Intelligent human-machine speech communication system

Maciej Majewski and Wojciech Kacalak

Technical University of Koszalin, Department of Mechanical Engineering
Raclawicka 15-17, 75-620 Koszalin, Poland

{maciej.majewski, wojciech.kacalak}@tu.koszalin.pl

Abstract

In this paper there is an intelligent human-machine speech communication system presented, which consists of the intelligent mechanisms of operator identification, word and command recognition, command syntax and result analysis, command safety assessment, technological process supervision as well as operator reaction assessment. In this paper there is also a review of the selected issues on recognition of voice commands in natural language given by the operator of the technological device. A view is offered of the complexity of the recognition process of the operator's words and commands using neural networks made of a few layers of neurons. The paper presents research results of speech recognition and automatic command recognition using artificial neural networks.

Keywords: Speech, Voice Communication, Command Recognition, Artificial Intelligence, Artificial Neural Networks, Human Machine Interface.

I. Intelligent Two-Way Speech Communication

Humans communicate with other humans in many ways, but surely voice communication is the most widely used. Speech is a singularly efficient way for humans to express ideas and desires. Nowadays technological devices can already be provided with enough intelligence to understand and act appropriately on voice commands. The voice communication with technological devices becomes a stronger challenge as technology becomes more advanced and complex. The obvious advantages of two-way communication by voice include the following:

1) Speech is the natural mode of communication for humans [1,5,6].
2) Voice control is particularly appealing when the human's hands or eyes are otherwise occupied.
3) The ubiquitous telephone can be an effective remote terminal for two-way voice communication with technological devices that can also speak, listen, and understand.

According to the new conception, the intelligent layer of two-way voice communication of the technological device with the operator presented in Figure 1, is equipped with the following intelligent mechanisms: operator identification, recognition of words and complex commands, command syntax analysis, command result analysis, command safety assessment, technological process supervision, and also operator reaction assessment [2,3].

