Санкт-Петербургский государственный университет

**ЗАТЕВАЛОВА Елена Сергеевна**

**Выпускная квалификационная работа**

**Проявление эффекта Мак-Гурка при восприятии**

**аудиально-визуальных стимулов носителями русского языка**

Уровень образования: бакалавриат

Направление 45.03.02 «Лингвистика»

Основная образовательная программа СВ.5051. «Теория и практика межкультурной коммуникации (английский язык)»

Научный руководитель:

к.ф.н., доцент, Кафедра общего языкознания имени Л.А. Вербицкой,

Риехакайнен Елена Игоревна

Рецензент:

к.ф.н., доцент, ФГАОУВО

«Пермский государственный национальный исследовательский университет»

Павлова Дарья Сергеевна

Санкт-Петербург

2023

Оглавление

[Введение 4](#_Toc136382852)

[Глава 1. Аудиовизуальная интеграция и эффект Мак-Гурка 7](#_Toc136382853)

[1.1. Объединение информации из разных каналов восприятия 7](#_Toc136382854)

[1.2. Многомодальное объединение в сфере образования и когнитивная теория мультимедийного обучения 10](#_Toc136382855)

[1.3. Аудиовизуальная интеграция 14](#_Toc136382856)

[1.4. Влияние фактора родного языка на восприятие кросс-модальных данных 20](#_Toc136382857)

[1.5. Другие факторы, влияющие на проявление эффекта Мак-Гурка 26](#_Toc136382858)

[1.6. Аудиовизуальная интеграция у детей 29](#_Toc136382859)

[1.7. Исследования на материале русского языка 32](#_Toc136382860)

[1.8. Выводы по главе 36](#_Toc136382861)

[Глава 2. Восприятие аудиовизуальных стимулов носителями русского языка 39](#_Toc136382862)

[2.1. Эксперимент 1 39](#_Toc136382863)

[2.1.1. Цель 39](#_Toc136382864)

[2.1.2. Стимулы 39](#_Toc136382865)

[2.1.3. Процедура эксперимента 40](#_Toc136382866)

[2.1.4. Участники 42](#_Toc136382867)

[2.1.5. Принципы обработки данных, аналитика 42](#_Toc136382868)

[2.1.6. Результаты: различия по группам 43](#_Toc136382869)

[2.1.7. Результаты: анализ количества ошибок 45](#_Toc136382870)

[2.1.8. Выводы по ошибкам 47](#_Toc136382871)

[2.1.9. Результаты: влияние предпочтительной модальности восприятия 48](#_Toc136382872)

[2.1.10. Общие выводы по Эксперименту 1 49](#_Toc136382873)

[2.2. Эксперимент 2 51](#_Toc136382874)

[2.2.1. Цель 51](#_Toc136382875)

[2.2.2. Стимулы 51](#_Toc136382876)

[2.2.3. Процедура эксперимента 52](#_Toc136382877)

[2.2.4. Участники 53](#_Toc136382878)

[2.2.5. Принципы обработки данных, аналитика 53](#_Toc136382879)

[2.2.6. Результаты 54](#_Toc136382880)

[2.2.7. Выводы по Эксперименту 2 56](#_Toc136382881)

[2.3. Эксперимент 3 56](#_Toc136382882)

[2.3.1. Цель 56](#_Toc136382883)

[2.3.2. Процедура 57](#_Toc136382884)

[2.3.3. Участники 57](#_Toc136382885)

[2.3.4. Принципы обработки данных, аналитика 58](#_Toc136382886)

[2.3.5. Результаты 58](#_Toc136382887)

[2.3.6. Выводы 60](#_Toc136382888)

[Заключение 62](#_Toc136382889)

[Список литературы 66](#_Toc136382890)

# Введение

Восприятие информации является процессом, влияющим на всю жизнь человека и во многом определяющим успех коммуникации. С раннего возраста мы учимся познавать мир через **каналы восприятия** с помощью органов чувств. Сегодня учёные активно изучают способы обработки различной информации человеком, а новейшие разработки влияют на медицину, образование, развитие систем синтеза и распознавания речи, которые облегчают пользователю работу, которая ранее подразумевала бы ввод информации вручную. Несомненно, методы получения, обработки и вывода информации рассматриваются многими науками с самых разных точек зрения и по сей день. Психолингвистика не стала исключением.

