



Факультет гуманитарных наук

Фундаментальная и
прикладная лингвистика

Нижний Новгород

Создание корпуса устной речи для диагностических целей с использованием автоматических алгоритмов распознавания речи

Докладчик: Жилина Полина Павловна 21ФиПЛ-2

Научный руководитель: к. филол. н., старший научный сотрудник Центра языка и мозга,
Хоменко Анна Юрьевна

Соруководитель: приглашенный преподаватель, Бадасян Александра Арсеновна



Актуальность

- Устная речь – источник когнитивной и психоэмоциональной информации.
- Точная транскрипция – основа лингвистического и диагностического анализа.
- Автоматическое распознавание речи (Automatic Speech Recognition, ASR) позволяет масштабировать обработку звучащей речи.
- Аннотированные корпуса – ключевой ресурс для анализа устной речи в том числе в диагностических и исследовательских целях.



Новизна

- В отличие от большинства русскоязычных работ, ограничивающихся использованием различных версий модели Whisper для автоматического распознавания речи (Мамаев, 2023; Sherstinova и др., 2024), **в настоящем исследовании сравнивается эффективность трёх современных ASR-моделей — Whisper, FastConformer и GigaAM**, две из которых пока редко применяются в лингвистических исследованиях, но при этом демонстрируют высокую точность распознавания русской устной речи на различных наборах данных.
- Проведён **лингвистически ориентированный анализ ошибок** распознавания с их **классификацией** и **выводами о пригодности моделей** для создания транскриптов в корпусах, ориентированных на диагностику речи.



Объект:

устная речь, содержащая достоверную и недостоверную информацию.

Предмет:

- особенности устной речи, которые либо не поддаются фиксации средствами автоматического распознавания речи (ASR), либо воспроизводятся неверно, приводя к ошибкам в транскриптах.
- особенности дискурсивной разметки автоматически созданных транскриптов.

Цель работы:

создание корпуса устной речи для диагностических целей, включающего как монологическую, так и диалогическую речь, содержащую как правдивые так и ложные высказывания, с применением современных алгоритмов автоматического распознавания речи



Задачи:

- проанализировать научную литературу по теме исследования;
- получить и сравнить автоматические транскрипции, сгенерированные выбранными ASR-моделями (Whisper, FastConformer, GigaAM-RNNT), с эталонными ручными расшифровками, подготовленными для анализа ошибок распознавания;
- классифицировать ошибки с точки зрения их лингвистической и диагностической значимости;
- выявить наиболее эффективный алгоритм распознавания речи для создания последующих транскриптов в корпусе;
- разработать протокол разметки транскриптов, учитывая дискурсивные и просодические характеристики речи;
- сформировать структуру корпуса и подготовить его к размещению в виде открытого репозитория.



Методы:

- **ASR-распознавание:** автоматическая транскрипция с использованием моделей на архитектурах Transformer (Whisper), FastConformer и RNN-T (GigaAM).
- **Ручная транскрипция:** создание эталонных текстов для последующего сравнения.
- **Сравнительный анализ:** сопоставление транскриптов, классификация и локализация ошибок.
- **Дискурсивная разметка:** разработка протокола аннотирования и структуры корпуса.



Материал исследования:

- **60 видеофонограмм (20 человек);**
- **возрастная группа: от 17 до 55+ лет;**
- **общая продолжительность материала составляет 231 минута 37 секунд;**
- В основу корпуса легли устные рассказы респондентов, собранные в три этапа:
 - **Этап 1.** Ложный рассказ:
 - Участник по заданному плану рассказывал о вымышленном событии, заведомо зная, что должен лгать; после монолога интервьюер задавал уточняющие вопросы.
 - **Этап 2.** Правдивый рассказ:
 - Респондент описывал реальное событие, также с опорой на план. Вопросы задавались после рассказа.
 - **Этап 3.** Свободный выбор:
 - Участник сам решал, будет ли его рассказ правдивым или вымышленным. Интервьюер также задавал уточняющие вопросы.

