

Московский государственный технический университет имени Н. Э. Баумана
(национальный исследовательский университет)

Выпускная квалификационная работа бакалавра

«Метод распознавания эмоций по звучащей речи на основе скрытой марковской модели»

Студент: Казаева Татьяна Алексеевна ИУ7-86Б

Научный руководитель: Строганов Юрий Владимирович

Москва, 2023 г.

Цель и задачи работы

Цель - разработать метод определения эмоций по звучащей речи на основе скрытой марковской модели

Задачи:

- проанализировать существующие эмоциональные корпуса и выбрать наиболее подходящий для обучения классификатора
- проанализировать информативные признаки, характеризующие эмоциональную речь
- проанализировать классификаторы, чаще всего используемые в анализе речевых эмоций
- спроектировать и реализовать метод детектирования эмоций
- рассчитать качественные характеристики классификатора

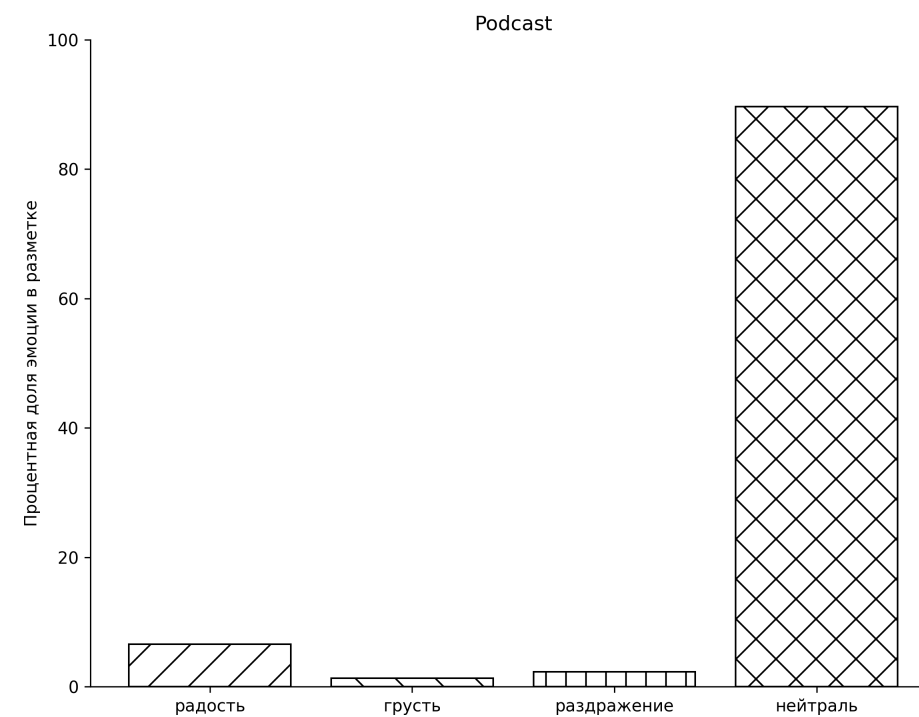
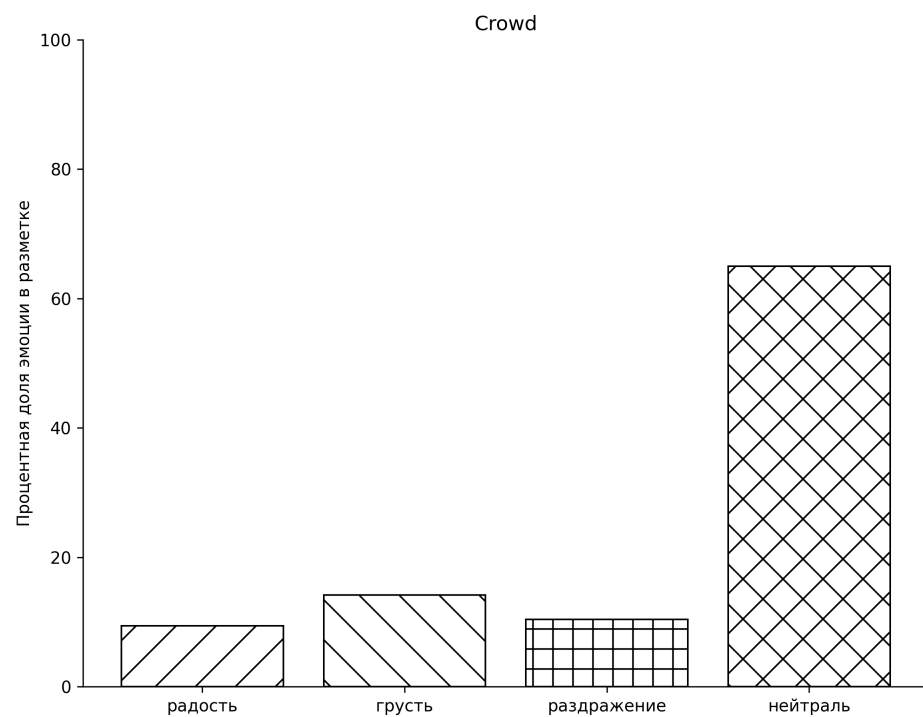
Определение эмоций

