Министерство образования и науки Российской Федерации

Государственное образовательное учреждение высшего профессионального образования «Московский физико-технический институт (государственный университет)»;

Центр «ПУСК» МФТИ

Программа «Науки о данных»

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

по направлению 01.04.02 «Прикладная математика и информатика»

Тема: «ML-моделирование перевода с русского жестового на русский звуковой»

(рабочее название)

Студент:

Кудрявцева П. Д.

Научный руководитель:

Каприелова М. С.

Москва 2024

**Содержание**

[Введение 3](#_Toc134374881)

[Глава 1. Обзор литературы и доступных данных по теме 8](#_Toc134374882)

[1.1 Особенности жестовых языков и РЖЯ в частности 8](#_Toc134374883)

[1.2 Наборы данных в области распознавания и перевода с жестовых языков 12](#_Toc134374884)

[1.3 Подходы к решению задачи в зарубежной и отечественной литературе 15](#_Toc134374885)

[Список литературы 20](#_Toc134374886)

Введение

Общение – важная составляющая в жизни любого человека. Все мы испытываем потребность делиться с другими своими мыслями, чувствами, точками зрения, хотим быть понятыми и услышанными. Также общение важно и для развития человека – мы многому учимся именно через общение с другими людьми: родными, близкими, учителями, друзьями, коллегами.

Глухие и слабослышащие люди зачастую сталкиваются с определёнными сложностями в общении с людьми слышащими. Основное средство коммуникации между глухими и слабослышащими – это жестовые языки. Далеко не каждый слышащий человек знаком с жестовым языком своих соотечественников. Для глухих же научиться понимать (посредством чтения по губам) и говорить (не слыша самих себя) – весьма сложная задача, которая требует кропотливой работы как с его стороны, так и со стороны преподавателя. Более того, в некоторых случаях (например, при общении с врачом в маске) человек оказывается лишённым даже возможности прочитать по губам.

Статистические данные свидетельствуют о постоянном росте числа новорожденных с нарушениями слуха. По данным Всемирной федерации глухих, более 300 жестовых языков используются более чем 700 миллионами людей по всему миру. По данным ВОЗ, количество глухих и слабослышащих людей в мире увеличивается год от года:

* порядка 278 миллионов человек в 2005 году,
* порядка 360 миллионов человек в 2015 году (+14%),
* порядка 466 миллионов человек в 2019 году (+13%),
* к 2050 году ВОЗ прогнозирует, что это число удвоится.

Традиционным посредником в общении между глухими/слабослышащими людьми и слышащими людьми выступает сурдопереводчик. По законодательству РФ, услуги сурдоперевода частично оплачиваются за счёт федерального бюджета. Однако объём услуги, которую глухой человек может получить за счёт государства относительно небольшой – максимум 40 часов в год. Кроме того, сурдопереводчики – достаточно редкая профессия: по данным Министерства просвещения РФ[[1]](#footnote-1), в 2020-м году в России насчитывалось около 900 сурдопереводчиков, при этом на каждого из них приходилось до 100 тысяч глухих. Также посредничество постороннего человека, может быть не всегда желательно для глухого человека: например, в вопросах, касающихся его здоровья.

Со стороны медицины также постоянно ведутся разработки технологических решений, помогающим глухим и слабослышащим людям ориентироваться в мире звуков. На настоящее время самым эффективным решением в этом направлении является кохлеарная имплантация. Однако она подходит не всем. Кроме того, слышат люди с имплантами несколько иначе, чем слышащие, и вынуждены какое-то время учиться воспринимать и различать звуки с помощью импланта.

Ещё одним мостиком для общения между глухими и слышащими людьми мог бы стать автоматический переводчик с жестового языка на звуковой и в обратном направлении. Несмотря на впечатляющие результаты, достигнутые учёными и разработчиками в этом направлении, необходимое качество такого перевода в настоящий момент не достигнуто. Глухие предъявляют следующие требования к такой системе автоматического перевода [Гриф и др., 2021]:

* точность перевода не менее 90%,
* качество визуализации должно быть максимально приближено к движениям человека-сурдопереводчика.