Продолжительность высказываний не регламентировалась; записи сохранялись в аудиоформате .mp3



Результаты

Результаты автоматической оценки качества транскриптов, созданных тремя ASR-алгоритмами в нормальных условиях (без зашумлений):

метрика (%) / ASR - модель	<i>nemo-fastconformer-ru-rnnt</i>	<i>openai/whisper-large-v3-turbo</i>	<i>gigaam-v2-rnnt</i>
WER (Word Error Rate)	11.09	16.46	6.18
CER (Character Error Rat)	4.4	9.94	2.62
MER (Match Error Rate)	10.94	16.76	6.10
WIL (Word Information Lost)	17.42	20.84	9.54
WIP (Word Information Preserved)	82.58	79.16	90.46

Таблица 1. Сравнение метрик качества автоматического распознавания речи для трёх моделей ASR в условиях без зашумлений

Оценка производилась на аудиозаписях из 12 аудиофайлов, в которых общая длительность монологической речи составила 31 минуту 21 секунду (15% от общего кол-ва незашумленного материала)



Результаты

Результаты автоматической оценки качества транскриптов, созданных тремя ASR-алгоритмами (зашумленные условия):

метрика (%) / ASR - модель	<i>nemo-fastconformer-ru-rnnt</i>	<i>openai/whisper-large-v3-turbo</i>	<i>gigaam-v2-rnnt</i>
WER (Word Error Rate)	22.48	24.77	13.18
CER (Character Error Rat)	9.11	16.63	6.28
MER (Match Error Rate)	22.25	24.48	12.94
WIL (Word Information Lost)	34.47	32.11	19.53
WIP (Word Information Preserved)	65.53	67.89	80.47

Таблица 2. Сравнение метрик качества автоматического распознавания речи для трёх моделей ASR в условиях с зашумлениями

Оценка проводилась на аудиозаписях из 6 аудиофайлов, в которых общая длительность монологической речи составила 24 минуты 16 секунд. Эта длительность соответствует всей монологической зашумленной речи, представленной в данном наборе данных.



Результаты (ошибки замен)

- **Ошибки, не искажающие смысл**

- Затрагивают лишь формальные грамматические характеристики слова (например, род, число, падеж у существительных и прилагательных):

эталонный транскрипт гипотеза	отвезти другой моей подруге отвезти другой моей подруги
----------------------------------	--

- **Ошибки, искажающие смысл.** В результате появляется полностью другая лексема:

- не связанная с контекстом, но схожую по звуковому облику:

эталонный транскрипт гипотеза	мне дают бланк я заполняю этот бланк не дают бланк я заполняю этот банк
----------------------------------	--

- несуществующий звуковой аналог:

эталонный транскрипт гипотеза	не было толкучки никто на пятки не наступал не было толпучки никто на пятки не наступал
----------------------------------	--



Результаты (ошибки замен)

Модель распознавания речи/вид ошибки	Замены, не влияющие на смысл лексемы (%)	Замены, искажающие смысл лексемы (%)	Замена на несуществующий звуковой аналог (%)	Замена на существующую лексему, не связанную с контекстом, но схожую по звуковому облику (%)
<i>whisper-large-v3-turbo</i>	40%	60%	28%	72%
<i>nemo-fastconformer-ru-rnnt</i>	49%	51%	31%	69%
<i>gigaam-v2-rnnt</i>	33%	67%	13%	87%

Таблица 3. Распределение ошибок замены в транскриптах моделей автоматического распознавания речи (в условиях отсутствия фоновых шумов)



Результаты (ошибки замен)

Модель распознавания речи/вид ошибки	Замены, не влияющие на смысл лексемы (%)	Замены, искажающие смысл лексемы (%)	Замена на несуществующий звуковой аналог (%)	Замена на существующую лексику, не связанную с контекстом, но схожую по звуковому облику (%)
<i>whisper-large-v3-turbo</i>	52%	48%	24%	76%
<i>nemo-fastconformer-ru-rnnt</i>	56%	44%	38%	62%
<i>gigaam-v2-rnnt</i>	40%	60%	25%	75%

Таблица 4. Распределение ошибок замены в транскриптах моделей автоматического распознавания речи (в условиях фоновых шумов)



Результаты (ошибки пропусков)

