

## Дистанционное обучение как современный формат совершенствования навыка чтения с губ (на базе корпуса омовизем)

М. А. Мясоедова\*, З. П. Мясоедова

ФГБУН «Институт проблем управления им. В.А. Трапезникова Российской академии наук», г. Москва, Российская Федерация

Адрес: 117997, Российская Федерация, г. Москва, ул. Профсоюзная, д. 65

\* mariamarfi@mail.ru

### Аннотация

Статья посвящена актуальной проблеме в области визуального восприятия и понимания устной речи человеком с ограниченными возможностями по слуху. Низкая вероятность правильного понимания визуальной информации объясняется соответствием одной виземы нескольким фонемам. Это влияет на образование в устной речи разнозначных слов со схожими артикуляционными рисунками. Для определения этих слов использован введённый авторами термин «омовиземы». Большое количество в русской речи таких слов приводит к их взаимозаменяемости и не позволяет различить их между собой не только вне контекста, но в отдельных случаях и при его наличии. Цель исследования заключается в анализе возможности и целесообразности использования корпуса омовизем для решения определённых трудностей в процессе обучения чтению с губ. В настоящее время одним из перспективных направлений в образовании является дистанционное обучение. В статье рассматривается использование дистанционной формы для анализа структуры чтения с губ на основе корпуса омовизем русского языка. Данный корпус разработан на основе подготовленного материала, сгруппированного в отдельные цепочки по принципу одинаковых артикуляционных оболочек слов, за исключением псевдослов. Он предназначен для ознакомления пользователя с наличием в устной русской речи омовизем. Это позволит ему пополнить свой словарный запас и упростить визуальное восприятие устного сообщения. Приведены примеры возможных вариантов распознанных слов со схожими виземами на основе контента разработанного корпуса. Полученные результаты исследования позволяют сделать вывод об актуальности выбранной темы. Они помогут спрогнозировать неопределённость при выборе правильного значения. Дистанционное обучение как современный формат направлено на развитие личностных способностей людей с ограниченными способностями по слуху, повышении их навыков в плане понимания устного высказывания, полноценной социальной адаптации в обществе.

**Ключевые слова:** визуальное распознавание устной речи, виземы, омовиземы, псевдослова, дистанционное обучение, корпусный подход

*Авторы заявляют об отсутствии конфликта интересов.*

**Для цитирования:** Мясоедова М. А., Мясоедова З. П. Дистанционное обучение как современный формат совершенствования навыка чтения с губ (на базе корпуса омовизем) // Современные информационные технологии и ИТ-образование. 2022. Т. 18, № 2. С.374-382. doi: <https://doi.org/10.25559/SITITO.18.202202.374-382>

© Мясоедова М. А., Мясоедова З. П., 2022



Контент доступен под лицензией Creative Commons Attribution 4.0 License.  
The content is available under Creative Commons Attribution 4.0 License.



Original article

## Remote Study as a Modern Format to Improve our Lip-Reading Skills (Based on our Homovisemes Corpus)

M. A. Myasoedova\*, Z. P. Myasoedova

V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russian Federation

Address: 65 Profsoyuznaya St., Moscow 117997, Russian Federation

\* mariamarfi@mail.ru

### Abstract

This article continues our work, where we describe possible barriers for a person with hearing impairments to visually perceive and understanding verbal speech by observing the speaker's articulations. It is difficult to correctly understand visual information since one viseme corresponds to several phonemes. This affects how different words with similar articulation patterns (homovisemes) form up in verbal speech. These words are defined by the term "homovisemes" introduced by the authors. There are many such words in Russian speech. So, these words are interchangeable and thus indistinguishable from each other, not only out of context, but sometimes even when a context is available. The purpose of this research is to analyze if it is possible and expedient to use our homovisemes corpus to solve certain difficulties in how we learn to read from lips. Currently, one of the promising areas in education is remote study. In the article, we investigate how to use a distance form to analyze a structure of lip-reading based on our homovisemes corpus of Russian. This corpus is based on our prepared material, grouped into separate chains according to the principle of identical articulatory shells of words; pseudo-words are excluded. It is intended to familiarize the user with homovisemes in oral Russian speech. This will allow him to replenish his vocabulary and easily visually perceive an oral message. We present examples of possible variants of recognized words with similar visemes based on our developed corpus. Our results allow us to conclude on how relevant the chosen topic is. The results will help us to predict uncertainty when we choose the right value. Remote study is a modern study format. It is aimed to develop personal abilities of people with hearing impairments, to improve their skills to understand oral utterance, and to socially adapt these people in our society.

**Keywords:** visual recognition of spoken language, visemes, homovisemes, pseudowords, remote study, corpus approach

*The authors declare no conflict of interest.*

**For citation:** Myasoedova M.A., Myasoedova Z.P. Remote Study as a Modern Format to Improve our Lip-Reading Skills (Based on our Homovisemes Corpus). *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. 2022; 18(2):374-382. doi: <https://doi.org/10.25559/SITITO.18.202202.374-382>



## Введение

В последнее время во многих странах наблюдается интерес к использованию новейших информационных технологий в образовательном процессе с целью развития личностных способностей пользователей, повышении их навыков и полноценной социальной адаптации [1].

