The Non-Force Interaction Theory for Reflex System Creation with Application to TV Voice Control

Iurii Teslia[1], Nataliia Popovych[1], Valerii Pylypenko[2] and Oleksandr Chornyi[1]
[1]*Kyiv National University of Constuction and Archtecture, Povitroflotsky Avenue, Kyiv, Ukraine*
[2]*International Research and Training Center for Information Technologies and Systems, National Academy of Sciences of Ukraine, Kyiv, Ukraine*

Keywords: Non-Force Interaction, Information Theory, Speech Recognition, Reflex Systems, Self-organization, Phoneme-by-Phoneme Recognizer.

Abstract: The paper presents the aspects and conclusions of the theory of non-force interaction, discloses the possibilities of its application to the creation of artificial intelligence systems. The method of calculation of the reaction on the non-force actions in the sphere of intellectual activity and the universal model of intellectual reflex system are proposed. On this basis the reflex voice system for control of technical devices is developed. The article describes the system and results of its usage for controlling the TV. In particular: the special features of controlling TV's functionality with voice commands; ignoring the commands, that are not addressed to the system; learning new commands and desired reactions on user's requests; adjusting system's behaviour based on user's speech. The work is aimed to demonstrate the possibilities of the theory of non-force interaction in the field of study of the mechanisms of the brain, and creation on this basis artificial systems that approach in terms of its "intelligence" to human intelligence.

1 PROBLEM STATEMENT

Despite of the investment of considerable amount of work and financial resources into the speech recognition research there are still no effective multipurpose tools for understanding of oral speech. Without solving of this problem it is difficult to talk about successful automation of all spheres of human activity. The lack of good solutions, which can be generally valid in many areas of human activity, requires searching for a fundamentally new approach to tackle this task. The theory of non-force interaction (Teslia, 2005) provides significant help for finding a solution of this problem. It gives a formal basis for describing processes of information exchange in biological systems in general and in the human brain in particular. This gives a prospect for the development of new tools for solving many problems in cybernetics.

1.1 Analysis of the Main Research and Publications

As examined by Anusuya and Katti (2009, p. 181), in the area of development of human speech processing systems a number of interesting methods and approaches were developed. Global giants of the software industry – Nuance Communications, Google, Apple, have implemented the most effective ones within commercially successful systems.

The most notable achievement in this area is a joint product of Apple and Nuance – the personal assistant and a question-and-answer system Siri. The application uses natural language processing for answering questions and offering advice. Developing Siri, Apple used the results of 40 years of research conducted by "Artificial Intelligence Centre" and the work of the research groups from the most famous universities in the world. This study is, perhaps, the largest artificial intelligence project to date. Google also offers a very high quality solution for speech recognition and voice control for the operating system Android – "Voice Actions" system. All of these solutions are designed for mobile platforms.

These companies have created a voice control system for TV – the company Nuance with its Dragon TV platform, and Google with Google TV. Both systems are a supplement to the TV, implemented as a separate hardware module with the

Teslia I., Popovych N., Pylypenko V. and Chornyi O..
The Non-Force Interaction Theory for Reflex System Creation with Application to TV Voice Control.
DOI: 10.5220/0004754702880296
In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence* (ICAART-2014), pages 288-296
ISBN: 978-989-758-015-4

corresponding software platform.

1.2 The Unsolved Part of the Problem

All of these systems have one distinguishing feature in common – they are built using an approach based on large amounts of training data and the use of "cloud" computing centres (Jyothi, Johnson, Chelba and Strope, 2012, p. 41). This requires a permanent Internet connection to transmit the speech to remote server for processing. Also, the systems listed above poorly support international languages, e.g. Ukrainian and Russian.

1.3 The Purpose of the Article

It is therefore necessary to develop speaker-independent speech recognition tools, which could be easily implemented and adapted to different languages and will not require large computing resources. This article is focused on presenting a new technology for creating systems able to respond to commands expressed in natural language.

