of the non-fuzzy decision trees classifier.

#### IV. CONCLUSION

The conducted research showed that the representation of emotional speech in the form of a scalogram is a good solution due to the possibilities of its processing. The proposed method of its modification allowed for the development of an emotional speech model, which included a scalogram obtained by a discrete wavelet transform, the level of speech signal decomposition, the threshold value allowing to obtain a binary form of the scalogram and the number of its sub-areas.

The correctness of the model has been proven by many classification tests. Simple classifiers such as the k-NN algorithm as well as more complex ones were used. It has also been shown that the application of fuzzy logic rules allows to significantly simplify the model (due to the lower level of signal decomposition), as well as to speed up the necessary calculations.

The classification efficiency not exceeding 94% may cause some concern. This may be due to the randomness of the learning data on the one hand and the number of samples of emotional speech recordings on the other. This will form the basis for further research into the model. The conducted research focuses on samples of recordings containing induced or acted emotions. Further research will also focus on attempts to test the correctness of the model in real-time applications, e.g. during a telephone conversation.

What is more, all analysed collections contain recordings of the same type, i.e. coming from the speech apparatus, therefore the physical parameters of all recordings are similar. The only difference is the presence of certain sounds characteristic of certain languages (Slavic, Germanic, etc.). Future research may help to obtain consistent signal processing parameters for specific language groups, not individual languages.

Additional research to build a set of classifiers allowing to increase the effectiveness of the identification of emotions in samples while maintaining a short processing time of a single file is being carried out.

It should be noted that individual emotions appear very rarely in real conversations. Most often, in one statement, emotions intertwine. Research is also being carried out on their proper separation and classification.

#### REFERENCES

- [1] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, nos. 1–2, pp. 227–256, 2003.
- [2] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech: A review," *Int. J. Speech Technol.*, vol. 15, no. 2, pp. 99–117, Jun. 2012.