Дистанционное обучение является одним из перспективных направлений в образовании, в том числе и в инклюзивном, предназначенном для людей с ограниченными возможностями по слуху, владеющих навыком визуального восприятия устной речи – чтения с губ говорящего. Для анализа функционирования языковых единиц активно используются компьютерные возможности, в частности, поисковые системы, значительно облегчающие доступ к необходимой объективной информации<sup>1</sup>.

Данное исследование является продолжением работы по выявлению, изучению и описанию возможных барьеров на пути визуального восприятия человеком устной речи по артикуляции говорящего, снижающих качество её распознавания. Одним из таких серьёзных барьеров является соответствие одной виземы нескольким фонемам, что приводит к образованию в устной речи разнозначных слов со схожими артикуляционными рисунками. Для определения этих слов использован введённый авторами термин «омовиземы» [2]. Большое количество в русской речи таких слов приводит к их взаимозаменяемости и не позволяет различить их между собой не только вне контекста, но в отдельных случаях и при его наличии.

Предоставление реципиенту наиболее полной информации о возможной взаимной замене этих речевых элементов позволит спрогнозировать результат их визуального восприятия. Вне контекста этот процесс часто приводит к неопределённости или полному непониманию смысла полученной им информации.

Этот факт требует глубокого и детального исследования и изучения особенностей визуального восприятия устной речи при наличии соответствующего речевого материала и использования разработок различных вспомогательных средств на базе новых технологий.

Цель данной работы состоит в проведении исследования возможности и целесообразности использования речевого корпуса для решения определённых трудностей в процессе обучения чтению с губ. В соответствии с поставленной задачей был разработан корпус омовизем на основе материала, полученного в результате сбора и анализа слов с использованием артикуляционной транскрипции.

Такой подход позволяет выявить слова с одинаковыми визуальными оболочками и объединить их в группы, что будет способствовать совершенствованию или получению реципиентом новой информации, необходимой для повышения уров-

ня владения им навыком чтения с губ и практических знаний в этой области,

Актуальность данной работы заключается в необходимости оказания помощи людям с ограниченными возможностями по слуху с целью интеграции их в общество слышащих людей.

Результаты исследования могут быть применены в практике обучения чтению с губ на материале разработанного корпуса омовизем, включающего более 5 тысяч единиц, сгруппированных по схожести артикуляционных рисунков.

## Концепция визуального распознавания устной речи

### Формирование и восприятие устной речи

Коммуникация людей происходит в основном благодаря речи, являющейся основным средством их общения и позволяющей им получать и передавать большое количество необходимой информации<sup>2</sup> [3]-[6]. Владея речью, человек способен формулировать свои мысли и правильно воспринимать чужие на понятном коммуникантам языке, будь то литературное, узко-профессиональное или иное его употребление (территориальный диалект, жаргон, просторечие и т.п.)

Человек с нарушением слуховой функции обладает феноменальной способностью визуально воспринимать устную речь по движению артикуляционных органов говорящего. Данный способ восприятия речи не всем доступен в равной степени и представляет собой довольно сложную и многогранную задачу, что обусловлено влиянием на этот процесс ряда факторов. Это, прежде всего, влияние фонетических процессов, характер сформированной и переданной коммуникатором информации и способность реципиента воспринимать её соответственно заложенному в ней смыслу.

Наличие плохо различимых визем во многом связано с процессом воспроизведения соответствующих им звуков в значительной мере скрыт внутри ротовой полости, что существенно усложняет их зрительное восприятие.

Следует уточнить, что указанные факторы, как правило, проявляются одновременно, усложняя в определённой степени процесс визуального восприятия речи. При этом, результат распознавания реципиентом устного высказывания часто будет отличаться от исходной его формы. Отсюда вполне естественно стремление исследователей к всестороннему комплексному изучению феномена визуального восприятия устной речи<sup>3</sup>.

Детальное изучение отличительных особенностей визуального восприятия устной речи от фонетического её восприятия требует наличия речевого материала, на которого будут определены отклонения в плане состава фонем и визем слов устной речи. Это позволит выявить возможные несоответствия характеристик их речевых элементов в плане их выражения

<sup>1</sup> Плунгян В. А. «Интегрум» и Национальный корпус русского языка в лингвистических исследованиях // Integrum: точные методы и гуманитарные науки ; под ред. Г. Никопорец-Такигава. М.: Летний сад, 2006. С. 76-84.

<sup>2</sup> Чистович Л. А. Речь. Артикуляция и восприятие. М.: Наука, 1965. 241 с.; Филатова Е. В. Порождение и восприятие речи (Авторская делимитация речи как способ представления её реальных единиц). Донецк: ДонНУ, 2017. 296 с. (Типологические, сопоставительные, диахронические исследования; Т. 14). URL: <http://repo.donnu.ru:8080/jspui/bitstream/123456789/4766/1/2298.pdf> (дата обращения: 28.05.2022).

<sup>3</sup> Мясоедова М. А., Мясоедова З. П., Петухова Н. В. Визуальное восприятие устной речи и обучающая программа «Читаем с губ». М.: ООО «Буки Веди», 2015. 153 с.



и содержания. Выполнение данной задачи реализовано на материале разработанного авторами корпуса омовизем, т.е. русских слов со схожими виземами.

