

Facial Expressions Analysis for Applications in the Study of Sign Language

Vladyslav Kuznetsov¹[0000-0002-1068-769X], Iurii Krak^{1,2}[0000-0002-8043-0785],
Olexander Barmak³[0000-0003-0739-9678], Anatolii Kulias¹[0000-0003-3715-1454]

¹Glushkov Cybernetics Institute, Kyiv, 40 Glushkov avenue, 03187, Ukraine,
kuznetsow.wlad@gmail.com, kulyas@nas.gov.ua

²Taras Shevchenko National University of Kyiv, 64/13 Volodymyrska str., 01601, Ukraine,
krak@univ.kiev.ua

³National University of Khmelnytsky, 11, Instytutska str., 29016, Ukraine,
alexander.barmak@gmail.com

Abstract. Elements of information technology for analysis of facial expressions to apply in interactive study of sign language are described. The main elements of information technology, its structure and experimental implementation has been discussed. Analysis was carried out in order to identify how the classifier error rate depends on type of classifier, number of feature and teach set. Optimal constructs of classifiers were proposed, giving appreciable improvement of existing algorithms.

Keywords: information technology, sing language, modeling, identification, facial expressions

1 Introduction

Sign language is a tool for communication data transmission in communication of deaf and hard of hearing people [1]-[3]. According to statistics, significant percent of the world population has congenital hearing problems that make it difficult or impossible communication with hearing people in everyday speech. A lot of countries has a government program that determines the main directions of improving the conditions of deaf people, including a broad plan to study sign language among the general population that often in contact with deaf people. One step that aims to solve this problem is creation of educational programs using modern information technology.

The most complete solution to this problem is approached educational system [4] using animated 3D models, integrated tools and user interfaces of interactive communication with computer (see. Fig. 1).

Despite the fullness of the means implementing interactive learning sign language environment development of other means of reproduction and recognition of sign language, including facial expressions that indicate a signed language intonation, accent, logical and emotional meaning of sentence, is required. Modeling and recognition of elements of sign language [5],[6] is important because these elements in

spoken language hearing people passed by voice instead in signed language these aspects of speech transmitted in part by modifying the gesture (amplitude and smoothness of movements) that studies hearing people are not quite acceptable because gestures at high speed badly perceived. For ease of reading, the gesture is accompanied by facial expressions, which are typically used in communication deaf in a sign language (which is the same grammar speaking to the sequence of grammar inclusive), or in carriers of sign language with a very high level of sign language (there are analogue expression in spoken language).

It is important to obtain the characteristics of displays of emotion based on their accurate detection [7],[8] instead classification ones. In previous studies we have worked on the algorithm and methods that can be used in the information technology for analysis (identification) of facial expressions on a human face; in particular, we considered a sequence of receiving video clips of facial expressions, involving optical sensors (markers), mounted on a human face [4], the application of computer vision algorithms for numerical data describing the state change facial expressions over time numerical data analysis tools, including a single-layer perception and methods for reducing data dimensionality [9].