- [3] V. Chernykh and P. Prikhodko, "Emotion recognition from speech with recurrent neural networks," in *Computation and Language*. Ithaca, NY, USA: Cornell Univ. Press, 2017, pp. 1–18.
- [4] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," in *Proc. ICASSP*, Montreal, QC, Canada, 2004, pp. 577–580.
- [5] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," *Speech Commun.*, vol. 48, no. 9, pp. 1162–1181, 2006.
- [6] F. Charles, D. Pizzi, M. Cavazza, T. Vogt, and E. Andre, "EmoEmma: Emotional speech input for interactive storytelling," in *Proc. AAMAS*, Estoril, Portugal, 2008, pp. 1381–1382.
- [7] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 829–837, Jul. 2000.
- [8] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Trans. Audio, Speech, Language Process.*, vol. 13, no. 2, pp. 293–303, Mar. 2005.
- [9] J. H. L. Hansen and D. A. Cairns, "ICARUS: Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments," *Speech Commun.*, vol. 16, no. 4, pp. 391–422, Jun. 1995.
- [10] D. Kaminska, "Emotional speech recognition based on the committee of classifiers," *Entropy*, vol. 21, no. 10, pp. 1–17, 2019.
- [11] D. Kaminska and A. Pelikant, "Recognition of human emotions from a speech signal based on Plutchik's model," *Int. J. Electron. Telecommun.*, vol. 58, no. 2, pp. 165–170, 2012.
- [12] S. Lalitha, D. Geyasruti, R. Narayanan, and M. Shravani, "Emotion detection using MFCC and cepstrum features," *Proc. Comput. Sci.*, vol. 70, pp. 29–35, Jan. 2015.
- [13] P. Jiang, H. Fu, and H. Tao, "Speech emotion recognition using deep convolutional neural network and simple recurrent unit," *Eng. Lett.*, vol. 27, no. 4, pp. 1–6, 2019.
- [14] N. Anithadevi, P. Gokul, S. M. Nandan, R. Magesh, and S. Shiddharth, "Automated speech recognition system for speaker emotion classification," in *Proc. ICCCS*, Patna, India, Oct. 2020, pp. 1–5.
- [15] A. Chamoli, A. Semwal, and N. Saikia, "Detection of emotion in analysis of speech using linear predictive coding techniques (L.P.C)," in *Proc. ICISC*, Coimbatore, India, Jan. 2017, pp. 1–4.
- [16] O. A. Mohammad and M. Elhadef, "Arabic speech emotion recognition method based on LPC and PPSD," in *Proc. ICCAKM*, Dubai, United Arab Emirates, Jan. 2021, pp. 31–36.
- [17] T. S. Gunawan, M. F. Alghifari, M. A. Morshidi, and M. Kartiwi, "A review on emotion recognition algorithms using speech analysis," *Indonesian J. Electr. Eng. Informat.*, vol. 6, no. 1, pp. 12–20, Mar. 2018.
- [18] A. M. Badshah, J. Ahmad, N. Rahim, and S. W. Baik, "Speech emotion recognition from spectrograms with deep convolutional neural network," in *Proc. PlatCon*, Busan, South Korea, Feb. 2017, pp. 1–5.
- [19] P. Powroznik and D. Czerwiński, "Spectral methods in POLISH emotional speech recognition," *Adv. Sci. Technol. Res. J.*, vol. 10, no. 32, pp. 73–81, 2016.
- [20] A. M. Badshah, B. N. Rahim, N. Ullah, J. Ahmad, K. Muhammad, M. Y. Lee, S. Kwon, and S. W. Baik, "Deep features-based speech emotion recognition for smart affective services," *Multimedia Tools Appl.*, vol. 78, no. 5, pp. 5571–5589, Mar. 2019.
- [21] V. V. R. Vegesna, K. Gurugubelli, and A. K. Vuppala, "Application of emotion recognition and modification for emotional Telugu speech recognition," *Mobile Netw. Appl.*, vol. 24, no. 1, pp. 193–201, Feb. 2019.
- [22] N. Hossain and M. Naznin, "Finding emotion from multi-lingual voice data," in *Proc. COMPSAC*, Madrid, Spain, Jul. 2020, pp. 408–417.
- [23] S. A. Al-aghha, H. H. Saleh, and R. F. Ghani, "Multi-model emotion expression recognition for lip synchronization," in *Proc. IT-ELA*, Baghdad, Iraq, Jul. 2020, pp. 171–177.
- [24] L. Cai, J. Dong, and M. Wei, "Multi-modal emotion recognition from speech and facial expression based on deep learning," in *Proc. CAC*, Shanghai, China, Nov. 2020, pp. 5726–5729.
- [25] R. A. Rammohan, J. Medikonda, and D. I. Pothiyil, "Speech signal-based modelling of basic emotions to analyse compound emotion: Anxiety," in *Proc. DISCOVER*, Udupi, India, Oct. 2020, pp. 218–223.
- [26] J. Deng, S. Frühholz, Z. Zhang, and B. Schuller, "Recognizing emotions from whispered speech based on acoustic feature transfer learning," *IEEE Access*, vol. 5, pp. 5235–5246, 2017.
- [27] C. Prakash, V. B. Gaikwad, R. R. Singh, and O. Prakash, "Time series classification using Gaussian mixture models of reconstructed phase spaces," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 6, pp. 779–783, Jun. 2004.
