

Обзорная статья

УДК 004.7

## СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ АРАБСКОЙ РЕЧИ И ЙЕМЕНСКОГО ДИАЛЕКТА

✉ **Радан Наим Хусейн.**

**Тверской государственный технический университет, г. Тверь, Россия**

✉ [naeem.radan@gmail.com](mailto:naeem.radan@gmail.com)

*Аннотация.* Статья посвящена анализу научно-исследовательских работ по системам автоматического распознавания речи арабского языка и системам автоматического распознавания речи основных диалектов арабского мира. Существует несколько методов реализации систем распознавания речи. Одними из новых методов являются использование нейронных сетей и скрытых Марковских моделей, применяемых для распознавания речи. Арабский язык является одним из самых распространенных языков и наименее исследуемым с точки зрения распознавания речи. Данная статья представляет собой краткий обзор по имеющимся исследованиям в области распознавания арабской речи и арабского йеменского диалекта. В работе проанализированы наборы инструментов, доступные для разработки систем распознавания арабской речи. Приведены методики и алгоритмы, использованные для классификации и идентификации арабских диалектов. На текущий момент систем автоматического распознавания речи йеменского диалекта разработано относительно мало по сравнению с системами автоматического распознавания речи для современного стандартного арабского языка и системами автоматического распознавания речи других арабских диалектов.

*Ключевые слова:* распознавание речи, распознавание арабской речи, арабский йеменский диалект, нейронные сети, скрытые Марковские модели, идентификация диалектов, классификация диалектов

**Для цитирования:** Радан Наим Хусейн. Системы автоматического распознавания арабской речи и йеменского диалекта // Науч.-аналит. журн. «Вестник С.-Петерб. ун-та ГПС МЧС России». 2023. № 2. С. 194–212.

Review article

## AUTOMATIC SPEECH RECOGNITION SYSTEMS FOR ARABIC SPEECH AND YEMENI DIALECT

✉ **Radan Naim Hussein.**

**Tver state technical university, Tver, Russia**

✉ [naeem.radan@gmail.com](mailto:naeem.radan@gmail.com)

*Abstract.* The article is devoted to the analysis of research works on automatic speech recognition systems of the Arabic language and automatic speech recognition systems of the main dialects of the Arab world. There are several methods for implementing speech recognition systems. One of the new methods is the use of neural networks and hidden Markov models used for speech recognition. Arabic is one of the most widely spoken languages and one of the least researched in terms of speech recognition. This article is a brief overview of the available research in the field of Arabic speech recognition and the arabic yemeni dialect. The paper analyzes the sets of tools available for the development of Arabic speech recognition systems. The methods and algorithms used for the classification and identification of Arabic dialects are given. To date, there has been relatively little development of automatic speech recognition systems for Yemeni dialect compared to automatic speech recognition systems for modern standard arabic and automatic speech recognition systems for other arabic dialects.

© Санкт-Петербургский университет ГПС МЧС России, 2023

**Keywords:** speech recognition, arabic speech recognition, arabic yemeni dialect, neural networks, hidden Markov models, dialect identification, dialect classification

**For citation:** Radan Naim Hussein. Automatic speech recognition systems for arabic speech and yemeni dialect // Scientific and analytical journal «Vestnik Saint-Petersburg university of State fire service of EMERCOM of Russia». 2023. № 2. P. 194–212.

### Введение

В настоящее время арабский язык является одним из самых распространенных языков во всем мире, он занимает четвертое место после китайского, испанского и английского языков. Число говорящих на арабском языке составляет около 310 млн (родной язык) и еще около 270 млн человек использует арабский в качестве второго языка. На текущий момент было проведено относительно мало исследований по распознаванию речи на арабском йеменском диалекте по сравнению с другими диалектами арабского языка. Существует много опубликованных исследований по моделям, методам и приложениям для распознавания речи на стандартном арабском языке, а также на других арабских диалектах. Проблеме изучения арабского йеменского диалекта уделено весьма мало исследований, отсутствуют системы распознавания данного диалекта. Литература по йеменскому арабскому диалекту ограничена. В данном обзоре проанализированы доступные научные источники по распознаванию йеменской арабской речи и арабского языка [1].

### Методология обзора

Основное внимание обзора уделяется следующим исследовательским вопросам: 1) методы разработки систем автоматического распознавания речи (САРР); 2) исследования по разработке систем автоматического распознавания арабской речи; 3) исследования по разработке систем автоматического распознавания арабского йеменского диалекта.

Анализируемые базы данных, в которых проводился поиск, следующие: журналы ACM, IEEE, Springer, Sage, Science Direct, King Saud University (KSU). Временный интервал обзора сосредоточен на опубликованных статьях с 2000 г. по настоящее время. На рис. 1 представлена количественная оценка использованных источников списка литературы.

### Количественная оценка использованных источников списка литературы

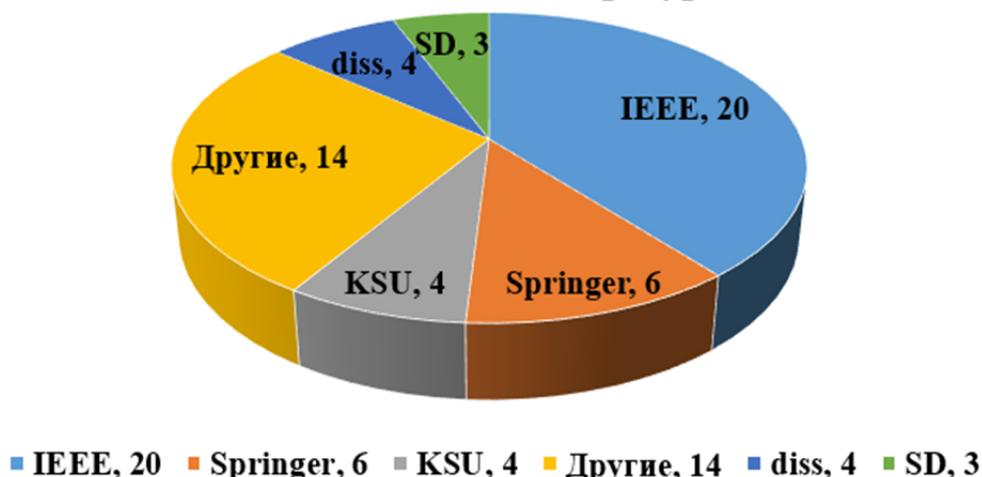


Рис. 1

## Системы автоматического распознавания арабской речи, системы классификации и идентификации арабских диалектов

В данном подразделе представлен краткий обзор систем распознавания арабской речи. В нем представлен обзор литературы по системам распознавания арабской речи, и обсуждаются некоторые проблемы арабского языка с точки зрения распознавания речи. Рассматривая работы, обращается внимание на три фактора: 1) размер исследовательской базы данных; 2) тип используемой системы распознавания; 3) полученные результаты. Ряд исследователей в последнее время занимались разработкой систем распознавания арабской речи. На рис. 2 показаны основные арабские диалекты, в табл. 1 продемонстрировано краткое описание перечисленных диалектов.



Рис. 2. Арабские диалекты

Таблица 1

### Основные арабские диалекты и их краткое описание

Диалект	Описание
Египетский арабский диалект (EGY)	Охватывает диалекты долины Нила: Египет и Судан
Левантйский арабский диалект (LEV)	Включает диалекты Ливана, Сирии, Иордании, Палестины
Арабский диалект стран Арабского залива (Gulf)	Включает диалекты Кувейта, Объединенных Арабских Эмиратов, Бахрейна и Катара. Саудовской Аравии и Омана
Североафриканский арабский диалект (NOR)	Охватывает диалекты Марокко, Алжира, Туниса и Мавритании. Иногда включается ливийский арабский диалект
Иракский арабский диалект (IRQ)	Содержит элементы как левантйского диалекта, так и арабский диалект стран Арабского залива
Йеменский арабский диалект (Yem)	Часто считается отдельным классом

## Системы распознавания арабской речи с использованием скрытых Марковских моделей. Системы распознавания непрерывной речи

В работе [2] реализована дикторонезависимая САРР непрерывной арабской речи с использованием Sphinx и скрытых Марковских моделей (hidden Markov model (HMM)). Извлечение признаков речи выполнено с использованием мел-частотных кепстральных коэффициентов (MFCC). Входные звукозаписи (415 от 20 мужчин и 20 женщин). Система обучена с использованием семи часов звукозаписи и протестирована с использованием одного часа звукозаписи. В результате получена точность системы распознавания слов, равная 92,67 % и 93,88 %. Частота ошибок в словах (word error rate (WER) равна 11,27 % и 10,07 % с диакритическими знаками и без них соответственно.