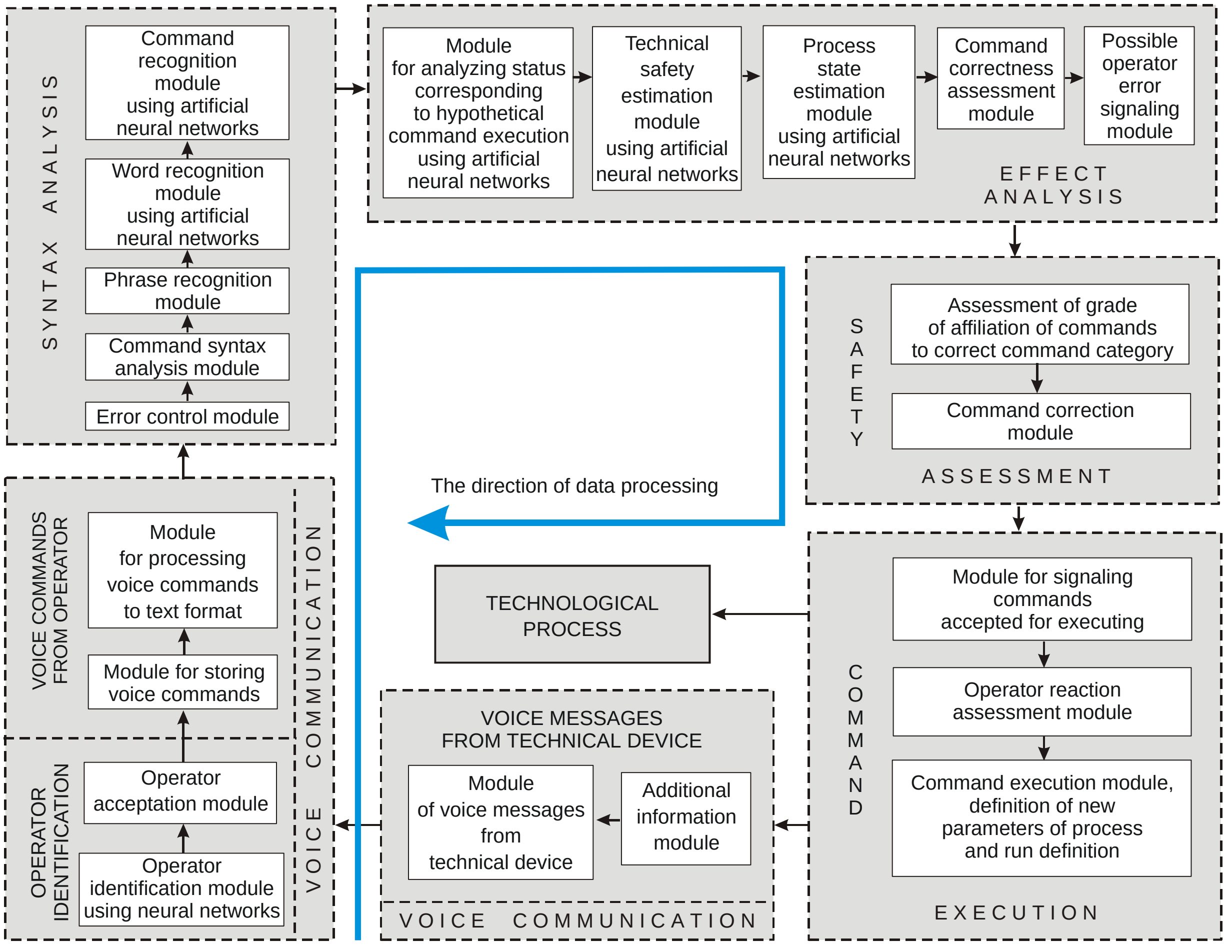

SYNTAX ANALYSIS
Command recognition module using artificial neural networks
Word recognition module using artificial neural networks
Phrase recognition module
Command syntax analysis module
Error control module
Module for analyzing status corresponding to hypothetical command execution using artificial neural networks
Technical safety estimation module using artificial neural networks
Process state estimation module using artificial neural networks
Command correctness assessment module
Possible operator error signaling module
EFFECT ANALYSIS
SAFETY ASSESSMENT
Assessment of grade of affiliation of commands to correct command category
Command correction module
The direction of data processing
VOICE COMMANDS FROM OPERATOR
Module for processing voice commands to text format
Module for storing voice commands
VOICE COMMUNICATION
TECHNOLOGICAL PROCESS
COMMAND EXECUTION
Module for signaling commands accepted for executing
Operator reaction assessment module
Command execution module, definition of new parameters of process and run definition
OPERATOR IDENTIFICATION
Operator acceptation module
Operator identification module using neural networks
VOICE MESSAGES FROM TECHNICAL DEVICE
Module of voice messages from technical device
Additional information module
VOICE COMMUNICATION

Figure 1: Scheme of the intelligent layer of two-way speech communication between the technological device and the operator

If the operator is identified and authorized by the intelligent speech communication system in Figure 2, a produced command in continuous speech is recognized by the speech recognition module and processed to the text format [2]. Then the recognized text is analyzed with the syntax analysis subsystem. The processed command is sent to the word and command recognition modules using artificial neural networks to recognize the command, which next is sent to the effect analysis subsystem for analyzing the status corresponding to the hypothetical command execution, consecutively assessing the command correctness, estimating the process state and the technical safety, and also possibly signaling the possible error caused by the operator. The command is also sent to the safety assessment subsystem for assessing the grade of affiliation of the command to the correct command category and making corrections. Next the command execution subsystem signalizes commands accepted for executing, assessing reactions of the operator, defining new parameters of the process and run directives [2,4]. The subsystem for voice communication produces voice commands to the operator.

The speech recognition engine is a continuous density mixture Gaussian Hidden Markov Model system which uses vector quantization for speeding up the Euclidean distance calculation for probability estimation. The system uses context dependent triphonic cross word acoustic models with speaker normalization based on vocal tract length normalization, channel adaptation using mean Cepstral subtraction and speaker adaptation using Maximum Likelihood Linear Regression.

