

**MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA**  
**Universitatea Tehnică a Moldovei**  
**Facultatea Calculatoare, Informatică și Microelectronică**  
**Departamentul Ingineria Software și Automatică**

**Admis la susținere**  
**Șef departament:**  
**Fiodorov Ion, conf. univ., dr.**

---

„\_” \_\_\_\_\_ 2022

**CERCETAREA SISTEMELOR DE RECUNOAȘTERE**  
**VOCALĂ**  
**Teză de master**  
**în Tehnologia informației**

**Masterand:** **Țurcanu Andrei, gr. TI-201M**

**Coordonator:** **Prof. univ. Bolun Ion**

**Consultant:** **Cojocarul Svetlana, lect. univ.**

**Chișinău 2022**

## Rezumat

Teza de master prezentată în acest raport de către studentul Țurcanu Andrei are ca scop familiarizarea cu tehnologiile de recunoaștere vocală, cercetarea metodologiilor și a algoritmilor implementați în acest domeniu. Acest scop a reușit din analiza domeniului recunoașterii vocale și observărilor precedente. În urma cercetărilor a fost stabilit faptul că aceste tehnologii iau o amploare mare atât în Republica Moldova cât și în întreaga lume. Modurile de utilizare a acestor tehnologii sunt diverse și pot aduce un beneficiu major atât pentru sfera economică a țării cât și pentru întreaga societate umană.

Unede din domeniile în care se poate aplica tehnologia respectivă sunt:

- Asistent personal digital (siri, alexa, cortana, etc);
- În domeniul sănătății publice;
- Învățarea limbilor noi;
- Online banking;
- Sporirea securității cu ajutorul biometriei vocale;
- Platforme de E-Comerce și validarea plăților cu ajutorul biometriei vocale;

Structura acestei lucrări este organizată pe capitole. În lista de mai jos, acestea sunt enumerate alături de o descriere succintă a fiecăruia.

**Capitolul 1** prezintă analiza situației curente în sfera recunoașterii vocale, cercetarea lucrărilor din domeniu respectiv, analiza subiectului și determinarea unei baze informaționale care urmează a fi desfășurată ulterior.

**Capitolul 2** prezintă proiectarea de ansamblu a algoritmilor utilizați în recunoașterea vocii, acest capitol determină o structură comună caracteristică prelucrării semnalelor, analizelor de date și structurarea pașilor comuni necesari pentru un algoritm antrenat în recunoașterea vocii.

**Capitolul 3** prezintă cercetarea în detaliu a algoritmilor destinați recunoașterii vocale, compararea și determinarea neajunsurilor și a părților forte din fiecare algoritm.

**Capitolul 4** oferă o vizualizare ale divizărilor fonetice, specifice unui limbaj anumit, tehnicile necesare pentru a separa un sunet, prelucrarea sunetelor de impurități, sinteza și descompunerea limbajului uman într-un semnal definit de spectograma sunetului.

**Capitolul 5** descrie problema divergenței sunetelor la nivel rasial și cultural, și anume cum se deosebesc sunetele pronunțate de diferite categorii sociale, aici vedem o problemă clară în recunoașterea vocii, după colectarea datelor de către mai multe instituții din SUA, putem observa o divergență a datelor.

## Abstract

The master's thesis presented in this report by the student Țurcanu Andrei aims at familiarization with voice recognition technologies, research of methodologies and algorithms implemented in this field. This goal emerged from the analysis of the field of speech recognition and previous observations. Research has established that these technologies are gaining momentum both in the Republic of Moldova and in the world. The ways in which these technologies are used are diverse and can bring major benefits to the economic sphere of countries as well as to human society as a whole.

Some of the areas in which this technology can be applied are:

- Personal digital assistant (siri, alexa, cortana, etc);
- In the field of public health;
- Learning new languages;
- Online banking;
- Increasing security with voice biometrics;
- E-Commerce platforms and validation of payments using voice biometrics;

The structure of this paper is organized into chapters. In the list below, they are listed along with a brief description of each.

**Chapter 1** analysis of the current situation in the field of voice recognition, research of works in the field, analysis of the subject and determination of an informational action will be carried out later.

