

# Conceptual Design of Innovative Speech Interfaces with Augmented Reality and Interactive Systems for Controlling Loader Cranes

Maciej Majewski and Wojciech Kacalak

**Abstract** The paper presents a concept of implementation of augmented reality, interactive systems and an operator's speech interface for controlling lifting devices. The aim of the experimental research is to design a prototype of an innovative system for controlling a mobile crane, equipped with a vision and sensorial system, interactive manipulators with force feedback, as well as a system for bi-directional voice communication through speech and natural language between an operator and the controlled lifting device. The system is equipped with several adaptive intelligent layers for human biometric identification, speech recognition, word recognition, analysis and recognition of commands and messages, sentence meaning analysis, command effect analysis and safety assessment, process supervision and human reaction assessment. The article presents the designed structure of an innovative system for interaction of lifting devices with their operators, which provides versatility in terms of application of the system when used for controlling and supervising modern machines and devices in conditions of difficulty or increased risk.

**Keywords** Intelligent interface · Interactive system · Speech communication · Intelligent control · Augmented reality · Applied neural networks

## 1 Introduction

In the upcoming era we will be facing rapid development of robotics and cybernetics. Implementation of achievements of those fields has all the potential of paving a path to producing the best results in terms of performance and safety of transshipment of materials and products. This is perfectly exemplified by innovative systems of control

---

M. Majewski (✉) · W. Kacalak  
Faculty of Mechanical Engineering, Koszalin University of Technology,  
Raclawicka 15-17, 75-620 Koszalin, Poland  
e-mail: maciej.majewski@tu.koszalin.pl

W. Kacalak  
e-mail: wojciech.kacalak@tu.koszalin.pl

designed for processes of precise positioning of objects and cargo, which make use of intelligent systems of interaction between lifting devices and their operators. The most up-to-date artificial intelligence-based technologies find their application in the process of designing modern systems for controlling and supervising machines. An example are vision systems—machine vision, augmented reality (AR), voice communication as well as interactive controllers providing force feedback.

The design and implementation of intelligent human-machine interfaces is an important field of applied research. A speech interface using the natural language is ideal because it is the most natural, flexible, efficient, and economical form of human communication. Application of intelligent human-machine speech interfaces offers many advantages. It ensures robustness against human errors and efficient supervision of cargo positioning processes with adjustable level of automated supervision. Speech interfaces also improve the cooperation between a human and a mobile crane in respect to the richness of communication. This eliminates the need for a human to be present near working lifting devices. Further, speech interaction allows for higher organization level of transport processes, which is significant for their efficiency and machine humanization. Transport decision and optimization systems can be remote elements of transport processes.

The presented research involves the development of a system for controlling a mobile crane, equipped with a vision and sensorial system, interactive manipulators with force feedback, as well as a system for bi-directional voice communication through speech and natural language between an operator and the controlled lifting device. The main aim of the experimental research is to investigate potential possibilities of using innovative operator-machine communication technologies to control lifting devices. The goal is to develop higher-level intelligent systems for supervision of cargo placement, and to make an attempt at combining results of the research into a uniform concept of an innovative system for controlling a crane as well as building its prototype.

## 2 The State of the Art

Recent advances in development of prototypes of human-machine speech-based interfaces are described in articles in [1–4]. The speech and natural language of these interfaces are spontaneous and their vocabularies are usually about tens of thousands of words. In many potential applications of spoken language understanding systems, the limiting factor may not be the error rates but an ability of the system to manage and recover from errors. The integration of speech recognition and natural language in applications faces many of the same challenges that each of its components face: accuracy, robustness, portability, speed, and size. The integration also gives rise to some new challenges [5] which include integration strategies, coordination of understanding components with system outputs, and handling of spontaneous speech effects.

With few exceptions, current research in human-machine speech-based interfaces has focused on understanding of spoken input [3, 4]. However, many if not most applications involve a collaboration between the human and the machine. In many cases, spoken language output is an appropriate means of communication that may or may not be taken advantage of, because of lack of coordination of understanding components with system outputs. This paper offers an approach using a concept of a complete speech communication system to deal with the above problems.