Почему же при текущем уровне развития машинного перевода с одного жестового языка на другой, для жестовых языков удовлетворительный результат пока ещё не достигнут? С какими сложностями сталкиваются исследователи, работающие с жестовыми языками?

Человек, говорящий на жестовом языке, использует для передачи своего сообщения не только движения рук (которые сами по себе допускают определённую вариативность), но также и движение корпусом (например, при сравнении), выражение лица (в РЖЯ, например, при вопросе необходимо нахмурить брови), «проговаривание» слов губами (чтение по губам может помочь различить похожие между собой жесты, однако не все глухие задействуют губы), движения головой (например, в РЖЯ при вопросе голова как правило движется немного вперед в сторону собеседника), позиции тела (например, при пересказе разговора двух и более лиц). Таким образом, каждый из указанных компонентов несёт в себе определённую смысловую нагрузку, а значит, при упрощении задачи только до движений руками, теряется существенная часть информации, а необходимость учитывать все факторы приводит к усложнению задачи.

С развитием ИИ, и особенно глубокого обучения, исследователи всё больше внимания стали уделять построению именно интеллектуальных систем перевода для жестовых языков. Среди работ последнего десятилетия чётко прослеживается популярность методов глубокого обучения. Например, работы на базе набора данных RWTH-PHOENIX-Weather 2014 чаще всего строят распознавание жестов на базе свёрточной нейронной сети (CNN). Те же работы, которые сфокусированы не на распознавании отдельных жестов, а на переводе непрерывной жестовой речи, чаще всего дополняют CNN лингвистической моделью (как правило, BLSTM) [Koller, 2020]. Тут следует отметить, что на данный момент большая часть работ всё-таки ориентирована на распознавание отдельных изолированных жестов, а не на перевод непрерывной жестовой речи.

В качестве входных данных сейчас наиболее популярны RGB-изображения или видео. Также есть варианты получения данных с сенсоров: электронные перчатки, технологии захвата движения, Kinect), но они существенно более дорогие, и сложно представить их массовое использование при работе с автопереводчиком.

Степень популярности того или иного языка у исследователей определяется в основном наличием общедоступного размеченного набора данных. Лидером по количеству публикаций в настоящий момент выступает американский жестовый язык. Но это верно только для небольших наборов данных (со словарём менее 200 жестов). На более крупных наборах лидирует уже китайский жестовый, а на наборах со словарём более 1000 жестов (RWTH-PHOENIX-Weather 2014) – немецкий жестовый [Koller, 2020].

Немануальные составляющие жестового языка (такие как: движения головы, корпуса, губ, бровей) чаще всего учитываются в работах именно по немецкому жестовому языку [Koller, 2020]. Исследователи РЖЯ также довольно часто принимают во внимание немануальные параметры. Однако в последние годы в мировой практике есть тренд к работе с полнокадровыми изображениями (когда говорящий на жестовом языке полностью попадает в кадр и все его движения рассматриваются как единое целое).

Подробнее о существующих в настоящий момент наборах данных, основных наработках и о трендах в подходах к распознаванию и переводу с жестовых языков (в том числе РЖЯ) см. главу 1.

Исследований по РЖЯ в настоящий момент существенно меньше, чем по американскому, немецкому или китайскому жестовым языкам. Одна из главных проблем тут – отсутствие общедоступных размеченных наборов данных, достаточно больших по объему и словарному запасу. Но это не значит, что нужно отказываться от исследований по тем наборам, которые доступны на данный момент.

**Объект исследования настоящей работы:** возможности автоматизации перевода с русского жестового языка (РЖЯ) на звуковой русский язык.

**Предмет исследования:** разработка подходов к автоматическому распознаванию непрерывной речи на РЖЯ и её переводу на звуковой русский язык методами глубокого обучения.

**Цель исследования:** обучить нейронную сеть (или набор нейронных сетей), которая будет показывать приемлемою точность перевода непрерывной жестовой речи в видеоформате на русский звуковой язык.

**Задачи исследования:**

1. Изучить международную практику в области интеллектуального машинного перевода с жестовых языков, а также наработки отечественных исследователей в отношении РЖЯ.
2. Выбрать несколько наиболее популярных/показывающих хорошую точность архитектур для перевода непрерывной жестовой речи и применить их к поставленной задачи. И применить их к выбранному набору данных.
3. Провести сравнительный анализ полученных результатов.
4. Выбрать модель с лучшим результатом и постараться определить, за счёт каких факторов результат получился лучше.

