



MINISTERUL EDUCAȚIEI NAȚIONALE
UNIVERSITATEA PETROL-GAZE DIN PLOIEȘTI
FACULTATEA DE INGINERIE MECANICĂ ȘI ELECTRICĂ

TEZĂ DE DOCTORAT

~ REZUMAT ~

CONTRIBUȚII LA DEZVOLTAREA UNUI SISTEM AUTOMAT DE RECUNOAȘTERE A VORBIRII

Conducător științific:

Prof. univ. dr. ing. Mihaela Oprea

Doctorand:

Mat. Daniela Gavrilă (Șchiopu)

Ploiești, 2014

CUPRINSUL TEZEI DE DOCTORAT

Lista figurilor	5
Lista tabelelor	9
Lista acronimelor	10
Introducere	12
Capitolul 1. Fundamente teoretice ale recunoașterii vorbirii	15
1.1 Considerații generale.....	15
1.2 Definirea problemei recunoașterii vorbirii.....	16
1.3 Arhitectura generală a unui sistem de recunoaștere a vorbirii	17
1.4 Prelucrarea și analiza semnalului vocal.....	19
1.4.1 Modelul acustic de producere a vorbirii.....	19
1.4.2 Parametrizarea semnalului vocal.....	19
1.4.3 Analiza prin predicție liniară	24
1.4.4 Analiza Fourier.....	29
1.4.5 Analiza homomorfică	29
1.4.6 Analiza liniară perceptuală.....	32
1.5 Metode și tehnici de recunoaștere a vorbirii	34
1.6 Structura generală a unui sistem automat de recunoaștere a vorbirii	35
1.7 Concluzii parțiale	38
Capitolul 2. Stadiul actual al cercetărilor în domeniul recunoașterii vorbirii.....	39
2.1 Metode utilizate în recunoașterea vorbirii	39
2.1.1 Metode statistice. Modelele Markov ascunse (MMA)	40
2.1.2 Metode neuronale. Rețelele neuronale artificiale.....	43
2.1.3 Metode sintactice. Modelarea lingvistică	55
2.1.4 Alinierea temporală dinamică (Dynamic Time Warping).....	57
2.1.5 Metode hibride MMA-RNA.....	60
2.1.6 Considerații sintetice asupra metodelor utilizate în recunoașterea vorbirii	61
2.2 Sisteme de recunoaștere automată a vorbirii.....	63
2.2.1 Sisteme dezvoltate pe plan internațional.....	64
2.2.2 Sisteme dezvoltate pe plan național	72
2.2.3 Criterii de evaluare a sistemelor de recunoaștere a vorbirii.....	75
2.2.4 Studiu al performanțelor sistemelor de recunoaștere automată a vorbirii	76

2.3	Concluzii parțiale	80
Capitolul 3. Contribuții privind dezvoltarea unor sisteme de recunoaștere a vorbirii pentru limba română		81
3.1	Recunoașterea deterministică a secvențelor în cazul cuvintelor rostite izolat	81
3.1.1	Alinierea temporală utilizată în recunoașterea vorbirii – studiu de caz.....	82
3.2	Aplicarea metodelor statistice în recunoașterea vorbirii.....	83
3.2.1	Extragerea parametrilor vocali utilizând metoda mel-cepstrală	83
3.2.2	SRV pentru recunoașterea cuvintelor rostite izolat utilizând modele Markov ascunse și analiza mel-cepstrală.....	84
3.2.3	SRV pentru recunoașterea cuvintelor conectate utilizând rețele Markov ascunse și analiza mel-cepstrală.....	91
3.3	Aplicarea metodelor neuronale în recunoașterea vorbirii.....	93
3.3.1	Sistem pentru recunoașterea cuvintelor rostite izolat utilizând rețele neuronale artificiale și coeficienți LPC.....	94
3.3.2	Sistem discriminant pentru recunoașterea cuvintelor rostite izolat utilizând rețele neuronale artificiale și coeficienți MFCC.....	98
3.4	Studiu comparativ al sistemelor implementate.....	102
3.5	Dezvoltarea unei aplicații web pentru construirea bazelor de date utilizate în recunoașterea comenzilor vocale.....	103
3.5.1	Caracteristicile bazelor de date utilizate în recunoașterea vorbirii.....	103
3.5.2	Realizarea aplicației web pentru construirea bazei de date BDCVR	104
3.6	Concluzii parțiale	106
Capitolul 4. Sistem hibrid pentru recunoașterea vorbirii în limba română		108
4.1	SRV_BC – Sistem de recunoaștere a vorbirii bazat pe cunoștințe	108
4.1.1	Arhitectura SRV_BC.....	109
4.1.2	Baza de cunoștințe a sistemului SRV_BC.....	110
4.2	Metrici pentru evaluarea performanțelor sistemului SRV_BC.....	115
4.2.1	Rata de recunoaștere la nivel de cuvânt	115
4.2.2	Rata de recunoaștere Good-Turing adaptată	115
4.2.3	Rata greșelilor ortoepice	117
4.3	Concluzii parțiale	117
Capitolul 5. Cercetări experimentale privind dezvoltarea sistemului automat SARCV de recunoaștere a vorbirii		118
5.1	Descrierea sistemului SARCV.....	118
5.1.1	Abordarea sistemică a procesului de recunoaștere a comenzilor vocale.....	119