### 3 The Design of an Innovative Speech Interface

The ARSC (Augmented Reality and Smart Control) prototype control system uses: intelligent visual-aid systems based on augmented reality, interactive manipulation systems providing force feedback, as well as natural-language voice communication techniques. Realization of the cargo processes is in conditions of uncertainty and unrepeatability of processes. We propose a new concept which consists of a novel approach to these systems, with particular emphasis on their ability to be truly flexible, adaptive, human error-tolerant, and supportive both of human-operators and data processing systems. A diagram depicting the ARSC system concept is presented in abbreviated form in Fig. 1. The concept specifies integration of a system for bi-directional natural-language communication with a visual and sensorial system. The research has dedicated special attention towards the possibility of partial or full commercialization of its results.

The proposed interactive system (Fig. 2) contains many specialized modules and it is divided into the following subsystems: a subsystem for voice communication between a human-operator and the mobile crane, a subsystem for natural language meaning analysis, a subsystem for operator's command effect analysis and evaluation, a subsystem for command safety assessment, a subsystem for command execution, a subsystem of supervision and diagnostics, a subsystem of decision-making and learning, a subsystem of interactive manipulators with force feedback, and a visual and sensorial subsystem.

The novelty of the system also consists of inclusion of several adaptive layers in the spoken natural language command interface for human biometric identification, speech recognition, word recognition, sentence syntax and segment analysis, command analysis and recognition, command effect analysis and safety assessment, process supervision and human reaction assessment.

#### 3.1 *Meaning Analysis of Messages and Commands*

The proposed method for meaning analysis of words, commands and messages uses binary neural networks for natural language understanding. The motivation behind using this type of neural networks for meaning analysis is that they offer an advan-