В качестве набора данных для исследования возьмём Корпус русского жестового языка [Электронный ресурс][[2]](#footnote-2), разработанный под руководством С. И. Бурковой в Новосибирском государственном техническом университете. Подробнее о характеристиках корпуса и его сравнение с другими доступными наборами данных, подходящими для решения поставленной задачи, см. главу 1.

* методологический аппарат исследования: TBD;
* основные пункты новизны, выносимой на защиту: TBD;
* теоретическая и/или практическая значимость исследования: TBD;
* апробация исследования (если таковая имеется) : TBD;
* краткая характеристика структуры диссертации: TBD.

Глава 1. Обзор литературы и доступных данных по теме

В этой главе более детально погрузимся в тему исследования и особенности поставленной задачи. Сначала рассмотрим особенности жестовых языков и их отношение к звуковым языкам. Затем исследуем доступные наборы данных, как по РЖЯ, так и по другим жестовым языкам, сравним их между собой, выделим основные преимущества и недостатке, а также подробнее остановимся на Корпусе русского жестового языка С. И. Бурковой, на основе которого будет вестись дальнейшее исследование. В последней части главы посмотрим, какие подходы применяют международные и отечественные исследователи к решению подобных задач.

* 1. Особенности жестовых языков и РЖЯ в частности

Жестовые языки выполняют ту же функцию средства коммуникации, что и звуковые языки. Это полностью самостоятельные языки со своей грамматикой, морфологией и фонологией: то есть жестовая речь – это совсем не пословесный перевод со звукового языка, а совершенно другой язык со своими собственными правилами. Когда глухие люди учатся читать на основном языке страны, где они живут, они, по сути, учатся читать на иностранном для них языке. В качестве иллюстрации возьмём предложение: «*Папа приготовил шашлык»*. На русском жестовом (если перевести отдельные жесты в слова) это предложение будет сформулировано так: *«Шашлык папа готовить всё»*.[[3]](#footnote-3)

Развиваются жестовые языки также независимо от звуковых. В результате бывает так, что в странах с общим звуковым языком, жестовые языки – разные. Например, американский жестовый язык (ASL) гораздо больше похож на французский жестовый (LSF), чем на британский жестовый (BSL), при том, что звуковой язык у США и Великобритании - один (английский).

В жестовых языках, подобно звуковым, формируются и свои диалекты. В случае с РЖЯ, например, исследователи отмечают различия между Москвой и Санкт-Петербургом[[4]](#footnote-4), между Москвой и Новосибирском[[5]](#footnote-5), также в литературе можно встретить казахско-русский жестовый язык (K-RSL). К сожалению, РЖЯ не так хорошо изучен лингвистами на данный момент, как звуковые языки, и какой-то формализованной информации по основным диалектам РЖЯ нет. Вероятно, это отчасти связано с тем, что официальное признание РЖЯ получил совсем недавно - 30 декабря 2012 (по ФЗ «О внесении изменений в статьи 14 и 19 Федерального закона “О социальной защите инвалидов в Российской Федерации”»).

В жестовой речи помимо собственно РЖЯ используется калькирующая жестовая речь (КЖР), дактиль. В дактильной азбуке каждой букве звучащего языка ставится в соответствие определённая конфигурация (и иногда движение) руки (дактилема). КЖР, как правило, используют для имён, названий, каких-то новых слов, для которых говорящий не знает жеста. Основная же часть коммуникации ведётся посредством жестов, а не дактилем.