Это исследование рассматривает некоторые принципы работы познавательных процессов и их связь с ведущими каналами восприятия. Речь идёт в том числе о различных аспектах восприятия, или **перцепции**, то есть процесса обработки сенсорной информации, в результате которой мы интерпретируем окружающий мир как совокупность предметов и событий[[1]](#footnote-1), формируя идеальную модель (субъективный образ) объективно существующей действительности.

В работе рассматривается такой феномен восприятия, как **эффект Мак-Гурка**, а также интерференция, многомодальное объединение и аудиовизуальная интеграция, связанные с ним. Первоочередной вопрос, поставленный нами, состоит в определении того, что оказывает влияние на силу и частоту проявлений этого эффекта. **Гипотеза**, выдвинутая в работе, заключается в том, что существуют определённые факторы (например, возраст и пол слушающего или диктора, а также определённые сочетания слогов), которые провоцируют наиболее сильные проявления эффекта Мак-Гурка.

**Актуальность** исследования обусловлена важностью изучения вопросов, связанных с мышлением и восприятием как для описания познавательных процессов человека, так и для развития современных речевых технологий, а также активным развитием этой области в последние годы. Несмотря на большое количество работ, посвящённых проявлению эффекта Мак-Гурка у носителей разных языков, до недавнего времени подобных исследований на материале русского языка не существовало. **Целью** данного исследования является изучение эффекта Мак-Гурка и выявление факторов, влияющих на его проявление.

Для её достижения необходимо решить следующие **задачи**:

* описать феномен аудиовизуальной интеграции и эффект Мак-Гурка как одно из его проявлений;
* изучить исследования эффекта Мак-Гурка, проведенные на материале разных языков;
* провести экспериментальное исследование восприятия неконгруэнтных аудиовизуальных стимулов носителями русского языка;
* выявить факторы, оказывающие влияние на проявление эффекта Мак-Гурка у носителей русского языка.

**Объектом** исследования является проявление эффекта Мак-Гурка у носителей русского языка, а **предметом** – факторы, влияющие на его проявление.

Основным **методом** исследования является психолингвистический эксперимент, включающий в себя задание на восприятие аудиовизуальных стимулов.Следовательно, **материалом** работы являются данные, полученные в ходе эксперимента.

**Новизна** работы в первую очередь обусловлена практически полным отсутствием подобных исследований на материале русского языка. В будущем данное исследование могло бы дополнить знания о процессах мультимодальной обработки в целом и эффекте Мак-Гурка в частности как в кросс-языковой, так и в сугубо русскоязычной среде, что определяет **теоретическую значимость** работы, и внести **практический вклад** в такие сферы, как дубляж фильмов и распознавание речи человека компьютером.

Промежуточные результаты исследования обсуждались на семинарах исследовательской группы «Спектр» (руководитель – Е.И. Риехакайнен), а также были представлены на XXIV международной конференции SPECOM в докладе «Should we believe our eyes or our ears? Processing incongruent audiovisual stimuli by Russian listeners» и на X Всероссийской научно-практической конференции «Центральные механизмы речи» имени проф. Н.Н. Трауготт в докладе «Проявление эффекта Мак-Гурка при восприятии речи носителями русского языка». Исследование выполняется при поддержке гранта №21-18-00429 Российского научного фонда (рук. – Т.В. Черниговская).