Подход	Основа подхода	Категории	Примеры решений
Дискретный	Выделение базовых эмоций	Некоторый набор дискретных эмоций	Affectiva, RAVDESS, SAVEE, EmoDB...
Многомерный	Координатное многомерное пространство	Валентность, активация, интенсивность	RECOLA, колесо эмоций Плутчика
Гибридный	Комбинация дискретного и многомерного подходов	Базовые эмоции и уровни силы эмоций	"Песочные часы эмоций"

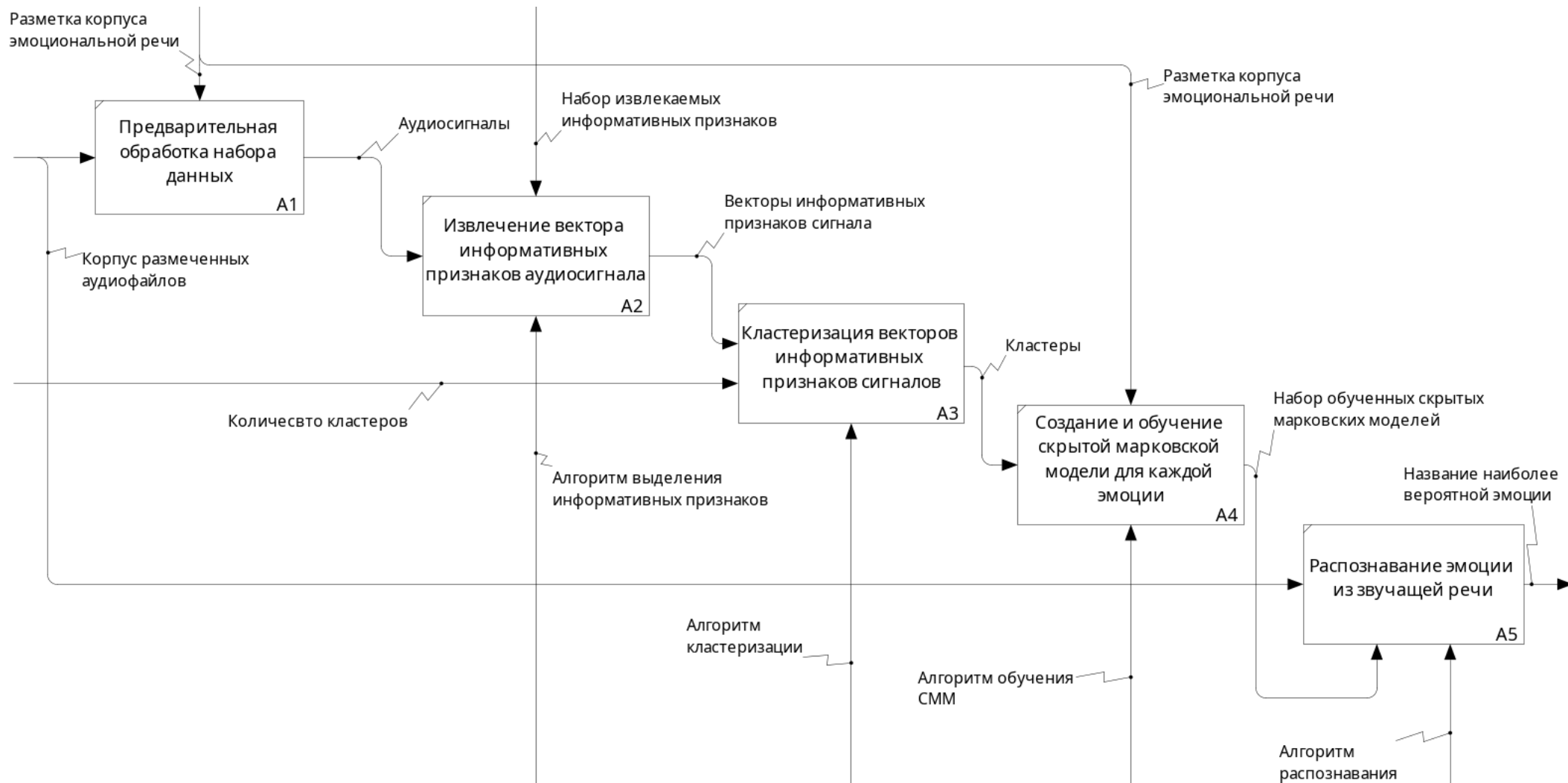
Корпуса звучащей речи

Название	Количество эмоций	Количество голосов		Лексикон	Публичный	Поддержка русского языка
		М	Ж			
RAVDESS	7	12	12	2 предл.	да	нет
SAVEE	6	4	0	15 предл.	да	нет
Emo-DB	6	5	5	10 предл.	да	нет
TESS	7	0	2	200 слов	да	нет
RUSLANA	4	12	49	10 предл.	нет	да
DUSHA	4	-	-	обширный	да	да
REC	-	-	-	обширный	нет	да

Распределение классов разметки корпуса DUSHA



Предлагаемый метод



Обучающий набор данных, составленный из корпуса DUSHA

В обучающий набор было включено 1500 аудиофайлов каждого класса разметки.

Подгруппа	Всего	Тренировочная выборка	Тестовая выборка
раздражение	1 ч. 02 мин. 47 сек.	50 мин. 05 сек.	12 мин. 41 сек.
нейтраль	1 ч. 02 мин. 23 сек.	50 мин. 02 сек.	12 мин. 20 сек.
радость	1 ч. 02 мин. 01 сек.	50 мин. 32 сек.	12 мин. 29 сек.
грусть	1 ч. 02 мин. 53 сек.	51 мин. 57 сек.	12 мин. 55 сек.

Шумочистка к аудиофайлам не применялась.

Просодические признаки речи

Признаки оцениваются в баллах: 1 – низший балл, 3 – высший балл.

	Устойчивость к шуму	Информативность	Емкость представления
Частота основного тона	3	1	1
Интенсивность	3	2	3
Темп речи	3	3	3
Паузация	1	1	3

Спектральные признаки речи

Признаки оцениваются в баллах: 1 – низший балл, 3 – высший балл.