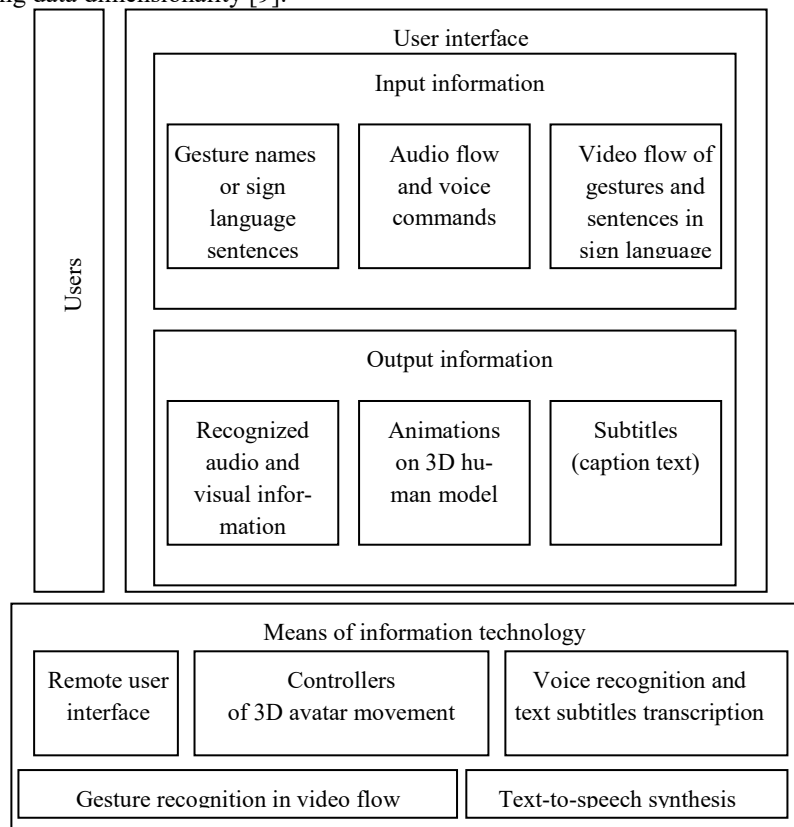


Fig. 1. Scheme information technology user interaction with the computer in the educational system to study sign language.

Results of the study [9] showed that there is a need to improve the efficiency of algorithms for problem of identification facial expressions changes over time; current implementation showed a low level of resolution of different classes of facial expressions. In order to improve current results suggest the following problem statement:

- propose a set of characteristics that can be used to identify facial expressions based on data obtained from the video stream;
- analyze existing means and methods of recognition of facial expressions on a set of characteristics that are applied in the analysis of facial expressions and other problems that reduce to identify changes in the characteristics of the object in time;
- develop a pilot information technology for identifying facial expressions and its software implementation;
- conduct a pilot test of the existing information technology to determine the best settings, algorithms, combinations thereof and the conditions under which it is possible to improve obtained previously results.

2 Methods of identification of the change over time characteristics

In the research [9] states that the problem of identifying the instantaneous state of the face does not require complex algorithms; perception or support vectors machines method are quite enough and give acceptable quality recognition - the order of 80-90% in the worst case - and uninformative small study sample small number of signs and to 99.5% at a rate sufficient training set (Table. 1). Instead, the task of examination of time dependent signal even for two classes of facial expressions is very complex and requires careful study. The same can be said when using the methods of reduction of data dimensionality while examining the single countdown (state of momentary) of facial expression. This indicates the similarity of sources of distortions that lead to identification errors.

Table 1 Identification accuracy for different samples of the algorithm in the existing technology.

Methods	Implementations and the number of samples in signals				
	Whole signal				1 countdown
	432	1080	2160	4320	36
PLA	64,8%	64,8%	64,8%	64,8%	98%
DCT+KLT	64,8%	64,8%	64,8%	64,8%	-
SVD+KLT	55%	58%	60%	62,5%	-
SVD	55%	58%	59%	61%	32%
KLT/FFT+KLT	52%	54%	57%	59,5%	-

In order to analyze changes of the object in time, during the duration of facial expression, the algorithm need to cope with varying intensity at the beginning and in the end, different amplitude, different waveform and signal distortion, different initial phase, the shape of the front and recession signal and varying duration.

In order to solve this problem (analysis of sequences of states) the most appropriate and expedient is to observe similar problems and algorithms for solving such problems which give a reasonable accuracy on a similar data.

Closest are several types of problems:

- phonograms-analysis (spectrograms);
- neural impulses, brain-waves, cardiograms, miohram analysis and so all;
- individual nerve cells process analysis.

In modern studies related to the analysis of phonograms [10] deep learning algorithms (with a teacher and without teacher), such as convolution neural networks, stacked denoised auto encoders, convolution auto encoders, deep belief networks, are used very often. Dynamic time warping method [11], which previously used to compare two close (similar) waveforms of varying duration, can also be taken in account as noticeable. Electrocardiograms and ion channels are analyzed using Karhunen-Loeve [12] and other integral transformations.

Methods of examination of time series studied in paper are:

- method of dynamic time warping, combined with some type of classifier (in particular, based on SVM, ANN and deep learning methods), using correlation between signals of corresponding sensors of different facial expressions and within correlation between signals from different sensors from each other in the same facial expression, as a net input. It also can help to test the hypothesis of generation of facial mimic signals by single nerve impulse (and their respective synchronization) through the evaluation of the correlation various mimic components;

- methods of analysis for spectrum signal including bringing output to the soundtrack-like input of defining the parameters of spectral analysis, constructing spectral characteristics of sample signals and build classifier based on convolution neural network. This will test the possibility of signal convolution with fixed window dimension compared to dynamic time warping.