В работе [3] рассмотрена адаптация САРР арабского языка к носителям с иностранным акцентом. Для обучения и тестирования системы автором разработан список высказываний из корпуса LDC-WestPoint, случайным образом выбраны четыре различных и несвязных списка. Таким образом, объём базы данных составил 600 высказываний. Адаптация достигается с помощью методов линейной регрессии максимального правдоподобия (maximum likelihood linear regression (MLLR) и апостериорного максимума (Maximum a posteriori estimation (MAP) и их комбинация MLLR и MAP. При реализации системы использованы корпус современного стандартного арабского языка (Modern Standard Arabic (MSA)). Для проектирования и тестирования систем распознавания речи во всех экспериментах использован набор инструментов скрытой Марковской модели (The Hidden Markov Model Toolkit (HTK) и MFCC. Точность неадаптированной системы составила 89,02 % при распознавании слов и 93,19 % при распознавании фонем. Наилучшее повышение точности системы составило 7,37 %.

Работа [4] посвящена проблеме автоматической идентификации диалектов арабского языка. Автором создана база данных звукозаписей для двух диалектов арабского языка EGY и Gulf. Система разработана с использованием HMM и признаков MFCC для построения моделей диалектов и их идентификации. Общая точность системы идентификации диалектов равна 96,7 %. На рис. 3 приведена зависимость точности САРР числа гауссианов при использовании MFCC и смещённый дельта кепстр (Shifted Delta Cepstral (SDC) и видно, что точность системы увеличивается при применении MFCC и MFCC+ SDC, также видно, что с ростом числа гауссианов улучшается точность системы.

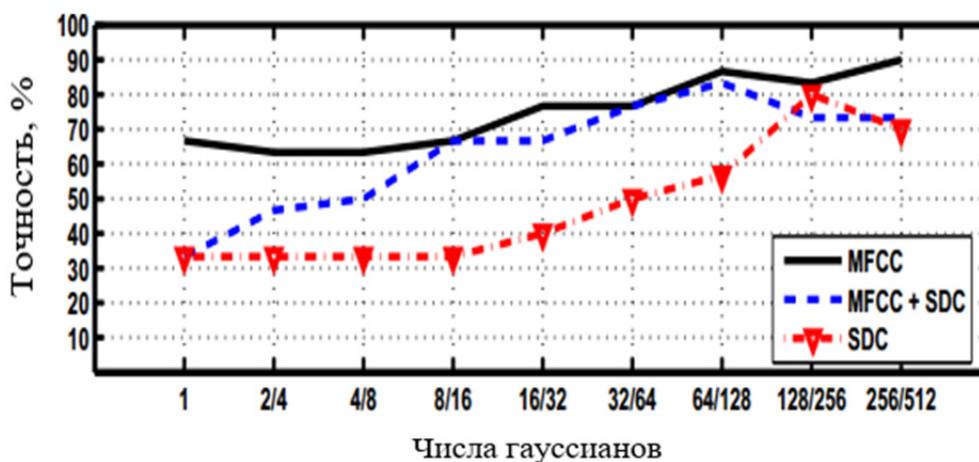


Рис. 3. Зависимость точности САРР при разновидности числа гауссианов и разных информативных признаков

В работе [5] разработана система распознавания непрерывной речи, независимая от говорящего со средним словарным запасом. Для проектирования системы использованы две языковых модели: биграмма и контекстно-свободная грамматика. Результаты системы получены с WER, равной 5,26 % и 2,72 % для языковых моделей, на основе биграммы и WER, равной 0,19 % и 0,99 %, на основе грамматики. Модели созданы с использованием НММ/MFCC.

В работе [6] создана система непрерывного распознавания речи арабского языка с использованием НММ/MFCC с моделью гауссовой смеси (Gaussian Mixture Model (GMM)). Точность системы распознавания увеличена до 94 %.

### **Системы распознавания изолированной речи**

В работе [7] разработано САРР арабских произнесенных цифр. Система предназначена для распознавания изолированной речи. При построении системы распознавания использованы НММ/НТК. Данные цифр собраны из телефонного арабского речевого корпуса (Saudi Accented Arabic Voice Bank (SAAVB)) для обучения и тестирования системы. В результате получена общая точность системы распознавания, равная 93,72 %.

В работе [8] разработана система для распознавания изолированных арабских цифр. В работе исследованы классификаторы-машины опорных векторов (Support Vector Machine (SVM)) и метод k-ближайших соседей (k-nearest neighbour (KNN)). Система основана на сочетании классификаторов GMM/НММ/SVM/MFCC, GMM/НММ/KNN/MFCC и GMM/НММ/MFCC. Входные звукозаписи получены от 92 дикторов (46 мужчин и 46 женщин), где 10/46 использованы для обучения и 36/46 для тестирования. Общее количество базы данных составлено 4 600 звукозаписей. Точность системы распознавания достигла 90,62 % и 88,26 % для гибридной системы GMM/НММ/KNN и базовой системы GMM/НММ/MFCC соответственно.

В исследовании [9] реализована система распознавания изолированных арабских цифр. Обучающая выборка состоит из звукозаписей от 50 дикторов, каждый из которых трижды произнес десять цифр. Система основана на сочетании методов векторного квантования (Vector Quantization (VQ)), НММ и кодирование с линейным предсказанием (The Linear Predictive Code (LPC)). Точность системы достигла 95 %.

В работе [10] исследованы методы идентификации арабских диалектов. В ходе проведения исследования, автором получена точность системы идентификации, равная 88 %.

В работе [11] исследована дикторозависимая САРР арабских слогов для разговорных цифр. Система разработана на основе НММ/MFCC в чистой и шумной среде. База данных, используемая для обучения и для тестирования, состоит из 44 носителей египетского диалекта. В чистой среде эксперименты показывают, что точность распознавания слогов превосходит точность распознавания монофонов, трифонов и слов на 2,68 %, 1,19 % и 1,79 % соответственно. В шумной среде – 2,09 %, 1,5 % и 0,9 % соответственно.

В работе [12] разработана система обучения арабскому произношению с использованием моделей НММ. Для обнаружения ошибок был разработан алгоритм идентификации, основанный на длительности фонемы. Оценка точности системы составила около 1,78 % неправильных речевых сегментов.

В работе [13] создана система автоматического распознавания арабской речи, записаны 24 арабских слова от 19 дикторов, с тремя повторениями (всего 1 368 слов). Автором разработана система с использованием НММ и 39 MFCC. Количество состояний MFCC от 4 до 10, каждое состояние имеет 8-гауссовое распределение. Система обучена с использованием 720 слов (10 дикторов) и протестирована с использованием 648 слов (9 дикторов). Результаты точности распознавания слов составляют 92,92 % и WER равна 7,08 %.

## Системы распознавания арабской речи с использованием Sphinx4

Одной из интересных систем распознавания арабской речи является Sphinx (CMU, университет Карнеги-Меллона), она находится в свободном доступе. Система позволяет исследовательским группам со скромным бюджетом быстро приступить к проведению исследований и разработке приложений.