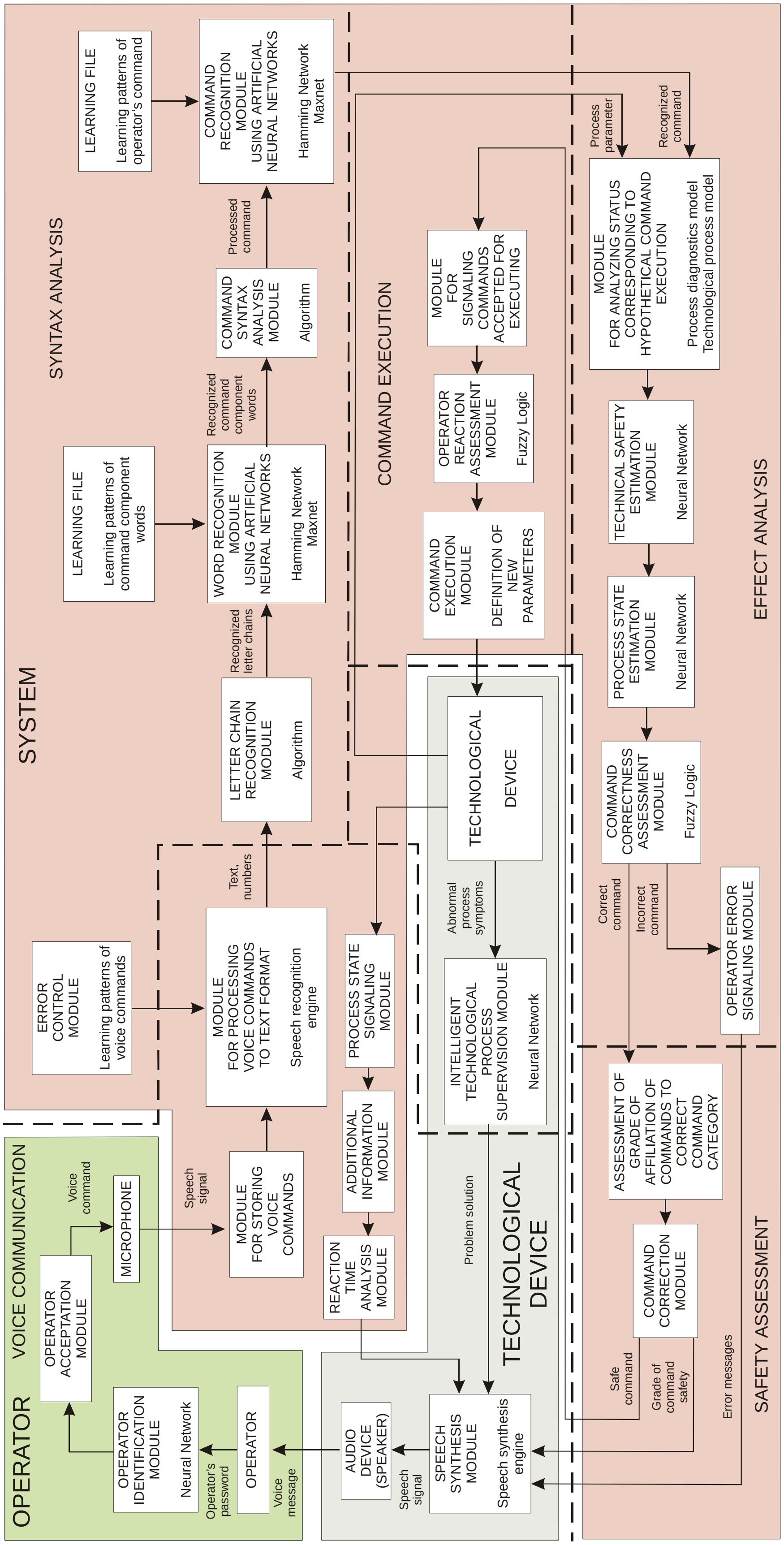

SYSTEM
SYNTAX ANALYSIS
COMMAND EXECUTION
EFFECT ANALYSIS
SAFETY ASSESSMENT
TECHNOLOGICAL DEVICE
OPERATOR
VOICE COMMUNICATION
LEARNING FILE
Learning patterns of operator's command
COMMAND RECOGNITION MODULE USING ARTIFICIAL NEURAL NETWORKS
Hamming Network Maxnet
Processed command
COMMAND SYNTAX ANALYSIS MODULE
Algorithm
Recognized command component words
LEARNING FILE
Learning patterns of command component words
WORD RECOGNITION MODULE USING ARTIFICIAL NEURAL NETWORKS
Hamming Network Maxnet
Recognized letter chains
LETTER CHAIN RECOGNITION MODULE
Algorithm
Text, numbers
ERROR CONTROL MODULE
Learning patterns of voice commands
MODULE FOR PROCESSING VOICE COMMANDS TO TEXT FORMAT
Speech recognition engine
Speech signal
MODULE FOR STORING VOICE COMMANDS
Voice command
MICROPHONE
OPERATOR ACCEPTATION MODULE
OPERATOR IDENTIFICATION MODULE
Neural Network
Operator's password
OPERATOR
Voice message
AUDIO DEVICE (SPEAKER)
Speech signal
SPEECH SYNTHESIS MODULE
Speech synthesis engine
MODULE FOR SIGNALING COMMANDS ACCEPTED FOR EXECUTING
OPERATOR REACTION ASSESSMENT MODULE
Fuzzy Logic
COMMAND EXECUTION MODULE
DEFINITION OF NEW PARAMETERS
TECHNOLOGICAL DEVICE
Abnormal process symptoms
INTELLIGENT TECHNOLOGICAL PROCESS SUPERVISION MODULE
Neural Network
Problem solution
PROCESS STATE SIGNALING MODULE
ADDITIONAL INFORMATION MODULE
REACTION TIME ANALYSIS MODULE
Process parameter
Recognized command
MODULE FOR ANALYZING STATUS CORRESPONDING TO HYPOTHETICAL COMMAND EXECUTION
Process diagnostics model
Technological process model
TECHNICAL SAFETY ESTIMATION MODULE
Neural Network
PROCESS STATE ESTIMATION MODULE
Neural Network
COMMAND CORRECTNESS ASSESSMENT MODULE
Fuzzy Logic
Correct command
Incorrect command
OPERATOR ERROR SIGNALING MODULE
ASSESSMENT OF GRADE OF AFFILIATION OF COMMANDS TO CORRECT COMMAND CATEGORY
COMMAND CORRECTION MODULE
Safe command
Grade of command safety
Error messages

Figure 2: Architecture of the intelligent two-way speech communication system

II. Automatic Command Recognition

In the automatic command recognition system as shown in Figure 3, the speech signal is processed to text and numeric values with the module for processing voice commands to text format. The separated words of the text are the input signals of the neural network for recognizing words. The network has a training file containing word patterns. The network recognizes words as the operator's command components, which are represented by its neurons in Figure 4. The recognized words are sent to the algorithm for coding words. Next the coded words are transferred to the command syntax analysis module.