Methods which include integral transformation, especially Karhunen-Loeve expansion and discrete cosine transform [13] for signal analysis of fixed length (normalized in amplitude and time) will also be implemented in order to find out possibility/necessity of reduction of dimensionality of data and their applicability in connection with classification algorithms in task of analysis of the instantaneous state facial expressions.

3 Information Technology of the Facial Expressions Analysis

Information technology (Fig. 2) describes the processing of data containing the facial expression. In order to obtain a formal description of the process as a generalized block diagram we need to identify all the components of this process - the objects, properties and states of objects, relationships and dependencies at various stages of processing.

Input information. As input video objects are the examples of facial expressions (video stream) containing the sequence of states of facial expression (facial gesture or

mimic morpheme) during an interval of muscle or mimic activity (contraction - relaxation within a range of muscle activity).

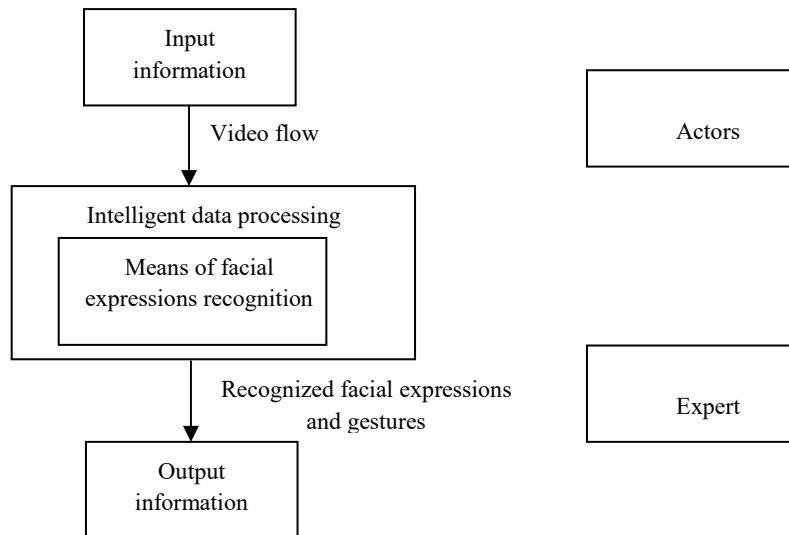


Fig. 2. Scheme information technology user interaction with the computer in the educational system to study sign language

Sampling frequency, signal number, information readout, signal number (ID), speaker identification are the main information signal characteristics.

The elements of input connections have a hierarchy: each countdown of a signal is part of a set of samples from a specific serial number, determining the order of appearance of a specific point in the video stream.

Processing the input data. The processing of the input data stream operates using algorithms to identify and track objects in video sequences to obtain coordinates of key points on each frame and the emerging key points coordinate data flow.

At this stage you must have to input constraints described by such terms, face orientation angle relative to the camera, sampling frequency, type of video compression, brightness of underlying objects on video, as well as object tracking algorithm restrictions, depending on the type and configuration of algorithms.

Getting the trajectory of the key points. The flow of data can be seen as a sampled characteristic points movement trajectory in some space formed by geometric transformations of space facial expressions with scale, rotation and shift of the base coordinate system.

Since each element (reference) of trajectory of movement of the key points is associated with a certain frame of video sequence, it is possible to group input on some characteristics derived by analyzing the trajectories of movement key points (for example, maximum activity, a minimum of activity, the mean interval values, etc.) on derived trajectory and original video sequence.

Data markup. This process describes the creation of some metadata that allow to group video frames and corresponding key points on the face. Automated processing

of video sequences and the corresponding data on the trajectory of motion was used to markup data. The output data is an ordered list of video sequence frame number and the corresponding notation (e.g. frame corresponding to the rest state and the frame corresponding to some activity) that allows to operate on specific instant face states that are most informative.

Normalization of data. For further processing of the ordered list of frames is necessary to bring all examples of movements to one scale. For this, all samples were ordered by a specific actor, thus data can be easily transferred to the new metric that will be comparable for all actors and allows changing scale of the mimic movements of each morpheme that is, obviously, specific to each actor.