## Корпусный подход к дистанционному обучению навыку чтения с губ

Наблюдаемый во всём мире интерес к использованию компьютерных возможностей для изучения интеллектуальных форм взаимодействия человека и компьютера, к которым относятся поисковые системы и электронные корпуса, привёл к прочному вхождению в нашу жизнь технологии по распознаванию речи, на основе которых разрабатываются обучающие программы [7].

Развитие и внедрение компьютерных технологий позволило существенным образом удовлетворить ожидаемый спрос на оказание услуг по обучению людей с потерей слуха чтению с губ путём предоставления им доступа к необходимой информации и программам для овладения этим навыком с использованием различных подходов к обучению. Популярность использования корпусов объясняется возможностью исследования языковых закономерностей и особенностей на обработанном и представленном в виде электронной платформы речевом материале<sup>4</sup> [7]-[12].

Определение навыка того или иного человека в считывании устной речи с губ существенно зависит от состава представленного ему речевого материала. Основной целью проекта является направление внимания обучаемых на некоторые отличительные особенности сходных речевых элементов в плане выражения

В соответствие с поставленной целью исследования образования омовизем в русской речи был разработан соответствующий корпус, как информационно-справочная система в электронной форме. Данный корпус, являющийся новой функцией разработанной ранее мультимедийной программы «Читаем с губ» [13, 14] содержит некую дополнительную информацию о свойствах, входящих в него групп слов с отличием между собой графическими и фонетическими характеристиками и воспринимаемыми зрительно одинаково. Это обстоятельство является преградой на пути выбора одного из значений и приводит к неопределённости. Данный материал предназначен для ознакомления с возможными преградами на пути их визуального распознавания.

Объём корпуса составляет более 5 тысяч слов с разным графическим оформлением и существенными различиями смысловых значений и непрерывно пополняется новыми словами, представляющими интерес в плане их визуального восприятия и распознавания.

При разработке данного онлайн-корпуса были реализованы несколько подсистем, основными из которых являются система представления фонетической и артикуляционной транскрипции и подсистема поиска данных по заданным параметрам.

## Проблема прогнозирования результата визуального восприятия устной речи

Процесс визуального восприятия реципиентом устной речи, помимо указанных выше факторов, существенно зависит от уровня его словарного запаса и от его когнитивной способности правильно понимать обращённую к нему речь, опираясь на свой предыдущий опыт [15].

В устной речи практически каждый её компонент может иметь одного или несколько «двойников» в плане артикуляционных образов и отличаться другими характеристиками в плане выражения и содержания.

При прогнозировании результата визуального восприятия речевых элементов могут быть использованы формальный и корпусный подходы. Оба подхода заключаются в поиске к определённому слову всех возможных омовизем.

Для исследования были использованы односложные слова разной длины «груз» и «хруст», имеющие в своём составе такие общие виземы как [г – г', к – к', х – х'], [л – л', р – р'], [з – з', с – с', ц] и ударную гласную [у]. Применительно к слову «хруст» будет дополнительная цепочка визем – [д – д', т – т', н – н']. Соответственно разработанной в ходе проведенного исследования В. И. Бельтюковым<sup>5</sup> классификации фонем и правилу комбинаторики по заданному слову было определено количество вариантов слов со схожими с ним виземами.

Важным моментом в данном исследовании является соответствие в словах числа графем и фонем.

### Формальный подход

Формальный подход позволяет на основе комбинаторики получить все возможные варианты слов со схожими артикуляционными рисунками, среди которых лишь малая часть из них будет иметь смысловое значение.

Так, к слову «груз» было получено число вариантов  $Q$ , равное 120 речевых элементов со схожими виземами. Однако чисто формальный подход к формированию слов из набора фонем приводит к большому количеству вариантов, среди которых смысловое значение  $q$  имеют лишь 2 слова – «груз» и «клюз». Остальные слова являются псевдословами, представляющими собой группы букв / фонем, которые не составляют слово и, следовательно, не имеют смыслового значения. Главная особенность псевдослов состоит в том, что по внешнему виду они выглядят как настоящие слова, поскольку могут произноситься на определённом языке. Таким образом, доля реальных омовиземных слов будет составлять  $q / Q = 1,67\%$ , а вероятность выбора одного из двух слов –  $P = 0,5$ .

Ко второму выбранному слову «хруст», согласно представленной выше записи фонем, было определено общее число возможных вариантов речевых единиц  $Q = 720$ , а слов в этом списке всего лишь 4 ( $q = 4$ ) – «груздь», «грусть», «хруст», «хлюст» (разг.). Доля этих слов будет составлять всего  $d = 0,56\%$ , а вероятность выбора каждого из слов –  $P = 0,25$ .

Общее число вариантов слов из двух выбранных составляет 6 единиц.

<sup>4</sup> Саженин И. И. Корпусные методы в лексикографии: опыт создания модели Словарного корпуса : дисс. ...канд. филол. наук : 10.02.01. Новосибирск: НГПУ, 2013. 192 с.

<sup>5</sup> Бельтюков В. И. Чтение с губ. М.: Педагогика, 1970. 184 с.



**Корпусный подход**

Корпусный подход даёт возможность предоставить пользователю по заданному им слову варианты омовиземных слов, перечень которых будет дополнен словами, заранее сгруппированных по артикуляционной транскрипции и хранимых в БД. Как показал результат распознавания, несмотря на разное число букв и фонем к приведённым выше исходным словам будут добавлены слова «Русь» и «руст».

