

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение высшего
образования**

«Нижегородский государственный университет им. Н.И. Лобачевского»

Институт информационных технологий, математики и механики

УТВЕРЖДЕНО
президиумом Ученого совета ННГУ
протокол от
«30» ноября 2022 г. № 13

Рабочая программа дисциплины

Технологии обработки речи

Уровень высшего образования

магистратура

Направление подготовки (специальность)

020402 Фундаментальная информатика и информационные технологии

Направленность образовательной программы

Искусственный интеллект

Форма обучения

очная

Нижний Новгород
2023

1. Место и цели дисциплины в структуре ОПОП

Дисциплина ФТД.02 «Технологии обработки речи» относится к факультативным дисциплинам направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии», направленность «Искусственный интеллект». Дисциплина преподается в 3 семестре.

№ Варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
1	ФТД. Факультативы	Дисциплина ФТД.02 «Технологии обработки речи» является факультативом в ООП направления подготовки 02.04.02 Фундаментальная информатика и информационные технологии

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции* (код, содержание индикатора)	Результаты обучения по дисциплине**	
ОПК-3. Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	ОПК-3.1. Знает методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей.	Знать теоретические основы методов и алгоритмов технологий обработки речи и их особенности в системах искусственного интеллекта (ИИ)	собеседование или доклад на семинаре
	ОПК-3.2. Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем	Уметь решать задачи подготовки исходных данных в виде аудиозаписей с заданными свойствами, выбирать алгоритмы обработки речи для решения конкретных задач и анализировать результаты методами технологий обработки речи.	доклад на семинаре, решение семинарских и домашних задач

3. Структура и содержание дисциплины

3.1. Трудоемкость дисциплины

Общая трудоемкость	1 ЗЕТ
Часов по учебному плану	36
в том числе:	
аудиторные занятия (контактная работа):	33
- занятия лекционного типа	32
- занятия семинарского типа	
- занятия лабораторного типа	
- текущий контроль (КСР)	1
самостоятельная работа	3
Промежуточная аттестация	зачет

3.2 Содержание дисциплины

№ п/п	Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Часов						
		В том числе						
		Всего	Контактная работа (работа во взаимодействии с преподавателем), часы из них				Самостоятельная работа обучающегося, часы	
Занятия лекционного типа	Занятия семинарского типа		Лабораторные работы	Консультации	Всего			
1	Введение в речевые технологии. История синтеза и распознавания. Какие существуют задачи в речевых технологиях. Как человек воспринимает звук. Сравнение с компьютерным зрением и обработкой естественных языков.	2	2				2	0
2	Цифровая обработка сигналов. Как звук дискретизируется для компьютеров. Характеристики аудиосигналов. Представления для работы со звуком. Дискретное преобразование Фурье. Спектрограмма, мелспектрограмма, мелкепстральные коэффициенты. Восстановление аудиосигнала из спектрограммы: алгоритм Гриффина-Лима.	2	21				2	0
3	Введение в распознавание речи. Обсуждение задачи распознавания. Сравнение различных представлений текста в качестве единиц речи. Проблема выравнивания единиц речи и акустических признаков: State-space models, Attention mechanism. Дискриминативная и генеративная постановки задачи распознавания. Метрики качества распознавания. Word error	4	4				4	1