Setting the functional converter. This process describes the iterative procedure for determining relationships (functional relationships) between input and some linguistic variables describing relation of inputs (video examples of facial expressions and its metadata) to a certain class. As for input, functional converter establish a priori class (beforehand known value of linguistic variable), which operates according to the motion trajectories key points on the face.

The most adaptive to this problem is classification algorithms and other algorithms that implement iterative procedure linkages.

Output data. The output data are linguistic variables corresponding to some input video sequence or reference throughout the interval samples. For such data we may need to set some attribute that describes the accuracy of functional dependence (identification) for some specific case (facial expression or mimic morpheme) by repeated execution of iterative procedure for multiple implementations of the data generated from the entire set of facial expressions video examples.

4 Experimental Realization of Information Technology for Identifying Facial Expressions

Database (DB) was created to implement storage means; a block diagram of this DB (Entity-Relationship model) is shown in Fig. 3.

We used database management system PostgreSQL [14] to implement database and the appropriate driver database to Oracle JVM (Java Virtual Machine) [15], which is connected to the executive files experimental program implementation.

Experimental software information technology was implemented in the environment IntelliJ IDEA [16] for Scala language [17] involving Java libraries and consists of several modules.

In order to provide the basic operations management we implemented by module and class *Main*. The module communicates with other modules to produce and output the converted data, including:

- input data* containing the trajectory of the markers on the actor's face;

- metadata* describing the characteristics of individual facial manifestations, including temporal range of activity, type of facial expression, identifier, actor who demonstrates mimic display, etc;

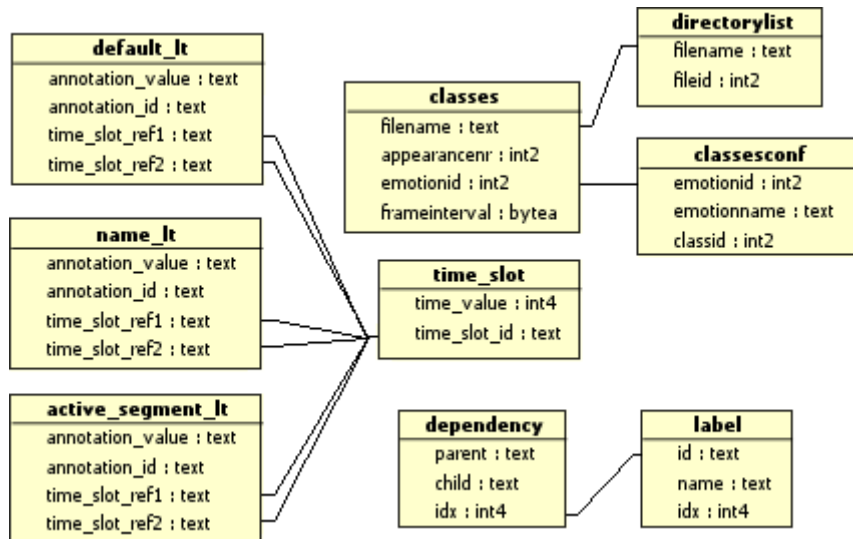


Fig. 3. Data base block diagram

- **transformed data** containing parameters obtained on the basis of input and removal of shift, scale and rotation distortions customized for each actor separately;

- **input and functional converter output data**, obtained by the rejection of the transformed data (may be rejected as fragments interval activity of facial expression, and the whole facial expression) and functional values of the dependent variable transformer;

- **coefficient setting of functional converter** that control facial expressions withdrawal rules that implements functional converter. To verify the configuration of functional converter (applicability coefficient settings for other data), its settings are checked for data that was rejected at the preliminary stage.

Object-oriented approach was offered to implement information technology; **InputReader** module, that implements reader input operations, consist of such classes: **Env** (file and means of reading the contents of the directory), **Marker** (information about the coordinates of each coordinate of the marker in the time and the means to import data containing the trajectory of motion markers), field **Data**, which contains an array unique samples of facial expressions, read from raw input, implemented as an array of objects of type **Marker**.