5.1.2	<i>Structura ierarhică a sistemului automat de recunoaștere a comenzilor vocale.....</i>	<i>121</i>
5.2	Instrumente utilizate pentru dezvoltarea sistemului SARCV	122
5.2.1	<i>Prezentarea robotului LEGO Mindstorms NXT, Robo_Arm</i>	<i>122</i>
5.2.2	<i>Instrumente software utilizate</i>	<i>124</i>
5.3	Achiziția datelor	124
5.4	Modelarea acustică	125
5.5	Modelarea lingvistică.....	125
5.6	Modulele de recunoaștere ale sistemului SARCV	126
5.6.1	<i>Modulul de recunoaștere a vorbirii bazat pe tehnici de aliniere temporală</i>	<i>128</i>
5.6.2	<i>Modulul de recunoaștere a vorbirii bazat pe tehnici neuronale</i>	<i>130</i>
5.7	Rezultate experimentale	133
5.7.1	<i>Experimentul 1.....</i>	<i>134</i>
5.7.2	<i>Experimentul 2.....</i>	<i>136</i>
5.8	Concluzii parțiale	138
Capitolul 6. Concluzii finale, contribuții și direcții viitoare de cercetare		140
6.1	Concluzii finale.....	140
6.2	Sinteză a contribuțiilor	141
6.3	Direcții viitoare de cercetare	145
Lista lucrărilor publicate.....		146
Bibliografie		148
Webografie		157
ANEXE.....		159
ANEXA 1.	Glosar de termeni	159
ANEXA 2.	Producerea și percepția vorbirii	163
ANEXA 3.	Fonemele limbii române.....	165
ANEXA 4.	Modulul SARCV bazat pe DTW (cod Matlab).....	169
ANEXA 5.	Modulul SARCV bazat pe rețele neuronale (cod Matlab)	175

Cuvinte cheie: recunoașterea vorbirii, sistem automat, robot, comandă vocală, semnal vocal, parametrizarea semnalului vocal

REZUMAT

Recunoașterea vorbirii (sau recunoașterea limbii vorbite) de către un sistem computerizat constituie una din problemele de actualitate pe care cercetătorii încearcă să le rezolve, problematica acestui domeniu fiind determinată de mulți factori care influențează performanța recunoașterii și, în final, posibilitatea de a dezvolta cu succes sisteme inteligente (sisteme de înțelegere a vorbirii, sisteme de dialog etc.). Recunoașterea vorbirii va fi folosită cel mai probabil într-un viitor nu foarte îndepărtat atât în sisteme de dialog om-mașină, cât și în sisteme electronice încorporate sau sisteme automatizate industriale.

În acest context, prezenta teză de doctorat se subscrie preocupărilor de cercetare în domeniul recunoașterii vorbirii și celor referitoare la dezvoltarea de sisteme automate de recunoaștere a vorbirii.

Principalele contribuții ale tezei sunt atât contribuțiile de *cercetare teoretică* (cum sunt: extinderea capabilităților unui sistem de recunoaștere a vorbirii utilizând un sistem expert și tehnici de data mining, propunerea unor metrici pentru evaluarea performanței unui astfel de sistem și a unei structuri pentru un sistem automat de recunoaștere a vorbirii; de asemenea, propunerea unui algoritm pentru ghidarea sistemului în alegerea celui mai potrivit modul de recunoaștere), dar și *cercetare experimentală* (dezvoltarea unui sistem automat de recunoaștere a vorbirii, testarea funcționalității sistemului, construirea unei baze de date pentru cuvinte în limba română, validarea sistemului). Ceea ce merită menționat este faptul că toate elementele propuse sunt generale, dar au fost testate și validate pentru limba română, putând fi însă adaptate cu ușurință și la altă limbă.