	Устойчивость к шуму	Информативность	Емкость представления
Мел-кепстральные коэффициенты	2	3	3
Частоты первых четырех формант	2	3	2
Джиттер, шиммер	1	1	1

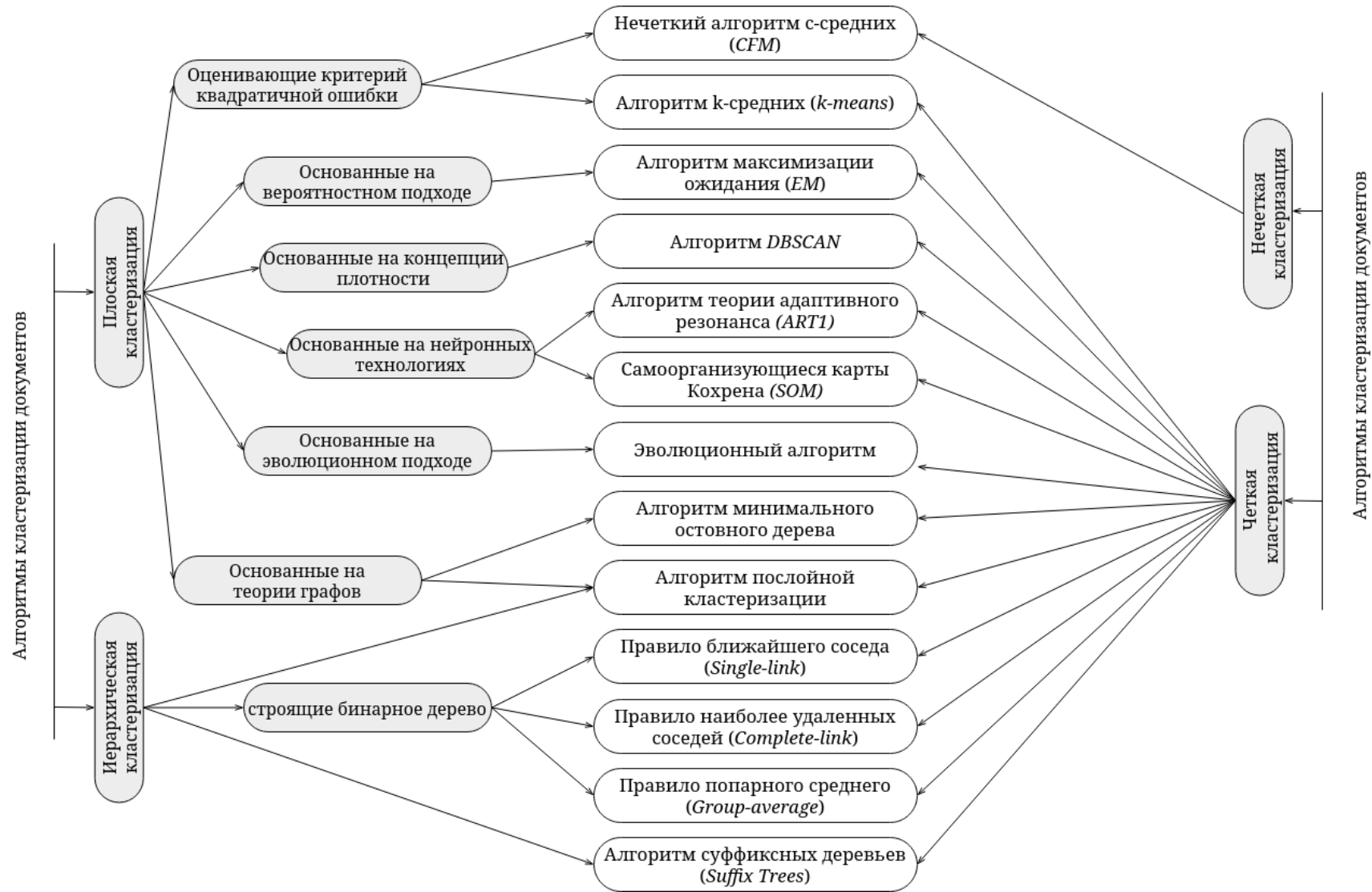
Мел-кепстральные коэффициенты

Мел-кепстральный коэффициент под номером n вычисляется согласно:

$$c_j(n) = \sum_{m=0}^{M-1} T_j(m) \cos \left(\pi n \left(m + \frac{1}{2} \right) / M \right), \quad 0 \leq n < M,$$

где M - размерность банка мел-фильтров из треугольных фильтров, $T_j(m)$ - логарифмическое значение энергии компонент спектра на выходе мел-фильтра m , j - номер кадра аудиозаписи.