Остановимся подробнее на составляющих жестовой речи. За счёт чего говорящий на жестовом языке коммуницирует смысл и своё отношение к сказанному? Современные исследователи жестовых языков выделяют тут 5 основных компонентов:

* *Форма (конфигурация)* руки: как расположены между собой пальцы, согнуты они или выпрямлены, расположены рядом или расставлены в стороны и т.п.
* *Локация* руки: где относительно тела говорящего расположен жест. Локацию подразделяют на место и сеттинг. Место – одна из крупных областей, где может исполняться жест (голова, лицо, шея, грудь, талия, нейтральное жестовое пространство). Сеттинг – детализация положения жеста внутри этой крупной области (если в качестве места выбрать лицо, сеттингами могут быть, напрмимер, правая щека, левый глаз, губы, подбородок и т.п.).
* *Ориентация* ладони – как расположена ладонь в пространстве. Например, ладонь может быть направлена от говорящего (или наоборот), вниз, в сторону и т.д.
* *Движение* руки: форма траектории, по которой движется рука, направление этого движения, поворот ладони (изменение ориентации), изменение конфигурации руки.
* *Немануальные* составляющие: движения корпусом (часто делаются при сравнении), движения головой, плечами, особенности мимики лица (например, нахмуренные брови при вопросе, улыбка при выражении радости), маусинг (проговаривание губами слов – не у всех, и не всегда есть связь со звуковым языком). Положение всего тела в пространстве также может использоваться говорящим, например, при рассказе о группе лиц.

Значение жеста может поменяться иногда при изменении даже одного из этих параметров.

В учебнике С.И. Бурковой [Буркова, 2019] всё многообразие жестов классифицируется по следующим признакам:

* Одноручные и двуручные жесты – в зависимости от того, участвует в исполнении жеста только ведущая рука или обе.
* Статичные и динамичные жесты – по признаку отсутствия или наличия движения руками.
* Симметричные и асимметричные жесты – это деление характерно для двуручных жестов (одинаково или по-разному работают руки в процессе исполнения).
* Синхронные и поочерёдные жесты – в зависимости от того одновременно задействуются руки или по очереди.

На исполнение жестов (кроме симметричных) оказывает влияние ведущая рука. У правшей – это правая рука, у левшей – левая. Соответственно, такие жесты в исполнении левши будут зеркальным отображением этих же жестов в исполнении правши. А значит, при обучении модели распознавания жестов, вероятно, имеет смысл учесть и зеркальные отображения исходных изображений, чтобы нивелировать возможное влияние фактора ведущей руки.

Рассмотрим, какие особенности жестовой речи, по мнению учёных ([Гриф и др., 2021], [Кагиров, Рюмин, 2022]), повышают сложность задачи распознавания и перевода жестовых языков и, в частности, РЖЯ:

* Отсутствие единого стандарта жестовой речи, наличие диалектов, одновременное использование РЖЯ и КЖР.
* Отсутствие достаточно больших по объёму наборов данных для распознавания непрерывной речи на жестовом языке.
* Разная скорость жестикуляции у говорящих.
* Фон влияет на то, как выглядит изображение.
* Эпентеза – дополнительные переходные движения между жестами, которые появляются в непрерывной речи.
* Комбинаторные изменения (вариативность) параметров жестов в непрерывной речи и вытекающая из этого проблема определения границ между жестами.

Остановимся подробнее на составляющих вариативности жестов в непрерывной речи и их аналогах в звуковых языках (табл. 1).

Таблица 1. Комбинаторные изменения в жестовых и звуковых языках

|  |  |
| --- | --- |
| Жестовые языки | Звуковые языки |
| Ассимиляция (уподобление параметров одного жеста параметрам соседнего жеста):   * антиципация - пассивная рука готовится заранее (одноручный жест перед двуручным), * персеверация - пассивная рука вместо возвращение в положение покоя сохраняет своё положение в предыдущем двуручном жесте (одноручный жест после двуручного), * эллипсис - исчезновение некоторых формантов из конфигураций и локализаций между жестами. | Ассимиляция (уподобление артикуляции звуков), аккомодация (частичное приспособление артикуляции смежных звуков), качественная редукция звуков. |
| Редукция жеста (исполнение движения с меньшей траекторией). | Редукция звука (сокращение длительности звука, изменение его качественных характеристик, и иногда даже выпадение). |
| Метатеза (взаимная замена начальной и конечной локализации при исполнении жестов с двойной локализацией). | Метатеза (взаимная перестановка звуков или слогов). |
| Редупликация, выпадение движения при использовании сочетания жестов или повторяющегося движения. | Диереза (выпадение звуков), гаплология (выпадение звуков по причине диссимиляции). |
| Варьируются положение выделенных пальцев (согнуты в разной степени), ориентация ладони, локализация и амплитуда движения. | Диссимиляция (расподобление артикуляции звуков, утрата ими общих фонетических признаков), качественная редукция звука. |
| Эпентеза (добавление жестов). | Эпентеза (вставка звуков). |