В работе [14] предложена система распознавания разговорной арабской речи. Исследование демонстрирует возможную адаптивность CMU Sphinx4 к арабскому языку с большим словарным запасом. Система использует 367 предложений, всего 14 232 слов. Разработанная система независима от диктора, обеспечивает точность, равную 91,6 % при использовании записанных файлов и 90,2 % при использовании микрофона.

В работе [15] рассмотрена система распознавания арабской речи на основе Sphinx4. Автором исследован автоматизированный подход к созданию фонетических словарей для Корана и стандартного арабского языка. В работе разработаны три корпуса: 1) корпус Корана, продолжительностью около 18,5 часов; 2) корпус управления и контроля, продолжительностью около 1,5 часов; 3) корпус арабских цифр, продолжительностью менее одного часа речи. Автором получена точность системы, равная 98,182 %, и WER, равная 1,818 %.

В работе [16] использован инструмент Sphinx для распознавания изолированных арабских цифр. Данные были записаны от шести дикторов. Точность системы распознавания равна 86,66 %.

В работе [17] исследована система распознавания арабской речи с использованием диакритического и недиакритического текста. Для выполнения работы использовалась CMU Sphinx. Входные данные представляют собой поток непрерывной речи MSA, содержащий 13,5 часов звукозаписей для обучения и 4,1 часа звукозаписей для тестирования. Экспериментальные результаты показывают, что текстовая система без диакритических знаков имеет WER 76,4 %, в то время как система на основе диакритического текста имеет WER 63,8 %. В табл. 2 приведен сравнительный анализ SAPP арабской речи с использованием скрытых Марковских моделей.

Таблица 2

### Сравнительный анализ SAPP арабской речи с использованием НММ

Источник	Система распознавания	База данных	Число дикторов	WER, %	Точность, %
[2]	НММ / GMM / MFCC	415 звукозаписей	40	7,33 6,12	92,67 93,88
[3]	НММ / MFCC	600	110	10,98 6,81	89,02 93,19
[4]	НММ / MFCC	30 звукозаписей	30	3,3	96,7
[5]	НММ / MFCC	1 340 и 306 слов	100	5,26 2,72	94,74 97,72
[6]	НММ / MFCC / GMM	300 звукозаписей	12	6	94
[7]	НММ / MFCC	15 450 звукозаписей	1033	6,28	93,72
[8]	GHMM / SVM/ MFCC и GHMM / KNN/ MFCC	4 600 звукозаписей	92	9,38	90,62 88,26
[9]	НММ / LPC	1 500	50	11,74	95 91
[10]	Языковые модели	100 000 высказываний	–	14	86
[11]	НММ / MFCC	–	44	9,25	90,75

Источник	Система распознавания	База данных	Число дикторов	WER, %	Точность, %
[12]	HMM/MLLR	11 600 звукозаписей	38	1,78	98,22
[13]	HMM / MFCC	1 368 слов	19	7,08	92,92
[14]	CMU Sphinx4 / MFCC	1 432 слов	40	8,4 9,8	91,6 90,2
[15]	HMM / CMUSphinx4	13,24 часа звукозаписей	342	1,818	98,182
[16]	HMM / MFCC / Sphinx	300 звукозаписей	6	13,34	86,66
[17]	HMM / CMU Sphinx	1 660 звукозаписей	19	36,2	63,8

### Системы распознавания арабской речи с использованием искусственных нейронных сетей

Искусственные нейронные сети (Artificial neural network (ANN) широко использованы в области автоматического распознавания речи (Automatic speech recognition (ASR) в течение последних трех десятилетий. Наиболее полезными характеристиками ANN при решении проблемы ASR являются отказоустойчивость и нелинейность. Одной из важных моделей нейронных сетей являются многослойные перцептроны (multi-layer perceptron (MLP), которые представляют собой сети прямой связи с нулем, одним или несколькими скрытыми слоями. Возможности MLP проистекают из нелинейностей. Структура любой сети зависит от сложности сети и требований исследуемого приложения [18, 19].

### Системы распознавания непрерывной речи

В работах [1, 19] исследована арабская речь многожанровой трансляции (Multi-Genre Broadcast (MGB) с целью идентификации основных арабских диалектов – EGY, LEV, NOR, персидский, а также MSA. Автором разработана систему MGB с использованием звукозаписей с длительностью 1 200 часов. WER системы равна 12,03 %. Системы обучены с использованием структуры моделирования максимальной взаимной информации без решетки (lattice-free maximum mutual information (LF/MMI) и рекуррентной нейронной сети с языковыми моделями соединений с максимальной энтропией (recurrent neural network with maximum entropy (RNN/ME), а также с использованием нейронной сети с временной задержкой (Time Delay Neural Network (TDNN) и рекуррентной нейронной сети (recurrent neural network (RNN) с долговременной памятью (Long short-term memory (LSTM).

В исследовании [20] показано использование нейронных сетей с языковыми моделями (recurrent neural network language modeling toolkit (RNN/LM) для распознавания корпуса новости арабской трансляции. WER системы равна 14 %. При проведении эксперимента использованы звукозаписи с длительностью 1 400 часов.

В работе [21] описана система для исследования влияния различных оконных функций на точность системы распознавания. Результаты показывают повышение точности системы распознавания (рис. 4) при малом изменении параметров речевого сигнала (с применением окна Хана).

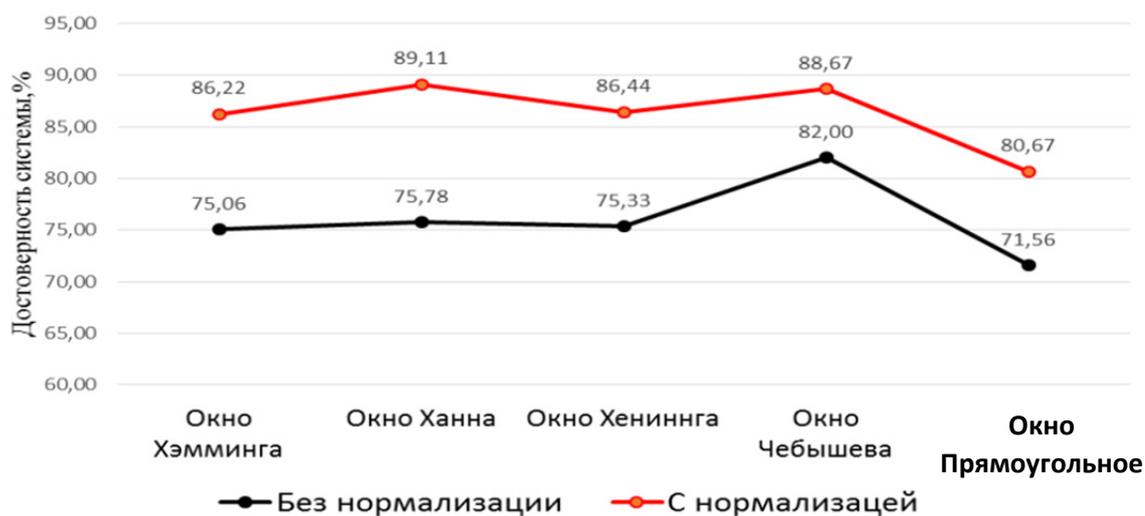


Рис. 4. Достоверность CAPR при исследовании разных оконных функций

### Системы распознавания изолированной речи

В работе [22] исследованы комбинации VQ и НММ для распознавания арабской речи. В результате разработана система на основе VQ/НММ для распознавания арабской речи. Автором получена точность системы, равная 95 %.

В работе [23] разработана система распознавания арабских цифр с использованием ANN и НММ. Система на основе ANN достигла 99,5 % правильного распознавания цифр в случае дикторозависимой системы и 94,5 % в случае дикторонезависимой системы. С другой стороны, система распознавания на основе НММ достигла 98,1 % правильного распознавания цифр в случае дикторозависимой системы и 94,8 % в случае дикторонезависимой системы.