2 THE MAIN MATERIAL OF RESEARCH

As has been shown and proven by biologists, development of reflexes to various influences lies at the basis of living beings functioning. Reflex (from Lat. reflexus – reflected) – is a stereotypical reaction of the living organism to a stimulus that passes involving the nervous system (Purves, Williams, White and Mace, 2004).The assumption of the reflex nature of the higher centres of the brain was first developed by scientist-physiologist I. Sechenov. Before him, physiologists and neurologists did not dare to raise the question of the possibility of a physiological analysis of the mental processes – this was the field of expertise of the psychologists. The ideas of I. Sechenov were further developed in the works of I. Pavlov, who discovered the methods of the objective experimental research of the brain cortex functions, developed the method of generation of conditioned reflexes and worked out a theory of higher nervous activity. A great contribution to the formation of the theory of reflexes was made by Charles S. Sherrington (Nobel Prize in Physiology and Medicine, 1932). He discovered reflexes coordination, mutual inhibition and facilitation.

At the level of simple biological objects the reflexes allow the production of the "right" response to the state of the environment. At the human level reflexes are developed not only as a reaction on the physical influence, but also as a reaction on the informational impact in the socio-political sphere, on actions of other people, such as a teacher at the school, colleagues, etc. Of course, these reflexes are very complex, ambiguous and cannot be represented by a simple stimulus–response model. This model can be rather presented as a "set of actions"–"the most favourable reaction from the standpoint of a positive attitude".

The principles of development of artificial intelligence systems, which are based on these ideas, are examined in the theory of non-force interaction (Teslia, 2013b). The practical output of this theory is creation of reflex intelligent systems. In particular, the systems for evaluation of investment proposals in development; natural language access to databases; assessment of the impact of harmful substances in the water resources of the region on the health of the population; predicting of the outcomes of sport events. These systems are described by Teslia in his work (2010). The main advantages of these systems are the ease of their creation and the effectiveness of the solutions of various intellectual tasks.

2.1 Fundamentals of the Theory of Non-force Interaction

The main idea of the theory of non-force interaction (NFI) is that any interaction in the Nature first leads to a change in the material object's internal organization (*introformation*), which in its turn leads to a change in object's behavior (motion) (Teslia, 2005). For one-dimensional motion the introformation is represented by the geometric model (Fig. 1) with two domains of displacements (DD).

As measures, which "generate" motion, the difference and sum of the sizes of DD are taken (Teslia, 2005). Let's define $d = i^+ - i^-$ as object's certainty about the displacement in the direction Z, $i = i^+ + i^-$ as object's awareness about the displacement in the direction Z.

Using the relativistic mechanics we obtain the ratio between velocity of drift (V), probability of displacement (p), certainty (d) and awareness (i) of the material objects (Teslia, 2013b):

$$V = (2p - 1) \cdot c \Rightarrow p = \frac{V + c}{2c} \quad (1)$$

$$i = \frac{1}{2\sqrt{p(1-p)}} \quad (2)$$

$$d = \pm 0.5\sqrt{\frac{p}{1-p} + \frac{1-p}{p} - 2} \quad (3)$$

$$i = \sqrt{d^2 + 1} \quad (4)$$

$$p = 0.5 + \frac{d}{2i} \quad (5)$$

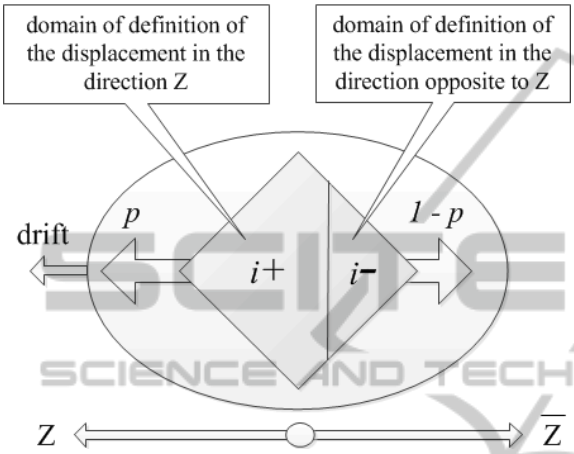

Figure 1: Geometrical interpretation of introformation.

Based on the laws of momentum conservation the sum of certainties is obtained (Teslia, 2013a):

$$d_\Sigma = \sum_i d_i \quad (6)$$

where d_Σ – the total certainty of objects in a closed system; d_i – certainty of material object M_i.