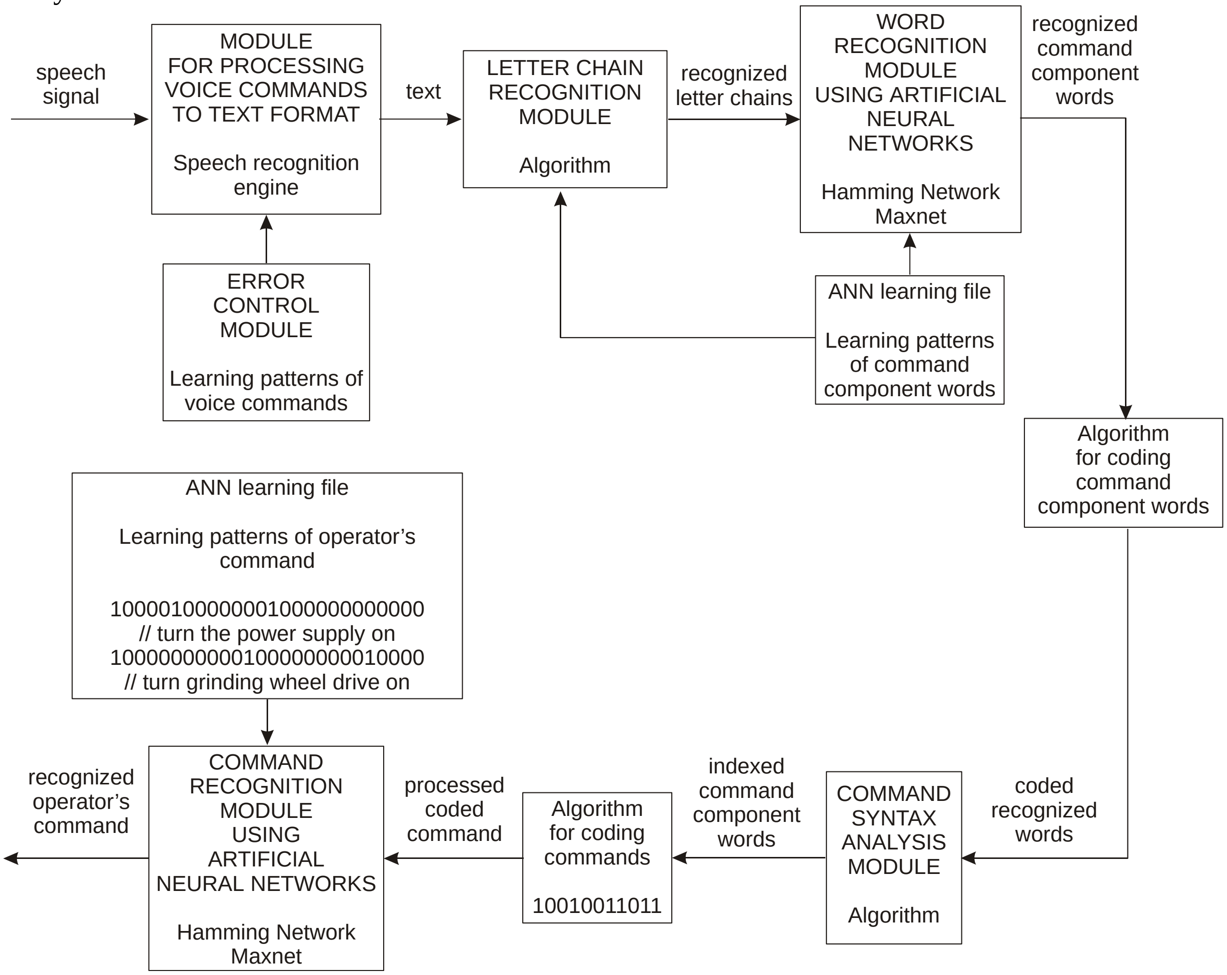

speech signal
MODULE FOR PROCESSING VOICE COMMANDS TO TEXT FORMAT
Speech recognition engine
text
LETTER CHAIN RECOGNITION MODULE
Algorithm
recognized letter chains
WORD RECOGNITION MODULE USING ARTIFICIAL NEURAL NETWORKS
Hamming Network Maxnet
recognized command component words
ERROR CONTROL MODULE
Learning patterns of voice commands
ANN learning file
Learning patterns of command component words
Algorithm for coding command component words
ANN learning file
Learning patterns of operator's command
10000100000001000000000000
// turn the power supply on
10000000000100000000010000
// turn grinding wheel drive on
recognized operator's command
COMMAND RECOGNITION MODULE USING ARTIFICIAL NEURAL NETWORKS
Hamming Network Maxnet
processed coded command
Algorithm for coding commands
10010011011
indexed command component words
COMMAND SYNTAX ANALYSIS MODULE
Algorithm
coded recognized words

Figure 3: Scheme of the automatic command recognition system

It is equipped with the algorithm for analyzing and indexing words. The module indexes words properly and then they are sent to the algorithm for coding commands. The commands are coded as vectors and they are input signals of the command recognition module using neural network. The module uses the 3-layer Hamming neural network in Figure 5, either to recognize the operator's command or to produce the information that the command is not recognized. The neural network is equipped with a training file containing patterns of possible operator's commands.

The 3-layer Hamming neural network with the additional layer named Maxnet is capable of optimal classification of binary images and represents a heteroassociative memory. The command recognition module using Hamming Maxnet neural networks is capable of recognizing different commands of the same meaning in natural language. The Hamming neural network has a high rate of processing signals and also works in parallel. The recurrent network Maxnet containing neurons connected in feedback performs functions of vector classifier.