	rate(WER). Расстояние Левенштейна и алгоритм Левенштейна.							
4	State-space модели распознавания речи. Inference и train треллисы. Жадное декодирование. Connectionist Temporal Classification (CTC) model. Неоднозначность отображения речевых единиц в текст. Необходимость специального “бланк”-символа. Треллисы с “бланк”-символом. Представление вероятности последовательности единиц речи. Функция потерь. Forward algorithm, backward algorithm, forward-backward algorithm. Мягкое выравнивание.	2	2				2	0
5	Контекстное моделирование при помощи языковых моделей. Проблемы жадного декодирования. Языковые модели, оценки качества - perplexity. N-gram, нейросетевое языковое моделирование. Beam Search decoding. Схема и алгоритм для CTC модели. Интеграция языковых моделей в префиксное декодирование.	2	2				2	0
6	Системы распознавания речи, основанные на механизмах внимания. Обусловливание языковых моделей на акустические признаки. Авторегрессионные энкодер-декодер модели с механизмом внимания. Схема. Декодер, его цели, схема, возможные реализации. Энкодер, его цели, схема, возможные реализации. Механизм внимания. Обучение и предсказание, функция потерь. Возможные проблемы такого моделирования и пути их решения.	4	4				4	0
7	Последние разработки в ASR. Masked Language Modelling. Semi-supervised learning. Noisy-Student training и Wav2Vec. Распознавание речи из аудиосигнала без промежуточного представления в виде спектрограммы или мелспектрограммы, путем скрытого представления модели.	2	2				2	0
8	Введение в синтез речи. Обсуждение задачи. Проблемы неопределенности “правильности” синтеза. Метрики качества (MOS, CrowdMOS, MUSHRA, SER, SBS, Robotness). Схема синтеза. Препроцессинг текста. Генерация аудио. Конкатенативные подходы: дифонный синтез и Unit selection. Параметрический синтез. Семинар с реализацией дифонного синтеза.	4	4				4	1
9	Вокодеры. Цели вокодеров. Авторегрессионные модели. WaveNet - нейросетевой вокодер. Схема, блоки сети. Mu-law embedding. Обусловливание на акустические признаки. Обучение и предсказание. Masked Autoregressive Flow (MAF). Вариационные автокодировщики. Semi-supervised training. Grokking.	4	4				4	0
10	Акустические модели. Скрытые марковские модели. Полносвязные сети. Рекуррентные сети. RNN with frame- and phoneme-wise subnetworks (upsampling models). Attention based сети: Char2Wav, Tacotron. Проблемы расходимости attention. Способы решения. Local-sensitive attention. Tacotron2. Upsampling + Attention: Fast Pitch, Fast Speech. Soft upsampling. Локальный attention.	4	4				4	1
11	Возможности акустических моделей. Какие типы информации содержатся в речи. Просодия.	2	2				2	0

	Моделирование просодии: вариационные автокодировщики, style tokens. Multi-speaker, multi-language синтез. Использование верификационных ASR моделей. Reversal gradient.							
	Текущий контроль	1						
	Промежуточная аттестация: зачет							
	Итого	33	32				33	3

Практические занятия (семинарские занятия) организуются, в том числе в форме практической подготовки, которая предусматривает участие обучающихся в выполнении отдельных семинарских и домашних работ, связанных с будущей профессиональной деятельностью.

На проведение практических занятий (семинарских и домашних работ) в форме практической подготовки отводится 16 часов.

Практическая подготовка направлена на формирование и развитие:

- практических навыков в соответствии с профилем ООП: Разработка, тестирование, оптимизация программного обеспечения (ПО).
- компетенций – ОПК-3: «Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования» на примере современных технологий обработки речи, в том числе с применением искусственного интеллекта.

Текущий контроль успеваемости реализуется в формах опросов на занятиях лабораторного типа. Промежуточная аттестация проходит в традиционных формах (зачет).

3. Учебно-методическое обеспечение самостоятельной работы обучающихся

А. Виды самостоятельной работы студентов

В качестве самостоятельной работы студенты выполняют задания семинарских и домашних работ, которые реализуют различные методы из курса технологий обработки речи:

1) Семинарские работы (по 10 баллов):

- a) 1-й семинар - знакомство со звуком, преобразование аудиосигнала до мелспектрограммы и обратно к аудиосигналу
- b) 2-й семинар - алгоритм Левенштейна, расстояние Левенштейна, визуализация преобразований из одного текста в другой.
- c) 3-й семинар - CTC forward-backward алгоритм, мягкое выравнивание, жадное декодирование.
- d) 4-й семинар - вариационный автокодировщик, доказательство формул функции потерь.