Объяснить это можно, во-первых, незаметным артикуляционным рисунком при формировании фонем [г], [к] и [х] в невидимой части ротовой полости, что не позволяет различить их между собой при зрительном восприятии, а также слабо выраженным артикуляционным рисунком сочетания фонем [ст] в конце слова.

Необходимо отметить, что определение слов среди общего числа вариантов комбинирования фонем, рассматривается только в рамках тех слов, которые не выходят за пределы контента корпуса.

Таким образом, при визуальном восприятии все перечисленные слова могут восприниматься одинаково (рис. 1).

В сравнении с формальным подходом корпусный подход позволяет получить более точное число слов «двойников» в большую сторону. Так при формальном подходе на приведённых примерах видно, что не все слова вошли в общий список. При заданном слове «груз» были «потеряны» слова – «груздь», «грусть», «хруст», «хлюст», а при слове «хруст» – «груз» и «хлюст». Объясняется это тем, что эти слова имеют разную длину за счёт отсутствия первых букв, с которых начинается поиск.

Общее число возможных вариантов будет составлять  $Q = 280$ , а слов в этом списке будет лишь 5. Доля реальных омовиземных слов в данном случае будет составлять около 1,8%, вероятность выбора одного из слов –  $P = 0,2$ .

Сравнение результатов доли реальных слов из общего количества вариантов речевых элементов показывает, что корпус-

ный подход даёт более полный набор омовизем в одной цепочке, чем формальный подход.

Таким образом, разработанный корпус позволяет пользователю при выборе им из списка определённого слова получить на экране из подготовленной БД полный перечень возможных вариантов слов – омовизем (рис. 2).

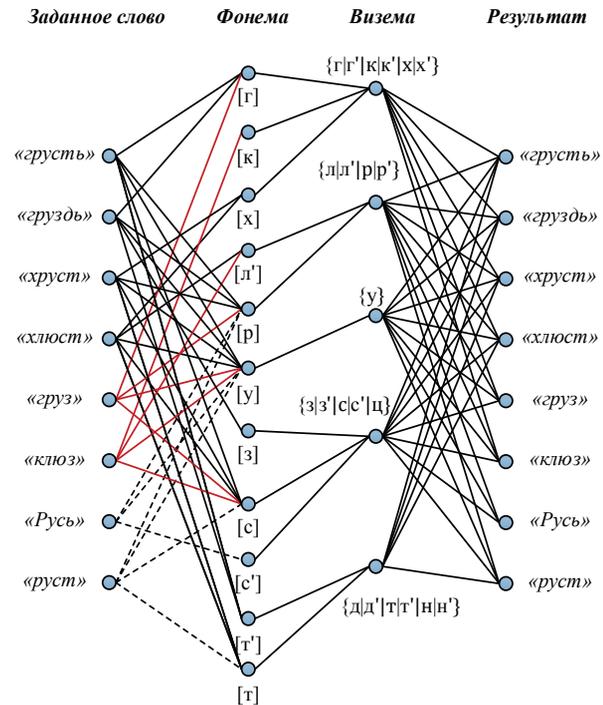


Рис. 1. Схема образования омовизем при корпусном подходе  
Fig. 1. The formation scheme of homovizems in the corpus approach

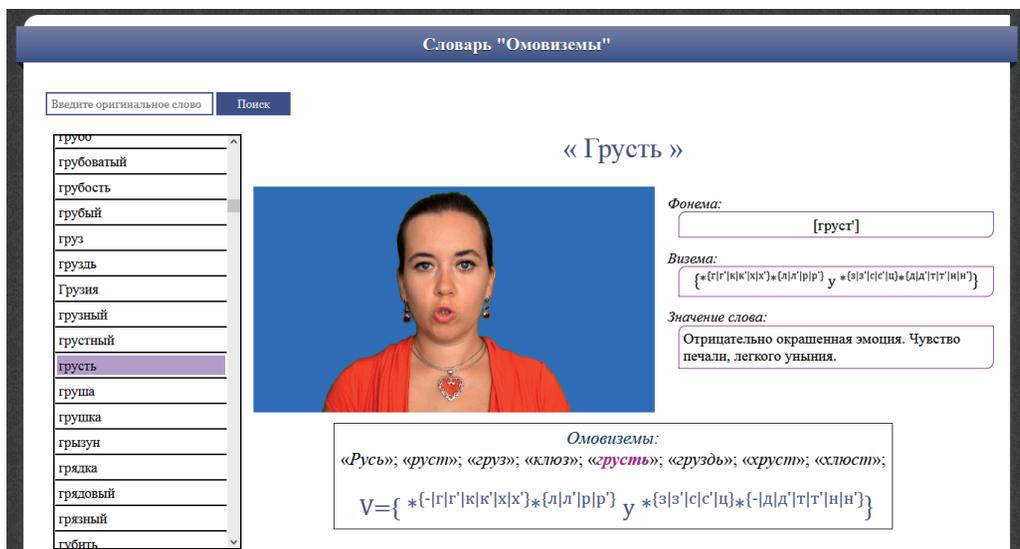


Рис. 2. Экранная форма поиска слов со схожими артикуляционными рисунками при корпусном подходе  
Fig. 2. On-screen form of word search with similar articulation patterns in corpus approach



Основная цель включения таких речевых элементов в БД состоит в направлении внимания пользователя при просмотре им представленного на экране материала на такие слова, совершенно различные с точки зрения орфографии, фонетики и их значения. Знакомство с такими «двойниками» исходного слова позволит обойти барьеры на пути визуального распознавания устной речи и снизить неопределённость при выборе одного из них.