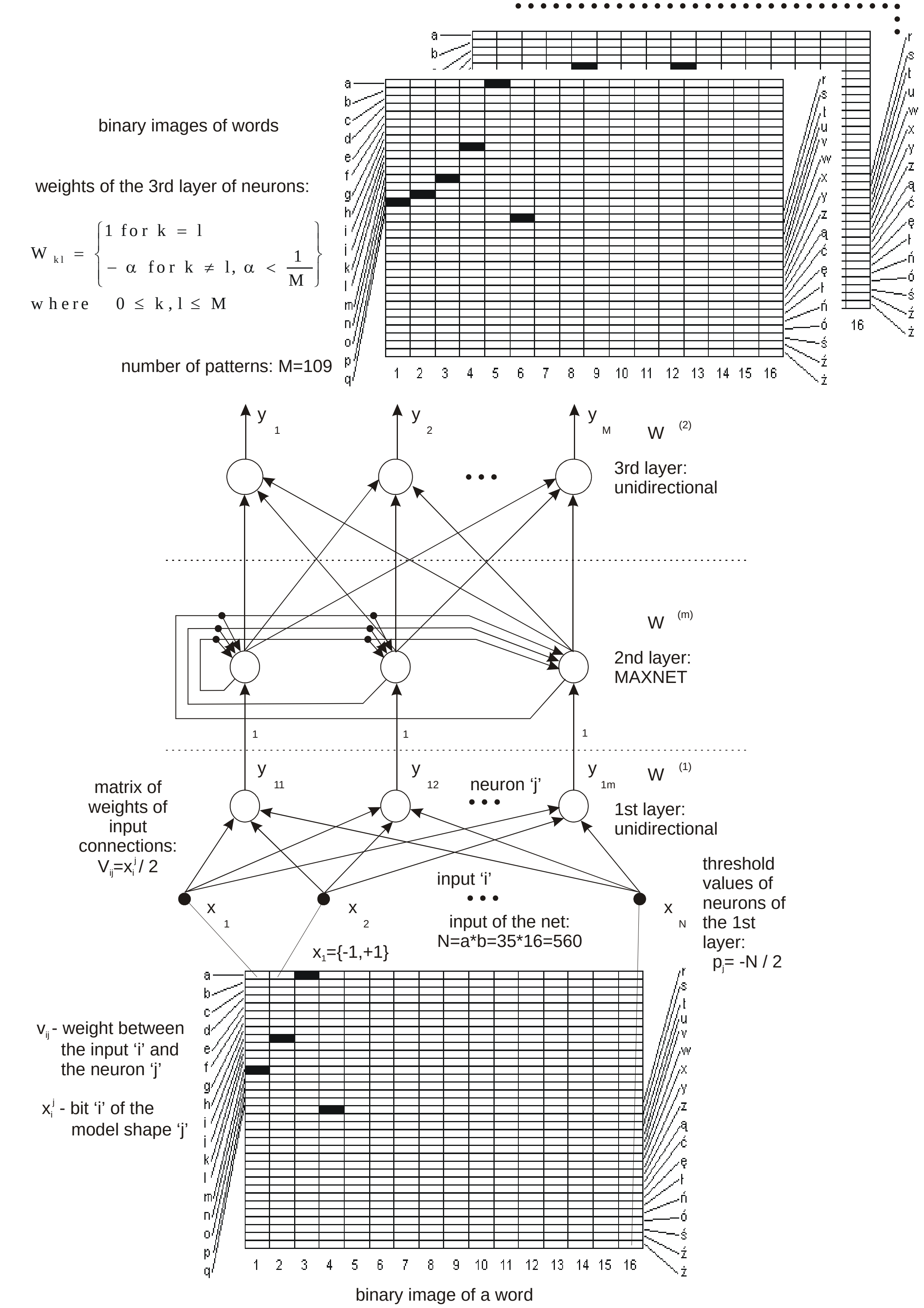

binary images of words
weights of the 3rd layer of neurons:
$W_{kl} = \begin{cases} 1 \text{ for } k = l \\ -\alpha \text{ for } k \neq l, \alpha < \frac{1}{M} \end{cases}$
where $0 \leq k, l \leq M$
number of patterns: M=109
a b c d e f g h i j k l m n o p q
r s t u v w x y z ą ć ę ł ń ó ś ź ż
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
16
y_1
y_2
y_M
$W^{(2)}$
3rd layer: unidirectional
$W^{(m)}$
2nd layer: MAXNET
1
1
1
y_{11}
y_{12}
neuron 'j'
y_{1m}
$W^{(1)}$
1st layer: unidirectional
matrix of weights of input connections: $V_{ij}=x_i^j / 2$
input 'i'
threshold values of neurons of the 1st layer: $p_j= -N / 2$
x_1
x_2
x_N
input of the net: N=a*b=35*16=560
$x_1=\{-1,+1\}$
v_{ij} - weight between the input 'i' and the neuron 'j'
x_i^j - bit 'i' of the model shape 'j'
binary image of a word