Таким образом, мы рассмотрели основные особенности жестовых языков, показали, что РЖЯ и русский звуковой – это разные (по сути иностранные друг для друга) языки, и подробнее остановились на факторах, представляющих определённую сложность с точки зрения задачи распознавания и перевода жестовых языков. Теперь перейдем к вопросу данных, доступных для решения подобных задач.

* 1. Наборы данных в области распознавания и перевода с жестовых языков

Большая часть работ на тему распознавания и перевода жестовых языков опирается в своих исследований на один из общедоступных «контрольных» наборов данных. Контрольными эти наборы называют потому, что благодаря их открытости, наличию разметки и популярности у других авторов, они используются сообществом для сравнения результатов исследований. В таблице 2 приведён сравнительный анализ основных из таких наборов данных, предназначенных именно для решения задач перевода. Для задач распознавания подходит большее количество наборов данных, их основное отличие - в отсутствии перевода на тот или иной звуковой язык.

Здесь важно затронуть вопрос транскрибирования жестовых языков. Первая система аннотации бала предложена Уильямом Стоуки [Stokoe, 1960] для описания американского жестового языка с помощью графических символов. До этого момента жесты рассматривали как неделимые единицы. Затем стали появляться и другие методы. В настоящий момент чаще всего для транскрибирования жестовых языков используют **глоссирование**.

Согласно Большому толковому словарю, **глосса** – это «перевод или объяснение непонятного слова, выражения, помещённое на полях или в самом тексте древних и средневековых рукописей; научный комментарий законов или судебных решений». В случае с аннотированием жестовых языков в соответствие каждому жесту в транскрипции записывается слово, которое он обозначает, иногда вместе с дополнительными общепринятыми символами для обозначения значимых лексико-грамматических особенностей (например, знак # используется для слов, произнесенных дактилем).

Таблица 2. Контрольные наборы данных для тренировки систем перевода с жестовых на звуковые языки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Набор данных** | **Язык** | **Тематика** | **Разметка** | **Объём** |
| RWTH-PHOENIX-Weather 2014 T | Немецкий жестовый | Прогноз погоды за 2009 – 2011 на немецком телеканале PHOENIX | Глосс-аннотация;  Письменный перевод на немецкий | выпусков прогноза погоды  глосс  2,887 слов  8,257 видео |
| CSL-Daily | Китайский жестовый | Повседневная жизнь (путешествия, покупки, медицинская помощь) | Глосс-аннотация;  Письменный перевод на китайский | 2,000 глосс  2,343 слов  20,654 видео |
| ASLG-PC12 | Американский жестовый (ASL) | Тексты на английском проекта «Гутенберг» преобразованы в глоссы американского жестового. | Глосс-аннотация;  Письменный перевод на английский | 800 млн. слов/глосс |
| How2Sign | Американский жестовый (ASL);  Английский текст | Перевод на американский жестовый язык видео-инструкций из набора данных How2 | Глосс-аннотация;  Письменный перевод на английский;  Глубина;  2D и 3D – скелеты; | 80 часов видео  16 тыс. слов |

Из таблицы 2 видно, что есть большая проблема с наличием качественных наборов данных для обучения интеллектуальных систем перевода с жестовых языков. Из 4-х популярных наборов данных только для китайского жестового языка собраны данные по темам повседневного общения – наиболее релевантная сфера для применения таких переводчиков. По большинству же жестовых языков, в том числе РЖЯ, отсутствуют даже такие, хоть и относительно ограниченные, но размеченные и общедоступные наборы данных. И в этом одна из основных причин, которая препятствует прогрессу в данном направлении.