Проведённое исследование показало сложность процесса визуального распознавания устной речи, требующего немалых затрат сил и времени для овладения этим навыком.

В последнее время во всём мире широко распространяется процесс автоматизации в различных областях, в том числе в области распознавания речи и образов. На сегодняшний день создано немало компьютерных программ для распознавания речи по артикуляции говорящего и перевода её в текстовую форму [16]-[25]. Это существенно облегчает людям со слуховой патологией общение и способствует их интеграции в мир слышащих.

Однако, как показывает опыт, полностью полагаться на вспомогательные устройства не стоит, поскольку каждый из них может выйти из строя временно или надолго по причине технической неисправности или быть вне зоны доступа. В данных случаях человек становится беспомощным в плане понимания обращённой к нему речи. Во избежание данной ситуации человек должен владеть навыком визуального восприятия устной речи, что поможет ему хоть в малой степени разобраться в сложившейся ситуации и понять смысл озвученного высказывания. Получение определённых знаний позволит ему в процессе тренировки ознакомиться с некоторыми возможными барьерами на пути распознавания устной речи с губ говорящего, приводящими к неверному или полному пониманию смысла высказывания.

## Заключение

Для интеграции людей со слуховой патологией в общество от них требуется владение навыком визуального восприятия устной речи, что определяет новые требования к их обучению в области дистанционного инклюзивного образования. Одним из важнейших аспектов оптимизации и совершенствования дистанционного учебного процесса является внедрение компьютерных технологий в практику визуального распознавания устной речи. Наличие информационной поддержки открывает широкие перспективы для реципиентов со слуховой патологией, позволяя им стать активными членами коммуникационного процесса. Дистанционное образование как его новая модель помогает расширить доступ к образованию и обучению всех желающих.

В настоящей статье рассматривается вопрос, связанный с визуальным восприятием устной речи, в процессе которого человек неизбежно сталкивается с проблемой различения слов со схожими артикуляционными рисунками (омовиземами), наличие которых в русской речи довольно велико, что подтверждено экспериментальным путём. Для проведения нашего исследования был разработан корпус омовизем на основе материала, сгруппированного в отдельные цепочки по принципу одинаковых артикуляционных оболочек, описанных согласно артикуляционной транскрипции. Создание и пополнение корпуса омовизем речевыми элементами русского языка на основе их общих визуальных характеристик является полезным, перспективным и престижным для нашей страны делом. Работа в этом направлении позволит получить новые результаты, создать полезные приложения и внесет свой вклад в решение важной социальной задачи помощи людям с ограниченными возможностями по слуху.

Полученные результаты исследования, проведённого на материале корпуса омовизем, могут найти применение в дальнейших лингвистических исследованиях и разработках программ данной тематики с привлечением людей, владеющих навыком чтения с губ и хорошо знающих все правила и тонкости данного процесса.

## Список использованных источников

- [1] Lip Reading Using Committee Networks With Two Different Types of Concatenated Frame Images / D. Jang [и др.] // IEEE Access. 2019. Vol. 7. P. 90125-90131. doi: <https://doi.org/10.1109/ACCESS.2019.2927166>
- [2] Мясоедова М. А., Мясоедова З. П. Межъязыковая омонимия как источник коммуникативного барьера при чтении с губ иноязычных слов (с позиции носителя русского языка) // Современные информационные технологии и ИТ-образование. 2020. Т. 16, № 2. С. 379-388. doi: <https://doi.org/10.25559/SITITO.16.202002.379-388>
- [3] Cutler A., Butterfield S. Word boundary cues in clear speech: A supplementary report // Speech Communication. 1991. Vol. 10, issue 4. P. 335-353. doi: [https://doi.org/10.1016/0167-6393\(91\)90002-B](https://doi.org/10.1016/0167-6393(91)90002-B)
- [4] Towards Pose-Invariant Lip-Reading / S. Cheng [и др.] // ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE Computer Society, 2020. P. 4357-4361. doi: <https://doi.org/10.1109/ICASSP40776.2020.9054384>
- [5] Ганзеева Е. О. Проблема речевосприятия и речепонимания в современной науке // Материалы ежегодной научной конференции преподавателей и аспирантов университета. Минск: МГЛУ, 2021. Ч. 1. С. 66-68. URL: <http://e-lib.mslu.by/handle/edoc/8507> (дата обращения: 28.05.2022).
- [6] Deformation Flow Based Two-Stream Network for Lip Reading / J. Xiao [и др.] // 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). Buenos Aires, Argentina: IEEE Computer Society, 2020. P. 364-370. doi: <https://doi.org/10.1109/FG47880.2020.00132>