Figure 4: Scheme of the 3-layer neural network for word recognition

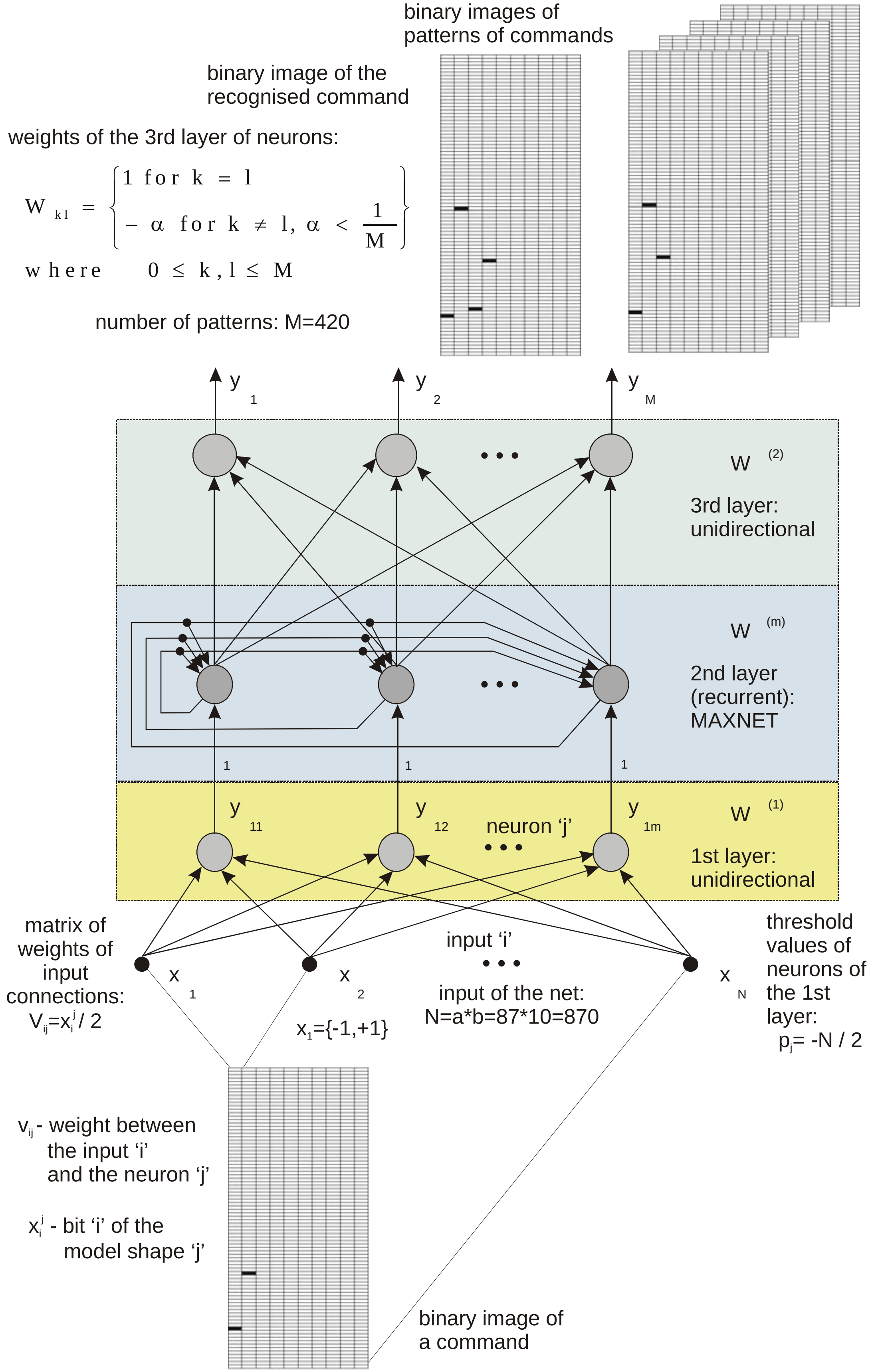

binary images of patterns of commands
binary image of the recognised command
weights of the 3rd layer of neurons:
$W_{kl} = \begin{cases} 1 \text{ for } k = l \\ -\alpha \text{ for } k \neq l, \alpha < \frac{1}{M} \end{cases}$
where $0 \leq k, l \leq M$
number of patterns: M=420
y_1
y_2
y_M
$W^{(2)}$
3rd layer: unidirectional
$W^{(m)}$
2nd layer (recurrent): MAXNET
1
1
1
y_{11}
y_{12}
neuron 'j'
y_{1m}
$W^{(1)}$
1st layer: unidirectional
matrix of weights of input connections: $V_{ij}=x_i^j / 2$
x_1
x_2
input 'i'
x_N
threshold values of neurons of the 1st layer: $p_j= -N / 2$
$x_1=\{-1,+1\}$
input of the net: N=a*b=87*10=870
v_{ij} - weight between the input 'i' and the neuron 'j'
x_i^j - bit 'i' of the model shape 'j'
binary image of a command

Figure 5: Scheme of the 3-layer neural network for automatic command recognition

III. Research Results of Automatic Command Recognition

For the evaluation of research results of the automatic speech recognition, it has to be defined how to calculate the command recognition rate. The calculation is done after performing each case of recognition event. The recognition rate R is calculated from the formula for the total number of errors T, and the error rate E. The total number of errors is the sum of the insertion errors and the out-of-context errors. The error rate equals to the total number of errors divided by the total number of commands in a case.

$$T = I + O$$

$$E = \frac{T}{N} * 100\%$$

$$R = 100\% - E$$

As shown in Figure 6a, the speech recognition module recognizes 85-90% of the operator's words correctly. As more training of the neural networks is done, accuracy rises to around 95%.