В работе [24] исследованы влияния диакритических знаков на точность системы распознавания арабской речи. Автором разработана система на основе сверточной нейронной сети (convolutional neural network (CNN), CNN/LSTM. Языковая модель обучена с использованием RNN/LM и LSTM/LM. Экспериментальные результаты показали, что удаление диакритических знаков увеличивает частоту WER по сравнению с диакритическими данными. Достигнутое среднее снижение WER составило 13,52 %.

В работе [25] показана система распознавания йеменских диалектов с использованием моделей НММ/MFCC, точность системы достигла 95 %.

В работе [26] исследованы НММ и глубоких нейронных сетей (Deep Neural Networks (DNN) и (end2end (E2E) CAPR человеческой речи (human speech recognition (HSR) на арабском языке и его диалектах. Модели системы обучены непосредственно таким образом, чтобы преобразовывать акустические характеристики речи в текстовые транскрипции. E2E ASR привлек внимание академических и промышленных кругов. WER работы 12,5 %, 27,5 %, 33,8 %. Результаты показывают, что человеческая производительность в арабском языке по-прежнему значительно лучше, чем производительность машины с абсолютным разрывом WER в среднем на 3,5 %. На рис. 5 приведена зависимость ошибки CAPR для различных объемов по обучающим базам данных.

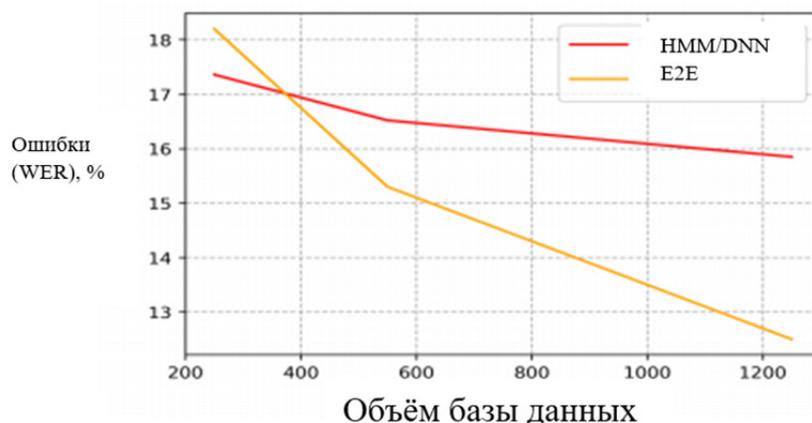


Рис. 5. Зависимость ошибки CAPP

В работе [27] разработана и протестирована система распознавания арабских цифр на основе RNN. Характеристики кадров извлекаются с использованием параметров MFCC для уменьшения количества информации во входном сигнале. Система распознавания обеспечивает правильное распознавание цифр в 99,5 % для многодикторного режима и в 94,5 % для независимого от говорящего режима. В работе [28] исследованы различные комбинированные архитектуры классификаторов на основе нейронных сетей, варьируя начальные веса, архитектуру, тип и обучающие данные линейного векторного квантования (linear vector quantization (LVQ) для распознавания арабского языка, точность системы распознавания достигла 96 %. Объём базы данных составлен 600 звукозаписей голоса, 300 звуков для обучения и 300 звуков для тестирования. В работе [29] описана система распознавания разговорных арабских цифр. Система проектирована с использованием MLP. Система работает в двух разных режимах – дикторозависимом и дикторонезависимом. База данных объемом 1 700 звукозаписей от 17 дикторов (носителей арабского языка). Общая точность системы составлена 99,47 % в первом режиме и 96,46 % во втором режиме. Цифра 9 не распознана системой в обоих режимах из-за большой несхожести цифры со всеми остальными. С другой стороны, цифры 1, 4 и 8 имеют относительно высокий уровень ошибок, особенно в дикторонезависимом режиме.

В работе [30] представлен новый подход к разработке системы распознавания и идентификации арабских фонем с использованием контуров интенсивности и формантных частот. Разработанная система распознавания основана на наборах данных от 40 дикторов. Система проектирована с использованием обобщенной регрессионной нейронной сети (Generalized Regression Neural Networks (GRNN)). Общая точность системы достигла 82,59 %.

В работе [31] исследована система распознавания разговорных арабских цифр с использованием двух методов: алгоритм определения основного тона (Pitch Detection Algorithm (PDA) и алгоритм кепстральной корреляции (Cepstrum Correlation Algorithm (CCA)). Голосовые команды «шесть» и «девять» распознаны с наихудшей точностью. В работе [32] описано сравнение трех подходов распознавания арабских цифр, в результате показано, что распознавание арабских разговорных цифр с использованием параметров MFCC лучше всех остальных. Сравнительный анализ провиден с использованием алгоритма динамической трансформации временной шкалы (dynamic time wrapping (DTW)).

В работе [33] автором исследована система распознавания арабской речи. Входные звукозаписи представляют собой десять арабских слов, и каждое из них повторялось 25 раз одним и тем же диктором. Подход извлечения признаков выполнялся путем применения стационарного бионического вейвлет-преобразования (Stationary Bionic Wavelet Transform (SBWT)), затем коэффициенты MFCC рассчитываются и объединяются для подачи на вход MLP. В результате получилась точность классификации, равная 98 %. В табл. 3

приведен сравнительный анализ САРП арабской речи с использованием искусственных нейронных сетей.

Таблица 3

### Сравнительный анализ САРП арабской речи с использованием ANN

Источник	Система распознавания	База данных	Число дикторов	WER, %	Точность, %
[18]	DNN / LSTM / MFCC	1 400 часов	450–500	14	86
[19]	Языковые модели	16 часов звукозаписей	–	12,03 37,5	87,97
[20]	Языковые модели и ANN	737 000 слов	–	12,4 20,28	87,6 79,72
[21]	HMM / MFCC	300 звукозаписей	6	3,5	96,5
[22]	HMM / VQ	1 500 звукозаписей	50	5	95
[23]	HMM / MFCC / ИНС	1 700 звукозаписей	17	0,5 3,5	99,5 94,5
[24]	CNN-LSTM / MFCC	4 372	1	14,92	85,08
[25]	HMM / MFCC	150 звукозаписей	6	4	95
[26]	HMM–DNN	1 200 часов	–	3,5	–
[27]	RNN / MFCC	1 700 звукозаписей	17	0,5 3,5	99,5 94,5
[28]	LVQ / ИНС / MFCC	600 звукозаписей	10	4	96
[29]	MLP / MFCC	1 700 звукозаписей	17	0,53	99,47
[30]	GRNN / LPC	28 звукозаписей	40	17,41	82,59
[32]	DTW / MFCC	780 звукозаписей	6	13	87
[33]	MLP / SBWT / MFCC	250 звукозаписей	1	2	98

Таким образом, вышеприведенная информация указывает на то, что арабский йеменский диалект почти не исследован, есть потребность в создании специализированной системы по автоматическому распознаванию данного диалекта.

### Системы классификации и идентификации арабских диалектов

В работе [34] приведены особенности тихамейского йеменского диалекта, а также решены три проблемы: 1) отсутствие кратких гласных; 2) морфологическая сложность; 3) расхождение между диалектным и формальным арабским языком. Результаты системы демонстрируют снижение частоты ошибок в словах. В работе [35] изучена дикторнезависимая изолированная система арабских цифр с использованием двух классификаторов – полиномиального классификатора (polynomial classifier (PNC)) и классификатора DTW. Получен вывод, что полиномиальный классификатор обеспечивает лучшую точность распознавания и гораздо более быстрый отклик при тестировании, чем алгоритм DTW.

В работе [36] исследована система идентификации арабских диалектов, в том числе кувейтский, йеменский и марокканский диалект с использованием частоты основного тона (F0). Сравнение значений выравнивания F0 проведено с использованием пяти носителей каждого диалекта. Явные различия, обнаруженные, влияющие на результаты классификации исследуемых диалектов, следующие: а) значения F0; б) влияние структуры слога на пики F0. Эффект незначителен для йеменского арабского диалекта и кувейтского арабского диалекта. Однако структуры слогов существенно влияют на результаты классификации марокканского арабского диалекта.