Obiectivul principal al cercetării aplicative a constat în dezvoltarea unui sistem automat de recunoaștere a vorbirii care să permită execuția unor comenzi vocale de către un robot.

Principale obiective specifice au fost:

- Analiza stadiului actual al realizărilor în domeniul recunoașterii vorbirii;
- Realizarea unor cercetări privind abordarea sistemică a procesului de recunoaștere a vorbirii și proiectarea structurii și arhitecturii unui sistem automat de recunoaștere a vorbirii;
- Identificarea celor mai potrivite metode statistice, neuronale și de recunoaștere a formelor pentru recunoașterea cuvintelor rostite izolat sau conectate (comenzi vocale);
- Sporirea capabilităților unui sistem automat de recunoaștere a vorbirii prin includerea unui sistem expert și a unor tehnici de data mining, cu rolul de a îmbunătăți performanța unui astfel de sistem;
- Dezvoltarea unui sistem automat de recunoaștere a vorbirii (analiza, proiectarea, implementarea și testarea în vederea validării sistemului, prin execuția de comenzi vocale de către un robot).

Teza de doctorat este structurată în șase capitole, după cum urmează.

Capitolul 1, Fundamente teoretice ale recunoașterii vorbirii, evidențiază aspecte teoretice ale recunoașterii limbii vorbite. În cadrul acestui capitol este prezentată arhitectura generală a unui sistem de recunoaștere a vorbirii, cât și principalele metode de prelucrare și analiză a semnalului vocal.

Ultima parte a capitolului este dedicată prezentării structurii și arhitecturii unui sistem automat de recunoaștere a vorbirii, propuse de autoare.

Capitolul 2, Stadiul actual al cercetărilor în domeniul recunoașterii vorbirii, descrie stadiul actual al cercetărilor și realizărilor în acest domeniu, atât pe plan național, cât și internațional. Sunt descrise cele mai utilizate metode pentru recunoașterea vorbirii, cât și principalele sisteme dezvoltate. În finalul capitolului este realizat un studiu comparativ al metodelor de recunoaștere a vorbirii și o analiză a principalelor caracteristici și a performanțelor sistemelor prezentate.

Capitolul 3, Contribuții privind dezvoltarea unor sisteme de recunoaștere a vorbirii pentru limba română, conține o serie de experimente efectuate prin aplicarea de tehnici neuronale, statistice și de aliniere temporală pentru realizarea unor sisteme de recunoaștere a cuvintelor rostite izolat sau conectate.

În ultima parte a capitolului este prezentată contribuția autoarei privind dezvoltarea unei aplicații web ce are rolul de a achiziționa în mod semi-automat datele acustice pentru crearea unei baze de date de vorbire.

Capitolul 4, Sistem hibrid pentru recunoașterea vorbirii în limba română, propune sistemul hibrid SRV_BC, care are la bază o abordare bazată pe inteligența artificială, abordare ce utilizează cunoștințe fonetico-acustice și lexicale, pentru a dezvolta o bază de cunoștințe care îmbunătățește performanța sistemului de recunoaștere a vorbirii, utilizând un sistem expert și tehnici de data mining. De asemenea, sunt propuse două metrici ce evaluează performanța unui astfel de sistem.

Capitolul 5, intitulat Cercetări experimentale privind dezvoltarea sistemului automat SARCV de recunoaștere a vorbirii, evidențiază contribuțiile autoarei referitoare la dezvoltarea unui sistem automat de recunoaștere a unor comenzi vocale de către un robot LEGO Mindstorms NXT. Astfel, au fost propuse spre aplicare două abordări: una conexionistă (prin rețele neuronale artificiale) și una bazată pe recunoașterea formelor (aliniere temporală). Pentru ghidarea automată a sistemului în selecția uneia din cele două metode a fost propus un algoritm ce utilizează corelația încrucișată între semnale, cât și reguli euristice.

Capitolul 6, Concluzii finale, contribuții și direcții viitoare de cercetare, prezintă principalele concluzii și contribuții ale tezei de doctorat, cât și unele direcții de continuare a cercetărilor în acest domeniu.