**Chapter 2** presents the overall design of the algorithms used in voice recognition, this chapter determines a common structure characteristic of signal processing, data analysis and structuring the common steps required for an algorithm involved in voice recognition.

**Chapter 3** detailed research of algorithms for speech recognition, comparison and determination of shortcomings and strengths of each algorithm.

**Chapter 4** provides an overview of the phonetic divisions specific to a particular language, the techniques needed to separate a sound, the processing of sound from impurities, the synthesis and decomposition of human language into a signal defined by the sound spectrum.

**Chapter 5** describes the problem of racial and cultural sound divergences, namely how the sounds pronounced by different social categories differ, here we see a clear problem in voice recognition, after collecting data to several US institutions, we can see a divergence of data. .

# CUPRINS

INTRODUCERE .....	1
1 ANALIZA SITUAȚIEI ÎN SFERA RECUNOAȘTERII VOCALE .....	2
1.1 Analiza domeniului.....	2
1.2 Tehnologiile similare din deceniile trecute .....	4
1.3 Spre o mașină care comunică .....	10
1.4 Cercetarea lucrărilor din domeniu .....	12
2 PROIECTAREA ALGORITMILOR DE RECUNOAȘTERE VOCALĂ .....	14
2.1 Fazele de antrenare .....	15
2.2 Faza de testare .....	15
2.3 Faza de procesare avansată.....	16
3 CERCETAREA ALGORITMILOR DE RECUNOAȘTERE A VOCII .....	20
3.1 Convertor analog-digital .....	21
3.2 Lingvistica si foneme.....	22
3.3 Modelul Hidden Markov în recunoașterea vorbirii.....	22
3.4 Rețele neuronale artificiale .....	23
3.5 Algoritmul KNN .....	24
3.6 Algoritmul Gaussian NB.....	25
4 IDENTIFICAREA ZONELOR VOCALICE ÎN SECVENȚE ROSTITE ÎN ROMÂNĂ .....	27
4.1 Etichetarea semnalelor vocale.....	27
4.1 Tehnici de identificare a zonelor vocalice în secvențe rostite in limba română. ....	28
5 DIFERENȚELE RASIALE IN RECUNOAȘTEREA AUTOMATA A VORBIRII.....	30
5.1 Rezultate.....	31
5.2 Materiale și metode.....	41
5.3 Potrivirea.....	43
5.4 Măsurarea densității dialectelor .....	44
Concluzii.....	45
Bibliografie.....	47

## INTRODUCERE

În era informatizării, multe produse high-tech au intrat treptat în viața noastră de zi cu zi și ne-au schimbat semnificativ obiceiurile. Pe de altă parte, tehnologiile informaționale continuă să evolueze către o abordare mai centrată pe om. Tehnologia de identificare biometrică, care ne oferă metode mai simple și mai convenabile de identificare a anumitor persoane, a înlocuit treptat unele dintre metodele de autentificare existente care ar trebui cercetate înainte ca oamenii să le poată gestiona corespunzător, sisteme de recunoaștere a feței utilizate în locuri publice, de către forțele de ordine, organizații [1] și asistentul mobil vocal Siri pe iPhone, Bixby Voice pe Galaxy [2], sunt exemple de rezultate de identificare biometrică.

Recunoașterea unei persoane după voce este una dintre formele de autentificare biometrică, care face posibilă identificarea printr-o combinație de caracteristici unice ale vocii și se referă la metode dinamice de biometrie. Recunoașterea vorbitorului este o tehnologie care poate identifica automat vorbitorul pe baza formei de undă a vorbirii, care reflectă caracteristicile fiziologice și comportamentale, parametrilor vorbirii a vorbitorului. La fel ca sistemele tradiționale de recunoaștere a vocii, există două etape principale, și anume, antrenament și testare.

Învățarea este procesul de extragere a caracteristicilor fonetice dintr-un vorbitor care a fost deja înregistrat sau salvat ca probă, stocarea lor într-o bază de date și familiarizarea sistemului cu caracteristicile vocii vorbitorului.

Testarea este procesul de comparare a sunetului și a caracteristicilor fonetice îndoielnice dintr-o bază de date de recunoaștere a vorbitorului. Două seturi populare de caracteristici, adesea utilizate în analiza semnalului de vorbire sunt coeficienții cepstrali de frecvență Mel (MFCC) și coeficienții cepstrali de predicție liniară (LPCC). Cele mai populare modele de recunoaștere sunt cuantizarea vectorială (VQ), deformarea dinamică a timpului (DTW) și rețeaua neuronală artificială (ANN) [3].