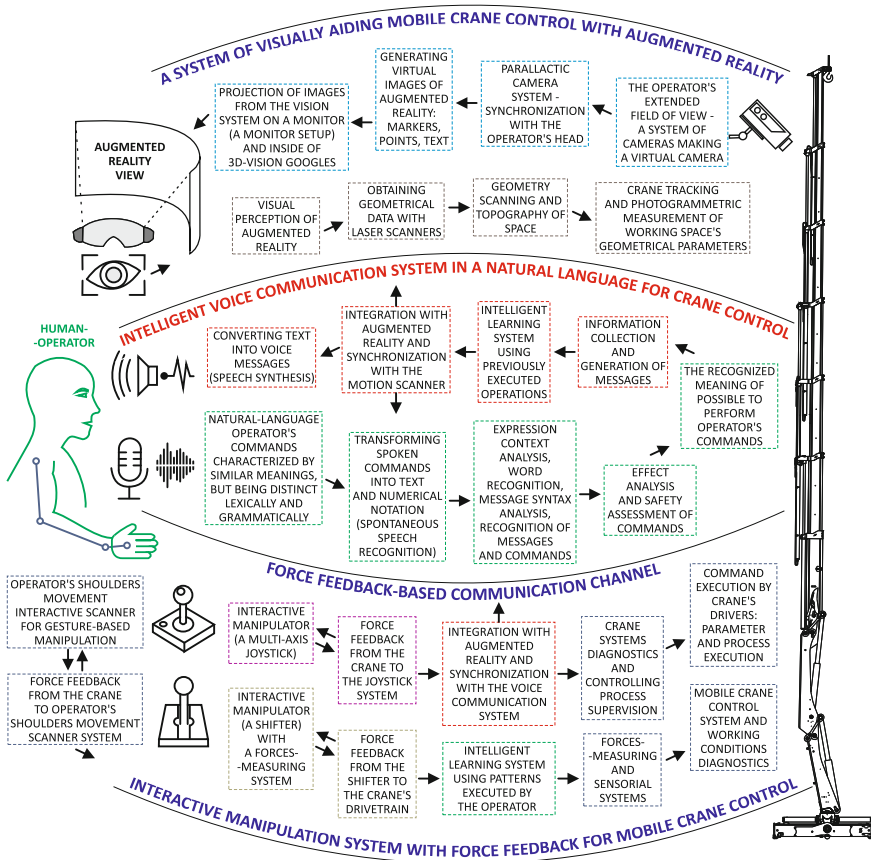
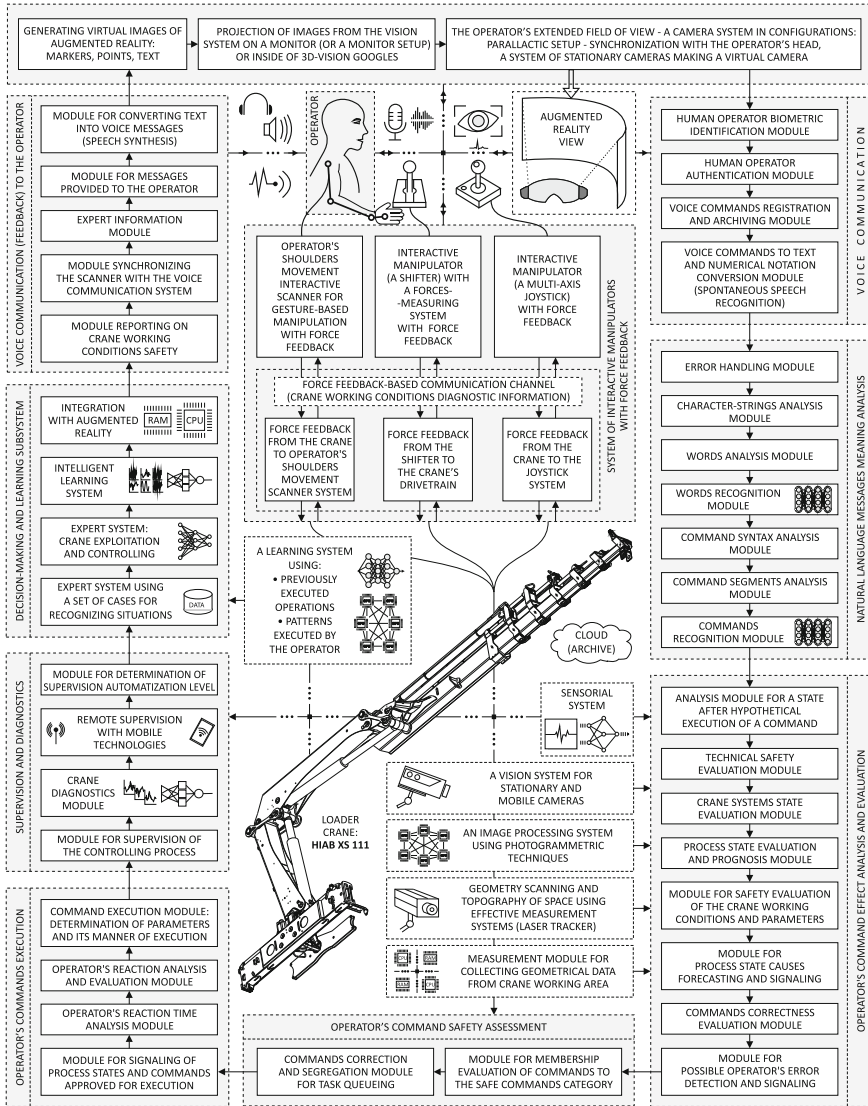


Fig. 1 A concept of the ARSC control systems for loader cranes (Hiab XS 111)

tage of simple binarization of words, commands and sentences, as well as very fast training and run-time response.