Abstract of doctoral thesis entitled

**CONTRIBUTIONS TO THE DEVELOPMENT OF AN AUTOMATIC SPEECH
RECOGNITION SYSTEM**

Doctoral student:
Math. GAVRILĂ (ȘCHIOPU) DANIELA

Doctoral advisor:
Prof. Dr. Eng. OPREA MIHAELA

TABLE OF CONTENTS

Lists of figures	5
Lists of tables	9
Lists of acronyms	10
Introduction.....	12
Chapter 1. Theoretical fundamentals of speech recognition	15
1.1 Introduction to speech recognition.....	15
1.2 Defining the problem of speech recognition	16
1.3 Architecture of a speech recognition system.....	17
1.4 Processing and analyzing of speech signal	19
1.4.1 <i>Acoustic model of speech production.....</i>	<i>19</i>
1.4.2 <i>Parametrization of speech signal.....</i>	<i>19</i>
1.4.3 <i>Linear Predictive Coding.....</i>	<i>24</i>
1.4.4 <i>Fourier analysis.....</i>	<i>29</i>
1.4.5 <i>Homomorphic analysis of speech.....</i>	<i>29</i>
1.4.6 <i>Perceptual Linear Predictive analysis of speech.....</i>	<i>32</i>
1.5 Methods and techniques of speech recognition.....	34
1.6 Structure of an automatic speech recognition system	35
1.7 Chapter conclusions	38

Chapter 2. State of the art in speech recognition.....	39
2.1 Methods for speech recognition	39
2.1.1 <i>Statistical methods. Hidden Markov Models</i>	40
2.1.2 <i>Neural methods. Artificial neural networks</i>	43
2.1.3 <i>Sintactic methods. Linguistic modeling</i>	55
2.1.4 <i>Dynamic Time Warping</i>	57
2.1.5 <i>Hybrid methods HMM-ANN</i>	60
2.1.6 <i>A synthesis of the methods for speech recognition</i>	61
2.2 Systems of automatic speech recognition	63
2.2.1 <i>Systems developed worldwide</i>	64
2.2.2 <i>Systems developed in Romania</i>	72
2.2.3 <i>Evaluation criteria for speech recognition systems</i>	75
2.2.4 <i>Performance analysis of automatic speech recognition systems</i>	76
2.3 Chapter conclusions	80
Chapter 3. Contributions to the development of some speech recognition systems for Romanian language	81
3.1 Deterministic recognition of sequences in case of isolated spoken words	81
3.1.1 <i>Temporal alignment used in speech recognition – case study</i>	82
3.2 Using statistical methods in speech recognition.....	83
3.2.1 <i>Extraction of vocal parameters using the mel-cepstral method</i>	83
3.2.2 <i>ASR system for recognition of isolated spoken words using hidden Markov models and mel-cepstral analysis</i>	84
3.2.3 <i>ASR system for recognition of conectate spoken words using hidden Markov models and mel-cepstral analysis</i>	91
3.3 Using artificial neural networks in speech recognition	93
3.3.1 <i>ASR system for recognition of isolated spoken words using artificial neural networks and LPC coefficients</i>	94
3.3.2 <i>Discriminant system for recognition of isolated spoken words using artificial neural networks and MFCC coefficients</i>	98
3.4 Comparative analysis of the proposed systems.....	102
3.5 Development of a web application for the construction of the databases used in voice commands recognition.....	103
3.5.1 <i>Characteristics of the databases used in speech recognition</i>	103
3.5.2 <i>Development of the web application for construction of the BDCVR database</i>	104
3.6 Chapter conclusions	106

Chapter 4. Hybrid system for Romanian speech recognition	108
4.1 SRV_BC – Knowledge-based ASR system for speech recognition.....	108
4.1.1 SRV_BC Architecture.....	109
4.1.2 SRV_BC Knowledge Base.....	110
4.2 Metrics for performance evaluation of the SRV_BC system.....	115
4.2.1 Word recognition rate	115
4.2.2 Adapted Good-Turing recognition rate	115
4.2.3 Orthoepic mistakes' rate	117
4.3 Chapter conclusions	117
Chapter 5. Experimental research for development of the SARCV automatic system for speech recognition.....	118
5.1 Presentation of the SARCV system.....	118
5.1.1 System theory based approach of the process of recognition of voice commands	119
5.1.2 The hierarchic structure of the automatic system of recognition of voice commands	121
5.2 Tools used for the development of the SARCV system.....	122
5.2.1 Presentation of Robo_Arm, a LEGO Mindstorms NXT Robot	122
5.2.2 Software tools	124
5.3 Data acquisition	124
5.4 Acoustic modeling	125
5.5 Linguistic modeling	125
5.6 The recognition modules of SARCV	126
5.6.1 The speech recognition module based on temporal alignment	128
5.6.2 The speech recognition module based on neural techniques	130
5.7 Experimental results.....	133
5.7.1 Experiment 1.....	134
5.7.2 Experiment 2.....	136
5.8 Chapter conclusions	138
Chapter 6. Conclusions, contributions, and future work	140
6.1 Conclusions.....	140
6.2 Summary of the contributions	141
6.3 Future work	145
Publication list	146
Bibliography	148