În această lucrare se vor cerceta algoritmi de clasificare pentru problema identificării persoanei prin voce folosind metode de învățare automată. O analiză comparativă a mai multor algoritmi de clasificare a fost executată. În primul experiment, a fost determinată metoda vectorului suport—0,90 și perceptronul multistrat—0,83, care a arătat cele mai bune rezultate. În cel de-al doilea experiment, a fost propus un perceptron multistrat cu o precizie de 0,93 utilizând metoda Scaler Robust pentru identificarea vocii persoanei. Prin urmare, pentru a rezolva această problemă, este posibil să se utilizeze un perceptron cu mai multe straturi, ținând cont de specificul semnalului de vorbire.

## Bibliografie

1. B.H. Juang# & Lawrence R. Rabiner\* # Georgia Institute of Technology, Atlanta \* Rutgers University and the University of California, Santa Barbara , Automatic Speech Recognition – A Brief History of the Technology.[citat07.09.2021]  
Disponibil:[https://folk.idi.ntnu.no/gamback/teaching/TDT4275/literature/juang\\_rabiner04.pdf](https://folk.idi.ntnu.no/gamback/teaching/TDT4275/literature/juang_rabiner04.pdf)
2. Abdelmajid Hassan Mansour. Voice recognition system using back propagation algorithm. [citat:09.10.2021]Disponibil:  
[https://www.researchgate.net/publication/282062349\\_Voice\\_recognition\\_Using\\_back\\_propagation\\_algorithm\\_in\\_neural\\_networks](https://www.researchgate.net/publication/282062349_Voice_recognition_Using_back_propagation_algorithm_in_neural_networks)
3. Orken Mammyrbayev. Voice identification using classification Algorithms [citat: 10.10.2021]  
Disponibil: <https://www.intechopen.com/chapters/68705>
4. Introduction to Various Algorithms of Speech,Recognition: Hidden Markov Model, Dynamic Time,Warping and Artificial Neural Networks,Pahini A. Trivedi ,V.V.P. Engineering College [citat:10.15.2021] Disponibil: <https://www.ijedr.org/papers/IJEDR1404035.pdf>
5. M. ZBANCIOC, H.N. TEODORESCU, M. FERARU, TEHNICI DE IDENTIFICARE A ZONELOR VOCALICE ÎN SECVENȚE ROSTITE ÎN LIMBA ROMÂNĂ M. ZBANCIOC, H.N.[citat:11.01.2021]Disponibil:  
[http://www.etc.tuiasi.ro/sibm/romanian\\_spoken\\_language/pdf/consilr2010\\_mz.pdf](http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/pdf/consilr2010_mz.pdf)
6. Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey , Racial disparities in automated speech recognition [citat11.01.2021] Disponibil: <https://www.pnas.org/content/117/14/7684>
7. Debnath Bhattacharyya, Rahul Ranjan, Farkhod Alisherov A. and Minkyu Choi, —Biometric Authentication: A Review , International Journal of u- and e- Service, Science and Technology Vol. 2, No. 3, September, 2009. [citat11.07.2021]
8. D. Harwell, B. Mayes, M. Walls, S. Hashemi, The accent gap. *The Washington Post*, 19 July 2018. Disponibil:<https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/>. [citat11.08.2021]
9. Judith A. Markowitz, —Voice Biometrics|, September 2000. Communications of the ACM Vol. 43, No.9, [citat11.12.2021]
10. M. Adda-Decker, L. Lamel, “Do speech recognizers prefer female speakers?” in *Proceedings of INTERSPEECH-2005* (International Speech Communication Association, 2005), pp .2205–2208 [citat11.01.2021]

11. S. Goldwater, D. Jurafsky, C. D. Manning, Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Commun* [citat11.12.2021]
12. [WinPitch by Instruments](https://www.winpitch.com) [citat11.12.2021], Disponibil: <https://www.winpitch.com>
13. GoldWave [citat11.12.2021], Disponibil GoldWave Inc. <https://www.goldwave.com>