XML [16] metadata reading module (and their corresponding classes and methods) **ConfigReader** and **SQLQuery**, that implements JDBC-interface to the database PostgreSQL, were implemented.

Through the interface, implemented in the module, metadata video files stored in XML-container, and recorded in the database table. Database interface separately read two types of file metadata - time characteristics and a cloud of tags that describe the video files and structural links between them, respectively.

After reading metadata file, containing the time activity characteristics of facial expressions, the data are entered in the table (Fig. 3); data on time intervals is stored in the table *time_slot*, the names of basic facial expressions – in table *default_lt*, facial expressions names – in table *name*, time values (countdowns) of maximum activity (saturation of facial expressions) – in table *active_segment*. All three tables are combined in two key fields: *annotation_id* and *time_slot*.

After reading the tag cloud metadata file, module creates following tables: *item*, which contains information about the connections of individual tags from the tag cloud from video files, table *label*, containing the tag names, tables *dependency* and *hierarchy* containing links tags in the tag cloud respectively. All three tables are combined on a key field *id*, which contains a unique hex SHA-256 cloud tag hash id.

File structure that describes a list of video files is stored in table *directorylist*, links between basic and derivative mimic expressions – in table *classesconf*. This information can be manually adjusted from the outside interface that provides software implementation or database administration tools *pgAdmin* provided by ODBC PostgreSQL.

Database queries that perform union and intersection tables, aggregates data derived from two different sources XML-data. Aggregate table *classes* keep separate tracks, with such table labels as: time intervals of activity, the names of facial expressions, file names, names of actors, classes of facial expressions and other service information used in the experimental program implementation.

The next module (*ParameterMaker*) implements the basic methods for processing the coordinates of markers and form the parameters, namely "rejection" of affine distortion calculation movements markers algorithms determine the quantitative characteristics of secondary facial expressions and more. It also implements methods of calculating the average, maximum and minimum quantitative and qualitative characteristics of facial expressions.

Affine distortion rejection algorithm subtracts flipping movement of fixed point on the horizontal and vertical movement from relative (namely, non absolute) coordinates of other points. To calculate estimates of quantitative characteristics of facial expressions, *ParameterMaker* also implements appropriate methods of calculating them based on the distance between the eyes of the actor. The result is a dimensionless ratio of this distance in the range from 0 to 1.

Appropriate scaling is performed to determine the quantitative characteristics of facial expressions. In order to take into account the differences in facial expressions forms in different people, each mimic expression determined as a percentage of the maximum values for the individual.

Data obtained by the module *ParameterMaker* and corresponding metadata transmitted to the input module *TrainTestDataMaker*. Depending on the type of transducer and the type of study facial expressions, of the total volume data is shown in part comprising representatives of facial expression, a particular actor or interval activity mimic activity (e.g. muscle contraction or retraction), which are divided into two sets of data - set that is used to configure the functional converter and to test coefficients set-functional converter. Elements from the original dataset are selected using

a random number generator; the amount of data in each of the data sets is defined by a certain constant that controls random number generator.

Next (*TrainTestParametersMaker*) module implements functional transformation rules that describe the output of facial expressions based on input data from the module *TrainTestDataMaker*. The transformation is implemented using algorithms library MLLib, with the involvement of these constructs: single- and multilayer perceptron (PLA) with hypotheses based on linear and radial basis functions (RBF), stacked denoising autoencoders (SAE), means reducing the dimension (convolution) data, including: cosine transform (DCT) and Karhunen-Loeve (KLT). The software implementation use different combinations of constructs, which were limited to the following algorithm: 1) single or multistep reduction the dimensionality of data and 2) reduced dimension processing by using the methods of linear and nonlinear classification.

Built-in library of algorithms MLLib was used to set the coefficients properly. After setting the coefficients, in order to verify coefficients of each functional configuration of the converter, its settings are checked for test data.

The influence of distortion of scale of a small amount of data was checked by performing iteratively a number of trials on test data applying coefficients of functional transformers on train on test data, and comparing values of errors, that describe the output on trained data comparing with known (a priori) output value.

5 Experimental Tests of Information Technology for Identifying Facial Expressions

Input data. Test information technology tests were carried out on a set of 300 facial expressions taken from five different faces of actors, which included from 2 to 14 representatives of basic facial expressions. In addition, for a number of experiments that included the study of a single countdown, we did not used whole interval of mimic activity of facial expressions; these intervals covered the mimic expression in a state of saturation.