Модель распознавания речи/вид ошибки	Служебные слова (предлоги, союзы, частицы)	Знаменательные слова	Целые синтаксические конструкции (клаузы)	Хезитационные паузы	Самоперебиван ия
<i>whisper-large-v3-turbo</i>	✓	✓	✓	✓	✓
<i>nemo-fastconformer-ru- rnnt</i>	✓	✓ (только личные местоимения)	х	✓	✓
<i>gigaam-v2-rnnt</i>	✓	✓ (только личные местоимения)	х	✓	✓

Таблица 5. Распределение видов пропусков по моделям автоматического распознавания речи

Промежуточные итоги

- Модель ***gigaam-v2-rnnt*** продемонстрировала **наилучшие результаты** среди протестированных систем и была выбрана для финальной обработки корпуса, включая ручную валидацию и аннотирование транскриптов.
- **Автоматическая транскрипция рассматривается как предварительный этап:** современные ASR-модели, не обученные на специализированных корпусах, не фиксируют паралингвистические элементы, критически важные для интерпретации высказываний.
- Наиболее обоснованным на текущем этапе развития технологий является **гибридный подход** (assisted transcription) (Bazillon, Esteve, Luzzati, 2004), сочетающий скорость автоматического распознавания с точностью и глубиной ручной корректировки.



Протокол разметки

Цель: разработка унифицированной системы аннотаций, обеспечивающей релевантность корпуса для задач диагностики правдивости/ложности устной речи.

Основные принципы:

- Разделение речевого потока на элементарные дискурсивные единицы (ЭДЕ).
- Фиксация речевых сбоев (паузы, повторы, фальстарты, самокоррекции) как индикаторов когнитивной нагрузки и коммуникативной неуверенности.

Теоретическая база:

- Исследования, выявляющие взаимосвязь речевых сбоев и восприятия лжи (Chen et al., 2020; Loy, Rohde, Corley, 2017; Solà-Sales и др., 2023; Loy, Rohde, Corley, 2018).
- Работа «Самоисправления говорящего и другие типы речевых сбоев как объект аннотирования в корпусах устной речи» (Подлесская В. И., Кибрик А. А., 2007) акцентирующая внимание на значимости системной фиксации речевых сбоев для изучения когнитивных и прагматических аспектов устного дискурса, подчеркивая их роль как ключевого объекта аннотирования в корпусах устной речи.

Практическая реализация:

- Протокол основан на системе разметки проекта «Рассказы о сновидениях...» (Кибрик А. и др., 2009), адаптированной для автоматического анализа (регулярные выражения, подсчет частот).