# Глава 1. Аудиовизуальная интеграция и эффект Мак-Гурка

## 1.1. Объединение информации из разных каналов восприятия

Человеческий мозг способен воспринимать и анализировать не только физические сигналы, которые поступают от органов чувств. Познавая, мы формируем целостную картину окружающего мира и положения человека в нем. Однако процессы, связанные с восприятием, до конца не изучены. Вот почему сегодня перед исследователями стоит целый комплекс задач, которые в той или иной степени подразумевают под собой изучение процессов восприятия информации. Обычно восприятие предполагает взаимодействие информации, поступающей по разным каналам (модальностям). Объединение информации, полученной из разных модальностей, — это процесс, с помощью которого данные из различных информационных каналов интегрируются в единое представление.[[2]](#footnote-2) Такое синергетическое использование и взаимное влияние друг на друга данных, получаемых из разных модальностей, называется **мультимодальным (многомодальным) объединением (интеграцией)**.[[3]](#footnote-3) Причём этот термин может относиться к любой стадии процесса, где присутствует реальная комбинация данных, поступающих из различных источников восприятия. Мультимодальное объединение, таким образом, зачастую становится неотъемлемой частью перцептивной обработки, и доказательства этому существуют практически для любой комбинации органов чувств. Так, Дж. Готтфрид и Р. Долан в своём исследовании[[4]](#footnote-4) в качестве стимулов предлагали участникам различные запахи и изображения, которые появлялись как вместе, так и по отдельности. Манипулируя степенью семантического соответствия между парами запах-изображение, они выявили, что процесс обонятельной перцепции облегчался в паре с конгруэнтным (т.е. соответствующим, передающим то же содержание) стимулом (по сравнению с неконгруэнтным). Таким образом, исследователи продемонстрировали существование в мозгу обонятельно-зрительной интеграции. Исследование М. Шанкар, К. Левитана и др.[[5]](#footnote-5) на примере манипуляций с цветом и этикеткой шоколадных конфет M&Ms показало, что эти визуальные факторы могут влиять на восприятие вкуса продукта. Другое исследование[[6]](#footnote-6) проверяло зависимость восприятия хрусткости и свежести картофельных чипсов от звуков, издаваемых во время надкусывания, и выявило корреляцию между слуховыми сигналами и оценкой пищи. Таким образом, существует множество доказательств существования различных вариантов многомодального объединения.

Однако несмотря на наличие исследований, посвящённых разным вариантам мультисенсорных взаимодействий, одним из наиболее изученных среди них остаётся взаимодействие между слухом и зрением. И неслучайно, учитывая, как часто в процессе коммуникации речь поступает от человека многомодально, то есть сразу по нескольким каналам (в нашем случае, например, в виде аудио- и визуальной информации)[[7]](#footnote-7). Сегодня мультимодальное объединение часто становится объектом исследований, посвящённых распознаванию речи компьютером, поскольку реализация полимодального представления информации оказалась эффективной в области искусственного интеллекта при проектировании многомодальных коммуникационных систем и человеко-машинных интерфейсов. Поэтому сегодня крайне актуален, в частности, анализ **стратегий объединения информации** из разных модальностей. Такой синтез может происходить на различных уровнях представления данных и с использованием следующих фундаментальных стратегий:

1) на уровне признакового описания (*feature level*), называемом *ранним объединением* (*early fusion*);

2) на семантическом уровне принятия решений (*decision level*), называемом *поздним объединением* (*late fusion*);

3) с использованием *гибридного подхода* (*hybrid approach*)[[8]](#footnote-8).

Работы, посвящённые методам многомодальной интеграции, сегодня позволяют специалистам программировать информационные системы таким образом, чтобы они могли распознавать данные, основываясь на сочетании и обработке различной информации. В дальнейшем, однако, многомодальное объединение нами будет рассматриваться именно с точки зрения восприятия информации человеком.

Итак, будучи одним из наиболее распространённых вариантов мультимодального объединения, аудиовизуальная интеграция (далее – АВИ) сегодня является неотъемлемой частью жизни каждого человека. Более того, случаи одновременного использования аудио- и визуальных данных встречаются не только в процессе межличностной коммуникации, а значит, при этом необязательно участие другого человека. Например, взаимодействие звуковой и зрительной информации (а иногда и данных, поступающих через другие каналы восприятия) происходит при вождении автомобиля или просмотре фильма. Хотя в некоторых случаях восприятие лишь через одну модальность (перцептивную) возможно, обработка данных становится точнее и происходит быстрее, если задействовано несколько перцептивных каналов. Это касается и восприятия речи: в ситуации непосредственного контакта с говорящим мы не только используем звуковую информацию, но и обрабатываем жесты и мимику собеседника и соотносим их с тем, что слышим[[9]](#footnote-9). Поэтому то, видим мы говорящего или нет, влияет и на восприятие его речи в целом. Вот почему зачастую процесс восприятия речи тесно связан с многомодальным объединением.