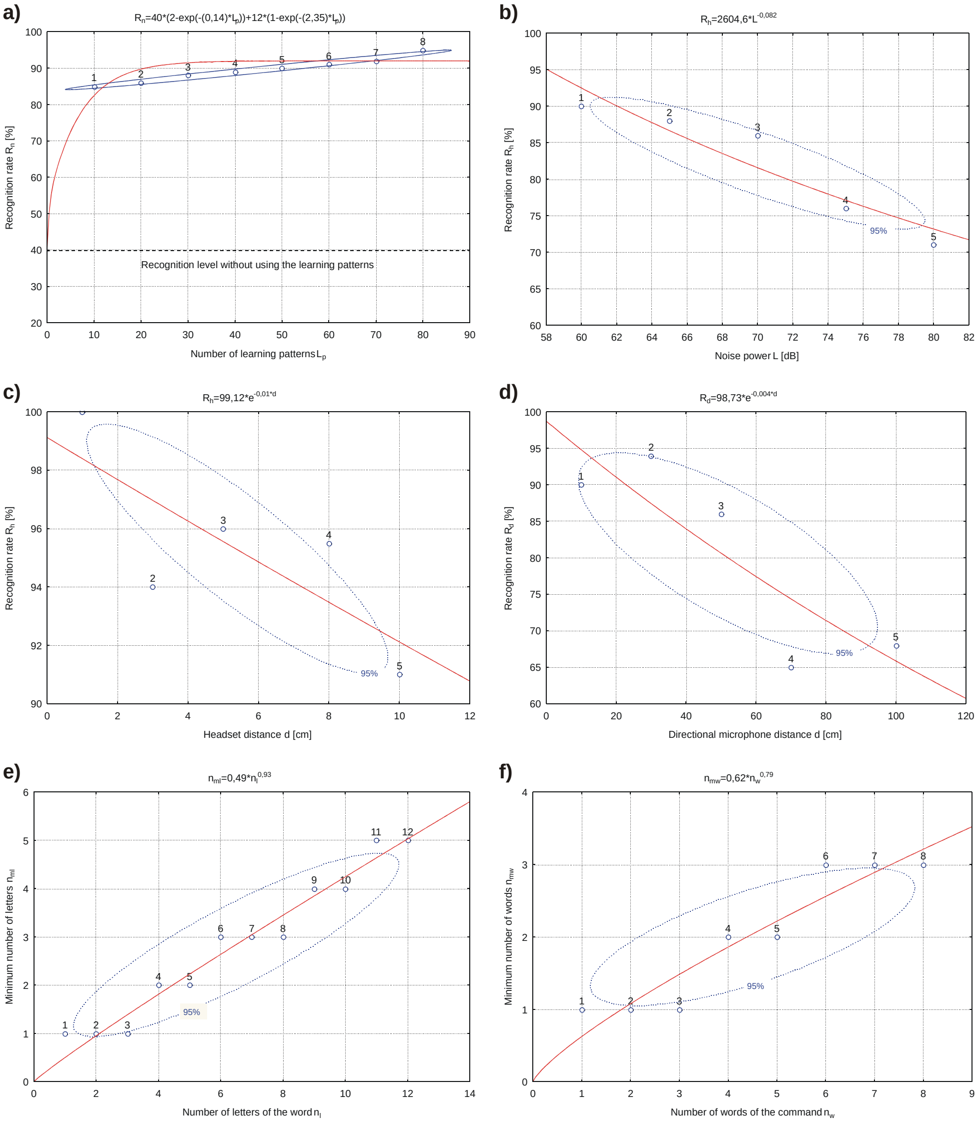


Figure 6: Speech and command recognition rate

For the research on command recognition at different noise power, the microphone used by the operator is the headset. As shown in Figure 6b, the recognition performance is sensitive to background noise. The recognition rate is about 86% at 70 dB and 71% at 80 dB. Therefore, background noise must be limited while giving the commands.

For the research on command recognition at different microphone distances, the microphone used by the operator is the headset. As shown in Figure 6c, the recognition rate decreases when the headset distance increases. The recognition rate has been dropped for 9% after the headset distance is changed from 1 to 10 cm. Also for the research on command recognition at different microphone distances, the microphone used by the operator is the directional microphone. As shown in Figure 6d, the recognition rate after 50 cm decreases reaching rate about 65%.

As shown in Figure 6e, the ability of the neural network to recognize the word depends on the number of letters. The neural network requires the minimal number of letters of the word being recognized as its input signals. As shown in Figure 6f, the ability of the neural network to recognize the command depends on the number of command component words. Depending on the number of component words of the command, the neural network requires the minimal number of words of the given command as its input signals.

The consecutive test measured the performance of the system by using 3 phases. In the first phase the slow version of giving commands and in the second phase the fast version of giving commands was used. The slow version describes the situation when the next command should be issued while the previous one is completed. In the fast version the next command can be issued while the previous one is still executing. In the third phase the average version of giving commands was used. The average version is the average of the slow and the fast version.