Классификация алгоритмов кластеризации



Классификаторы, наиболее часто используемые в аффективных вычислениях

Скрытая марковская модель (СММ)

- используется для моделирования *последовательностей данных*
- данные преобразуются в *последовательность наблюдений*

Искусственная нейронная сеть (ИНС)

- состоят из соединенных и взаимодействующих искусственных нейронов
- данные передаются через слои искусственных нейронов

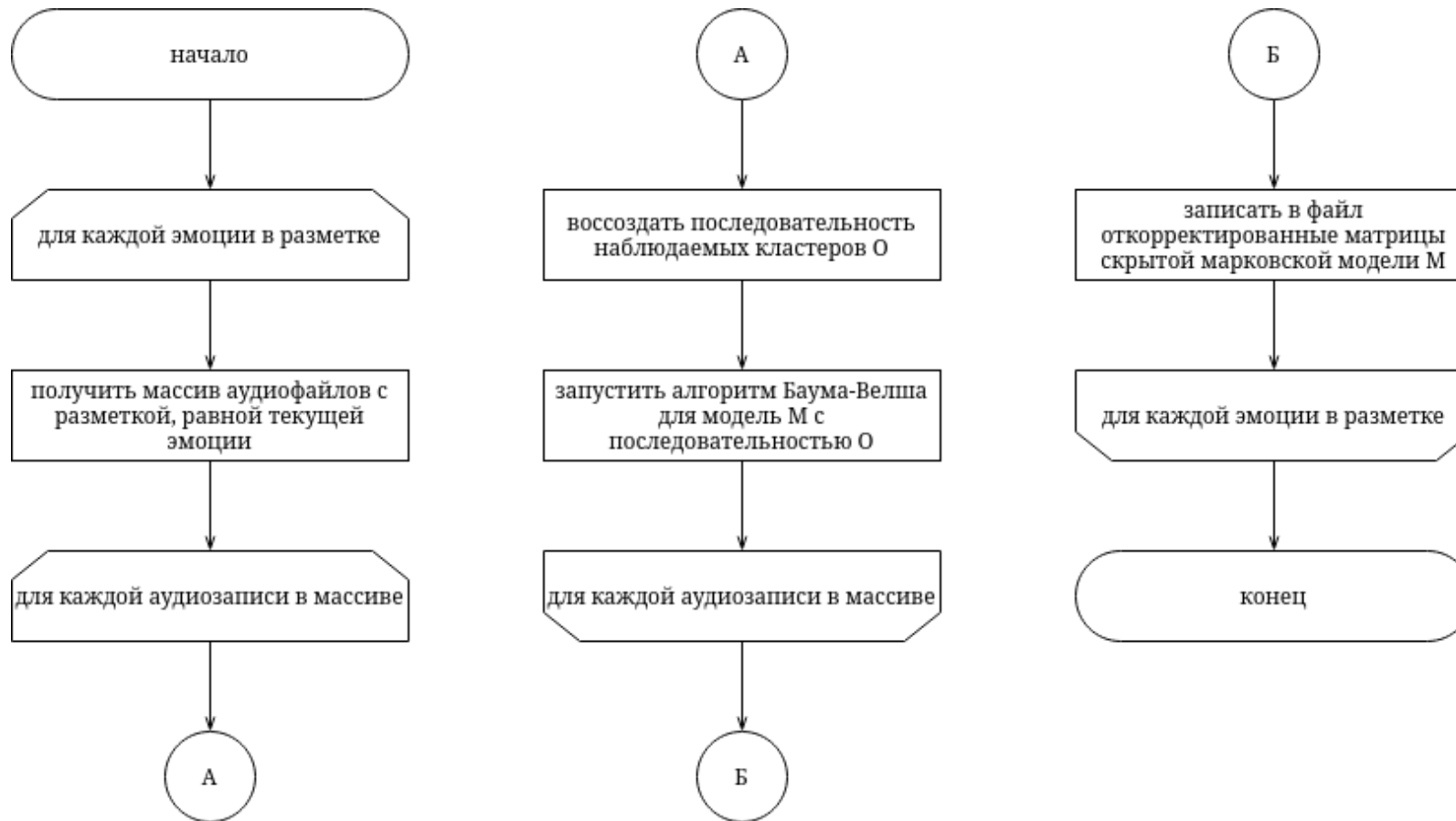
Скрытая марковская модель

Можно описать как двойной стохастический процесс:

проявление эмоции –
скрытый стохастический процесс, который
невозможно наблюдать напрямую