## 1.2. Многомодальное объединение в сфере образования и когнитивная теория мультимедийного обучения

Вопросы взаимодействия различных модальностей часто затрагиваются в исследованиях, посвящённых обучению. Так, многие специалисты в связи с изменением общественного сознания и внедрением новых технологий отмечали важность использования в процессе обучения данных из различных модальностей, а не только из одной. В частности, Т. Олдаковски[[10]](#footnote-10) подчеркивает, что система, в которой детей учат воспринимать информацию (чаще всего тексты) лишь через один канал восприятия (зрительный), не может подготовить их к тому, что в реальном мире за пределами класса от них будут ждать совершенно иного. Соответственно, использование многомодальных средств обучения может существенно улучшить систему и обучить детей таким образом, чтобы они соответствовали современным запросам рынка труда. Однако и он, и исследователи из других стран, например, Е. Свардемо Оберг и А. Окерфельдт[[11]](#footnote-11) из Швеции, проанализировали результаты представителей разных школ и пришли к выводу, что ни учителя, ни ученики пока не умеют эффективно использовать мультимодальные средства обучения. Так что, хотя технологии позволяют сделать это, образовательная система по-прежнему не приспособлена к применению мультимодальной интеграции в сфере образования. И всё же учёные, в том числе и российские[[12]](#footnote-12), продолжают разрабатывать новые методы с использованием, в частности, АВИ для обеспечения эффективных результатов в классе. Вероятно, такое разнообразие исследований, связанных с обучением, связано и с данными о корреляции между эффективностью восприятия и совмещением информации, полученной из разных модальностей. Так, в работах А. Бэдли и Гр. Хитча в 1974 г. была предложена модель рабочей памяти, позволяющая анализировать процесс восприятия речевых сигналов как в визуальной, так и акустической форме[[13]](#footnote-13). В этой связи нельзя не упомянуть когнитивную теорию мультимедийного обучения (CTML) Р. Майера[[14]](#footnote-14), в основе которой лежит три общих принципа, относящихся к теории познания. Эти принципы опираются на результаты исследований физиологов и психологов и на собственные экспериментальные исследования автора теории. Они заключаются в следующем:

1) Dual-channel: дидактическое воздействие имеет два компонента: вербальную (auditory/verbal) и визуальную (visual/pictorial); они обрабатываются раздельно через два разных канала обработки, соответственно слуховой и зрительный;

 2) Limit capacity, т.е. ограниченная пропускная способность этих каналов: за определённое время можно успешно воспринять лишь ограниченное количество когнитивных единиц;

3) Learning is an active process: обучение как процесс взаимодействия ментальной системы с потоком дидактического материала включает фильтрацию, сепарацию, организацию и интеграцию поступившей информации на основе имеющихся знаний.

Схематично усвоение мультимедийных данных по Р. Майеру можно представить следующим образом:

Рис. 1. Усвоение мультимедийных данных по Р. Майеру

Когнитивная теория мультимедийного обучения до сих пор активно изучается и используется при подготовке дидактических материалов разными специалистами. Хотя традиционно при подаче материала основным компонентом считается проговаривание текста вслух, нельзя недооценивать важность визуализации. Более того, представляется сложным, если не невозможным, полностью отказаться от одной составляющей в пользу другой. Современные средства обучения позволяют повысить его эффективность с помощью одновременного воздействия сразу на несколько органов чувств, а проблемы и задачи, возникающие при этом, активно обсуждаются и постоянно находят новые решения. Большинство педагогов и учёных сходятся в том, что для эффективного усвоения информации очень важно грамотно сочетать данные разных типов.

Это подтверждают и недавние исследования с опорой на CTML, проведённые на материале русского языка. Так, изучая способы более эффективного представления информации, Е.И. Риехакайнен и Л.Л. Скоробогатько провели исследование[[15]](#footnote-15), в котором предлагали четыре различных варианта изложения одного и того же материала: 1) просто устный текст, 2) устный текст и слайды презентации, на которых он был записан дословно, 3) устный текст и презентация с кратким изложением материала и 4) устный текст с его кратким письменным изложением и иллюстрациями. Затем 132 участника (по 33 человека на каждый из четырёх возможных вариантов изложения материала) оценивали чёткость изложения, формат изложения и интересность материала. В результате выяснилось, что сочетание визуального и аудиального форматов являлось более эффективным способом подачи материала, чем просто аудиоформат, а краткое изложение воспринималось лучше, чем дословная запись. Таким образом выяснилось, что полимодальный (т.е. сочетающий в себе письменное и устное изложение) текст воспринимается лучше.

Т.Е. Петрова[[16]](#footnote-16) в своём исследовании, также опирающемся на CTML, на материале биографий четырёх русских писателей изучала, как читатели обрабатывают и осмысляют информацию в разных форматах (инфографика, аудиотекст, инфографика в сочетании с аудиотекстом и письменный формат). В этом исследовании участвовали как носители русского языка, так и изучающие русский язык как иностранный. Результаты показали, что аудиотекст воспринимался хуже всего в обеих группах. Иностранцы лучше всего усваивали письменный текст, тогда как в группе носителей русского статистически значимых различий в обработке остальных вариантов текста (письменного или полимодального) выявить не удалось. Тем не менее, автор предполагает, что понимание текста больше зависит от читателя, чем от формата самого текста. В статье Д.В. Иванько и др.[[17]](#footnote-17) утверждается, что сочетание модальностей повышает надёжность как когнитивной, так и автоматической системы в случае ошибки или сбоя. Таким образом, в целом данные экспериментальных исследований говорят в пользу применения полимодальных текстов для лучшего усвоения образовательного материала.