- [7] Плунгян В. А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. № 2(16). С. 7-20. URL: <https://elibrary.ru/item.asp?id=15127694> (дата обращения: 28.05.2022).
- [8] Савчук С. О., Сичинава Д. В. Обучающий корпус русского языка и его использование в преподавательской практике // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 317-334. URL: <https://elibrary.ru/item.asp?id=18933892> (дата обращения: 28.05.2022).
- [9] Нагель О. В. Корпусная лингвистика и ее использование в компьютеризированном языковом обучении // Язык и культура. 2008. № 4. С. 53-59. URL: <https://elibrary.ru/item.asp?id=11990991> (дата обращения: 28.05.2022).
- [10] Иргизова К. В. Корпусная лингвистика в отечественном и зарубежном языкознании на современном этапе // Огарёв-online. 2019. № 6(127). С. 1-9. URL: <https://elibrary.ru/item.asp?id=39195626> (дата обращения: 28.05.2022).
- [11] Corpus Linguistics and Corpus-Based Research and Its Implication in Applied Linguistics: A Systematic Review / A.M.S. Al-Hamzi, A. Gougui, Y. Sari Amalia, T. Suhardijanto // PAROLE: Journal of Linguistics and Education. 2020. Vol. 10, no. 2. P. 176-181. doi: <https://doi.org/10.14710/parole.v10i2.176-181>
- [12] An improved automatic lipreading system to enhance speech recognition / E. Petajan [и др.] // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'88). Association for Computing Machinery, New York, NY, USA, 1988. P. 19-25. doi: <https://doi.org/10.1145/57167.57170>
- [13] Myasoyedova M. A., Myasoyedova Z. P. Computer assessment of how well a person visually recognizes verbal Russian speech // 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT). Kazakhstan, Almaty: IEEE Computer Society, 2018. P. 1-5. doi: <https://doi.org/10.1109/ICAICT.2018.8747072>
- [14] Myasoyedova M. A., Myasoyedova Z. P., Farkhadov M. P. Articulatory Uncertainty as a Result of Visual Recognition of Modified Sounds in Russian Speech // Proceedings of the 11th IEEE International Conference on Application of Information and Communication Technologies (AICT2017). Russian Federation, Moscow: IEEE Computer Society, 2017. Vol. 1. P. 169-172. doi: <https://doi.org/10.1109/ICAICT.2017.8686938>
- [15] Матвеева Л. Ю. Вероятностное прогнозирование звучащей речи: к постановке проблемы (обзор) // Саратовский научно-медицинский журнал. 2015. Т. 11, № 2. С. 216-220. URL: <https://elibrary.ru/item.asp?id=27121628> (дата обращения: 28.05.2022).
- [16] Ibrahim M. Z., Mulvaney D. J. Geometrical-based lip-reading using template probabilistic multi-dimension dynamic time warping // Journal of Visual Communication and Image Representation. 2015. Vol. 30. P. 219-233. doi: <https://doi.org/10.1016/j.jvcir.2015.04.013>
- [17] Lip reading using external viseme decoding / J. Peymanfard, M. Reza Mohammadi, H. Zeinali, N. Mozayani // 2022 International Conference on Machine Vision and Image Processing (MVIP). IEEE Computer Society, 2022. P. 1-5. doi: <https://doi.org/10.1109/MVIP53647.2022.9738749>
- [18] Adaptive Semantic-Spatio-Temporal Graph Convolutional Network for Lip Reading / C. Sheng, X. Zhu, H. Xu, M. Pietikäinen, L. Liu // IEEE Transactions on Multimedia. 2022. Vol. 24. P. 3545-3557. doi: <https://doi.org/10.1109/TMM.2021.3102433>
- [19] Qu L., Weber C., Wermter S. LipSound2: Self-Supervised Pre-Training for Lip-to-Speech Reconstruction and Lip Reading // IEEE Transactions on Neural Networks and Learning Systems. 2022. P. 1-11. doi: <https://doi.org/10.1109/TNNLS.2022.3191677>
- [20] Training Strategies for Improved Lip-Reading / P. Ma [и др.] // 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE Computer Society, 2022. P. 8472-8476. doi: <https://doi.org/10.1109/ICASSP43922.2022.9746706>
- [21] Deng M., Xiong S. Phoneme-based lipreading of silent sentences // 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). IEEE Computer Society, 2022. P. 206-210. doi: <https://doi.org/10.1109/IPEC54454.2022.9777317>
- [22] Prajwal K., Afouras T., Zisserman A. Sub-word Level Lip Reading With Visual Attention // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE Computer Society, 2022. P. 5152-5162. doi: <https://doi.org/10.1109/CVPR52688.2022.00510>
- [23] Wang H., Pu G., Chen T. A Lip Reading Method Based on 3D Convolutional Vision Transformer. *IEEE Access*. 2022; 10:77205-77212. doi: <https://doi.org/10.1109/ACCESS.2022.3193231>
- [24] Zhang X. Boosting Lip Reading with a Multi-View Fusion Network / X. Zhang [и др.] // 2022 IEEE International Conference on Multimedia and Expo (ICME). Taipei, Taiwan: IEEE Computer Society, 2022. P. 1-6. doi: <https://doi.org/10.1109/ICME52920.2022.9859810>
- [25] Learning from the Master: Distilling Cross-modal Advanced Knowledge for Lip Reading / S. Ren [и др.] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE Computer Society, 2021. P. 13320-13328. doi: <https://doi.org/10.1109/CVPR46437.2021.01312>