Поскольку ни в одном из контрольных наборов данных РЖЯ не представлен, подходящие для исследования данные необходимо искать где-то ещё. Для этого обратимся к публикациям исследователей, работающих над системами распознавания и перевода собственно РЖЯ, а не абстрактного жестового языка. Такими системами достаточно активно в настоящий момент занимаются:

* Новосибирский государственный технический университет (НГТУ);
* Санкт-Петербургский Федеральный исследовательский центр  
  Российской академии наук (СПБ ФИЦ РАН), Лаборатория речевых и многомодальных интерфейсов;
* Центром развития технологий AI в Sber AI.

Соберём основные характеристики наборов данных, на которые они ссылаются в своих публикациях в аналогичную таблицу (табл. 3).

Таблица 3 Данные, упоминаемые в публикациях по теме распознавания жестов и перевода с РЖЯ

|  |  |  |  |
| --- | --- | --- | --- |
| **Набор данных** | **Тематика** | **Разметка** | **Объём** |
| Корпус РЖЯ С.И. Бурковой (НГТУ)  <http://rsl.nstu.ru/> | Повседневная | Глоссы для правой руки;  Глоссы для левой руки;  Перевод на русский | 175 видео |
| База данных РЖЯ поликлинического предназначения | Первичный приём у врача-терапевта | Сегменты движения и удержания  Морфологическая (глоссы)  Перевод на русский | 85 фраз;  7 часов видео |
| TheRuSLan | Покупки в супермаркете | Скелет (25 точек);  Перевод на русский язык;  Глоссы (по словам)  Классы жестов; | 154 жеста  8 часов 56 минут видео |
| Сурдосервер 2.0  http://surdoserver.ru/ | Видеоуроки | Нет |  |
| Датасет Sber AI | Повседневная бытовая | Глоссы (по словам) | 70 часов видео  2600 жестов  7500 фраз |

Таким образом, из пяти потенциальных источников, наиболее доступным и подходящим для исследования в рамках данной работы представляется Корпус русского жестового языка, разработанный в НГТУ под руководством Бурковой С.И.: датасеты Sber AI, TheRuSLan и база РЖЯ поликлинического назначения закрыты для внешних пользователей, а на Сурдосервере отсутствует разметка. Главные преимущества Корпуса РЖЯ Бурковой – в доступности данных (по запросу открывают доступ) и в наличии профессиональной разметки. Недостатки – в его малом размере (с точки зрения машинного обучения) и в отсутствии возможности выгрузить данные единым архивом.

* 1. Подходы к решению задачи в зарубежной и отечественной литературе

Теперь исследуем публикации на предмет того, какие методы и подходы применяются к обучению интеллектуальных систем распознавания и перевода с жестовых языков, каковы тренды и достигнутые результаты.

**[Koller, 2020]**

Начало работы над темой – 1983 год ([Grimes, 1983])

Смена подходов к решения с 2015 года

Переход от интрузивных к неинтрузивным методам

До 2015 не было больших наборов данных. После 2015 работа ведётся преимуществоенно с двумя базовыми корпусами: 1080 и 178 жестов.

Большинство работ нацелены на распознавание изолированных жестов (почти в два раза больше, чем для непрерывной жестовой речи).

Входные данные:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Метод | Описание | Интрузивность | Популярность | Языки | Стоимость | Ключевые даты |
| RGB изображения | Зрение: фичи, относящиеся к ладоням, положению пальцев и углов между ними оцениваются и потом используются для распознавания. Здесь необходимы изображения или видео, а также обработка этих изображений. | - | Самый популярный | Американский жестовый,  Китайский жестовый  Немецкий жестовый |  | Набирает популярность с 2005 года |
| RGB-D изображения |  | - |  | Китайский жестовый |  | Глубина стала популярна после выхода сенсоров Kinect в 2010 году. |
| разноцветные перчатки |  | + | Почти не используется |  |  |  |
| электронные перчатки |  | + |  |  |  |  |
| захват движения |  | + |  |  |  |  |

С 2005 года переход от интрузивных к неинтрузивным. Всё больше и больше используются изображения.

Мануальные параметры:

* Форма руки – самый популярный параметр (на маленьких датасетах, до 1000)
* Движение – 2 популярноть (на маленьких датасетах, до 1000)
* Локация – 2 популярность (на маленьких датасетах, до 1000)
* Ориентация

Немануальные параметры:

* Голова
* Губы
* Глаза
* Брови
* Моргание и взгляд

Полное представление говорящего на жестовом языке (суставы тела, RGB изображение, захватывающее всего человека целиком, такие же полнорамочные изображения с глубиной, полнорамочные движущиеся изображения с использованием оптического потока).