From the formula of the relativistic addition of velocities we obtain operation of the addition of certainties (Teslia, 2005):

$$d_Y = d_X \cdot i_{YX} + d_{YX} \cdot i_X \Rightarrow \\ d_{YX} = d_Y \cdot i_X - d_X \cdot i_Y \quad (7)$$

where d_{YX} – difference of certainties of the objects M_X and M_Y; d_Y – certainty of the object M_Y; d_X – certainty of the object M_X; i_Y – awareness of the object M_Y; i_X – awareness of the object M_X.

In the basis of NFI theory lays an assumption that our brain functioning is based on the same laws. The given interpretation can be used to create artificial intelligent systems, whose computing elements interact with each other under the same laws as everything interacts in nature!

Consider the example. Let the ball M_X is moving with speed V_X in the direction of Z (probability of displacement is p_X). The other ball M_Y catches it up and hits it. As a result the ball's M_X speed changes and is V_{YX} (probability of displacement is p_{YX}). If it is being hit by the ball M_Z, which moves with speed V_Z, than ball's M_X speed gets the value V_{ZX} (probability of displacement is p_{ZX}). As the punches are absolutely elastic we can calculate the speed of the ball M_X after collision with two balls M_Y and M_Z and obtain the following:

$$p_{ZYX} = f(p_{YX}, p_{ZX}) \quad (8)$$

Let's now consider the non-force impact on the intellectual system. Let the probability of some behavior (response) is p_X. But upon exposure of M_Y the probability of reaction becomes equal $p_{YX} = p(M_X/M_Y)$. And upon exposure of M_Z the probability of reaction becomes equal $p_{ZX} = p(M_X/M_Z)$. Then the probability of the reaction of M_X upon exposure of both M_Y and M_Z:

$$p_{ZYX} = p\left(\frac{M_X}{M_Y M_Z}\right) = \\ f\left(p(M_X/M_Y), p(M_X/M_Z)\right) \quad (9)$$

And what is the probability of reaction upon many exposures? Here can be used an algorithm that comes up from the law of conservation of momentum. Namely on this basis the reflexes in living organisms can be generated.

The only criterion of truth is practice. To begin with a theoretical model has been confirmed in a series of computer experiments on natural-language texts (Teslia, 2005). It turned out that the statistical regularities in the lyrics in Russian correspond to the given equations. In addition, a several reflex intelligent systems were developed on this theory (Teslia, 2013a). Therefore it is tempting to apply NFI theory for creation reflex system with application to TV voice control.

2.2 Introformation Method for Solving the Problem of Speech Understanding

The question arises whether it is possible to present the process of voice interaction using the models that are developed in the theory of non-force interaction? Traditionally, the speech recognition systems are based on the principle: "spoken language" → "representation of speech as a set of linguistic constructions" → "speech understanding". Based on the theory of non-force interaction the other model of natural language recognition can be suggested: "spoken language" → "estimation of non-force (informational) impact on the reaction" → "reaction (understanding or behaviour)" = REFLEX.

In a system that is built upon the NFI theory, the

recognition process of an informative part of a voice command could be built in a different way. Let us assume that any repetitive acoustical phenomenon of speech (and not just speech) can be associated with some symbol of a finite alphabet. This correspondence can be established for quite similar sounds, in order to reduce the size of the alphabet. Thus any voice command can be associated with a phrase consisting of alphabet symbols. This is very similar to phonetic transcription, but with this approach we are not limited by the phonemes of a particular language or by the acoustic phenomena of speech. Then from the voice command system's perspective, each symbol from a phrase like that, especially their combinations – is an "impact" on the system, and consequentially the appropriate action of the system would be a "reaction".

At the learning stage a voice command will be transformed into a phrase consisting of alphabet symbols and this phrase will be associated with the desired system's reaction. Thus, using the proposed introformation method, it will be possible to evaluate an impact of symbol's combinations (which were taken from the phrase) on the selected reaction. So it is possible to build so-called "base of reflexes" that will store a magnitude of the impact of some sequence of symbols on the selection of a specific system's reaction.

At the recognition stage combinations of symbols will be used to determine the most likely reaction. The magnitude of the impact of symbols' combinations will be taken from the already trained "base of reflexes".

This paper proposes to build a voice command understanding system using the above-stated principles.

Let's adapt the proposed by Teslia (2013b) introformation method for solving the problem of speech understanding:

1. Formation of the base of reflexes (RB) showing the statistical information on the pair "external influence (the utterance)" → "the correct response".