В работе [37] рассмотрены различные модели глубокого обучения для автоматической классификации арабских диалектов. Работа провидена с использованием CNN, LSTM, (Bi-directions long short-term memory (BiLSTM)) и (Convolution long short-term memory

(CLSTM). В работе [38] описана система идентификации арабского диалекта с использованием CNN с несколькими входами. Дополнительно описаны результаты по сети CNN-biLSTM. Система создана для исследования следующих арабских диалектов: EGY, Gulf, LEV, MSA и NOR.

В работе [39] отображена система распознавания и идентификации алжирского диалекта, система спроектирована с использованием основания на просодической особенности речевого сигнала. Особенности речи извлекаются после грубой сегментации согласных и гласных звуков. Диалектные модели построены с использованием DNN и SVM.

В работах [40–42] изучена роль просодических особенностей (интонации и ритма) при идентификации четырех арабских диалектов: персидского, IRQ, LIV и EGY. Просодические особенности могут значительно улучшать идентификацию по сравнению с чисто фонотактическим подходом с точностью идентификации 86,33 % для двух миллионных звукозаписей. Для выполнения работы автором использовано распознавание фонем с последующим моделированием языка (Phone Recognition followed by Language Modeling (PRLM)), используется фонотактическая информация для идентификации языков только по акустическому сигналу. Также исследована лексическая связь между MSA и диалектным арабским языком (dialectal Arabic (DA)). Результаты информативны и указывают на то, что левантийские диалекты очень похожи друг на друга и, кроме того, палестинский диалект кажется наиболее близким к MSA.

В работе [43] исследована проблема идентификации арабских диалектов. Автором собран собственный набор данных из-за отсутствия существующих наборов данных. Также сосредоточились на двух конкретных диалектах, а именно йорданский и египетский. Исследовано несколько методов извлечения признаков и построения классификаторов. Результаты были удивительно хорошими, поскольку классификатор байесовской сети достиг почти идеальной точности. В работе исследованы следующие классификаторы: простой вероятностный классификатор, использующий теорему (Байеса NaiveBayes (NB), простой классификатор таблицы решения (DecisionTable (DT), который использует таблицу решений для обобщения набора данных, реализация алгоритма Repeated Incremental Pruning to Produce Error Reduction (RIPPER (Jrip), классификатор (One Rule (OneR) – генерирует одно правило для каждого предиктора данных, затем выбирает правило с минимальной ошибкой, классификатор дерева решений (RandomTree (RT), Java-реализация алгоритма C4.5 (J48) с открытым исходным кодом для построения дерева решений, MLP классификатор искусственной нейронной сети с прямой связью, который использует обратное распространение для обучения и сигмовидную функцию в каждом узле, реализация алгоритма последовательной минимальной оптимизации (Sequential Minimal Optimization (SMO) для обучения SVM-классификатора, реализация классификатора К ближайших соседей (IBk), классификатор байесовской сети (BayesNet (BNet), Бэггинг (Bagging) подход к повышению стабильности и точности классификатора и уменьшению дисперсии и подход адаптивного повышения (AdaBoostM1) к созданию сильного классификатора из набора слабых классификаторов.

В табл. 4 проанализированы оценки систем классификаторов арабского алжирского диалекта.

В работе [44] описаны возможности классификации текстов арабских диалектов с использованием алгоритмов машинного обучения. Авторы исследовали пять групп арабских диалектов: MSA, IRQ, EGY, LEV и NOR. Входные данные следующие: 587 тыс. образцов; 16 179 помечены как DA и 7 821 как MSA. В работе [45] исследованы три подхода автоматического распознавания арабской речи. Экспериментальные результаты показывают, что предложенный подход к модели произношения снижает WER примерно на 1 %. Акустическое моделирование снижает WER на 1,2 %, а адаптированное языковое моделирование снижает WER на 1,9 %.

Таблица 4

## Системы классификаторов арабского алжирского диалекта

Алгоритм	Оценка системы			
	Точность	Повторения	F-Оценка	AUC
NB	0,859	0,857	0,85	0,828
DT	0,75	0,714	0,723	0,722
JRip	0,796	0,8	0,797	0,803
OneR	0,582	0,571	0,576	0,515
RT	0,808	0,714	0,724	0,767
J48	0,714	0,714	0,714	0,67
MLP	0,889	0,829	0,835	0,958
SMO	0,918	0,914	0,915	0,913
IBk	0,824	0,6	0,695	0,695
<b>BayesNet</b>	<b>0,974</b>	<b>0,971</b>	<b>0,972</b>	<b>1</b>

В работе [46] исследован подход иерархической классификации для идентификации разговорного арабского алжирского диалекта (Hierarchical classification approach for spoken Arabic algerian Dialect Identification (HADID)). В HADID применяется иерархическая классификация с использованием глубоких нейронных сетей. HADID обеспечивает повышение точности на 63,5 %.

В работе [47] разработана система распознавания арабских фонем на основе (Echo State Networks (ESN)), также используются и сравниваются методы MFCC и LPC. Система оценивается с использованием звукозаписи шести носителей арабской фонетики Саудовской Аравии (King Abdulaziz Arabic Phonetics Database (KAPD)) и 34 носителей из базы данных Центра понимания разговорного языка (CSLU2002) носителей с различной диалектикой из 12 арабских стран. Результаты тестирования арабских баз данных, KAPD и CSLU2002 показывают эффективность распознавания фонем 72,31 % и 38,20 % соответственно.

В работе [48] разработаны модели автоматического распознавания речи для суданского диалекта. Каждый стандартный блок модели сформирован с использованием CNN и Connectionist Temporal Classification (CTC). Средний уровень ошибок достиг 73,67 %. В работе [49] представлены проектирование и разработки системы автоматического распознавания арабских диалектов. Сетевая архитектура состоит из комбинации CNN и RNNs. Предложенная система достигла коэффициента ошибок в 14 %. В табл. 5 проиллюстрирован сравнительный анализ CAPR арабской речи с использованием искусственных нейронных сетей, скрытых Марковских моделей.

Таблица 5

## Сравнительный анализ CAPR арабской речи с использованием HMM, ANN и других классификаторов

Источник	Система распознавания	База данных	Число дикторов	WER, %	Точность, %
[35]	DTW / PNC/ MFCC	3 000 звукозаписей	30	25,3	74,7
[37]	LSTM / CNN / BLSTM / CLSTM	33к звукозаписей	–	–	83,8
[38]	CNN / CNN-biLSTM	22 994 звукозаписей	–	–	F1=0.52
[39]	DNN / SVM / MFCC	42 часов	–	–	60,09
[40–42]	PRLM / MFCC	145 часов	2194	18,4	81,60
[43]	BayesNet / MFCC / DWT	35 звукозаписей	2	–	100
[46]	DNNs / SVM / Prosodic	1 892 звукозаписей	41	–	63,5

Источник	Система распознавания	База данных	Число дикторов	WER, %	Точность, %
[47]	LPC / MFCC / ESN / MLP	640 звукозаписей	40	–	72,31
[49]	HMM / MFCC	3,870 звукозаписей	1033	6,28	93,72

### Факторы, влияющие на разработку и точность САР арабской речи

В результате данного исследования можно перечислить следующие проблемы и факторы, влияющие на разработку и точность САР при распознавании арабской речи: в арабском языке существуют короткие гласные знаки, которые обычно игнорируются при написании в тексте. Поэтому при разработке САР ASR увеличиваются ошибки распознавания из-за разновидности произношения слов. В арабском языке много диалектов, где слова произносятся по-разному в каждой стране и иногда в каждом регионе, например, в Йемене существует несколько диалектов в каждом регионе, отсюда есть разновидности фонетических характеристик арабской речи. Также диакритические знаки представляются различными возможными конкретными словами. Поскольку современный стандартный арабский язык пишется в недиакритизированных сценариях, из-за этого при произношении конкретного слова вводится много неясностей. Следовательно, при создании акустических моделей, используемых для обучения системы, встречается конкретное слово, но с разными моделями и разными значениями. Такие разновидности отрицательно влияют на точность системы распознавания. Морфологическая сложность арабского языка демонстрируется большим количеством аффиксов (префиксов и суффиксов).