In the natural language meaning analysis process, the speech signal is converted to text and numeric values by the spontaneous speech recognition module. After a successful utterance recognition, a text command in a natural language is further processed. Individual words treated as isolated components of the text are subsequently processed with the character-strings analysis module. The letters grouped in segments are then processed by the word analysis module. In the next stage, the analyzed word segments are inputs of the neural network for recognizing words (Fig. 3). The network uses a training file containing also words and is trained to recognize words as command components, with words represented by output neurons.



**Fig. 2** Designed structure of an innovative system for interaction of lifting devices with their operators equipped with a speech interface, vision and sensorial systems, and interactive manipulators with force feedback

In the meaning analysis process of text messages in a natural language, the meaning analysis of words as command or message components is performed. The recognized words are transferred to the command syntax analysis module which uses command segment patterns. It analyses commands and identifies them as segments with regards to meaning, and also codes commands as vectors. They are sent to the com-

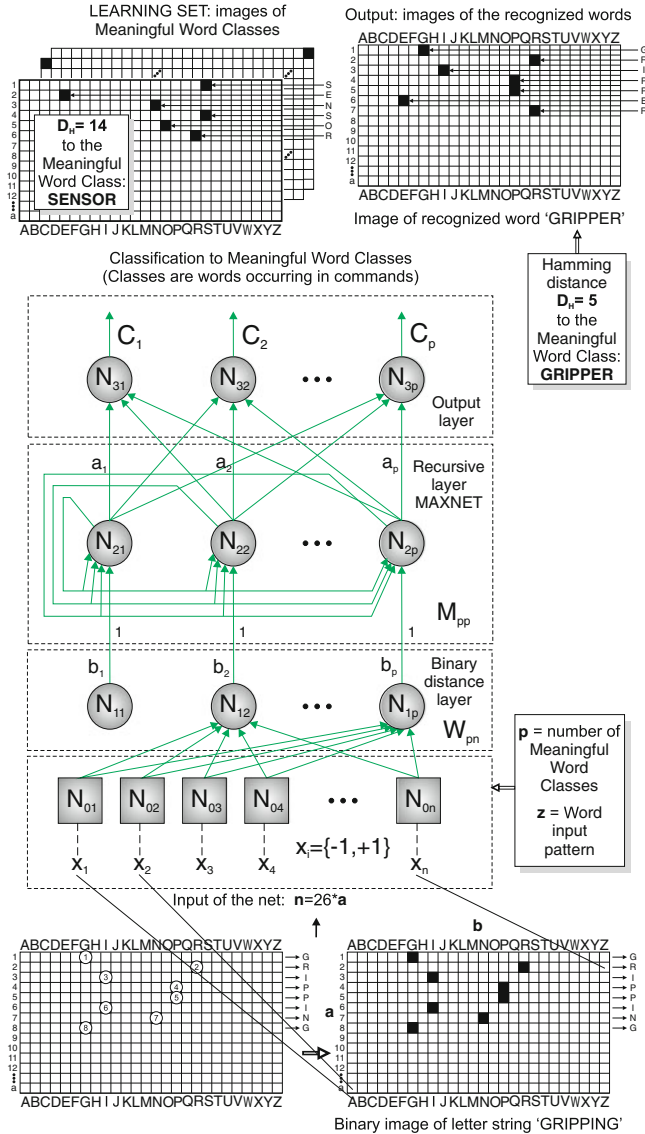


Fig. 3 Illustrative example of word recognition using neural networks

mand segment analysis module with Hamming networks with encoded command segment patterns. The commands become inputs of the command recognition module. The module uses a 3-layer Hamming network to classify the command and find its meaning (Fig. 4). The neural network of this module uses a training file with possible meaningful commands.