2. Based on the probabilities of reactions, stored in the RB, the certainty (Teslia, 2010) of reflex voice-activated control system (RVCS) is calculated relatively to these reactions (3). Let's use the following notations:

p_0 – unconditional probability of reaction x;

$p_j = p(x/y_j)$ – probability of reaction x providing that there was an action y_j (some fragment of human speech):

$$d_j = \begin{cases} +0.5\sqrt{\frac{p_j}{1-p_j}+\frac{1-p_j}{p_j}-2}, p_j \geq 0.5 \\ -0.5\sqrt{\frac{p_j}{1-p_j}+\frac{1-p_j}{p_j}-2}, p_j < 0.5 \end{cases}, \quad j = \overline{0,n} \tag{10}$$

where $d_j, j = \overline{1,n}$ – is the definiteness of reaction x providing that there was an action on the RVCS y_j (d_0 – is the definiteness of reaction in the case if there was no action on the RVCS).

3. The information awareness of RVCS in relation to these reactions is calculated based on the known probabilities (2):

$$i_j = \frac{1}{2\sqrt{p_j(1-p_j)}}, j = \overline{0,n} \tag{11}$$

where i_j – awareness of the system about the reaction x, upon the influence y_j (i_0 – the system's awareness regarding the reaction x in case of the absence of action on the system).

4. Using Teslia's (2013b) method the total increment of the certainty of the system's action based on all the impacts on the system can be calculated using (6) and (7):

$$\Delta d = \sum_{j=1}^{n}(d_j \cdot i_0 - d_0 \cdot i_j) = \sum_{j=1}^{n} d_j \cdot i_0 - \sum_{j=1}^{n} d_0 \cdot i_j = i_0 \sum_{j=1}^{n} d_j - d_0 \sum_{j=1}^{n} i_j \tag{12}$$

where Δd – the total increment of certainty of the RVCS's reaction.

5. The calculation of the increment of the RVCS's awareness using (4):

$$\Delta i = \sqrt{\Delta d^2 + 1} \tag{13}$$

where Δi – the increment of the system's awareness.

6. The calculation of the new certainty of the reaction x using (7):

$$d_\Sigma = \Delta d \cdot i_0 + d_0 \cdot \Delta i \tag{14}$$

where d_Σ – a new certainty of the reaction x.

7. The calculation of a new awareness of the RVCS using (4):

$$i_\Sigma = \sqrt{d_\Sigma^2 + 1} \tag{15}$$

where i_Σ – a new awareness of the RVCS.

8. Estimation of the probability of the reaction x using (5):

$$p_\Sigma = p(x/Y) = 0.5 + \frac{d_\Sigma}{2i_\Sigma} \quad (16)$$

where $p_\Sigma = p(x/Y)$ – estimation of the probability of the reaction x under actions $Y = \{y_j\}, j = \overline{1,n}$.

The idea of the above method is that it points to the expected "response" on impact, the adequacy of which complies with the well-known and experimentally verified physical laws. By assumption, the interaction of neurons is based on the same laws and works in accordance with the proposed model of non-force interaction. On this basis it is possible to create artificial introformation processors that work as neurons do. Not as classic and well known in cybernetics formal neurons that are not more similar to natural neurons, as a paper boat is similar to an ocean vessel. More sophisticated and complex structures can be developed on this basis, the structures that respond to stimulation (actions). All this is embodied in reflex intelligent systems capable of accumulating information about the operational environment and developing an adequate response (reflexes) on everything in this environment.

The result of the work of the authors' research team in 2012 is the reflex voice control system, created using the above method. Let's review its features.

3 REFLEX VOICE-ACTIVATED CONTROL SYSTEM FOR TECHNICAL DEVICES

Reflex voice-activated control system (RVCS) is designed to work using free-language input (Pylypenko, 2007) of control commands and content into the technical device. RVCS increases the efficiency of the technical device and provides voice interface, which relieves the operator's hands.

The system consists of two main modules: an automatic phoneme-by-phoneme recognizer and the Kernel module.

Phoneme-by-phoneme recognizer (Pylypenko, 2009) is an external binary application. The phoneme-by-phoneme recognizer carries out transformation of a voice command into a sequence of symbols from the finite alphabet.

RVCS implements the introformation method of reflex generation mentioned above.

Input data for the system is voice command represented as sound wave. Output data is a control action on a control object e.g. execution of an identified command according to the parameters specified by voice.