*Поступила 28.05.2022; одобрена после рецензирования 05.07.2022; принята к публикации 14.07.2022.*



**Об авторах:**

**Мясоедова Мария Александровна**, научный сотрудник, ФГБУН «Институт проблем управления им. В.А. Трапезникова Российской академии наук» (117997, Российская Федерация, г. Москва, ул. Профсоюзная, д. 65), **ORCID: <https://orcid.org/0000-0002-0317-4781>**, mariamarfi@mail.ru

**Мясоедова Зинаида Павловна**, научный сотрудник, ФГБУН «Институт проблем управления им. В.А. Трапезникова Российской академии наук» (117997, Российская Федерация, г. Москва, ул. Профсоюзная, д. 65), **ORCID: <https://orcid.org/0000-0001-6022-3503>**, mzinap@mail.ru

*Все авторы прочитали и одобрили окончательный вариант рукописи.*

## References

- [1] Jang D., Kim H., Je C., Park R., Park H. Lip Reading Using Committee Networks With Two Different Types of Concatenated Frame Images. *IEEE Access*. 2019; 7:90125-90131. (In Eng.) doi: <https://doi.org/10.1109/ACCESS.2019.2927166>
- [2] Myasoedova M.A., Myasoedova Z.P. Interlingual Homonymy Hinders Communication when a Person Reads Foreign Words from the Lips (from the Position of a Native Russian Speaker). *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2020; 16(2):379-388. (In Eng.) doi: <https://doi.org/10.25559/SITITO.16.202002.379-388>
- [3] Cutler A., Butterfield S. Word boundary cues in clear speech: A supplementary report. *Speech Communication*. 1991; 10(4):335-353. (In Eng.) doi: [https://doi.org/10.1016/0167-6393\(91\)90002-B](https://doi.org/10.1016/0167-6393(91)90002-B)
- [4] Cheng S., et al. Towards Pose-Invariant Lip-Reading. *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE Computer Society, Barcelona, Spain; 2020. p. 4357-4361. (In Eng.) doi: <https://doi.org/10.1109/ICASSP40776.2020.9054384>
- [5] Ganzeyeva E.O. *Problema rechevospriyatyi i recheponimaniya v sovremennoy nauke* [The problem of speech perception and speech understanding in modern science]. *Materialy yezhegodnoy nauchnoy konferentsii prepodavateley i aspirantov universiteta* = Proceedings of the annual scientific conference of teachers and graduate students of the university. Minsk State Linguistic University, Minsk; 2021. Part 1. p. 66-68. Available at: <http://e-lib.mslu.by/handle/edoc/8507> (accessed 28.05.2022). (In Russ., abstract in Eng.)
- [6] Xiao J., Yang S., Zhang Y., Shan S., Chen X. Deformation Flow Based Two-Stream Network for Lip Reading. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE Computer Society, Buenos Aires, Argentina; 2020. p. 364-370. (In Eng.) doi: <https://doi.org/10.1109/FG47880.2020.00132>
- [7] Plungyan V.A. *Korpus kak instrument i kak ideologiya: o nekotorykh urokakh sovremennoy korpusnoy lingvistiki* [Corpus as a tool and as an ideology: on some lessons of modern corpus linguistics]. *Russkij jazyk v nauchnom osveshchenii* = Russian Language and Linguistic Theory. 2008; (2):7-20. Available at: <https://elibrary.ru/item.asp?id=15127694> (accessed 28.05.2022). (In Russ.)
- [8] Savchuk S.O., Sichinava D.V. *Obuchayushchiy korpus russkogo yazyka i yego ispol'zovaniye v prepodavatel'skoy praktike* [Russian Educational Corpus and its use in teaching practice]. *Natsional'nyy korpus russkogo yazyka: 2006-2008. Novyye rezultaty i perspektivy* = Russian National Corpus: 2006-2008. New results and prospects. Saint Petersburg, Nestor-Istoriya Publ.; 2009. p. 317-334. Available at: <https://elibrary.ru/item.asp?id=18933892> (accessed 28.05.2022). (In Russ.)
- [9] Nagel' O.V. *Korpusnaya lingvistika i yeye ispol'zovaniye v komp'yuterizirovannom yazykovom obuchenii* [Corpus linguistics and its use in computer-based language teaching]. *Jazyk i kul'tura* = Language and Culture. 2008; (4):53-59. Available at: <https://elibrary.ru/item.asp?id=11990991> (accessed 28.05.2022). (In Russ., abstract in Eng.)
- [10] Irgizova K.V. *Korpusnaya lingvistika v otechestvennom i zarubezhnom yazykoznanii na sovremennoy etape* [Current state of Russian and international corpus linguistics]. *Ogarev-online*. 2019; (6):1-9. Available at: <https://elibrary.ru/item.asp?id=39195626> (accessed 28.05.2022). (In Russ., abstract in Eng.)
- [11] Al-Hamzi A.M.S., Gougui A., Sari Amalia Y., Suhardijanto T. Corpus Linguistics and Corpus-Based Research and Its Implication in Applied Linguistics: A Systematic Review. *PAROLE: Journal of Linguistics and Education*. 2020; 10(2):176-181. (In Eng.) doi: <https://doi.org/10.14710/parole.v10i2.176-181>
- [12] Petajan E., Bischoff B., Bodoff D., Brooke N.M. An improved automatic lipreading system to enhance speech recognition. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'88)*. Association for Computing Machinery, New York, NY, USA; 1988. p. 19-25. (In Eng.) doi: <https://doi.org/10.1145/57167.57170>
- [13] Myasoyedova M.A., Myasoedova Z.P. Computer assessment of how well a person visually recognizes verbal Russian speech. *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE Computer Society, Almaty, Kazakhstan; 2018. p. 1-5. (In Eng.) doi: <https://doi.org/10.1109/ICAICT.2018.8747072>
- [14] Myasoyedova M.A., Myasoyedova Z.P., Farkhadov M.P. Articulatory Uncertainty as a Result of Visual Recognition of Modified Sounds in Russian Speech. *Proceedings of the 11th IEEE International Conference on Application of Information and Communication Technologies (AICT2017)*. IEEE Computer Society, Moscow, Russian Federation; 2017. Vol. 1. p. 169-172. (In Eng.) doi: <https://doi.org/10.1109/ICAICT.2017.8686938>