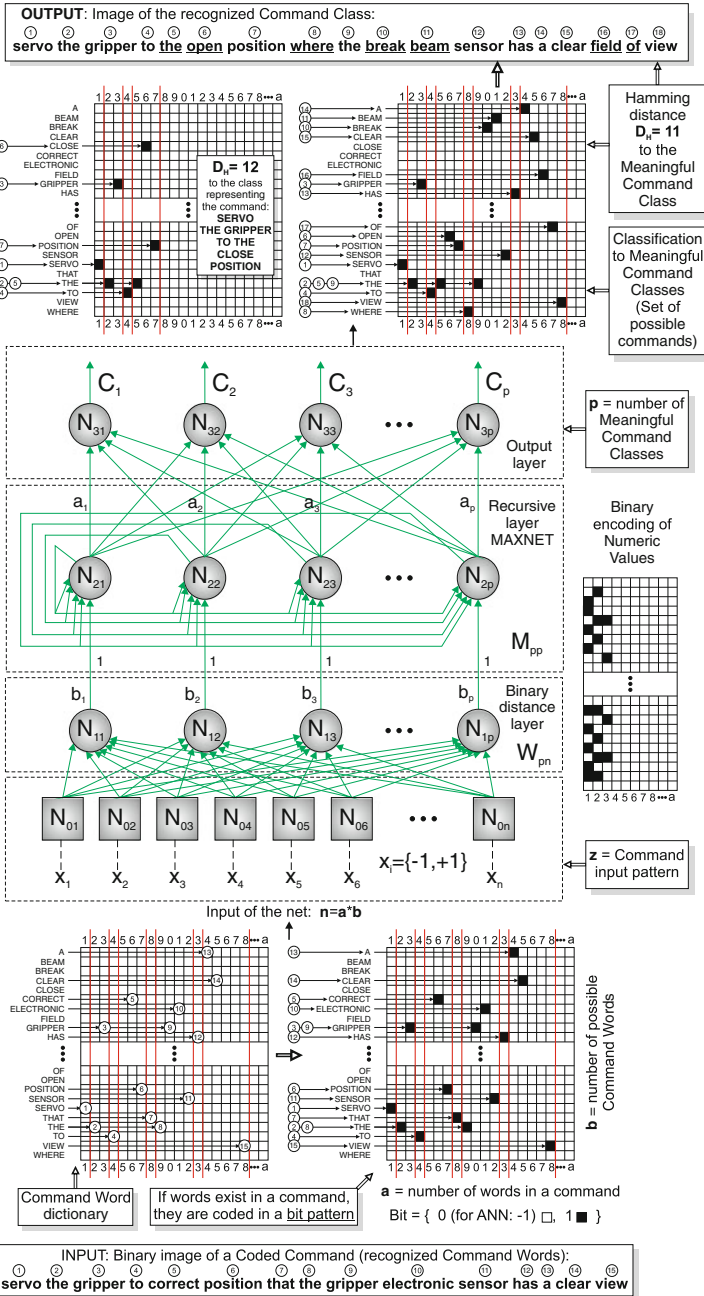


Fig. 4 Illustrative example of recognition of commands using neural networks

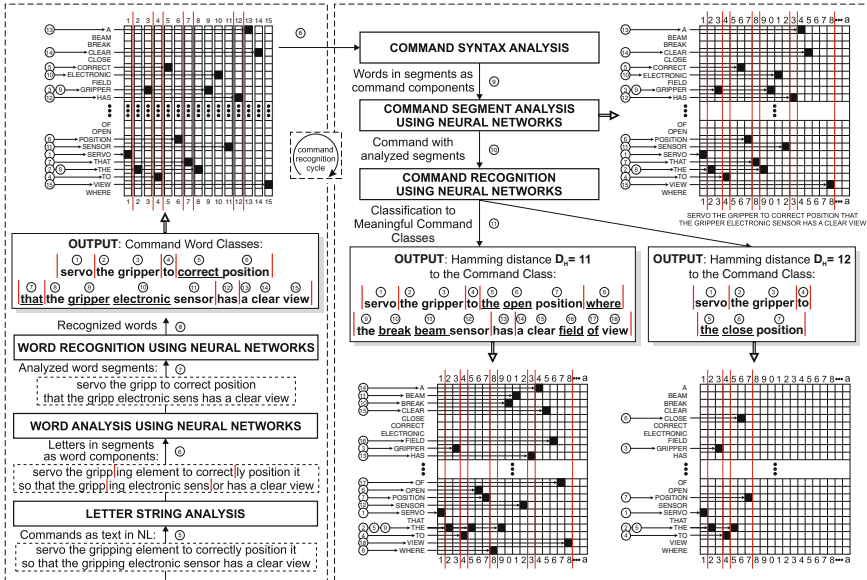


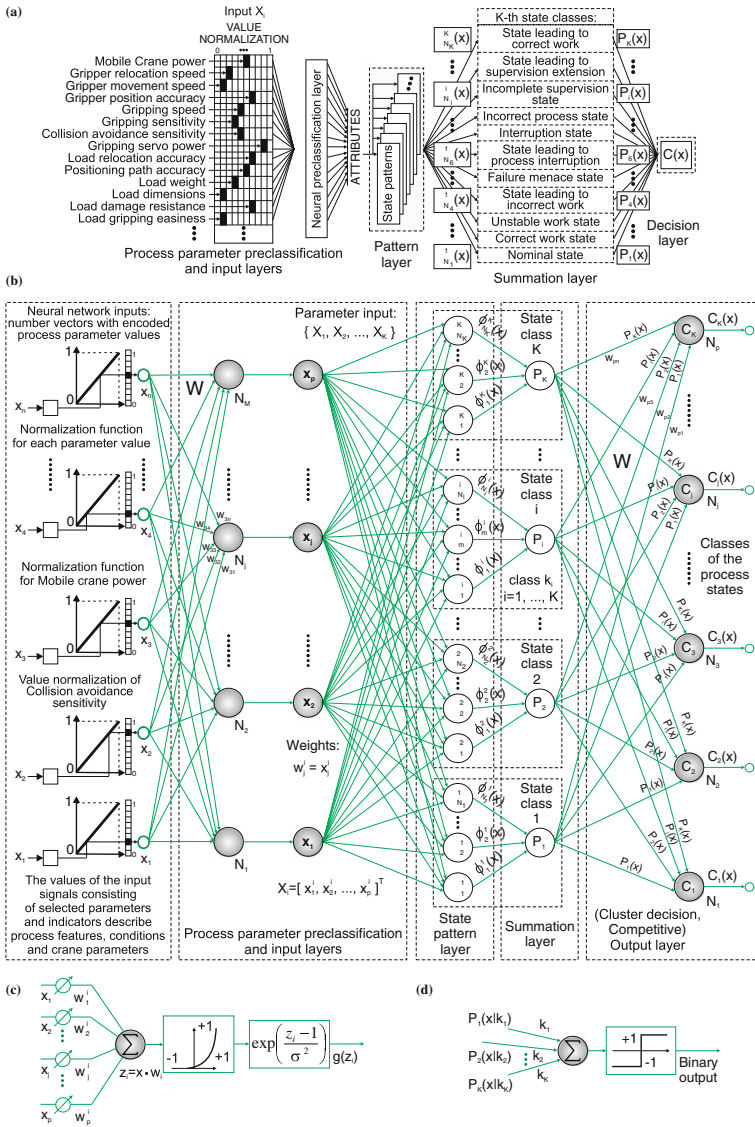
Fig. 5 Block diagram of exemplary command meaning analysis cycle

The Hamming network is chosen for both the word recognition and command recognition module as shown in Figs. 3 and 4. The network allows for simple binarization of words and sentences. The cycle of exemplary command meaning analysis is presented in Fig. 5. The structure and features of the Hamming network as a classifier-expert module for word and sentence recognition were described in detail in [6]. The network implements the nearest-neighbor classification rule. Each training data vector is assigned a single class and during the recognition phase only a single nearest vector to the input pattern  $x$  is found and its class  $C_i$  is returned. There are two main phases of the operation of the network: training (initialization) and classification.