While operating, the system generates the necessary visual and sound informational messages what gives the possibility to trace the process of commands identification, responses to them and, besides this, if necessary to change system's behaviour in real time.

Let's examine main parts of the system and its operation scheme with application to TV voice control in Fig. 2.

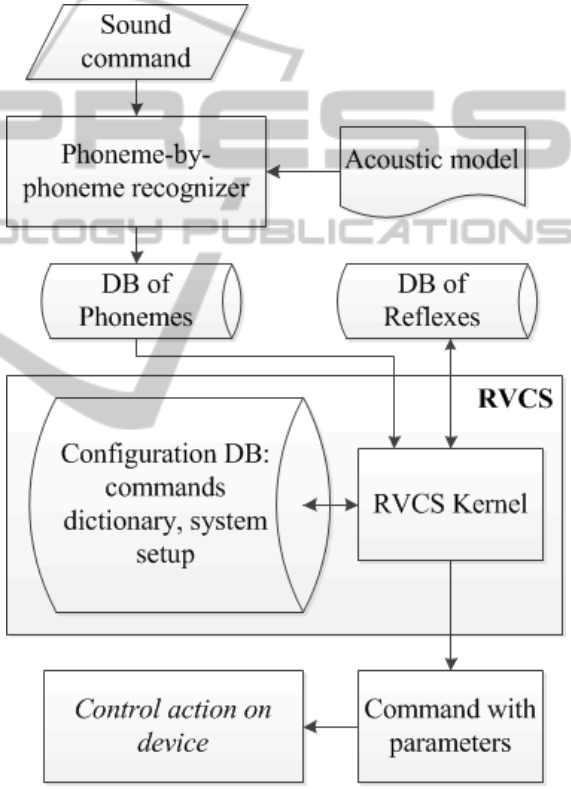

Figure 2: Structure of RVCS.

The Phoneme-by-Phoneme Recognizer. The phoneme transcribing algorithm developed by Pylypenko (2009) builds a phoneme sequence for a speech signal regardless to the dictionary. The constructed phoneme generative automata (Fig. 3) can synthesize all possible continuous speech model signals for any phoneme sequence. Then the phoneme-by-phoneme recognition of an unknown speech signal is applied.

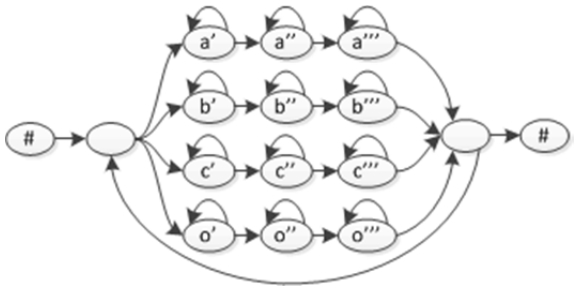

Figure 3: The phoneme generative automata.

First step of the phonemes recognition is **a feature extraction**. The speech signal is converted into a sequence of vector parameters with a fixed 25 ms frame and a frame rate of 10 ms. Then each parameter is pre-emphasized with filter:

$$P(z) = 1 - 0.97z^{-1} \quad (17)$$

Hamming window is applied. A fast Fourier transform is used to convert time domain frames into frequency domain spectra. These spectra are averaged into 26 triangular bins arranged at equal mel-frequency intervals. 12 dimensional mel-frequency cepstral coefficients (MFCCs) are obtained from cosine transformation and lifter. The log energy is also added as the 13th front-end parameter.

These 13 front-end parameters are expanded to 39 front-end parameters by appending first and second order differences of the static coefficients.

Cepstral mean normalization was applied to deal with the constant channel assumption.

Input to the feature extraction algorithm: a digitized sound wave.

Output from the feature extraction algorithm: a feature vector.

Also the phoneme-by-phoneme recognizer uses an **acoustic model**. This model uses hidden Markov Models (HMMs) with 64 mixtures Gaussian probability density function for acoustic modelling. Diagonal covariances Gaussians are used. All units are modelled by 3 left-to-right states with skip transition. 56 Russian context-free phonemes with pause unit are chosen. For pattern matching the Viterbi algorithm is used.

The experimental accuracy of finding phoneme at the right place for known utterance equals to approximately 70%.

Input to the recognizer: the digitized sound wave, which carries the voice command.