**наблюдаемый набор мел-кепстральных
коэффициентов** –
процесс, который создает
последовательность наблюдений

Обучение скрытой марковской модели



Распознавание эмоций в речи

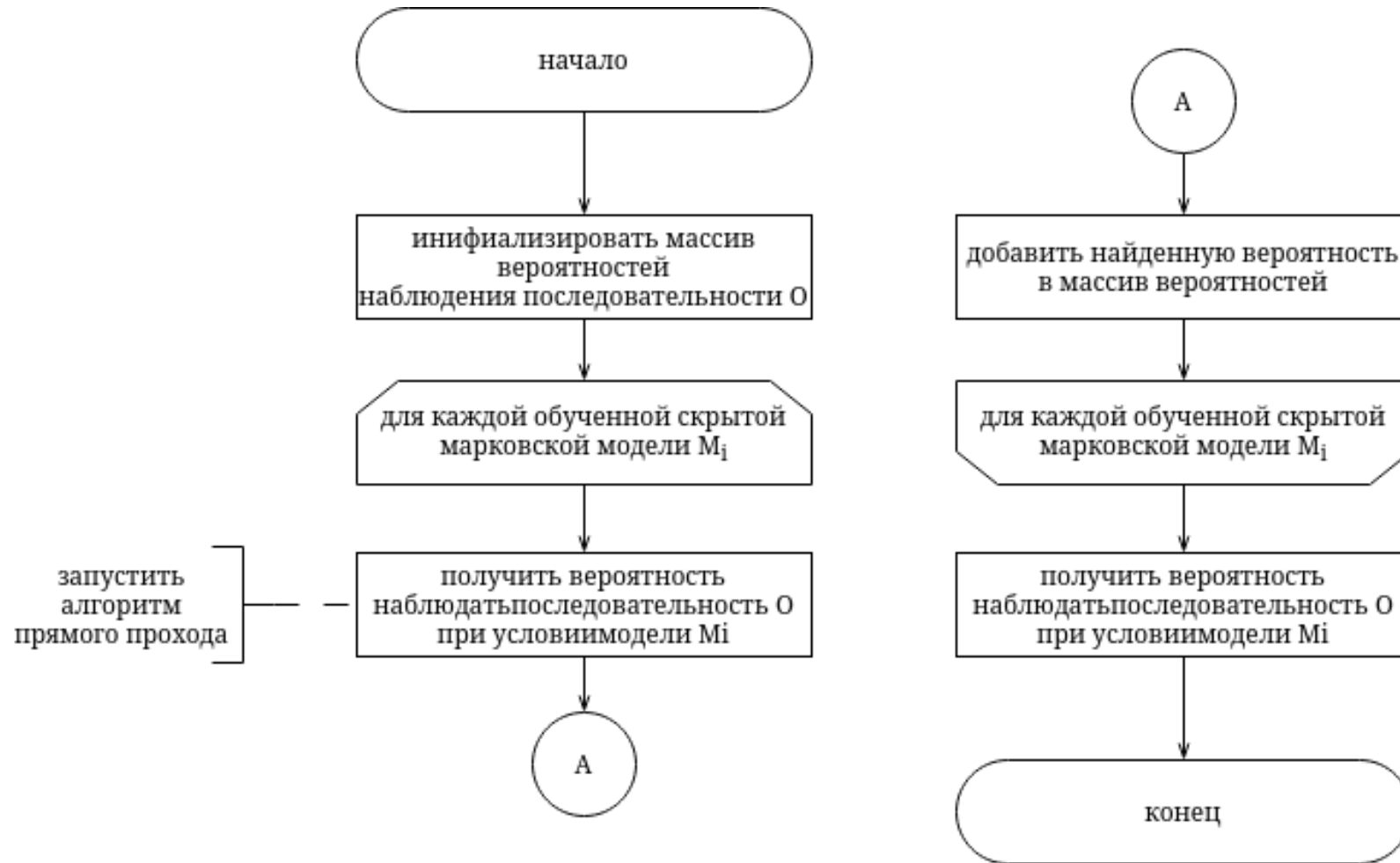
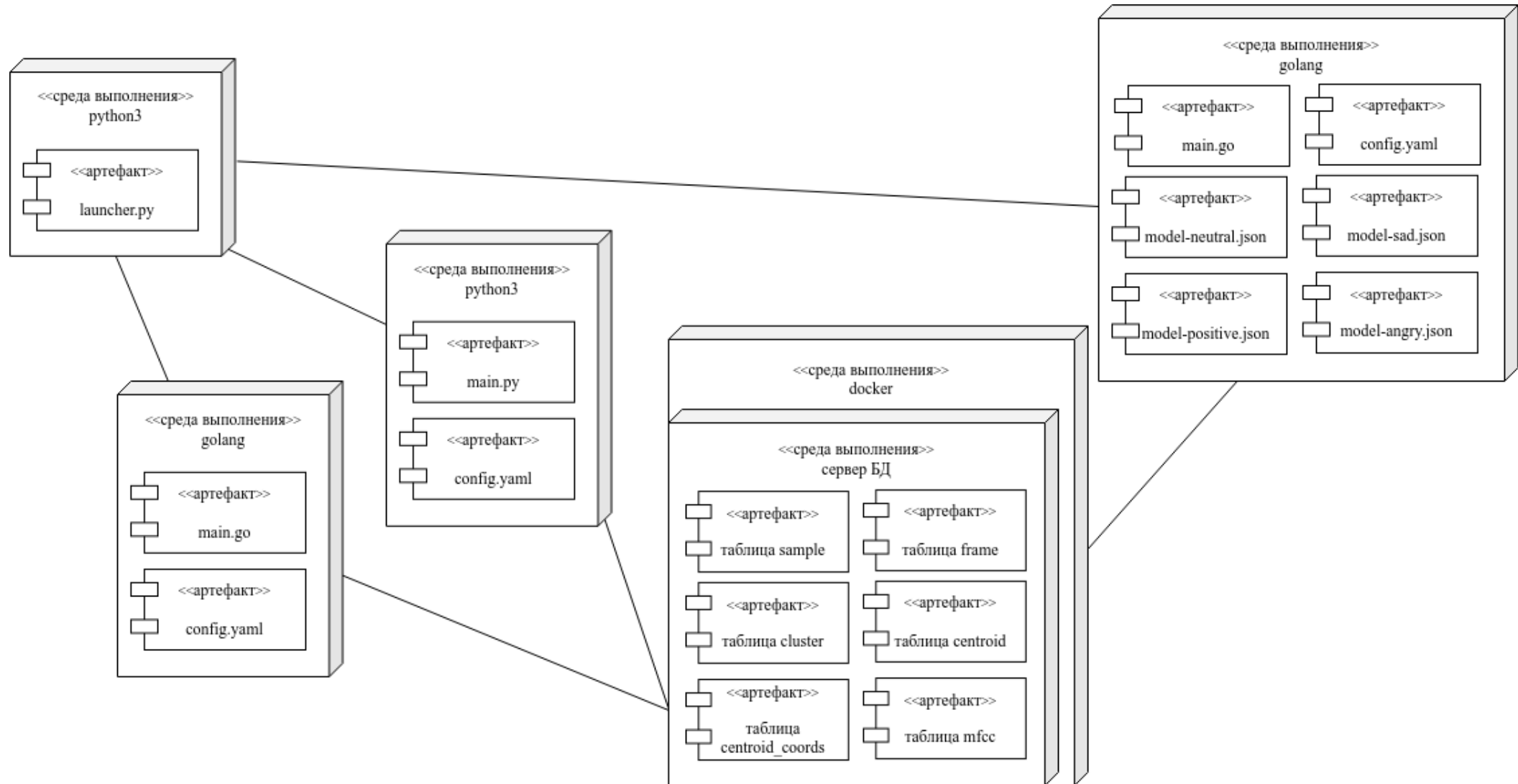