- [28] R. Povinelli, M. Johnson, A. C. Lindgren, and J. Ye, "Analysis of emotion recognition system through speech signal using KNN, GMM and SVM classifier," *J. Electron. Commun. Eng.*, vol. 10, no. 2, pp. 55–61, 2015.
- [29] V. Fernandes, L. Mascarehnas, C. Mendonca, A. Johnson, and R. Mishra, "Speech emotion recognition using mel frequency cepstral coefficient and SVM classifier," in *Proc. SMART*, Moradabad, India, Nov. 2018, pp. 200–204.
- [30] B. C. Kamble, "Speech recognition using artificial neural network," *Int. J. Comput., Commun. Instrum. Eng.*, vol. 3, no. 1, pp. 1–4, 2016.
- [31] K. H. Lee and D. H. Kim, "Design of a convolutional neural network for speech emotion recognition," in *Proc. ICTC*, Jeju, South Korea, Oct. 2020, pp. 1332–1335.
- [32] D. Kaminska and T. Sapinski, "Polish emotional speech recognition based on the committee of classifiers," *Przegląd Elektrotech.*, vol. 93, pp. 101–106, Jun. 2017.
- [33] D. Kaminska, T. Sapinski, and G. Anbarjafari, "Efficiency of chosen speech descriptors in relation to emotion recognition," *EURASIP J. Audio, Speech Music Process.*, vol. 2017, no. 1, pp. 1–9, 2017.
- [34] T. Sapinski and D. Kaminska, "Emotion recognition from natural speech—Emotional profiles," *Logistyka*, vol. 4, pp. 9712–9721, Sep. 2015.
- [35] V. A. Petrushin, "Emotion in speech: Recognition and application to call centers," in *Proc. ANNIE*, St. Louis, MO, USA, 1999, pp. 7–10.
- [36] S. Hareli and U. Hess, "The social signal value of emotions," *Cogn. Emotion*, vol. 26, no. 3, pp. 385–389, Apr. 2012.
- [37] I. S. Engberg and A. V. Hansen, "Documentation of the Danish emotional speech database (DES)," in *Proc. EUROSpeech*, Rhodes, Greece, 1997, pp. 1–31.
- [38] P. Rajmic, "Method for real-time signal processing via wavelet transform," in *Proc. NOLISP*, Barcelona, Spain, 2005, pp. 368–378.
- [39] V. K. Yadav, A. Jain, and L. Bhargav, "Analysis and comparison of audio compression using discrete wavelet transform," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 1, pp. 310–313, 2015.
- [40] J. Enders, W. Geng, P. Li, M. Frazier, and D. Scholl, "The shift-invariant discrete wavelet transform and application to speech waveform analysis," *J. Acoust. Soc. Amer.*, vol. 117, no. 4, pp. 2122–2133, 2005.
- [41] S. Hemalatha, U. D. Acharya, and A. Renuka, "Wavelet transform based steganography technique to hide audio signals in image," *Proc. Comput. Sci.*, vol. 47, pp. 272–281, Jan. 2015.
- [42] R. Szweczyk, K. Grabowski, M. Napieralska, W. Sankowski, M. Zubert, and A. Napieralski, "A reliable iris recognition algorithm based on reverse biorthogonal wavelet transform," *Pattern Recognit. Lett.*, vol. 33, pp. 1019–1026, Jun. 2012.
- [43] D. Vasisht, S. Debasis, and D. V. Rao, "Fuzzy graph modelling of anonymous networks," in *Soft Computing Applications*, vol. 1222. Cham, Switzerland: Springer, 2018, pp. 432–444.
- [44] D. Krleza and K. Fertalj, "Graph matching using hierarchical fuzzy graph neural networks," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 892–904, Aug. 2017.
- [45] J. Wang, P. Neskovic, and L. N. Cooper, "Improving nearest neighbor rule with a simple adaptive distance measure," *Pattern Recognit. Lett.*, vol. 28, no. 2, pp. 207–213, Jan. 2007.
- [46] R. D. King, C. Feng, and A. Sutherland, "StatLog: Comparison of classification algorithms on large real-world problems," *Appl. Artif. Intell.*, vol. 9, no. 3, pp. 289–333, May 1995.
- [47] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights Imag.*, vol. 9, no. 4, pp. 611–629, 2018.
- [48] M. Witczak, J. Korbicz, M. Mrugalski, and R. J. Patton, "A GMDH neural network-based approach to robust fault diagnosis: Application to the DAMADICS benchmark problem," *Control Eng. Pract.*, vol. 14, no. 6, pp. 671–683, Jun. 2006.
- [49] I. Etbetaj, H. Bonakdari, A. H. Zaji, H. Azimi, and F. Khoshbin, "GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs," *Eng. Sci. Technol., Int. J.*, vol. 18, no. 4, pp. 746–757, Dec. 2015.
- [50] J. V. Marcos, R. Hornero, D. Álvarez, F. D. Campo, C. Zamarrón, and M. López, "Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry," *Comput. Methods Programs Biomed.*, vol. 92, no. 1, pp. 79–89, Oct. 2008.