Output from the recognizer: a set of phonemes, which represents the voice command

Example of an input phrase: "Quickly, turn on channel eighteen for me".

Example of an output: kwɪklitərnɑntʃænəletinfɔrmi.

The result of acoustic waves recognition is stored in **DB of Phonemes**.

RVCS Kernel contains program implementation of:

- The introformation method;
- The algorithm of allocation of combinations of phonemes;
- The learning algorithm, which accumulate statistics.

The Kernel uses DB of reflexes to store magnitudes of influences of different phonemes' sets for choosing a reaction.

Configuration DB includes:

- The commands dictionary;
- Work protocol;
- Setup configuration.

Input to the Kernel:

- from the recognizer – the set of phonemes;
- from the DB of reflexes – magnitudes of influences of each phonemes' combinations (from the input phrase) for choosing of all possible reactions where these combinations were used;
- from Configuration DB – a formal description of an interfacing protocol for interaction with TV and format of a command, which should be sent to the TV for execution.

Output from the Kernel: an executable **command with parameters.**

Example of output: Speaker (Teslia); Command (Change channel); Channel name (); Channel number: Tens (10), Units (8).

When a voice command is transformed into "a command with parameters" it can be executed by an *adapter module,* which performs **control actions**.

Adapter module contains program implementation of the algorithm of control of the technical device.

Input to the adapter: the command with parameters transformed into the formula form.

Result: change of the parameters of the technical device.

Example: turning on the 18th TV channel.

RVCS system is very simple. It implements *reflex behaviour model*. It functions in two modes – training and control. The base of reflexes is formed in training mode (with a teacher). In control mode RVCS produces a reaction to the speaker's appeal. Also, in this mode the self-training is implemented (if the speaker was not satisfied with system's reaction).

Every reflex class is implemented in separate

system component. Classes of reflexes: announcer, command, channel name, level of number, level of tens, level of units.

Every system component can be represented as separate "introformation neuron". Input data – a complete input set of phonemes and/or the reaction of other "introformation neurons". Output data – the reaction that is passed to other "introformation neurons" or to the technical device.

A structure of each component (reflex class) can be represented as a combination of three tables (Fig. 4). The tables are required for storing combinations of phonemes (P), which were found in all commands earlier; all possible reactions of the system (R); and relationships between combinations of phonemes and reactions (L).

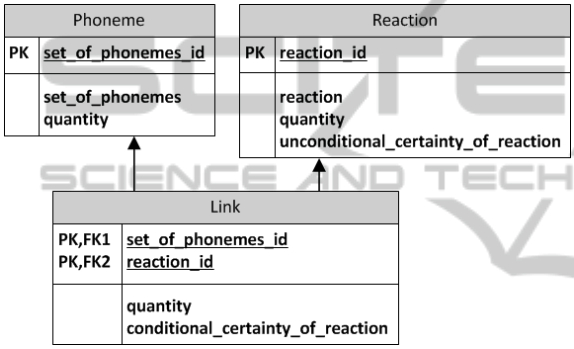

Figure 4: Structure of a component/reflex.

DB of reflexes can be represented as a combination of such components (Fig. 5).

Let's consider each table one-by-one. We note that the tables are filled during the training phase.

1. *Table of phonemes (P).* This table stores all *combinations of phonemes* from 2 to 10 symbols length, which were found in all voice commands during the training. Also it stores *a number of times* that a combination was encountered in all commands.

So, during the training, a set of phonemes from the example (kwɪklitərnɑntʃænəletinfərmi) will be split on set of combinations consisting of 2 symbols (kw, wɪ, ɪk, kl, li, etc.), 3 symbols (kwɪ, wɪk, ɪkl, etc.), 4 symbols and so on up to 10 symbols. Further, information about absolute quantity of occurrences of such n-gramms in all commands will be stored into this table (Table 1).

2. *Table of reactions (R). It stores system's reactions*. It contains: all possible reactions of RVCS; how many times each reaction was found in the training sample; *certainty* of each reaction without considering what phonemes' combination has influenced this reaction. Reaction "I don't know" ensures the system openness (Table 2).