- е) 5-й семинар - local sensitive attention для tacotron2, в результате получаем готовую систему синтеза речи.

II) Большие домашние работы (по 20 баллов):

- а) Домашняя работа 1 - классификация цифр (Audio-mnist) с использованием нейронных сетей.
- б) Домашняя работа 2 - сравнение различных архитектур CTC моделей (DNN, RNN, BiRNN), в результате получаем готовую систему распознавания речи.

В. Образовательные материалы для самостоятельной работы студентов

- 1) Тампель И.Б., Карпов А.А. Автоматическое Распознавание Речи. Учебное пособие. – СПб: Университет ИТМО, 2016. – 138 с. (<https://books.ifmo.ru/file/pdf/1921.pdf>)
- 2) Кипяткова И.С., Ронжин А.Л., Карпов А.А., “Автоматическая обработка разговорной русской речи”. – СПб.: ГУАП, 2013. – 314 с.
- 3) Карпов А.А., “Реализация автоматической системы многомодального распознавания речи по аудио- и видеоинформации” // Автоматика и телемеханика. 2014, Т. 75, № 12, С. 125-138.
- 4) Schwarz P., "Phoneme recognition based on long temporal context", Ph.D. thesis, Brno University of Technology, 2008. <http://www.fit.vutbr.cz/~schwarzp/publi/thesis.pdf>
- 5) Холоденко А.Б., “О построении статистических языковых моделей для систем распознавания русской речи” // Интеллектуальные системы, 2002. Т.6. Вып. 1-4. С. 381-394.
- 6) Made in Future: Речевые технологии для создания новых ценностей в бизнесе/ Группа компаний ЦРТ. 2022 (<https://www.speechpro.ru/media/news/made-in-future-rechevye-tehnologii-dlya-sozdaniya-novyh-cennostej-v-biznese>)

5. Фонд оценочных средств для промежуточной аттестации по дисциплине, включающий:

5.1. Описание шкал оценивания результатов обучения по дисциплине

Уровень сформированности компетенций (индикатора достижения компетенций)	Шкала оценивания сформированности компетенций						
	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено		зачтено				

<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки.	Минимально допустимый уровень знаний. Допущено много негрубых ошибки.	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки.	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме.	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме.	Продemonстрированы все основные умения,. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие владения материалом. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки.	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов.	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов.	Продemonстрирован творческий подход к решению нестандартных задач

Шкала оценки при промежуточной аттестации

Оценка (баллы)		Уровень подготовки
	Превосходно (60 и более)	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»

зачтено (20 и более)	Отлично (50)	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
	Очень хорошо (45)	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
	Хорошо (40)	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «хорошо»
	Удовлетворительно (20-30)	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено (< 20)	Неудовлетворительно (10)	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно», ни одна из компетенций не сформирована на уровне «плохо»
	Плохо (0)	Хотя бы одна компетенция сформирована на уровне «плохо»

5.2 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения

5.2.1 Контрольные вопросы

Вопросы	Код формируемой компетенции
1.Введение в речевые технологии. История синтеза и распознавания. Какие существуют задачи в речевых технологиях. Как человек воспринимает звук. Сравнение с компьютерным зрением и обработкой естественных языков. Цифровая обработка сигналов. Как звук дискретизируется для компьютеров. Характеристики аудиосигналов. Представления для работы со звуком. Дискретное преобразование Фурье. Спектрограмма, мелспектрограмма, мелкепстральные коэффициенты. Восстановление аудиосигнала из спектрограммы: алгоритм Гриффина-Лима.	ОПК-3.1
2.Введение в распознавание речи. Обсуждение задачи распознавания. Сравнение различных представлений текста в качестве единиц речи. Проблема выравнивания единиц речи и акустических признаков: State-space models, Attention mechanism. Дискриминативная и генеративная постановки задачи распознавания. Метрики качества распознавания. Word error rate(WER). Расстояние Левенштейна и алгоритм Левенштейна.	ОПК-3.1

