

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION

**Federal State Autonomous Educational Institution of Higher Education
«National Research Lobachevsky State University of Nizhny Novgorod»**

Институт информационных технологий, математики и механики

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Working programme of the discipline

Speech processing technologies

Higher education level

Master degree

Area of study / speciality

02.04.02 - Fundamental Informatics and Information Technology

Focus /specialization of the study programme

Artificial Intelligence and Data Analysis

Mode of study

full-time

Nizhny Novgorod

Year of commencement of studies 2024

1. Место дисциплины в структуре ОПОП

Дисциплина ФТД.02 Технологии обработки речи является факультативом в образовательной программе.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

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	Для текущего контроля успеваемости	Для промежуточной аттестации
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 ОПК-3.2: Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем ОПК-3.3: Имеет практический опыт применения разработки программного обеспечения и тестирования программных продуктов	ОПК-3.1: Знать теоретические основы методов и алгоритмов технологий обработки речи и их особенности в системах искусственного интеллекта (ИИ). ОПК-3.2: 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. ОПК-3.3: Имеет практический опыт применения и разработки программного обеспечения обработки речи	Задачи	Зачёт: Контрольные вопросы

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	1
Часов по учебному плану	36

в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	32
- занятия семинарского типа (практические занятия / лабораторные работы)	0
- КСР	1
самостоятельная работа	3
Промежуточная аттестация	0 Зачёт

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Введение в речевые технологии.	2	2		2	
Цифровая обработка сигналов.	2	2		2	
Введение в распознавание речи	5	4		4	1
State-space модели распознавания речи.	2	2		2	
Контекстное моделирование при помощи языковых моделей.	2	2		2	
Системы распознавания речи, основанные на механизмах внимания.	4	4		4	
Последние разработки в ASR.	2	2		2	
Введение в синтез речи.	4	4		4	
Вокодеры.	5	4		4	1
Акустические модели.	4	4		4	
Возможности акустических моделей.	3	2		2	1
Аттестация	0				
КСР	1			1	
Итого	36	32	0	33	3

Contents of sections and topics of the discipline

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. 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 Гриффины-Лима.

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 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.

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.

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.

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.

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.

9. Вокодеры.

Цели вокодеров. Авторегрессионные модели. WaveNet - нейросетевой вокодер. Схема, блоки сети. Multi-scale embedding. Обусловливание на акустические признаки. Обучение и предсказание. Masked Autoregressive Flow (MAF). Вариационные автокодировщики. Semi-supervised training. Grokking.

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.

11. Возможности акустических моделей.

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. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

- 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. Assessment tools for ongoing monitoring of learning progress and interim certification in the discipline (module)

5.1 Model assignments required for assessment of learning outcomes during the ongoing monitoring of learning progress with the criteria for their assessment:

5.1.1 Model assignments (assessment tool - Tasks) to assess the development of the competency ОПК-3:

1. Семинарские работы (с выступлением на семинаре):

- а) 1-й семинар - знакомство со звуком, преобразование аудиосигнала до мелспектрограммы и обратно к аудиосигналу.
- б) 2-й семинар - алгоритм Левенштейна, расстояние Левенштейна, визуализация преобразований из одного текста в другой.
- с) 3-й семинар - CTC forward-backward алгоритм, мягкое выравнивание, жадное декодирование.
- д) 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.

2. Большие домашние работы:

- 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.

Assessment criteria (assessment tool — Tasks)

Grade	Assessment criteria
pass	Выполнены все или большая часть этапов решения задачи или задача решена с незначительными недочетами. Результаты работы представлены преподавателю в срок.
fail	Выполнены не все задания или выполнены не в полном объеме (представлено не полное описание этапов выполнения заданий, получен неверный ответ, результаты работы не представлены преподавателю).

5.2. Description of scales for assessing learning outcomes in the discipline during interim certification

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения 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 ошибок	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. Ошибок нет.	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 базовых навыков. Невозможность оценить наличие	При решении стандартных задач не продемонстрированы базовые	Имеется минимальный набор навыков для	Продemonстрированы базовые навыки при решении	Продemonстрированы базовые навыки при решении	Продemonстрированы навыки при решении	Продemonстрирован творческий подход к решению

	навыков вследствие отказа обучающегося от ответа	навыки. Имели место грубые ошибки	решения стандартных задач с некоторым и недочетами	стандартных задач с некоторым и недочетами	стандартных задач без ошибок и недочетов	нестандартных задач без ошибок и недочетов	нестандартных задач
--	--	-----------------------------------	--	--	--	--	---------------------

Scale of assessment for interim certification

Grade		Assessment criteria
pass	outstanding	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "outstanding", the knowledge and skills for the relevant competencies have been demonstrated at a level higher than the one set out in the programme.
	excellent	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "excellent",
	very good	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "very good",
	good	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "good",
	satisfactory	All the competencies (parts of competencies) to be developed within the discipline have been developed at a level no lower than "satisfactory", with at least one competency developed at the "satisfactory" level.
fail	unsatisfactory	At least one competency has been developed at the "unsatisfactory" level.
	poor	At least one competency has been developed at the "poor" level.

5.3 Model control assignments or other materials required to assess learning outcomes during the interim certification with the criteria for their assessment:

5.3.1 Model assignments (assessment tool - Control questions) to assess the development of the competency ОПК-3

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, мелкеспектральные коэффициенты. Восстановление аудиосигнала из спектрограммы: алгоритм Гриффина-Лима.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Assessment criteria (assessment tool — Control questions)

Grade	Assessment criteria
pass	Студент ответил на большую часть вопросов возможно с незначительными недочетами.
fail	При ответе студент допускает грубые ошибки в основном материале и решении стандартных задач.

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

Основная литература:

1. Карпов А. А. Проектирование речевых интерфейсов для информационно-управляющих систем : учеб. пособие / Карпов А. А., Кипяткова И. С., Ронжин А. Л. - Санкт-Петербург : СПб ФИЦ РАН, 2012. - 76 с. - Книга из коллекции СПб ФИЦ РАН - Инженерно-технические науки. - ISBN 978-5-8088-0698-6., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=864765&idb=0>.

Дополнительная литература:

1. Бессмертный И. А. Интеллектуальные системы : учебник и практикум / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. - Москва : Юрайт, 2023. - 243 с. - (Высшее образование). - ISBN 978-5-534-01042-8. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=849292&idb=0>.
2. Чернышев С. А. Основы программирования на Python : учебное пособие / С. А. Чернышев. - 2-е изд. ; пер. и доп. - Москва : Юрайт, 2023. - 349 с. - (Высшее образование). - ISBN 978-5-534-17139-6. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=870851&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

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-zdravoohranenii>)
3. Нестор.BRIEF. Система протоколирования совещаний. Проект (<https://www.speechpro.ru/product/sistemy-audio-i-videoprotokolirovaniya/nestor>)

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)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 02.04.02 - Fundamental Informatics and Information Technology.

Author(s): Турлапов Вадим Евгеньевич, доктор технических наук, доцент.

Заведующий кафедрой: Мееров Иосиф Борисович, кандидат технических наук.

Программа одобрена на заседании методической комиссии от 13.12.2023, протокол № 3.