## 1.3. Аудиовизуальная интеграция

Ещё в прошлом веке выдвигалось множество предположений о существовании прямой корреляции между звучанием языковых единиц с одной стороны и их визуальным представлением с другой. Кроме того, существует ряд исследований, которые подтверждают улучшение точности распознавания данных при условии, что тот, кто воспринимает речь, может видеть и слышать говорящего. Так, К. Энджен, Ж. Фэлпс и др. наблюдали за взаимодействием между факторами, предположительно повышающими разборчивость речи (семантическим контекстом, чёткостью воспроизведения речи и визуальной информацией) в различных условиях. Участниками исследования были люди без нарушений слуха, которым предлагалась только аудиальная или аудиально-визуальная информация. Учёные выявили, что все три фактора действительно улучшили разборчивость текста в различных ситуациях. Более того, было установлено, что контекст, чёткость речи и визуальный ряд по-разному взаимодействуют и в каждом случае дают разную эффективность[[18]](#footnote-18). Ещё одним доказательством важности одновременной обработки звуковых и зрительных данных является исследование Н. Эрбера, который выяснил, что при восприятии речи люди во многом полагаются на визуальные подсказки и редко опираются лишь на аудиальную информацию, если при этом доступна зрительная, особенно если появляются шумы, которые становятся помехами для звукового восприятия информации. Чем сильнее был шум, тем больше участники эксперимента полагались на то, что видели[[19]](#footnote-19). Другой эксперимент, проведённый и описанный Н. Тай-Мюррей и др., исследовал распознавание речи у людей в возрасте от 22 до 92 лет и продемонстрировал, что аудиальные и визуальные данные дополняют друг друга и повышают эффективность восприятия одинаково на протяжении всей взрослой жизни[[20]](#footnote-20).

Таким образом, большинство учёных сходятся в том, что визуальная информация о звуках влияет на их слуховое восприятие. Такое сочетание и взаимное влияние зрительной и звуковой информации называется **аудиовизуальной интеграцией**. Важно понимать, что это взаимодействие происходит не только на высоких уровнях восприятия (таких как обработка текстов). Например, К. Грант, Б. Уолден и Ф. Сейтз в своём исследовании изучали, как люди с нарушениями слуха воспринимают разные типы информации на примере слогов (людям предлагали только звук, только артикуляцию или их сочетание). Большая часть участников эксперимента лучше воспринимала именно сочетание аудиальной и визуальной информации. Более того, в результате учёные выяснили, что обучение чтению по губам и АВИ могут быть полезны для некоторых людей: распознавание согласных у таких участников улучшилось на 26%[[21]](#footnote-21).

Артикуляция – не единственный тип визуальных данных, который влияет на восприятие речи при АВИ. Так, С. Кэлли и др. изучали, есть ли корреляция между пониманием речи и жестами рук в ситуациях естественной коммуникации. В эксперименте пятнадцать участников наблюдали за жестами говорящего при воспроизведении речи. Жест повлияли на восприятие речи: в частности, имела место семантическая интеграция, когда жесты не соответствовали речи. Кроме того, результаты исследования предполагают, что жесты могут интегрироваться с речью на ранних и поздних стадиях её обработки[[22]](#footnote-22).

И всё же, несмотря на значительное количество исследований в этой области, которые ведутся ещё с середины XX века, многие вопросы, связанные с тем, как именно происходит АВИ, остаются неизученными. В частности, не до конца ясно, какие факторы могут усиливать или ослаблять влияние одних данных на другие, зависит ли это от конкретных звуков или от индивидуальных особенностей говорящего и слушателя. Кроме того, неясно, является ли аудиовизуальная интеграция отдельной стадией процесса обработки речи или лишь следствием процесса восприятия в целом[[23]](#footnote-23), одинаков ли механизм интеграции у разных людей в общем, а также в случаях с конгруэнтными (совпадающими) и неконгруэнтными (различающимися) аудиовизуальными стимулами и т. п.[[24]](#footnote-24)