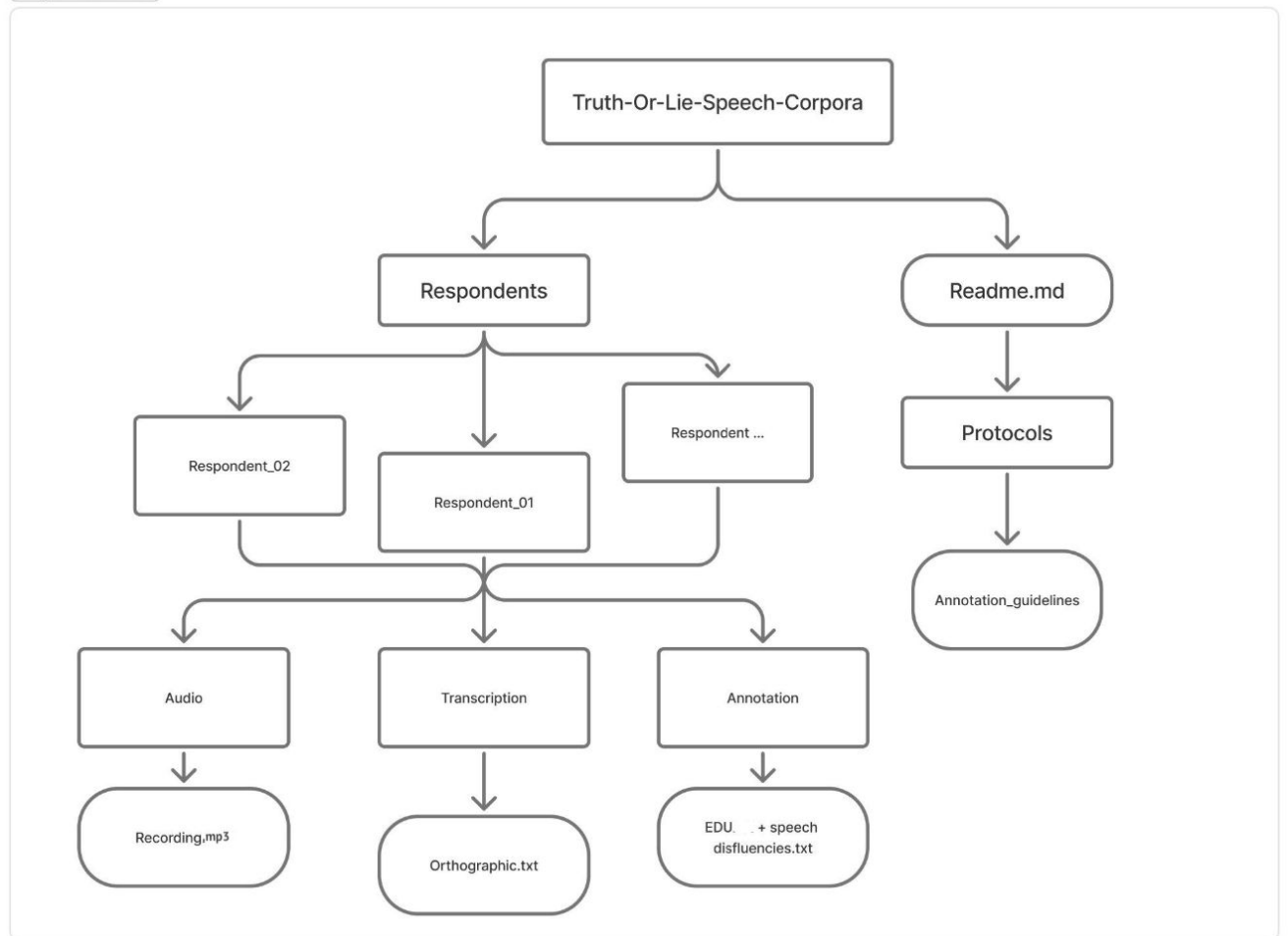
Значимость: создание инструмента, сочетающего традиции корпусной лингвистики и требования к машинной обработке, для поддержки последующих исследований в области диагностики речевой лжи.

Обозначение	Явление
...	Незаполненные пограничные и внутренние паузы
э	Заполненные паузы (с определенным вокалическим содержанием: случаи гласного заполнения)
м	Заполненные паузы (с определенным вокалическим содержанием: случаи когда пауза представлена губным сонорным или комбинацией сонорного с гласным)
а-а с-с я-я й-й	Удлиненная реализация фонем
=	Обрыв слова
	Слабый фальстарт (точка прерывания при самоисправлении внутри ЭДЕ)
Е	а) просодическая интеграция двух ЭДЕ; б) случае неполных ЭДЕ без самоисправления; в) случаи сильного фальстарта ВНУТРИ предложения Обобщенно: всё, что не является полноценной финальной/нефинальной ЭДЕ
S	Граница предложения в случае, когда оно обрывается, т.е.. заканчивается не одним из финальных пунктуационных знаков
ES	а) Одновременно незакончено ЭДЕ и предложение (упрощенная форма записи Е S) Частный случай: сильный фальстарт в случае, когда он находится в НАЧАЛЕ предложения. б) Любые другие ситуации. Например, ЭДЕ обрывается вместе с окончанием эпизода. Или ЭДЕ обрывается по внешним причинам и говорящий уже не заканчивает предложение, которое начал.
отсутствие пунктуации в конце ЭДЕ + Е или ES	Сильный фальстарт (точка прерывания при самоисправлении на границе ЭДЕ).
Е - обрыв ЭДЕ/ S - обрыв предложения	Сплит (разрыва ЭДЕ)
знак обрыва ЭДЕ + ()/ () - если ЭДЕ является полноценной	Парентеза
S - обрыв предложения	Односторонняя парентеза
Весь арсенал знаков завершения иллокуции	Иллокутивно неоднородные предложения
..	Иллокуция многоточия