На больших датасетах (более 1000 знаков) чаще всего анализируются полнорамочные изображения.

С 2005 полнорамочный подход набирает популярность и приближается к масштабам исследований формы руки.

С ростом датасетов тренд от мануальных параметров к глобальным фичам. Возможные причины:

- Доступность суставов тела и полнорамочных изображений с выходом Kinect в 2010

- Сдвиг в сторону глубокого обучения и тренд к вводу полнорамочных изображений вместо ручного фиче инжиниринга.

[Гриф]

Препятствия:

неполнота описания грамматической системы РЖЯ

отсутсвие "плавности" показа жестов аватаром-переводчиком РЖЯ, система управления которого использует систему нотаций языков глухих

перевод осуществляется преимущественно на калькирующую жестовую речь, а не на РЖЯ, обладающий выразительными возможностями

высокий процент ошибок при переводе многозначных слов и омонимов на жесты РЖЯ (более 20%)

* Основные технологи основаны на: компьютерном зрении и на сенсорных-перчатках (и их аналогов)
* Перчатки: механические или оптические сенсоры, приреплённые к перчаткам, которые надевает пользователь. движения пальцев трансформируются в электрический сигнал, почзволяющий определь положение рук для распознавания.

[Гриф]

Необходимо отделять эпентезу от собственно жестов.

312 предложений.

5 типов движений:

* нет жеста
* идёт жест
* начальное движение
* переходное движение
* конечное движение

Разметка видов эпентезы - [supervise.ly](https://supervise.ly/)

Реккурентная сеть (LSTM) - точность правильного распознавания эпентезы 95%

Для выделения эпентезы:

* основная идея - выделение признаков начала следующего жеста или паузы в речи.
* приёмы вычленения "быстрых" и "медленных" движений в видео потоке
* предполагается введение функции принадлежности (вероятности) каждого события к началу нового жеста.
* выделение всех элементов жестовой речи (жестов, состояния покоя, комбинаторных изменений и эпентезы) - требует расширения датасета

Список литературы

**Нормативно-правовые акты**

Постановление Правительства РФ от 25 сентября 2007 г. N 608 "О порядке предоставления инвалидам услуг по переводу русского жестового языка (сурдопереводу, тифлосурдопереводу)".

Федеральный закон от 30 декабря 2012 г. N 296-ФЗ "О внесении изменений в статьи 14 и 19 Федерального закона "О социальной защите инвалидов в Российской Федерации".

**Книги**

Большой толковый словарь русского языка. / ред.: С.А. Кузнецов. - Первое издание: СПб.: Норинт - 1998.

Введение в лингвистику жестовых языков. Русский жестовый язык: учебник / ред.: С.И. Буркова, В.И. Киммельман. – Новосибирск: Изд-во НГТУ, 2019. – 356 с.

**Статьи**

Гриф М.Г., Элаккия Р., Приходько А.Л., Бакаев М.А., Раджалакшми Е. Распознавание русского и индийского жестовых языков на основе машинного обучения // Системы анализа и обработки данных. - 2021. - №3(83). - С. 53-74.

URL: <https://journals.nstu.ru/vestnik/catalogue/contents/view_article?id=27523>

Гриф М.Г., Королькова О.О., Приходько А.Л. Распознавание жестовой речи с учётом комбинаторных изменений жестов // Материалы XXI Международной научно-методической конференции «Информатика: проблемы, методы, технологии». - В.: Вэлборн, 2021. С. 1387-1393.

Кагиров И.А., Рюмин Д.А. База данных русского жестового языка поликлинического предназначения: лингвистические особенности материала и аннотирования. // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. – 2022 - №20(3). – С. 90-108.

URL: <https://doi.org/10.25205/1818-7935-2022-20-3-90-108>

Adeyanju I.A., Bello O.O., Adegboye M.A. Machine learning methods for sign language recognition: A critical review and analysis // Intelligent Systems with Applications – 2021. – Vol.12. – P.2667-3053.