Анализ литературных источников показал, что при распознавании арабской речи встречается много трудностей, влияющих на производительность систем распознавания. Количество научно-исследовательских работ по системе распознавания йеменского диалекта сравнительно мало по сравнению с другими арабскими диалектами. Данная ситуация сложилась на настоящий момент из-за малочисленности квалифицированных специалистов по текущему направлению, политической обстановки в стране и отсутствия базы данных арабского йеменского диалекта [50, 51].

### Заключение

Из приведенного обзора научной литературы, очевидно, что при разработке систем идентификации арабских диалектов использовались различные методы, начиная от просьбы испытуемых слушать диалекты и заканчивая системами автоматической идентификации диалектов. При разработке системы распознавания арабской речи в большинстве случаев используются HMM, Sphinx и различные варианты ANN. При извлечении информативных признаков используются функции MFCC, LPC, просодические характеристики речевого сигнала и языковые модели. Изменчивость произношения слов связана с тем, что в повседневном общении арабы используют местные диалекты. Кроме того, сложная задача идентификации диалекта усложняется, когда речь идет о диалектах с ограниченными ресурсами, принадлежащими к одному и тому же региону. Следовательно, для повышения точности системы распознавания, необходимо внедрить алгоритм классификации или идентификации диалекта в состав системы САР.

В последнее время системы ASR широко используются во многих приложениях, таких как преобразование речи в текст, системы домашней безопасности и военные приложения. Область распознавания речи для арабского йеменского диалекта еще «не созрела» и нуждается в развитии, поэтому в дальнейшем будет разрабатываться специализированная система автоматического распознавания йеменского диалекта с использованием ANN и скрытых Марковских моделей. Предусмотрена возможность создания алгоритмов идентификации диалекта для повышения точности системы. Кроме

того, будет учтено влияние шумов на работу системы автоматического распознавания йеменского диалекта. При этом существует потребность в формировании базы данных арабской речи, в которой будут представлены звукозаписи носителей арабского языка из нескольких городов Республики Йемен. Данная база данных получит название Банк голоса арабского языка с йеменским диалектом (Yemeni Accented Arabic Voice Bank) и будет применяться при разработке и исследовании системы автоматического распознавания арабского йеменского диалекта.

#### Список источников

1. Arabic Speech Recognition with Deep Learning: A Review. Lecture Notes in Computer Science / W. Alghibab [et al.]. 2019. P. 15–31. DOI: 10.1007/978-3-030-21902-4\_2.
2. Natural speaker-independent Arabic speech recognition system based on HMM using Sphinx tools / Abushariah MAM, R.N. Ainon [et al.]. 2010. ICCCE.
3. Selouani S.-A., Alotaibi Y.A. Adaptation of foreign accented speakers in native Arabic ASR systems. Appl Comput Informat. 2011. № 9 (1). P. 1–10.
4. Alorfi F.S. Automatic Identification of Arabic Dialects Using Hidden Markov Models, Ph.D. thesis, USA: University of Pittsburgh, 2008.
5. The development of acoustic models for command and control Arabic speech recognition system / M. Nofal [et al.]. 2004. ICEEC'04.
6. Sa'Ed Abed, Hanem Ellethy, Mohammad H Alshayegi. Continuous formal arabic speech recognition system based on hidden Markov model. 3rd ICCSNIT on 26th–27th August 2017, in Montreal, Canada. 2017. ISBN: 9780998900032.
7. Alotaibi Y.A., Alghamdi M., Alotaiby F. Speech Recognition System of Arabic Digits based on A Telephony Arabic Corpus // Comput. Speech and Lang. – Department of Electrical Engineering, King Saud Univ., Riyadh, Saudi Arabia. 2008. P. 4.
8. New hybrid system (supervised classifier/HMM) for isolated Arabic speech recognition / H. Bourouba [et al.]. 2nd. ICTTA'06.
9. Bahi H., Sellami M. Combination of vector quantization and hidden Markov models for Arabic speech recognition. Proceedings in: ACS/IEEE ICCSA. 2001.
10. Omar F. Zaidan Arabic dialect identification // Computational Linguistics. 2014. Vol. 40. Iss. 1. P. 171–202.
11. Syllable-based automatic Arabic speech recognition in noisy-telephone channel. In: WSEAS transactions on signal processing proceedings, WSEAS / M. Azmi [et al.]. 2008. Vol. 4. Iss. (4). P. 211–220.
12. Computer aided pronunciation learning system using speech recognition techniques, NTERSPRECH 2006, ICSLP / SM. Abdou [et al.]. P. 249–252.
13. Elharati H.A., Alshaari M., Kępuska V.Z. Arabic Speech Recognition System Based on MFCC and HMMs. jcc. 2020. № 08 (03). P. 28–34.
14. Eman Z.E., PhD thesis. Arabic Continuous Speech Recognition System using Sphinx-4. 2012. P. 87.
15. Hyassat H., Abu Zitar R. Arabic speech recognition using SPHINX engine // International Journal of Speech Technology. 2006. № 9 (3-4). P. 133–150. DOI: 10.1007/s10772-008-9009-1.
16. Satori H., Harti M., Chenfour N. Introduction to Arabic speech recognition using CMU Sphinx system. ICTIS07, 2007.
17. Al-Anzi F.S., AbuZeina D. The effect of diacritization on Arabic speech recognition. 2017. IEEE Jordan Conference on AEECT. DOI: 10.1109/aeect.2017.8257758.
18. Ali A.R. Multi-Dialect Arabic Speech Recognition // 2020 International Joint Conference on Neural Networks (IJCNN). DOI: 10.1109/ijcnn48605.2020.9206658.
19. Ali A., Vogel S., Renals S. Speech recognition challenge in the wild: Arabic MGB-3. 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). P. 316–322.