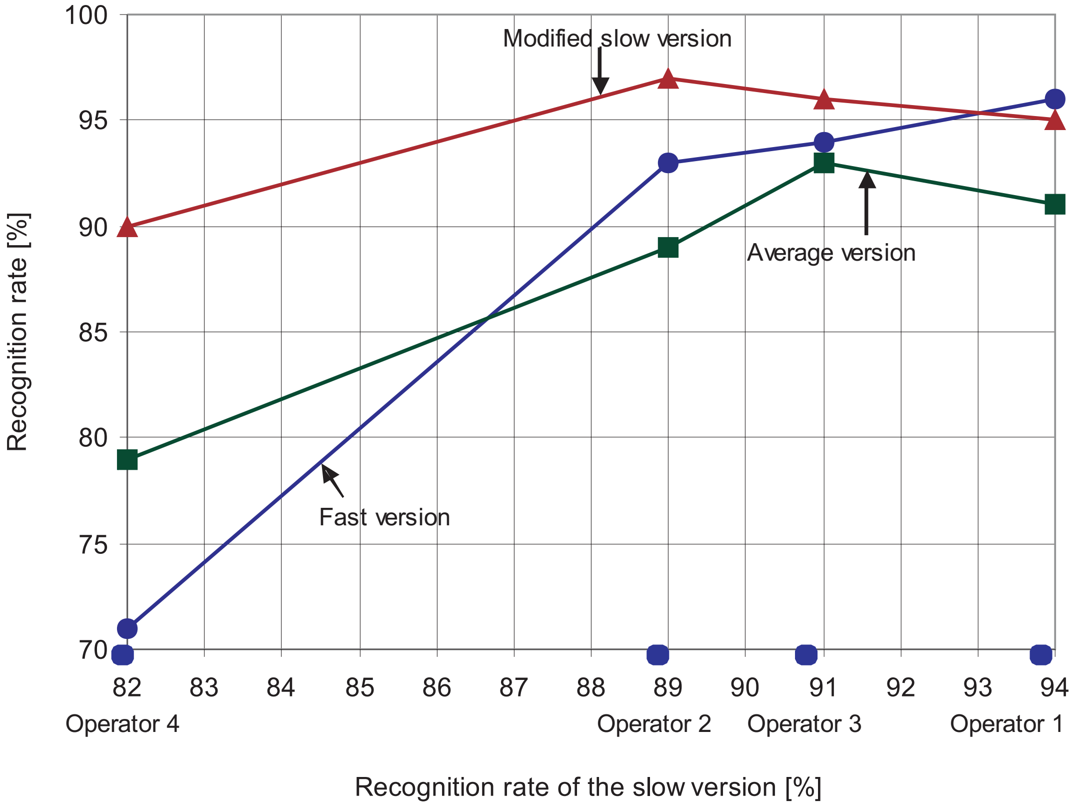


Figure 7: Recognition rate of the slow version of giving commands in the function of the recognition rate of the slow, fast and average version of giving commands for each operator

The choice for these 3 versions is based on the fact that some operators may speak faster and some more slowly with the system. The question would be then if the system performance remains the

same. In the modified slow version of giving commands, the syntax of some of the commands is changed to improve the performance of the system.
Figure 7 describes the research results for slow, fast and modified average version of giving voice commands by the operators.

IV. Conclusions and Perspectives

In the future, voice messages in natural language will undoubtedly be the most important way of communication between humans and machines. Great progress is made in many fields of science, where communication between the technological devices and the operator is an important task, e.g. motorization, road traffic, etc. The condition of the effectiveness of the presented intelligent two-way voice communication system between the technological device and the operator is to equip it with mechanisms of command verification and correctness. In the automated processes of production, the condition for safe communication between the operator and the technological device is analyzing the state of the technological device and the process before the command is given and using artificial intelligence for assessment of the technological effects and safety of the command. In operations of the automated technological processes, many process states and various commands from the operator to the technological device can be distinguished. A large number of combined technological systems characterize the realization of that process. In complex technological processes, if many parameters are controlled, the operator is not able to analyze a sufficient number of signals and react by manual operations on control buttons. The research aiming at developing an intelligent layer of two-way voice communication is very difficult, but the prognosis of the technology development and its first use shows a great significance in efficiency of supervision and production humanization.

References

[1] Jurafsky, D., Martin, J. H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice Hall, New Jersey 2000.

[2] Kacalak, W., Majewski, M.: Automatic recognition and safety estimation of voice commands in natural language given by the operator of the technical device using artificial neural networks, Proceedings of the ANNIE 2004 Conference, Artificial Neural Networks in Engineering ANNIE 2004, Vol. 14: Smart Engineering Systems Design, St. Louis, ASME Press, New York 2004, 831-836.

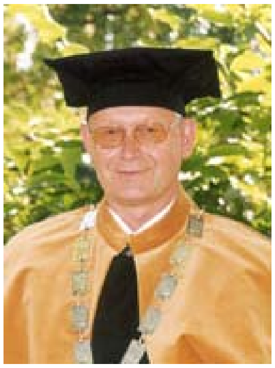
[3] Kacalak, W., Majewski, M.: Intelligent Layer of Two-Way Voice Communication of the Technological Device with the Operator, Lectures Notes in Artificial Intelligence 3070, Subseries of Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg New York 2004, 610-615.

[4] Kacalak, W., Majewski, M.: Selected problems of effect analysis and safety assessment of commands given by the operator of the technological device using artificial neural networks, International Industrial Simulation Conference ISC 2004, 7-9 June 2004, Malaga, Spain, Eurosis Ghent 2004, 35-39.

[5] O'Shaughnessy, D.: Speech Communications: Human and Machine, IEEE Press, New York 2000.

[6] Weinschenk, S., Barker, D. T.: Designing Effective Speech Interfaces, John Wiley \& Sons Inc., New York 2000.

Maciej Majewski was born in Polczyn-Zdroj, Poland. He studied Applications of Computers in Engineering at the Technical University of Koszalin. He also studied at the University of Granada in Spain and was given professional training by the European Commission DG in Luxembourg in 2001. He was conferred a Ph.D. degree in 2004 and has been working as a lecturer at the Technical University of Koszalin.



Wojciech Kacalak was born in Zdunska Wola, Poland. He obtained his doctor's degree in 1974 and as of 1989 he has been professor of technical sciences. Since 1970 he has been working at the Mechanical Faculty of the Technical University of Koszalin doing research on optimization of manufacturing processes using artificial neural networks.