Samples of facial expressions, or single samples of the proposed sets during processing module *TrainTestDataMaker* were broken at a ratio of 20% and 80% using the integrated random number generator.

Description of trials. Variants of functional implementations converters were studied under different conditions. There were 5 different tests conducted on experimental implementation.

During the *first trial* we had to reveal if there is some specific information, that appear only when observing time changes of mimic expression. This test was conducted on individual reference data, and on the interval of activity of facial expressions over time. Tests showed that the available data on individual samples (countdowns) and key expansion coefficients, obtained by reducing the dimensionality of data (direct cosine transform), correlated. Therefore, for the data type "countdown" we observed relatively low value of the error for both proposed methods. However, these algorithms have limits in creation of complex hypotheses based on the input data, that,

possibly may incorrectly perceive facial expression of low intensity (e.g. 10-30% of power in respect to “weakest” sample in teach set).

During the *second trial* we had been tested whether the reduction of the dimension of the data (both single point, and at intervals of integral activity of facial expressions over time) is applicable. This test was necessary to establish the causes of great recognition errors obtained in previous studies [9], while using methods of dimensionality reduction of the data.

Tests showed (Fig. 4): 1) in order to classify the data obtained from the dimensionality reduction methods using classification methods, they require a large number of major expansion coefficients (much greater than 3), indicating that the eigenvectors with low energy have significant informational contribution applied to classification algorithms (Fig. 4.a); 2) individual samples (countdowns) require a smaller amount of additional vectors than for the entire interval of mimic activity - at least 4 vs 20-40 respectively.

This indicates that 1) the effectiveness of classification algorithms strongly depends on the number of expansion coefficients and methods of reduction of dimensionality of data respectively 2) large variation in the data and its phase-frequency characteristics makes unclear the dependency of the value of error on quantity (and even order) of basis functions.

During the *third trial* we conducted an experiment in order to figure out the minimum amount of sample data (teach data) that is used to configure the function converter and, in particular, to test the effectiveness of each constructs (e.g. DCT+SVM, SVD+ANN and so all) only case of "single signal" (as showed the previous test, the individual samples values and key ratios of discrete cosine transform correlated and, therefore, the inputs of classifiers were relatively homogeneous).

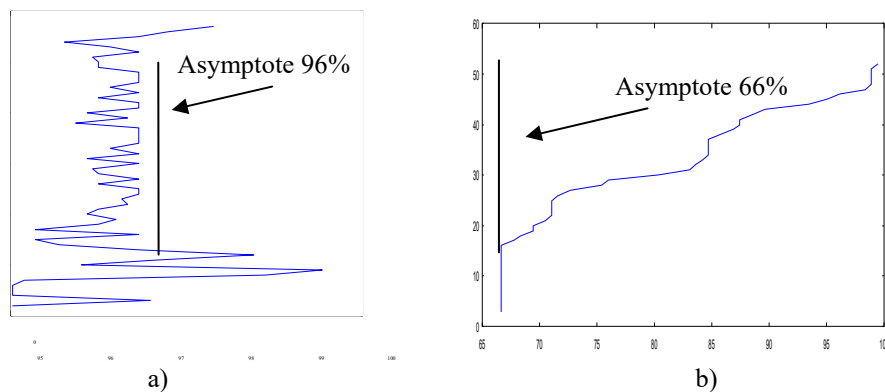


Fig. 3. Dependency of value of error on the number of basis functions for: SVM method - a) and single-layer neural network - b). Here vertical axis depict the number of basis functions (3-63), horizontal axis - the size of error (lower limit for errors 99.5%)

Tests have shown that even for the 103 order of data, one can find abnormal emissions (Fig. 5), reflecting one of the many bad implementations (e.g. iterations) of training on specific algorithm on exact data set. One can argue that the size of the

training set, sufficient to obtain results of satisfactory accuracy recognition (95%) in the range of 70-700 items, with no sample size is crucial in size accuracy, and most items are included in each class training and test samples and adaptability features to the elements of learning sample, which builds hyper resolution. This indicates a significant variance within the class that causes class transitions errors (e.g. element of one class appears to be element of another class).