20. Emami A., Mangu L. Empirical study of neural network language models for Arabic speech recognition. *IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*. 2007. P. 147–152.
21. Levin E., Al-Dhaibani A. Research of Window Function Influence on the Result of Arabic Speech Automatic Recognition. *USBREIT 2019*. IEEE 2019. DOI: 10.1109/usbreit.2019.873657.
22. Bahi H., Sellami M. A hybrid approach for Arabic speech recognition. *Proceedings in: ACS/IEEE ICCSA*. 2003.
23. Alotaibi Y.A. Comparative Study of ANN and HMM to Arabic Digits Recognition Systems // *Journal of King Abdulaziz University: Engineering Sciences*. 2008. Vol. 19. № 1. P. 43–60.
24. Alsayadi H.A., Abdelaziz A., Hegazy I., Fayed T. Non-diacritized Arabic speech recognition based on CNN-LSTM and attention-based models. *Nov. 2021*. № 41 (1). P. 1–13. DOI: 10.3233/JIFS-202841.
25. Аль-Дайбани А. Исследование методов и разработка алгоритмов обработки сигналов для систем автоматического распознавания телефонной речи в республике Йемен. *Владимир: Влгу*, 2019. 150 с.
26. Hussein A., Watanabe S., Ali A. Arabic speech recognition by end-to-end, modular systems and human. *Computer Speech&Language*. 2022. № 71. P. 101272. DOI: 10.1016/j.csl.2021.101272.
27. Alotaibi Y.A. Spoken Arabic digits recognizer using recurrent neural networks. In: *Proceedings of the fourth IEEE ISSPIT*. 2004. P. 195–199.
28. Essa E.M., Tolba A.S., Elmougy S. A comparison of combined classifier architectures for Arabic Speech Recognition. *International Conference on Computer Engineering & Systems*. 2008. DOI: 10.1109/icces.2008.4772985.
29. Alotaibi Y.A. Investigating spoken Arabic digits in speech recognition setting. *Information Sciences*. 2005. Vol. 173. № 1-3. P. 115–139.
30. A novel approach to increase the robustness of speaker independent Arabic speech recognition / M. Shoaib [et al.]. *IEEE. 7th. INMIC*. 2003. № 8-9. P. 371–376.
31. Anas A., Nura Z., Ali G. *Speech Recognition of Arabic Spoken Digits // University of Tripoli, Faculty of Engineering, P.O. Box 13589, Tripoli, Libya*. 2013. P. 8.
32. Ali Ganoun, Ibrahim Almerhag, Performance Analysis of Spoken Arabic Digits Recognition Techniques // *Ganoun Ali, Almerhag Ibrahim – Journal of electronic science and technology*. 2012. Vol. 10. № 2. P. 5.
33. Mourad T. Arabic Speech Recognition by Stationary Bionic Wavelet Transform and MFCC Using a Multi-layer Perceptron for Voice Control. In: *SCT*. Springer. 2022. P. 69–81.
34. Abdulghani M.A. The Phonology and Morphology of Yemeni Tihami Dialect: An Autosegmental Account. URL: <https://core.ac.uk/display/32600773> (дата обращения: 15.03.2023).
35. The application of polynomial discriminant function classifiers to isolated Arabic speech recognition. In: *Proceedings of the IJCNN / M. Khasawneh [et al.]*. 2004. P. 3077–3081.
36. F0 alignment patterns in arabic dialects / M. Yeou [et al.] // *Proc. of the 16th. ICPHS XVI*. Saarbrücken, Germany. 2007.
37. Lulu L., Elnagar A. Automatic Arabic Dialect Classification Using Deep Learning Models. *Procedia Computer Science*. № 142. P. 262–269. DOI: 10.1016/j.procs.2018.10.489.
38. Neural Network Architectures for Arabic Dialect Identification. *Association for Computational Linguistics*. In *Proc. of the Fifth Workshop on NLP for Similar Languages, VarDial / M. Elise [et al.]*. Santa Fe, New Mexico, USA. 2018. P. 128–136.
39. Soumia Bougrine H.C., Abdelali A. Spoken Arabic Algerian dialect identification. *Second ICNLS*. 2018. DOI: 10.1109/icnls.2018.8374383.
40. Biadisy F., Hirschberg J., Habash N. Spoken Arabic Dialect Identification Using Phonotactic Modeling, *Proceedings in: EACL*. 2009. P. 53–61.

41. Biadisy F., Hirschberg J., Using Prosody and Phonotactics in Arabic Dialect Identification, Proceedings in: Interspeech. 2009. P. 208–211.
42. A Lexical Distance Study of Arabic Dialects. Procedia Computer Science / K.A. Kwaik [et al.]. 2018. № 142. P. 2–13. DOI: 10.1016/j.procs.2018.10.456.
43. Spoken Arabic dialects identification: The case of Egyptian and Jordanian dialects. International Conference on Information and Communication Systems (ICICS) / M. Al-Ayyoub [et al.]. 2014. 5th. DOI: 10.1109/iacs.2014.6841970.
44. Alshutayri A., Atwell E. Arabic dialects annotation using an online game // 2nd International Conference on Natural Language and Speech Processing. (ICNLSPIEEE). 2018. DOI: 10.1109/icnls.2018.8374371.
45. Ahmed B.H.A., Ghabayen A.S. Arabic Automatic Speech Recognition Enhancement // Palestinian International Conference on Information and Communication Technology (PICICTIEEE'017). 2017. DOI: 10.1109/picict.2017.12.
46. Soumia Bougrine, Hadda Cherroun, Djelloul Ziadi. Hierarchical Classification for Spoken Arabic Dialect Identification using Prosody: Case of Algerian Dialects. Computation and Language. 2017.
47. Nadia H., Tony A. Echo State Networks for Arabic Phoneme Recognition. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering. 2013. Vol. 7. № 7.
48. Abdelaziz A.E.M. A Proposed Automatic Speech Recognition model for the Sudanese Dialect, MSc thesis, Sudan University of science and techno. 2020.
49. Ali A.R. Multi-Dialect Arabic Speech Recognition. 2020 IJCNN. DOI: 10.1109/ijcnn48605.2020.92066.
50. Modern standard Arabic based multilingual approach for dialectal Arabic speech recognition. In: Eighth SNLP / M. Elmahdy [et al.]. 2009. IEEE.
51. AbuZeina Dia, Elshafei Moustafa. Cross-Word Modeling for ASR // Elshafei, SpringerBriefs in in Electrical and Computer Engineering. 2012. 74 p. DOI: 10.1007/978-1-4614-1213-7\_2.

## References

1. Arabic Speech Recognition with Deep Learning: A Review. Lecture Notes in Computer Science / W. Alghibab [et al.]. 2019. P. 15–31. DOI: 10.1007/978-3-030-21902-4\_2.
2. Natural speaker-independent Arabic speech recognition system based on HMM using Sphinx tools / Abushariah MAM. R.N. Ainon [et al.]. 2010. ICCCE.
3. Selouani S.-A., Alotaibi Y.A. Adaptation of foreign accented speakers in native Arabic ASR systems. Appl Comput Informat. 2011. № 9 (1). P. 1–10.
4. Alorfi F.S. Automatic Identification of Arabic Dialects Using Hidden Markov Models, Ph.D. thesis, USA: University of Pittsburgh, 2008.
5. The development of acoustic models for command and control Arabic speech recognition system / M. Nofal [et al.]. 2004. ICEEC.
6. Sa'Ed Abed, Hanem Ellethy, Mohammad H Alshayegi. Continuous formal arabic speech recognition system based on hidden Markov model. 3rd ICCSNIT on 26th–27th August 2017, in Montreal, Canada. 2017. ISBN: 9780998900032.
7. Alotaibi Y.A., Alghamdi M., Alotaiby F. Speech Recognition System of Arabic Digits based on A Telephony Arabic Corpus // Comput. Speech and Lang. – Department of Electrical Engineering, King Saud Univ., Riyadh, Saudi Arabia. 2008. P. 4.
8. New hybrid system (supervised classifier/HMM) for isolated Arabic speech recognition / H. Bourouba [et al.]. 2nd. ICTTA'06.
9. Bahi H., Sellami M. Combination of vector quantization and hidden Markov models for Arabic speech recognition. Proceedings in: ACS/IEEE ICCSA. 2001.
10. Omar F. Zaidan Arabic dialect identification // Computational Linguistics. 2014. Vol. 40. Iss. 1. P. 171–202.