Последний из упомянутых вопросов является следствием того, что исследования в этой области можно условно разделить на две группы: 1) с использованием конгруэнтных стимулов, в которых аудиальная и визуальная информация дополняют друг друга, и 2) с использованием неконгруэнтных стимулов, в которых аудиальная и визуальная информация не совпадают, противореча друг другу. Одним из преимуществ второго подхода можно считать то, что он позволяет оценить «вклад», который вносит информация, поступающая по каждому из каналов, в процесс и результат интегрального восприятия в целом. Исходя из этого можно предположить, что у разных людей процесс аудиовизуальной интеграции может происходить по-разному, и в будущем это, возможно, позволит получить более объективные данные о том, существуют ли так называемые предпочтительные модальности восприятия. Однако традиционно в подобных исследованиях используются опросники, что ставит под сомнение объективность получаемых результатов[[25]](#footnote-25).

Ещё одним нерешённым вопросом, касающимся АВИ, является то, происходит объединение информации из двух модальностей автоматически или требует дополнительных когнитивных затрат. Это противоречие может быть связано с разницей в материалах, которые берутся в основу исследований, поскольку существует множество работ, которые защищают полярные точки зрения, приводя соответствующие доказательства. Некоторые учёные, среди которых, например, К. Колин и др.[[26]](#footnote-26), С. Сото-Фарако и др.[[27]](#footnote-27), исходят из того, что АВИ является быстрой автоматической мультисенсорной интеграцией. Однако многие другие исследователи считают, что АВИ всё же требует затрат внимания. Среди них – Э. Ван дер Бюрг и др.[[28]](#footnote-28), которые использовали так называемую attentional blink paradigm (феномен заключается в том, что людям трудно наблюдать за поставленной целью, если она следует за нецелью, если время между стимулами составляет от 0 до 300 мс), чтобы доказать роль внимания в АВИ и тем самым опровергнуть точку зрения о том, что этот процесс происходит автоматически.

В тех случаях, когда стимулы в разных модальностях не совпадают, возникает вопрос, требуют ли они разных затрат на обработку и, соответственно, могут ли разные стимулы заменять друг друга в исследованиях, оценивающих когнитивные затраты при АВИ. В работе В. Браун и Дж. Стрэнд[[29]](#footnote-29) проводится прямое сравнение затрат, связанных с интеграцией конгруэнтных и неконгруэнтных стимулов, и, хотя в процессах обработки двух типов стимулов явно присутствуют различия, они, по-видимому, не различаются по затратам времени и усилий.

Аудиовизуальная интеграция применяется во многих сферах. Так, Б.М. Лобанов и др.[[30]](#footnote-30) описывают с её помощью модель автоматического синтеза речи, а Р. Тэзе и др. не просто синтезировали речь, но и задействовали 3-D персонажей с реалистичными движениями губ, демонстрируя, что компьютерные речевые стимулы потенциально могут исключить фактор диктора в ряде исследований, обеспечивая чёткие временные рамки и повторяющееся при необходимости содержание стимулов[[31]](#footnote-31). Н. Альмейда и др. также предложили модель синтеза речи, которая, в отличие от большинства других, работающих по принципу «чёрного ящика» без прямой связи с фактическим процессом производства речи, предлагает артикуляционный подход к аудиовизуальному синтезу речи, который поддерживается структурой артикуляционной фонологии[[32]](#footnote-32). Кроме того, АВИ активно используется и для распознавания речи: например, Д. Иванько и др. предлагают решение для определения команд водителя по движениям его губ и дальнейшее внедрение этой системы в приложения для распознавания речи[[33]](#footnote-33).

Хотя АВИ часто изучается в контексте восприятия отдельных слов или даже цельных мыслей, этим её изучение не ограничивается. Большая часть экспериментальных исследований, рассматривающих взаимодействие слуховой и зрительной информации на более низких лингвистических уровнях, основана на эффекте Мак-Гурка[[34]](#footnote-34). Это мультисенсорное явление, при котором несовпадение звуковых и визуальных сигналов ведёт к иллюзорному восприятию вследствие интерференции. В оригинальном эксперименте 1976 года, позже получившем множество вариаций, участникам предлагалась видеозапись человека, многократно артикулировавшего слог /*ga*/, в то время как фонограмма воспроизводила звук /*ba*/. Когда респонденты закрывали глаза и слушали фонограмму или смотрели на движения губ без звука, они точно распознавали звучавшие/артикулируемые слоги. Однако если предъявлялись неконгруэнтные стимулы, то иногда респонденты слышали звуки, которых не было ни в одном из них. Например, когда людям показывали на экране человека, артикуляция губ которого соответствовала слогу /*ga*/, и одновременно давали послушать акустический сигнал /*ba*/, большинство из них слышали совсем другой звук – /*da*/, и не осознавали несоответствия. В более поздних исследованиях[[35]](#footnote-35) участники предлагали еще одну интерпретацию – /gba/, т.е. слог, включающий как произносимый, так и артикулируемый согласный. Как показывают исследования, эффект устойчив и проявляется вне зависимости от того, знает ли участник экперимента о нём[[36]](#footnote-36).