URL: <https://doi.org/10.1016/j.iswa.2021.200056>

Achraf Othman and Zouhour Tmar. “English-ASL Gloss Parallel Corpus 2012: ASLG-PC12, The Second Release”. Fourth International Conference On Information and Communication Technology and Accessibility ICTA’13, Hammamet, Tunisia, October 24-26, 2013.

Duarte, Amanda and Palaskar, Shruti and Ventura, Lucas and Ghadiyaram, Deepti and DeHaan, Kenneth and Metze, Florian and Torres, Jordi and Giro-i-Nieto, Xavier. A Large-scale Multimodal Dataset for Continuous American Sign Language / Conference on Computer Vision and Pattern Recognition (CVPR). – 2021 URL: https://arxiv.org/pdf/2008.08143v2.pdf

Koller O. Quantitative survey of the state of the art in sign language recognition // [https://arxiv.org/abs/2008.09918v2 - 2020](https://arxiv.org/abs/2008.09918v2%20-%202020).

Yutong Chen, Ronglai Zuo, Fangyun Wei, Yu Wu, Shujie Liu, Brian Mak. Two-Stream Network for Sign Language Recognition and Translation // https://arxiv.org/pdf/2211.01367v2.pdf

**Интернет-источники**

«В России отмечают День сурдопереводчика» / Пресс-служба [Минпросвещения России](https://edu.gov.ru/), 2020. URL: <https://edu.gov.ru/press/3084/v-rossii-otmechayut-den-surdoperevodchika/>

Онлайн-курс русского жестового языка (РЖЯ) / РООИ «Перспектива».

URL: <https://edu.perspektiva-inva.ru/courses/kurs-russkogo-zhestovogo-yazyka/>

Разработка модели машинного перевода с русского жестового языка / AIRI Seminars.

URL: https://youtu.be/OHQ8MmIJmEM

Сайт негосударственного образовательного частного учреждения дополнительного профессионального образования «Учебно-методический центр Всероссийского общества глухих» (НОЧУ ДПО «УМЦ ВОГ»). URL: https://nochuvog.ru/

Сайт общероссийской общественной организации инвалидов «Всероссийское общество глухих» (ВОГ). URL: <https://voginfo.ru/>

**Наборы данных**

Корпус русского жестового языка [Электронный ресурс] / Руководитель проекта С. И. Буркова – Новосибирск – 2012–2015 – Режим доступа: [http://rsl.nstu.ru/](http://rsl.nstu.ru/site/signlang)

CSL-Daily: Chinese Sign Language Corpus / University of Science and Technology of China (USTC) // Multimedia Computing & Communication, SLR Group.

URL: <http://home.ustc.edu.cn/~zhouh156/dataset/csl-daily/>

English-ASL Gloss Parallel Corpus 2012: ASLG-PC12

URL: <https://achrafothman.net/site/english-asl-gloss-parallel-corpus-2012-aslg-pc12/>

How2Sign (A Large-scale Multimodal Dataset for Continuous American Sign Language) URL:  <http://how2sign.github.io/>

WTH-PHOENIX-Weather 2014 T: Parallel Corpus of Sign Language Video, Gloss and Translation / RWTH Aachen University // Human Language Technology & Pattern Recognition Group

URL: <https://www-i6.informatik.rwth-aachen.de/~koller/RWTH-PHOENIX-2014-T/>

1. «В России отмечают День сурдопереводчика» / Пресс-служба [Минпросвещения России](https://edu.gov.ru/), 2020. URL: https://edu.gov.ru/press/3084/v-rossii-otmechayut-den-surdoperevodchika/ [↑](#footnote-ref-1)
2. http://rsl.nstu.ru//site/signlang [↑](#footnote-ref-2)
3. Пример взят из онлайн-курса русского жестового языка (РЖЯ) РООИ «Перспектива»: https://edu.perspektiva-inva.ru/lessons/3-vvedenie-v-grammatiku/ [↑](#footnote-ref-3)
4. https://nochuvog.ru/russkij-zhestovyj-yazyk-rzhya/ [↑](#footnote-ref-4)
5. https://voginfo.ru/vog/2015/05/30/rslidioms/ [↑](#footnote-ref-5)