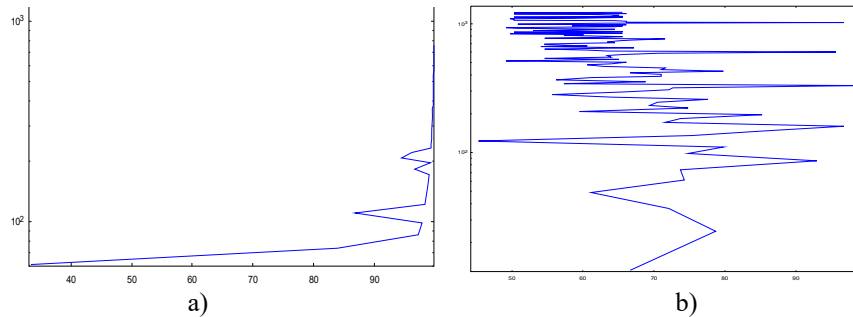


Fig. 4. The learning curve for the two implementation of classifiers: single layer neural network - a) and SVM based on tanh kernel - b). Logarithmic scale: horizontal axis depicts the volume of training set (101-103), the vertical axis - the value of errors (right asymptote ~ 99.5%, left asymptote ~ 40%)

During the *fourth trial* we had to suggest one way to improve the recognition accuracy of the neural network (which showed a relatively small number of emission value errors) on net input, with not applied dimension reduction. One good solution is good initial weights selection; deep learning methods answer this requirements. As a result of testing, we had found that the stacked autoencoders are able to improve learning accuracy of neural network with one hidden layer neurons compared with vanilla neural network of same structure.

During the *fifth trial* we had to figure out how to reduce the abnormal emissions in learning curves of learning algorithms enrolled on data, obtained after reduction of data dimension. One way to improve the algorithms is eliminating the time scale, frequency and phase mismatch between signals from different samples of facial expressions, without removing information about temporal characteristics. In order to do this, we compared the characteristics of each of the facial expressions together using Dynamic Time Warping (DTW). Values of independent pair DTW-correlation form (i.e. fill) the characteristic vector signal in the form of a square matrix (Fig. 5 a)). If necessary, the thinning is applied on matrix (Fig. 6, b). As a result of experiments, we had found that the most applicable algorithm that can be used in conjunction with thinned characteristic DTW vectors is SVM.

Therefore, as a result of all trials described above, we obtained the optimal set of constructs (combinations of algorithms) that give relatively high rates of accuracy of identification. Table 2 lists these constructs and optimistic value of precision out-of-sample for the best scenario (the largest sample size, the best set of basis functions,

the best implementation of a specific algorithm, etc.); numerator depicts the accuracy, denominator – number of eigen functions and size of teach vector respectively.



Fig. 5. Matrix of DTW pair correlation for samples of different facial expressions without thinning - a) and two-dimensional thinning using DCT - b).

Table 2 Identification accuracy for various implementations of the algorithms

Classifier type	Signal dimensions	
	1x20 (1 countdown)	20,30,60,120,...x20 (whole signal)
SVM (RBF)	99,5%/-	-
ANN MLP	99,5%/-	75%/-
KLT(SVD) + SVM (RBF)	99,5%/1x4	99,5%/1x18
DCT + SVM (RBF)	-	99,5%/3x20
SDAE (400x20x2) + ANN MLP	-	99,5%/-
DBN (400x20x2)	-	75%
Convolutional NN 12c-2s-6c-2s	-	75%
DCT+SVM (RBF)	-	99,5%/7x20
DCT+SVM (Linear)	-	99,5%/2x20
DTW + SVM	-	87,5%/-
DTW + ANN/DBN/SDAE/CNN	-	87,5%/-
DTW + DCT + ANN/DBN/SDAE/CNN		87,5%/10x19
DTW + DCT + SVM (RBF)	-	99,5%/7x19
DTW + DCT + SVM (Linear)	-	99,5%/2x19

6 Conclusion

As a result of experiments we greatly improved existing technology [9], using relatively the same data examples - facial expressions and facial expression components captured by means of optical motion capture of control points using optical markers, particularly for methods of identification changes facial expressions over time. Several different structures or combinations of the most effective methods of feature extraction and classification of facial expressions were applied, making possible to implement an effective system for identifying facial expressions. We analyzed the impact of the number of inputs, the training set on amount and type of classification errors of 1 and 2 race, which allowed to offer new constructs of multiple classifiers.