Что касается интерференции, она является одним из феноменов, которые до сих пор широко изучаются, и не только в связи с эффектом Мак-Гурка. На данный момент причина её возникновения неизвестна, хотя существует несколько теорий, выдвигающих предположения об этом[[37]](#footnote-37). Например, существует мнение о том, что интерференция возникает из-за наличия более и менее привычных человеку действий, и чем менее привычно действие, тем больше времени на него уходит. В случае с эффектом Мак-Гурка эту теорию можно было бы интерпретировать таким образом, что интерференция объяснялась бы наличием ведущей перцептивной модальности. Так, предпочтительный канал был бы более привычным для восприятия, однако ранее проведённые исследования эту теорию опровергли. Согласно другой теории, интерференция возникает в случае, если при наличии двух задач одна должна игнорироваться, и об этом предупреждают заранее. В таком случае больших когнитивных затрат стоит игнорировать одну информацию в пользу другой, и получается, что большая часть внимания уходит как раз на игнорируемое, так что его становится сложно не замечать. Задача усложняется ещё и постоянной проверкой участником самого себя: не выполнил ли он задачу случайно? Однако такую гипотезу пока нельзя ни подтвердить, ни опровергнуть, поскольку она связана с работой сознания, тоже ещё не изученного[[38]](#footnote-38).

Таким образом, психолингвистика – не единственная сфера, изучающая Эффект Мак-Гурка и понятия, связанные с ним. Существуют, например, работы, посвящённые принципам работы мозга, в которых также затрагивается эта тема. Так, этот феномен рассматривается в ряде исследований, направленных на описание нейронных механизмов обработки слуховой и зрительной речевой информации. Например, в работе Дж. Бэсл и др. с помощью эффекта исследовалось, как слуховые и визуальные особенности бимодальных событий представлены в сенсорной памяти[[39]](#footnote-39).

## 1.4. Влияние фактора родного языка на восприятие кросс-модальных данных

Восприятие речи лингвоспецифично[[40]](#footnote-40), поэтому исследования, посвящённые эффекту Мак-Гурка, проводятся на разных языках. В частности, существуют работы на базе английского[[41]](#footnote-41) языка.

В исследовании 1993 года были выявлены межъязыковые и/или культурные различия в аудиовизуальном восприятии речи среди носителей английского и японского языков. Эксперимент показал, что японские слушатели менее подвержены влиянию визуальных сигналов, чем американцы. Было четыре группы тестируемых: 10 японцев, которым предлагались японские слоги; 10 японцев, которые воспринимали английские слоги; 10 американцев, воспринимавших японские слоги, и 10 американцев, работавших с английскими слогами. Ответы носителей английского языка показали гораздо большее проявление эффекта Мак-Гурка. Кроме того, в обеих группах наблюдалось значительное влияние языка, на котором предъявлялись стимулы (каждая группа показала больший эффект визуальных сигналов на стимулы из неродного языка). Результаты также показали некоторые акустические и артикуляционные различия между [r] и [w] в двух языках.[[42]](#footnote-42) Базируясь на этом исследовании, в результате которого выяснилось, что носители японского языка гораздо меньше ориентируются на визуальную информацию при восприятии речи, нежели носители английского языка, учёные с помощью эффекта Мак-Гурка проверяли, будет ли то же относиться к носителям китайского, живущим в Японии, сопоставляя их результаты с результатами участников-американцев, и обнаружили, что у слушателей – носителей китайского языка (возможно, в силу их культурного сходства с японцами) эффект проявлялся гораздо меньше, чем у американцев. Более того, было обнаружено влияние фактора продолжительности жизни слушающих в Японии[[43]](#footnote-43). Однако существует и другое исследование, изучавшее проявления эффекта среди носителей китайского, проживавших в США менее трёх лет, носителей китайского, проживавших в США более трёх лет, и носителей английского языка. В первом эксперименте с помощью стимулов, сочетавших звук /ba/ и видео /ga/, были исследованы поведенческие и мозговые реакции, характерные для проявления эффекта Мак-Гурка. Во втором и третьем экспериментах исследовалась роль семантики и культурной связанности в восприятии, при котором проявлялся эффект. Были выявлены различия в активности в левом полушарии головного мозга при восприятии конгруэнтных и неконгруэнтных стимулов (когда эффект проявлялся). Результаты также показали наличие электрофизиологических доказательств взаимодействия эффекта Мак-Гурка и семантической/когнитивной обработки. Однако, вопреки гипотезе авторов, между двумя китайскими группами не было обнаружено значимых различий, что противоречит результатам ранее упомянутого исследования[[44]](#footnote-44).