Создание архитектуры корпуса

Corpora Structure





README

Code

Blame

1.92 MB

Raw



[View raw](#)

Truth-Or-Lie-Speech-Corpora

Аудиофайл	Транскрипт	Разметка
Ж18 (1) 3.mp3	ссылка	Ж18 (1) 3_2.txt
Ж18 (1) врёт.mp3	ссылка	Ж18 (1) врёт.mp3
Ж18 (1) правда.mp3	ссылка	Ж18 (1) правда.txt
Ж18 (2) 3.mp3	ссылка	Ж18 (2) 3.txt
Ж18 (2) врёт.mp3	ссылка	Ж18(2) врёт.txt
M19 (1) говорит правду.mp3	ссылка	M19(1)врёт.txt
M19 (2) говорит правду.mp3	ссылка	M19_3.txt
M19 (2) 3.mp3	ссылка	M19(2)3.txt
M18(1) врёт.mp3	ссылка	M18(1)_врёт.txt
M18(2) правда.mp3	ссылка	M18(2)_правда.txt

Code

Blame

32 lines (31 loc) · 2.73 KB

Raw



```
1  Время начала - чч:мм:сс.мс  Время окончания - чч:мм:сс.мс  Длительность - чч:мм:сс.мс  Рассказчик
2  00:00:00.080  00:00:01.690  00:00:01.610  Меня зовут Алексеева Анна,
3  00:00:02.320  00:00:04.500  00:00:02.180  я ходила на концерт со своими друзьями [E]
4  00:00:04.510  00:00:06.340  00:00:01.830  на концерт Рамиля. ...(0.4)
5  00:00:06.340  00:00:09.410  00:00:03.070  Инициатором концерта была я. ...(0.5)
6  00:00:09.412  00:00:15.451  00:00:06.039  А(0.3) так...(0.5) мне очень понравились эмоции от этого концерта,
7  00:00:15.451  00:00:18.686  00:00:03.235  а(0.2) времяпровождение,
8  00:00:18.706  00:00:19.608  00:00:00.902  а(0.1) очень круто. [S]
9  00:00:24.784  00:00:27.529  00:00:02.745  Я бы порекомендовала своим друзьям сходить на такой концерт,
10 00:00:27.529  00:00:32.588  00:00:05.059  а(0.2) чтобы отвлечься ...(0.3) от своих проблем,
11 00:00:32.608  00:00:35.824  00:00:03.216  а(0.2) чтобы провести круто время,
12 00:00:35.843  00:00:36.882  00:00:01.039  ну и всё.
```

Code

Blame

4 lines (2 loc) · 2.92 KB

Raw



```
1  Я Алексеева Анна. Аа, Я н= совсем.. ну вот [наверное] на этой неделе, аа, лазила в инстаграме и наткнулась на шоу-рум. Мне там очень сильн
2
3
4  Угу. Аа, какого цвета было платье? - Черное. - Черное. Ам, так, второй вопрос. Эм, в каком районе Челябинска, да, в Челябинске же это про
```



Итоги

- Сравнены три модели ASR по качеству транскрипции и фиксации речевых сбоев.
- Модель *gigaam-v2-rnnt* выбрана для финальной работы с корпусом.
- Современные ASR-системы не фиксируют паралингвистические элементы (паузы, фальстарты, самоперебивы) и требуют дообучения на специализированных данных.
- Все модели допускают как ошибки, искажающие смысл высказываний, так и ошибки, затрагивающие лишь грамматические категории, однако их частотность и выраженность различаются.
- Разработан протокол аннотирования, основанный на существующих принципах разметки устных корпусов, фиксирующий речевые сбои как потенциальные маркеры лжи и адаптированный для автоматического анализа.
- Корпус создан как открытый ресурс для дальнейшего расширения и исследований в области диагностики правдивости/ложности устной речи.

Спасибо за внимание!