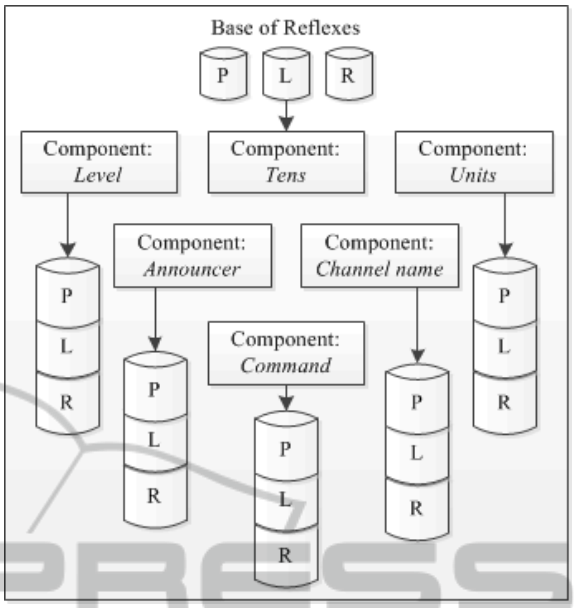

Figure 5: Structure of the base of reflexes.

Table 1: Phonemes sets table fragment.

P. Id	Set of Phonemes	Quantity
1	kw	699
2	wɪ	500
3	kwɪ	412
4	wɪk	388
5	itər	155
6	tərn	123

Table 2: Reaction tables fragment.

R. Id	Quantity	Reaction	Certainty
1	0	** I don't know **	-63,0178547
2	21	What do you need	-13,4087956
3	24	I'm talking to you	-12,5749908
4	106	Bless you	-6,0310190
5	736	*Turn on channel N*	-2,1567512
6	722	Turn on	-7,3254428
7	131	Forward	-5,4169182
8	39	Harmful	-9,9272670

So at the training stage a user indicates that combinations of phonemes from the example will lead to choosing the "Turn on channel N" reaction, a certainty of this reaction will be recalculated. Also a counter of occurrences of this reaction (in the entire training set) will be incremented.

3. *Table of relationship (L)* links table P and table R. It contains information about:

- *Quantity* – how many times which reaction was needed in case of some set of phonemes

as input data;
- *Certainty* of the reaction conditioned by the presence of a combination of phonemes, connected with this set of phonemes (Table 3).

In training mode the mentioned above tables accumulate information about what input phrases led to which reactions. In control mode this information is used for making the appropriate reactions to the speaker's appeals using the method, mentioned at the beginning of the section.

Table 3: Connections table fragment.

P. Id	R. Id	Quantity	Certainty
1 (kw)	2 (what do you n.?)	1	1,16683067
1 (kw)	5 (*turn on ch.*)	2	0,32652037
1 (kw)	6 (turn on)	2	3,62813851
1 (kw)	1 (don't know)	0	25,6847898
2 (wɪ)	5 (*turn on ch.*)	1	0,98438145
2 (wɪ)	1 (don't know)	0	44,4985958
3 (kwɪ)	5 (*turn on ch.*)	1	0,98438145
3 (kwɪ)	1 (don't know)	0	44,4985958
4 (wɪk)	5 (*turn on ch.*)	1	0,98438145
4 (wɪk)	1 (don't know)	0	44,4985958

The complete set of phonemes comes as input into every system component. In the reflex voice-activated control system it is not required:
- Create dictionaries;
- Execute morphological, syntactical, and semantic analysis of the text;
- Highlight the words and the commands.

The system reacts on audio stream and knows how to "extract" its informative part (based on the maximum certainty). The same way as the human does it.

To test these ideas RVCS was embodied in the TV voice control system – GUT (Teslia, 2013a).

For describing the procedure of the system's training it is necessary to emphasize that the system consists of two separated modules:
- The phoneme-by-phoneme recognizer, which translates a sound wave into a set of phonemes;
- The reflex voice control system, which recognizes a command based on this set.

Both systems are built on principles of supervised learning, thus two separate training data sets were used for their learning.

For the training of the phoneme-by-phoneme recognizer, namely for the acoustic model construction, we used a set of 500 sound files with words along with their phonetic transcriptions. The phonetic transcription was labeled manually. This set consisted of the most frequently used words for TV control.

For the training of the RVCS, namely for the construction of the reflex' DB, we used a set of 2000 different commands to TV. This training data set is comprised of sets of phonemes along with the corresponding reactions of the system. The already trained phoneme-by-phoneme recognizer for each command produced each set of phonemes.