In particular, the proposed algorithms and the neural network in connection with stacked denoising autoencoders, using eigenvalues of decomposition of output data, in particular, in task of identification of the temporal changes of micro expressions allowed to achieve higher efficiency (of the order of 99.5%) .

Research, that was carried out on the proposed algorithms, helped to improve the preliminary results and to established causes of poor quality recognition. Some solutions were proposed as well - a set of structures of multiple classifiers, that are alternative for methods for dimensionality reduction, based on three simple steps: 1) calculating DTW inner-correlation matrix; 2) applying thinning; 3) using classifier on thinned matrix, which can be applied in classification of other discrete time-dependent and correlated multiple channel signals that have different duration (number of countdowns).

References

1. Stokoe. W.C. Jr.: Sing language structure: An outline of the visual communication systems of the American deaf. Univ. of Buffalo (1960)
2. Sing Languages. Ed. by D. Brentari. Cambridge: Cambridge University Press (2010)
3. Smith, R., Morrissey, S., Somers, H.: HCI for the Deaf community: developing human-like avatars for sign language synthesis. In: Proceedings of the 4th Irish Human Computer Interaction Conference. Dublin (iDCI2010), Dublin, Ireland, 2-3 September 2010, pp.129-136 (2010)
4. Kryvonos, I.G., Krak, I.V.: Modeling human hand movements, facial expressions, and articulation to synthesize and visualize gesture information. *Cybernetics and Systems Analysis*. 47(4): 501-505 (2011)
5. Kryvonos, I.G., Krak, I.V., Barmak, O.V., Shkilniuk, D.V.: Construction and identification of elements of sign communication. *Cybernetics and Systems Analysis* 49(2): 163-172 (2013)
6. Krak, I.V., Kryvonos, I.G., Barmak, O.V., Ternov, A.S.: An Approach to the Determination of Efficient Features and Synthesis of an Optimal Band-Separating Classifier of Dactyl Elements of Sign Language. *Cybernetics and Systems Analysis* 52(2): 173-180 (2016)
7. Brunelli, R., Poggio, T.: Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(10): 1042-1052 (1993)

8. Martinez, A., Du, S.: A model of perception of facial expressions of emotion by human: research overview and perspectives. *Journal of Machine Learning Research* 13: 1589-1608 (2012)
9. Kryvonos, I.G., Krak, I.V., Barmak, O.V., Ternov, A.S., Kuznetsov, V.O.: Information Technology for the Analysis of Mimic Expressions of Human Emotional States. *Cybernetics and Systems Analysis* 51(1): 25-33 (2015)
10. Lee, H., Largman, Y., Pham, P., Ng, A.Y.: Unsupervised Feature Learning for Audio Classification Using Convolutional Deep Belief Networks, In: *Proceedings of Conference on Neural Information Processing Systems (NIPS 2009)*, Vancouver, Canada, 7-10 December 2009, pp. 1096-1104 (2009)
11. Al-Naymat, G., Chawla, S., Taheri, J.: Sparse DTW: A novel approach to speed up Dynamic Time Warping, *Australasian Data Mining*, Melbourne, Australia, ACM Digital Library, pp. 117-127 (2009)
12. Chumakov, A.G., Kurashov, V.N.: Karhunen-Loeve Basis Synthesis for Grate Capacity Signal Performance in Optical Processors, *SPIE Proc.*, 21-08, pp. 338-342 (1993)
13. Loeffler, C., Ligtenberg, A., Moschytz, G.: Practical Fast 1-D DCT Algorithms with 11 Multiplications, In: *International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, UK, 23-26 May 1989, pp. 988-991 (1989)
14. PostgreSQL: The World's Most Advanced Open Source Relational Database, <https://www.postgresql.org>, last accessed 2019/04/09.
15. Java SE at a Glance <http://www.oracle.com/technetwork/java/javase/overview/index.html>, last accessed 2019/04/10.
16. JetBrains, <https://www.jetbrains.com/idea/>, last accessed 2019/04/10.
17. The Scala Programming Language, <http://www.scala-lang.org>, last accessed 2019/04/10.