3.State-space модели распознавания речи. Inference и train треллисы. Жадное декодирование. Connectionist Temporal Classification (CTC) model. Неоднозначность отображения речевых единиц в текст. Необходимость специального “бланк”-символа. Треллисы с “бланк”-символом. Представление вероятности последовательности единиц речи. Функция потерь. Forward algorithm, backward algorithm, forward-backward algorithm. Мягкое выравнивание.	ОПК-3.1
4.Контекстное моделирование при помощи языковых моделей. Проблемы жадного декодирования. Языковые модели, оценки качества - perplexity. N-gram, нейросетевое языковое моделирование. Beam Search decoding. Схема и алгоритм для CTC модели. Интеграция языковых моделей в префиксное декодирование.	ОПК-3.1
5.Системы распознавания речи, основанные на механизмах внимания. Обусловливание языковых моделей на акустические признаки. Авторегрессионные энкодер-декодер модели с механизмом внимания. Схема. Декодер, его цели, схема, возможные реализации. Энкодер, его цели, схема, возможные реализации. Механизм внимания. Обучение и предсказание, функция потерь. Возможные проблемы такого моделирования и пути их решения.	ОПК-3.1
6.Последние разработки в ASR. Masked Language Modelling. Semi-supervised learning. Noisy-Student training и Wav2Vec. Распознавание речи из аудиосигнала без промежуточного представления в виде спектрограммы или мелспектрограммы, путем скрытого представления модели.	ОПК-3.1
7.Введение в синтез речи. Обсуждение задачи. Проблемы неопределенности “правильности” синтеза. Метрики качества (MOS, CrowdMOS, MUSHRA, SER, SBS, Robotness). Схема синтеза. Препроцессинг текста. Генерация аудио. Конкатенативные подходы: дифонный синтез и Unit selection. Параметрический синтез. Семинар с реализацией дифонного синтеза.	ОПК-3.1
8.Вокодеры. Цели вокодеров. Авторегрессионные модели. WaveNet - нейросетевой вокодер. Схема, блоки сети. Mu-law embedding. Обусловливание на акустические признаки. Обучение и предсказание. Masked Autoregressive Flow (MAF). Вариационные автокодировщики. Semi-supervised training. Grokking.	ОПК-3.1
9.Акустические модели. Скрытые марковские модели. Полносвязные сети. Рекуррентные сети. RNN with frame- and phoneme-wise subnetworks (upsampling models). Attention based сети: Char2Wav, Tacotron. Проблемы расходимости attention. Способы решения. Local-sensitive attention. Tacotron2. Upsampling + Attention: Fast Pitch, Fast Speech. Soft upsampling. Локальный attention.	ОПК-3.1
10.Возможности акустических моделей. Какие типы информации содержатся в речи. Просодия. Моделирование просодии: вариационные автокодировщики, style tokens. Multi-speaker, multi-language синтез. Использование верификационных ASR моделей. Reversal gradient.	ОПК-3.1

5.2.2 Задачи (семинарские и домашние работы) для контроля компетенции ОПК-3 в целом

Задача/Задание	Баллов	Комп-я
I. Семинарские работы (с выступлением на семинаре):		
a) 1-й семинар - знакомство со звуком, преобразование аудиосигнала до мелспектрограммы и обратно к аудиосигналу.	10	ОПК-3.1 ОПК-3.2
b) 2-й семинар - алгоритм Левенштейна, расстояние Левенштейна, визуализация преобразований из одного текста в другой.	10	ОПК-3.1 ОПК-3.2
c) 3-й семинар - CTC forward-backward алгоритм, мягкое выравнивание, жадное декодирование.	10	ОПК-3.1 ОПК-3.2
d) 4-й семинар - вариационный автокодировщик, доказательство формул функции потерь.	10	ОПК-3.1 ОПК-3.2
e) 5-й семинар - local sensitive attention для tacotron2, в результате получаем готовую систему синтеза речи.	10	ОПК-3.1 ОПК-3.2
II. Большие домашние работы:		