The training samples for the reflex system also had some specific features:
- Spontaneous speech with different levels of sound and noise was used;
- The sentences were pronounced at a different pace.

The reliability of the developed system was evaluated on the control sample, which included 600 commands:
- 200 of simple commands;
- 200 short sentences (3 to 5 words) that contained commands;
- 200 long sentences (at least 8 words) that also contained commands.

Experimental results:
- For simple commands – 98% correctly recognized actions;
- For short sentences – 90%;
- For long sentences – 86%;
- False alarm (reaction to natural background voice in the room) – 81 occurrences for 1 hour of work.

Experimental results show good performance of the system.

4 CONCLUSIONS AND PROSPECTS FOR FUTURE RESEARCH

The paper shows that based on the non-force model of interaction a fundamentally new artificial intelligence system can be created in many areas of human activity. The theory of non-force interaction reveals the root causes and the laws of interaction including the laws of interaction of the basic elements of the human brain – neurons. Also it may give a different view of the known physical laws. Thus a probabilistic interpretation of mechanical motion was proposed by Klapchenko and Teslia (2011).

Reflex voice-activated control system for technical devices, presented in this article, is a bright example of practical application of NFI theory for

building of systems that adequately respond to the information given in natural language. Reflex system's main advantages are:

Independence from dictionaries as the system responds to sounds, not words. Therefore, you can control it by clapping (one, two, etc.), whistling, slurred speech (for the deaf people), etc. To prove this the system has been trained to respond properly to commands in German without any readjustments and changes in software.

Economy of resources – its database includes only combinations of phonemes which are met in the input stream and estimation of their influence on system's response.

Ease of teaching. The system will be delivered trained in one or more languages. Without reconfiguring of the stenographer the user, if necessary, can further train the system. For training user utters a phrase, or portrays a sound (it will be represented with some set of "phonemes") and presses the key combination, which is a response to this command. Repeating of this step but using a command in other words will lead to better memorizing. Generally, we do the same while teaching a child.

Openness. In the case of conflicting instructions, the system will choose the one with stronger non-force impact. In the case of commands with approximately equal impact the system's reaction will be "I don't know" (see the first row in Table 2.). The reaction will be the same in the case of an entirely new command, as a reaction to this command has not been developed yet.

REFERENCES

Anusuya, M. A. and Katti, S. K. 2009, 'Speech Recognition by Machine: A Review', *International Journal of Computer Science and Information Security*, vol. 6, no. 3, pp. 181-205.

Jyothi, P., Johnson, L., Chelba, C. and Strope B. 2012, 'Large-scale discriminative language model reranking for voice-search', *Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT,* Association for Computational Linguistics, Stroudsburg, pp. 41-49.

Klapchenko, V.I. and Teslia, I. 2011, 'Probabilistic interpretation of mechanical motion', *Cornell University Library*, [online] Available at: *arxiv.org/pdf/1102.0441* [Accessed: 20 Sep 2013].

Purves, D., Williams, S., White, L. and Mace, A. 2004, *Neuroscience,* Sunderland, Mass: Sinauer.

Pylypenko, V. 2007, 'Extra Large Vocabulary Continuous Speech Recognition Algorithm based on Information Retrieval', *Proceedings of the 8th Annual Conference of the International Speech Communication Association,* Antwerp, pp. 1461-1464.

Pylypenko, V. 2009, 'Raspoznavanie klyuchevykh slov v potoke rechi pri pomoshchi foneticheskogo stenografa' [Recognition of keywords in the flow of speech using phoneme-by-phoneme recognizer], *Rechevye tekhnologii*, no. 1, pp. 75-79, Russian.

Teslia, I. 2005, *Nesilovoe vzaimodeystvie [Non-force interaction],* Kondor, Kiev (Ukraine), Russian.

Teslia, I. 2010, *Vvedenie v informatiku prirody [Introduction to Informatics of Nature],* Maklaut, Kiev (Ukraine), Russian.

Teslia, I. 2013a, *Prezentatsiya real'nogo primeneniya TNV [Presentation of the actual use of the Non-force interaction theory]*, [online] Available at: http://www.youtube.com/watch?v=m_pYXVndpbc [Accessed: 20 Sep 2013].

Teslia, I. 2013b, 'Theory of non-violent interaction', *International Journal "Information Theories and Applications"*, vol. 20, no. 1, pp. 88-99.