11. Syllable-based automatic Arabic speech recognition in noisy-telephone channel. In: WSEAS transactions on signal processing proceedings, WSEAS / M. Azmi [et al.]. 2008. Vol. 4. Iss. (4). P. 211–220.
12. Computer aided pronunciation learning system using speech recognition techniques, NTERSPREECH 2006, ICSLP / SM. Abdou [et al.]. P. 249–252.
13. Elharati H.A., Alshaari M., Kėpuska V.Z. Arabic Speech Recognition System Based on MFCC and HMMs. jcc. 2020. № 08 (03). P. 28–34.
14. Eman Z.E., PhD thesis. Arabic Continuous Speech Recognition System using Sphinx-4. 2012. P. 87.
15. Hyassat H., Abu Zitar R. Arabic speech recognition using SPHINX engine // International Journal of Speech Technology. 2006. № 9 (3-4). R. 133–150. DOI: 10.1007/s10772-008-9009-1.
16. Satori H., Harti M., Chenfour N. Introduction to Arabic speech recognition using CMU Sphinx system. ICTIS07, 2007.
17. Al-Anzi F.S., AbuZeina D. The effect of diacritization on Arabic speech recognition. 2017. IEEE Jordan Conference on AEECT. DOI: 10.1109/aeect.2017.8257758.
18. Ali A.R. Multi-Dialect Arabic Speech Recognition // 2020 International Joint Conference on Neural Networks (IJCNN). DOI: 10.1109/ijcnn48605.2020.9206658.
19. Ali A., Vogel S., Renals S. Speech recognition challenge in the wild: Arabic MGB-3. 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). P. 316–322.
20. Emami A., Mangu L. Empirical study of neural network language models for Arabic speech recognition. IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU). 2007. P. 147–152.
21. Levin E., Al-Dhaibani A. Research of Window Function Influence on the Result of Arabic Speech Automatic Recognition. USBEREIT 2019. IEEE 2019. DOI: 10.1109/usbereit.2019.873657.
22. Bahi H., Sellami M. A hybrid approach for Arabic speech recognition. Proceedings in: ACS/IEEE ICCSA. 2003.
23. Alotaibi Y.A. Comparative Study of ANN and HMM to Arabic Digits Recognition Systems // Journal of King Abdulaziz University: Engineering Sciences. 2008. Vol. 19. № 1. P. 43–60.
24. Alsayadi H.A., Abdelaziz A., Hegazy I., Fayed T. Non-diacritized Arabic speech recognition based on CNN-LSTM and attention-based models. Nov. 2021. № 41 (1). P. 1–13. DOI: 10.3233/JIFS-202841.
25. Al'-Dajbani A. Issledovanie metodov i razrabotka algoritmov obrabotki signalov dlya sistem avtomaticheskogo raspoznavaniya telefonnoj rechi v respublike Jemen. Vladimir: Vlgu, 2019. 150 s.
26. Hussein A., Watanabe S., Ali A. Arabic speech recognition by end-to-end, modular systems and human. Computer Speech&Language. 2022. № 71. P. 101272. DOI: 10.1016/j.csl.2021.101272.
27. Alotaibi Y.A. Spoken Arabic digits recognizer using recurrent neural networks. In: Proceedings of the fourth IEEE ISSPIT. 2004. P. 195–199.
28. Essa E.M., Tolba A.S., Elmougy S. A comparison of combined classifier architectures for Arabic Speech Recognition. International Conference on Computer Engineering & Systems. 2008. DOI: 10.1109/icces.2008.4772985.
29. Alotaibi Y.A. Investigating spoken Arabic digits in speech recognition setting. Information Sciences. 2005. Vol. 173. № 1-3. R. 115–139.
30. A novel approach to increase the robustness of speaker independent Arabic speech recognition / M. Shoaib [et al.]. IEEE. 7th. INMIC. 2003. № 8-9. P. 371–376.
31. Anas A., Nura Z., Ali G. Speech Recognition of Arabic Spoken Digits // University of Tripoli, Faculty of Engineering, P.O. Box 13589, Tripoli, Libya. 2013. P. 8.

32. Ali Ganoun, Ibrahim Almerhag, Performance Analysis of Spoken Arabic Digits Recognition Techniques // Ganoun Ali, Almerhag Ibrahim – Journal of electronic science and technology. 2012. Vol. 10. № 2. P. 5.
33. Mourad T. Arabic Speech Recognition by Stationary Bionic Wavelet Transform and MFCC Using a Multi-layer Perceptron for Voice Control. In: SCT. Springer. 2022. P. 69–81.
34. Abdulghani M.A. The Phonology and Morphology of Yemeni Tihami Dialect: An Autosegmental Account. URL: <https://core.ac.uk/display/32600773> (data obrashcheniya: 15.03.2023).
35. The application of polynomial discriminant function classifiers to isolated Arabic speech recognition. In: Proceedings of the IJCNN / M. Khasawneh [et al.]. 2004. P. 3077–3081.
36. F0 alignment patterns in arabic dialects / M. Yeou [et al.] // Proc. of the 16th. ICPHS XVI. Saarbrücken, Germany. 2007.
37. Lulu L., Elnagar A. Automatic Arabic Dialect Classification Using Deep Learning Models. Procedia Computer Science. № 142. R. 262–269. DOI: 10.1016/j.procs.2018.10.489.
38. Neural Network Architectures for Arabic Dialect Identification. Association for Computational Linguistics. In Proc. of the Fifth Workshop on NLP for Similar Languages, VarDial / M. Elise [et al.]. Santa Fe, New Mexico, USA. 2018. P. 128–136.
39. Soumia Bougrine H.C., Abdelali A. Spoken Arabic Algerian dialect identification. Second ICNLSP. 2018. DOI: 10.1109/icnlsp.2018.8374383.
40. Biadisy F., Hirschberg J., Habash N. Spoken Arabic Dialect Identification Using Phonotactic Modeling, Proceedings in: EACL. 2009. P. 53–61.
41. Biadisy F., Hirschberg J., Using Prosody and Phonotactics in Arabic Dialect Identification, Proceedings in: Interspeech. 2009. P. 208–211.
42. A Lexical Distance Study of Arabic Dialects. Procedia Computer Science / K.A. Kwaik [et al.]. 2018. № 142. P. 2–13. DOI: 10.1016/j.procs.2018.10.456.
43. Spoken Arabic dialects identification: The case of Egyptian and Jordanian dialects. International Conference on Information and Communication Systems (ICICS) / M. Al-Ayyoub [et al.]. 2014. 5th. DOI: 10.1109/iacs.2014.6841970.
44. Alshutayri A., Atwell E. Arabic dialects annotation using an online game // 2nd International Conference on Natural Language and Speech Processing. (ICNLSPIEEE). 2018. DOI: 10.1109/icnlsp.2018.8374371.
45. Ahmed B.H.A., Ghabayen A.S. Arabic Automatic Speech Recognition Enhancement // Palestinian International Conference on Information and Communication Technology (PICICTIEEE'017). 2017. DOI: 10.1109/picict.2017.12.
46. Soumia Bougrine, Hadda Cherroun, Djelloul Ziadi. Hierarchical Classification for Spoken Arabic Dialect Identification using Prosody: Case of Algerian Dialects. Computation and Language. 2017.
47. Nadia H., Tony A. Echo State Networks for Arabic Phoneme Recognition. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering. 2013. Vol. 7. № 7.
48. Abdelaziz A.E.M. A Proposed Automatic Speech Recognition model for the Sudanese Dialect, MSc thesis, Sudan University of science and techno. 2020.
49. Ali A.R. Multi-Dialect Arabic Speech Recognition. 2020 IJCNN. DOI: 10.1109/ijcnn48605.2020.92066.
50. Modern standard Arabic based multilingual approach for dialectal Arabic speech recognition. In: Eighth SNLP / M. Elmahdy [et al.]. 2009. IEEE.
51. AbuZeina Dia, Elshafei Moustafa. Cross-Word Modeling for ASR // Elshafei, SpringerBriefs in in Electrical and Computer Engineering. 2012. 74 p. DOI: 10.1007/978-1-4614-1213-7\_2.

**Информация о статье:**

Статья поступила в редакцию: 14.04.2023; одобрена после рецензирования: 28.04.2023;  
принята к публикации: 29.04.2023

**The information about article:**

The article was submitted to the editorial office: 14.04.2023; approved after review: 28.04.2023;  
accepted for publication: 29.04.2023

*Информация об авторах:*

**Радан Наим Хусейн**, Тверской государственный технический университет (170026, г. Тверь, наб. Афанасия Никитина, д. 22), e-mail: [naeem.radan@gmail.com](mailto:naeem.radan@gmail.com)

*Information about the authors:*

**Radan Naim Hussein**, Tver state technical university (170026, Tver, Afanasy Nikitin emb., 22), e-mail: [naeem.radan@gmail.com](mailto:naeem.radan@gmail.com)