МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра математической кибернетики

Волощик Василий Васильевич

Система распознавания речи на базе искусственной нейронной сети

Дипломная работа

Научный руководитель:
кандидат физико-математических наук, доцент кафедры математической кибернетики С.Е. Бухтояров

Допущен к защите

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2018 г.

заведующий кафедрой

математической кибернетики, профессор Гладков А. Л.

Минск, 2018

## РЕФЕРАТ

В дипломной работе страниц 64, рисунков 23, источников 35.

***Ключевые слова*:** нейронные сети; feature extractor; recognizer; шаблоны; траектории; обучение; признаки.

Хотя системы распознавания речи уже доступны на рынке в настоящее время, они либо основаны на статистических методах, либо являются платными и закрытыми продуктами крупных корпораций. В этой работе исследуется возможность альтернативных подходов решения проблемы. Представлена ​​система распознавания речи, состоящая из двух отдельных блоков, Feature Extractor и Recognizer. Блок Feature Extractor использует стандартный кодер LPC Кепстр, который преобразует входящую речь в траекторию в пространстве признаков LPC Кепстр, а затем самоорганизующуюся карту, которая адаптирует исход кодера для получения оптимальных траекторий представления слов в уменьшенные пространственные области измерений. Сравниваются конструкции блоков Recognizer на основе трех разных подходов. Производительность распознавателей типа «Шаблоны», «Многослойные перцептроны» и «Рекуррентные нейронные сети» тестируется на небольшой изолированной проблеме распознавания речи, зависящей от оратора. Экспериментальные результаты показывают, что траектории на таких пространствах с уменьшенной размерностью могут обеспечить надежные представления произносимых слов, одновременно уменьшая сложность обучения и работу Recognizer-а. Сравнение между различными подходами к дизайну Recognizer-ов, проводимых здесь, дает лучшее понимание проблемы и ее возможных решений. Также представлена процедура обучения, которая оптимизирует использование набора учебных материалов.

## РЭФЕРАТ

У дыпломнай працы старонак 64, малюнкаў 23, крыніц 35.

***Ключавыя словы***: нейронныя сеткі; Feature Extractor; Recognizer; шаблоны; траекторыі; навучанне; прыкметы.

Хаця сістэмы распазнаванне маўлення ўжо даступныя на рынку ў цяперашні час, яны ці заснаваныя на статыстычных падыходах, ці з’яўляюцца платнымі і закрытымі прадуктамі буйных карпарацый. У гэтай рабоце даследваецца магчымасць альтэрнатыўных падыходаў рашэння праблемы. Прадстаўлена сістэма распазнавання маўлення, якая складаецца з двух асобных блокаў, Feature Extractor і Recognizer. Блок Feature Extractor выкарыстоўвае стандартны кодар LPC Кепстр, які пераўтварае уваходнае маўленне ў траекторыю ў прасторы прыкметаў LPC Кепстр, а затым у самаарганізавальную мапу, якая адаптуе зыход кодара для атрымання аптымальных траекторый уяўляення слоў у паменшаныя прасторавыя вобласці вымярэнняў. Параўноўваюцца канструкцыі блокаў Recognizer на заснаванні трох розных падыходаў. Вытворчасць распазнавальнікаў тыпу “Шаблоны”, “Шматслойныя перцэптроны” і “Рэкурэнтныя нейронныя сеткі” тэстуецца на невялікай ізаляванай праблеме распазнавання маўлення, якая залежыць ад аратара. Эксперыментальныя вынікі паказваюць, што траекторыі на такіх прасторах з паменшанай памернасцю могуць забяспечыць надзейныя ўяўленні вымаўленых слоў, адначасова памяншаючы складанасць навучання і працу Recognizer-а. Параўнанне паміж рознымі падыходамі да дызайну Recognizer-аў, якія праводзяцца тут, дае найлепшае разуменне праблемы і яе магчымых рашэнняў. Таксама прадстаўлена працэдура навучання, якая аптымізуе выкарыстанне набора вучэбных матэрыялаў.

## ABSTRACT

In graduate work the number of pages 64, pictures 23, sources 35.

***Key words*:** neural links; feature extractor; recognizer; templates; trajectories; learning; features.

Although speech recognition products are already available in the market at present, their development is mainly based on statistical techniques. The work presented in this thesis investigates the feasibility of alternative approaches for solving the problem more efficiently. A speech recognizer system comprised of two distinct blocks, a Feature Extractor and a Recognizer, is presented. The Feature Extractor block uses a standard LPC Cepstrum coder, which translates the incoming speech into a trajectory in the LPC Cepstrum feature space, followed by a Self Organizing Map, which tailors the outcome of the coder in order to produce optimal trajectory representations of words in reduced dimension feature spaces. Designs of the Recognizer blocks based on three different approaches are compared. The performance of Templates, Multi- Layer Perceptrons, and Recurrent Neural Networks based recognizers is tested on a small isolated speaker dependent word recognition problem. Experimental results indicate that trajectories on such reduced dimension spaces can provide reliable representations of spoken words, while reducing the training complexity and the operation of the Recognizer. The comparison between the different approaches to the design of the Recognizers conducted here gives a better understanding of the problem and its possible solutions. A new learning procedure that optimizes the usage of the training set is also presented. Optimal tailoring of trajectories, new insights into the use of neural networks in this class of problems, and a new training paradigm for designing learning systems are the main contributions of this work.