Webography	157
APPENDIXES.....	159
APPENDIX 1. Term glossary	159
APPENDIX 2. Speech production and perception.....	163
APPENDIX 3. Phonemes of Romanian language.....	165
APPENDIX 4. The SARCV module based on DTW (Matlab code).....	169
APPENDIX 5. The SARCV module on neural networks (Matlab code).....	175

Keywords: speech recognition, automatic, robot, voice commands, speech signal, speech parameterization

ABSTRACT OF THE THESIS

Speech recognition (or recognition of the spoken language) by a computerized system is still one of the open research problem that researchers try to solve nowadays. The still unsolved issues are due to a variety of factors that influence the recognition performance, and further the possibility to build successfully intelligent systems (for speech understanding, for dialog systems etc.). Speech recognition will most probably be used in not so far future both in dialog systems human-machine and embedded systems, or in industrial automatic systems.

In this alive context, this doctoral thesis is particularly concerned with the research of speech recognition in all its aspects and with the development of automatic speech recognition systems. The main contributions of the thesis are both of *theoretical research* (such as enhancing the capabilities of an automatic recognition system by using an expert system and data mining techniques, proposing metrics for performance evaluation for such systems, and also introducing the structure and the architecture of an automatic system for speech recognition; moreover, an algorithm for guiding the system in choosing the most appropriate recognition module has been proposed), and of *experimental research* (the development of the automatic system for speech recognition, testing the system's functionality with various data sets, the construction of databases with Romanian words, the validation of the system). What it is worth mentioning is that all the proposed elements are general, they have been tested and validated for Romanian, but they can be easily adapted to any other language.

The main objective of *the applied research* in this thesis has been the development of an automatic speech recognition system that provides for the execution of voice commands on a robot.

The main specific objectives are:

- Overviewing the state of the art in speech recognition;

- Researching the system theory-based of the process of speech recognition and designing the structure and the architecture of an automatic speech recognition system;
- Identifying the most appropriate methods, i.e. statistical, neural, and for pattern recognition to recognize isolated or connected spoken words (voice commands);
- Enhancing the capabilities of an automatic recognition system by including an expert system (for improving the performance of the system);
- The development of an automatic system for speech recognition from analysis, design, implementation and testing to its validation by executing voice commands on a robot.

This doctoral thesis is structured in six chapters as follows:

Chapter 1, Theoretical fundamentals of speech recognition, highlights theoretical aspects of spoken language recognition. In this chapter is presented also the architecture of a speech recognition system, along with the main methods for processing and analyzing the voice signal. The last part of this chapter is dedicated to the presentation of both the structure and the architecture of an automatic system for speech recognition proposed by the thesis's author.

Chapter 2, State of the art in speech recognition, overviews the current state in research of this field worldwide and for Romanian language as well. The most utilized methods and the main developed systems for speech recognition are also overviewed here. The chapter ends with a comparison of the recognition methods, followed by an analysis of the main characteristics and performances of these systems.

Chapter 3, Contributions to the development of some speech recognition systems for Romanian language, contains several experiments performed by applying various techniques, namely neural, statistic, and of temporal alignment, for building systems for recognition of isolated or connected spoken words. The web application that has been developed by the author for the semi-automatic acoustic data acquisition is also presented here. It enables the storage of the data in a speech database.

In **Chapter 4, Hybrid system for Romanian speech recognition**, the author proposes a hybrid system SRV_BC, an approach rooted in artificial intelligence, which uses phonetic, acoustic, and lexical knowledge to develop a knowledge base that enhance the system for speech recognition with expert system capabilities. Two metrics that evaluate the performance of such systems are also proposed here.

In **Chapter 5, Experimental research for development of the SARCV automatic system for speech recognition**, are included the author's contributions with regard to the development of an automatic system for recognition of voice commands by a LEGO Mindstorms NXT robot. Two approaches have been taken: a connectional one (with artificial neural networks) and one based on pattern recognition (temporal alignment). To automatically guide the system to select one of the two methods, the author has devised an algorithm that uses both cross correlation between signals and heuristic rules.

Chapter 6, Conclusions, contributions, and future work, highlights the main conclusions and contributions of this doctoral thesis, and some future work ideas that could be followed to further this contributions.