- [15] Matveeva L.Y. *Veroyatnostnoye prognozirovaniye zvuchashchey rechi: k postanovke problemy (obzor)* [Probable prognosis of speech: the issue considering (review)]. *Saratovskiy nauchno-meditsinskiy zhurnal = Saratov Journal of Medical Scientific Research*. 2015; 11(2):216-220. Available at: <https://elibrary.ru/item.asp?id=27121628> (accessed 28.05.2022). (In Russ., abstract in Eng.)
- [16] Ibrahim M.Z., Mulvaney D.J. Geometrical-based lip-reading using template probabilistic multi-dimension dynamic time warping. *Journal of Visual Communication and Image Representation*. 2015; 30:219-233. (In Eng.) doi: <https://doi.org/10.1016/j.jvcir.2015.04.013>
- [17] Peymanfard J., Reza Mohammadi M., Zeinali H., Mozayani N. Lip reading using external viseme decoding. *2022 International Conference on Machine Vision and Image Processing (MVIP)*. IEEE Computer Society; 2022. p. 1-5. (In Eng.) doi: <https://doi.org/10.1109/MVIP53647.2022.9738749>
- [18] Sheng C., Zhu X., Xu H., Pietikäinen M., Liu L. Adaptive Semantic-Spatio-Temporal Graph Convolutional Network for Lip Reading. *IEEE Transactions on Multimedia*. 2022; 24:3545-3557. (In Eng.) doi: <https://doi.org/10.1109/TMM.2021.3102433>
- [19] Qu L., Weber C., Wermter S. LipSound2: Self-Supervised Pre-Training for Lip-to-Speech Reconstruction and Lip Reading. *IEEE Transactions on Neural Networks and Learning Systems*. 2022. p. 1-11. (In Eng.) doi: <https://doi.org/10.1109/TNNLS.2022.3191677>
- [20] Ma P., Wang Y., Petridis S., Shen J., Pantic M. Training Strategies for Improved Lip-Reading. *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE Computer Society; 2022. p. 8472-8476. (In Eng.) doi: <https://doi.org/10.1109/ICASSP43922.2022.9746706>
- [21] Deng M., Xiong S. Phoneme-based lipreading of silent sentences. *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*. IEEE Computer Society; 2022. p. 206-210. (In Eng.) doi: <https://doi.org/10.1109/IPEC54454.2022.9777317>
- [22] Prajwal K., Afouras T., Zisserman A. Sub-word Level Lip Reading With Visual Attention. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, New Orleans, LA, USA; 2022. p. 5152-5162. (In Eng.) doi: <https://doi.org/10.1109/CVPR52688.2022.00510>
- [23] Wang H., Pu G., Chen T. A Lip Reading Method Based on 3D Convolutional Vision Transformer. *IEEE Access*. 2022; 10:77205-77212. (In Eng.) doi: <https://doi.org/10.1109/ACCESS.2022.3193231>
- [24] Zhang X., et al. Boosting Lip Reading with a Multi-View Fusion Network. *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE Computer Society, Taipei, Taiwan. 2022. p. 1-6. (In Eng.) doi: <https://doi.org/10.1109/ICME52920.2022.9859810>
- [25] Ren S., Du Y., Lv J., Han G., He S. Learning from the Master: Distilling Cross-modal Advanced Knowledge for Lip Reading. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Nashville, TN, USA. 2021. p. 13320-13328. (In Eng.) doi: <https://doi.org/10.1109/CVPR46437.2021.01312>

Submitted 28.05.2022; approved after reviewing 05.07.2022; accepted for publication 14.07.2022.

#### About the authors:

**Maria A. Myasoedova**, Researcher, V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences (65 Profsoyuznaya St., Moscow 117997, Russian Federation), ORCID: <https://orcid.org/0000-0002-0317-4781>, mariamarfi@mail.ru

**Zinaida P. Myasoedova**, Researcher, V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences (65 Profsoyuznaya St., Moscow 117997, Russian Federation), ORCID: <https://orcid.org/0000-0001-6022-3503>, mzinap@mail.ru

All authors have read and approved the final manuscript.