### 3.2 Command Effect Analysis and Safety Assessment

In the innovative speech interface, the problem of effect analysis and safety assessment of commands can be solved with hybrid probabilistic neural networks. The proposed method (Fig. 6a) uses developed hybrid multilayer neural networks consisting of a modified probabilistic neural network combined with a single layer classifier. The probabilistic neural network is interesting, because it is possible to implement and develop numerous enhancements, extensions, and generalizations of the original model [7]. The presented approach can be suitable for many automated cargo manipulation processes.





**Fig. 6** a Hybrid neural model of effect analysis and safety assessment of commands in a cargo manipulation process. b The architecture of the hybrid neural network used. c Neuron of the pattern layer. d Neuron of the output layer

The effect analysis and safety assessment of commands is based on information on features, conditions and parameters of the cargo positioning process. The input signals of the network can include: mobile crane power, gripper relocation speed, gripper movement speed, gripper position accuracy, gripping speed, gripping sensi-

tivity, collision avoidance sensitivity, gripping servo power, load relocation accuracy, positioning path accuracy, load weight, load dimensions, load damage resistance coefficient, and load gripping easiness coefficient. The values of the input signals are subject to normalization.

The effect analysis and safety assessment is performed by the developed hybrid network that works as a classifier of the cargo manipulation process state. Its output computes the following classes: nominal state, correct work state, unstable work state, state leading to incorrect work, failure menace state, state leading to process interruption, process interruption state, incorrect process state, incomplete supervision state, state leading to supervision extension, state leading to correct work. The architecture of the hybrid network used is shown in Fig. 6b–d. It is composed of interconnected neurons organized in successive layers. The hybrid network consists of five layers: preprocessing, input, pattern, summation and output layers.

## 4 Conclusions and Perspectives

The designed interaction system is equipped with the most modern artificial intelligence-based technologies: vision systems, augmented reality, voice communication and interactive manipulators with force feedback. Modern control and supervision systems allow to efficiently and securely transfer, and precisely place materials, products and fragile cargo.

The proposed design of the innovative speech interface for controlling lifting devices has been based on hybrid neural network architectures. They serve as flexible engines for development, experimentation and validation of the presented design. The design can be considered as an attempt to create a standard intelligent system for execution, control, supervision and optimization of cargo handling processes using communication by speech and natural language. This is important for development of effective and flexible cargo manipulation methods.

**Acknowledgments** This project is financed by the National Centre for Research and Development, Poland (NCBiR), under the Applied Research Programme—Grant agreement No. PBS3/A6/28/2015.

## References

1. Kacalak, W., Majewski, M., Zurada, J.M.: Intelligent e-learning systems for evaluation of user's knowledge and skills with efficient information processing. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L., Zurada, J.M. (eds.) ICAISC 2010. LNCS, vol. 6114, pp. 508–515. Springer, Heidelberg (2010)
2. Kacalak, W., Majewski, M., Budniak, Z.: Interactive systems for designing machine elements and assemblies. *Manag. Prod. Eng. Rev.* **6**(3), 21–34 (2015). De Gruyter Open
3. Kumar, A., Metzke, F., Kam, M.: Enabling the rapid development and adoption of speech-user interfaces. *Computer* **47**(1), 40–47 (2014). IEEE

4. Ortiz, C.L.: The road to natural conversational speech interfaces. *IEEE Internet Comput.* **18**(2), 74–78 (2014)
5. Kacalak, W., Majewski, M., Budniak, Z.: Intelligent automated design of machine components using antipatterns. In: Jackowski, K., Burduk, R., Walkowiak, K., Wozniak, M., Yin, H. (eds.) *IDEAL 2015. LNCS*, vol. 9375, pp. 248–255. Springer, Heidelberg (2015)
6. Majewski, M., Zurada, J.M.: Sentence recognition using artificial neural networks. *Knowl. Based Syst.* **21**(7), 629–635 (2008). Elsevier
7. Specht, D.F.: Probabilistic neural networks. *Neural Netw.* **3**(1), 109–118 (1990)