[51] W. Feki-Sahnoun, H. Njah, A. Hamza, N. Barraj, M. Mahfoudi, A. Rebai, and M. B. Hassen, "Using general linear model, Bayesian networks and Naive Bayes classifier for prediction of *Karenia selliformis* occurrences and blooms," *Ecol. Informat.*, vol. 43, pp. 12–23, Jan. 2018.

[52] J. Bogaert, R. Ceulemans, and D. S.-V. Eysenrode, "Decision tree algorithm for detection of spatial processes in landscape transformation," *Environ. Manage.*, vol. 33, no. 1, pp. 62–73, Jan. 2004.

[53] D. F. Specht, "Probabilistic neural network," *Neural Netw.*, vol. 3, no. 1, pp. 109–118, 1990.

[54] T. K. Ho, "Random decision forests," in *Proc. ICDAR*, Montreal, QC, Canada, 1995, pp. 278–282.

[55] T. Dash, S. K. Nayak, and H. S. Behera, "Hybrid gravitational search and particle swarm based fuzzy MLP for medical data classification," in *Computational Intelligence in Data Mining*, vol. 1. New Delhi, India: Springer, 2015, pp. 35–43.

[56] F. Amato, M. Tonini, B. Murgante, and M. Kanevski, "Fuzzy definition of rural urban interface: An application based on land use change scenarios in Portugal," *Environ. Model. Softw.*, vol. 104, pp. 171–187, Jun. 2018.

[57] N. Allahverdi, S. Torun, and I. Saritas, "Design of a fuzzy expert system for determination of coronary heart disease risk," in *Proc. CompSysTech*, New York, NY, USA, 2007, pp. 1–8.

[58] P. Srivastava and A. Srivastava, "A note on soft computing approach for cardiac analysis," *J. Basic Appl. Sci. Res.*, vol. 2, no. 1, pp. 376–385, 2012.

[59] L. Kuncheva, *Fuzzy Classifier Design* (Studies in Fuzziness and Soft Computing), vol. 49, 1st ed. Berlin, Germany: Springer, 2000.



research interests include neural networks, classification, and prediction issues as well as signal processing topics.

**PAWEL POWROZNIK** received the B.S. and M.S. degrees in computer science from Maria Curie-Skłodowska University (UMCS), Lublin, Poland, in 2009 and 2011, respectively, and the Ph.D. degree in electrical engineering from the Lublin University of Technology (LUT), Lublin, in 2018. He joined the Department of Computer Science, LUT, in 2016, and currently works as an Assistant Professor. Since 2011, he has been working as a Mobile Application Developer. His



His main research interests include pattern recognition and machine learning, but also robotics, the IoT, and applied computer science.

**PIOTR WOJCICKI** received the M.Eng. degree in mechatronics from the Lublin University of Technology, Lublin, Poland, in 2014, where he is currently pursuing the Ph.D. degree. From 2014 to 2016, he worked as an Engineer at the Division of Silicon Microsystem and Nanos-structure Technology, Institute of Electron Technology, Piaseczno. Since 2016, he has been a Research Assistant with the Department of Computer Science, Lublin University of Technology.



Development, Poland. He is an author of more than 50 articles published in communication journals and presented at the national and international conferences. He participated in several research and development projects funded by the national authority as well as by the European Union. His scientific research interests include IP traffic engineering, wireless networks, sensor networks, and video streaming systems.

**SLAWOMIR W. PRZYLUCKI** received the M.Sc. and Ph.D. degrees in electrotechnology from the Lublin University of Technology (LUT), Poland, in 1991 and 1999, respectively. Since 1991, he has been with the LUT, where he is currently an Assistant Professor with the Department of Computer Science. From 2013 to 2016, he was an Expert in the field of IT projects for the Government of the Czech Republic. Since 2016, he has been an IT Expert at the National Center for Research and

• • •