Диаграмма развертывания



Результат классификации на тренировочной выборке

Матрица неточностей для классификации на тренировочной выборке.

экспертная оценка	оценка классификатора			
	злость	нейтраль	радость	грусть
злость	534	74	427	165
нейтраль	210	142	540	308
радость	369	72	603	156
грусть	333	88	416	363

Оценка результата классификации (1/2)

"нейтраль"

экспертная оценка

оценка классификатора

Полож.	Отриц.	
142	234	Полож.
1058	3366	Отриц.

$$\text{Precision}_{\text{нейтраль}} = 38\%$$

$$\text{Recall}_{\text{нейтраль}} = 12\%$$

$$F_{\text{нейтраль}} = 18\%$$

"грусть"

экспертная оценка

оценка классификатора

Полож.	Отриц.	
363	629	Полож.
837	2971	Отриц.

$$\text{Precision}_{\text{грусть}} = 36\%$$

$$\text{Recall}_{\text{грусть}} = 30\%$$

$$F_{\text{грусть}} = 33\%$$

Оценка результата классификации (2/2)

"злость"

оценка классификатора

экспертная оценка

Полож.	Отриц.	
534	666	Полож.
912	2688	Отриц.

$$\text{Accuracy}_{\text{злость}} = 67\%$$

$$\text{Precision}_{\text{злость}} = 36\%$$

$$\text{Recall}_{\text{злость}} = 44\%$$

$$F_{\text{злость}} = 40\%$$

"радость"

оценка классификатора

экспертная оценка

Полож.	Отриц.	
603	1383	Полож.
597	2217	Отриц.

$$\text{Precision}_{\text{радость}} = 30\%$$

$$\text{Recall}_{\text{радость}} = 50\%$$

$$F_{\text{радость}} = 37\%$$

Выводы

На выборке из **6000 элементов** с разнномерным распределением классов:

Точность	злость	нейтраль	радость	грусть	Σ
	67%	73%	58%	69%	67%

общая точность $\approx 67\%$

Класс, распознанный *наиболее* точно - "нейтраль" ($\approx 73\%$),
наименее точно - "радость" ($\approx 58\%$)

Заключение

Цель работы достигнута: был разработан и реализован метод распознавания эмоций по звучащей речи. Все поставленные задачи были выполнены:

- проанализированы русскоязычные и иностранные корпуса эмоциональной речи, для обучения классификатора был выбран корпус **DUSHA**
- проанализированы признаки, характеризующие эмоцию в речи, для классификации были использованы **мел-кепстральные коэффициенты**
- проведен обзор классификаторов, используемых в анализе речевых эмоций
- спроектирован и реализован метод детектирования эмоций
- с помощью качественных метрик (*F-мера, точность, полнота*) **оценен результат** классификации.

Дальнейшее развитие

сбор собственного корпуса звучащей речи,
содержащего аудиозаписи **студийного**
качества, озвученные профессиональными
актерами

расширение объема информации в разметке:
учет **интонационного контура** (ИК) для
каждой аудиозаписи