а) Домашняя работа 1 - классификация цифр (Audio-mnist) с использованием нейронных сетей.	20	ОПК-3.2
б) Домашняя работа 2 - сравнение различных архитектур СТС моделей (DNN, RNN, BiRNN), в результате получаем готовую систему распознавания речи.	20	ОПК-3.2

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

1) основная литература:

1. Тампель И.Б., Карпов А.А. Автоматическое Распознавание Речи. Учебное пособие. – СПб: Университет ИТМО, 2016. – 138 с. (<https://books.ifmo.ru/file/pdf/1921.pdf>)
2. Кипяткова И.С., Ронжин А.Л., Карпов А.А., “Автоматическая обработка разговорной русской речи”. – СПб.: ГУАП, 2013. – 314 с.

2) дополнительная литература:

1. Карпов А.А., “Реализация автоматической системы многомодального распознавания речи по аудио- и видеоинформации” // Автоматика и телемеханика. 2014, Т. 75, № 12, С. 125-138.
2. Schwarz P., "Phoneme recognition based on long temporal context", Ph.D. thesis, Brno University of Technology, 2008. <http://www.fit.vutbr.cz/~schwarzp/publi/thesis.pdf>
3. Холоденко А.Б., “О построении статистических языковых моделей для систем распознавания русской речи” // Интеллектуальные системы, 2002. Т.6. Вып. 1-4. С. 381-394.

3) Интернет-ресурсы:

1. Made in Future: Речевые технологии для создания новых ценностей в бизнесе/ Группа компаний ЦРТ. 2022 (<https://www.speechpro.ru/media/news/made-in-future-rechevye-tehnologii-dlya-sozdaniya-novyh-cennostej-v-biznese/>)
2. Технологии распознавания речи в здравоохранении. Проект (<https://tele-med.ai/proekty/tehnologii-raspoznavaniya-rechi-v-zdravooohranenii/>)
3. Нестор.BRIEF. Система протоколирования совещаний. Проект (<https://www.speechpro.ru/product/sistemy-audio-i-videoprotokolirovaniya/nestor>)

4) Программное обеспечение:

1. MS Windows установленная на компьютере обучающегося
2. MS Visual Studio Community 2017 – бесплатная версия.
3. Установка языка Python [<http://www.python.org/>].
4. Библиотека автоматизации GUI тестирования pywinauto [<http://pywinauto.github.io/>]
5. ПО визуализации фильтров и выходов слоев в Caffe [<http://nbviewer.jupyter.org/github/BVLC/caffe/blob/master/examples/00-classification.ipynb>].
6. ПО визуализации фильтров и выходов слоев в Torch [<https://github.com/facebook/iTorch>].

7. Материально-техническое обеспечение дисциплины (модуля)

Имеются в наличии учебные аудитории для проведения занятий лекционного типа, занятий

семинарского типа, промежуточной аттестации, а также помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет». Учебная и научная литература, учебно-методические материалы, представленные в библиотечном фонде, в электронных библиотеках и на кафедре математического обеспечения и суперкомпьютерных технологий.

При занятиях в компьютерном классе установлены: операционная система Windows (лицензия по подписке Microsoft Imagine), Microsoft Visual Studio 2013 (лицензия по подписке Microsoft Imagine), библиотека OpenCV (open source, <http://opencv.org/>).

Программа составлена в соответствии с требованиями ОС ВО ННГУ с учетом рекомендаций ФГОС ВО по направлению 02.04.02. – Фундаментальная информатика и информационные технологии.

Автор: д.т.н., проф. В.Е. Турлапов,

Зам. зав. кафедрой МОСТ И.Б.Мееров

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики от 30 ноября 2022 года, протокол № 3.