В другом исследовании эффект Мак-Гурка использовался для изучения языковых различий между японским и английским в аудиовизуальном восприятии речи. Участников просили идентифицировать аудио- и видеостимулы, а также конгруэнтные и неконгруэнтные слоги в аудиовизуальной презентации при различных уровнях отношения сигнал/шум. В первом эксперименте с двумя группами взрослых (носители японского и английского языков) результаты подтвердили данные о более слабом проявлении эффекта среди японцев как по проценту условно верных ответов, так и по времени реакции. Во втором эксперименте было проведено тестирование трёх возрастных групп (6, 8 и 11 лет) в каждой языковой группе. Результаты показали, что степень визуального влияния была одинаково низкой для всех шестилетних детей и увеличивалась с возрастом для англоязычных участников, особенно между 6 и 8 годами, но оставалась неизменной для японских участников. Это может быть связано с тем, что англоязычные взрослые и дети старшего возраста обрабатывали визуальную речевую информацию быстрее, чем слуховую, тогда как во времени реакции японских участников таких межмодальных различий обнаружено не было[[45]](#footnote-45).

Существует также исследование, сравнивающее проявления эффекта Мак-Гурка у носителей нидерландского языка и кантонского диалекта китайского языка. В этой работе авторы полагают, что фонология родного языка влияет на процесс обработки речи слушателем, и, соответственно, объясняют различия в восприятии стимулов у разных языковых групп специфическими для каждого конкретного языка процессами. И хотя авторы допускают возможность языковых различий у разных групп слушателей, наличие культурных перцептивных различий им представляется маловероятным. В эксперименте существовало три возможных оценки аудиовизуальных стимулов: слияние, смесь или слуховая реакция. Слияние – это такой ответ, в котором зрительная информация о месте артикуляции объединяется со слуховой информацией в один слог (например, звук /ba/ + артикуляция /da/ давала ответ /da/), смесь – это ответ, в котором визуальная информация о месте артикуляции смешивалась со аудиальными данными в двухфонемную композицию (например, /bda/), а слуховая реакция – это такая реакция, на которую визуальный стимул не оказывал влияния (в нашем примере это будет /ba/). Когда аудиальный губно-губной сочетался с визуальным переднеязычным (например, аудиальный /ba/ с визуальным /da/), носители кантонского диалекта китайского языка давали 26% смешанных ответов (по типу /bda/), 49% слитных ответов (по типу /da/) и 24% слуховых реакций (по типу /ба/). Носители же нидерландского языка дали лишь 9% смешанных ответов, 56% слияний и 35% слуховых реакций. Существенной разницы между двумя группами в случаях сочетания визуального билабиального со звуковым стимулом, соответствующим переднеязычному согласному, выявлено не было: носители кантонского диалекта дали 69% смешанных ответов, 17% слияний и 14% слуховых реакций, тогда как носители нидерландского языка дали 66% смешанных ответов, 22% слияний и 12% слуховых реакций. Таким образом, участники-кантонцы чаще (чем нидерландцы) давали смешанные ответы на стимулы, когда артикуляция переднеязычного сочеталась с аудиальным стимулом с билабиальным согласным[[46]](#footnote-46).

В эксперименте Д. Массаро и др. носителям разных языков предлагалось идентифицировать синтетические одномодальные или бимодальные речевые слоги, образованные из синтетических слуховых и зрительных /bа/-/dа/, которые повторялись пять раз. В первом эксперименте носители нидерландского языка идентифицировали проверочные слоги как /ba/ или /da/. Чтобы проверить надежность результатов, носителям нидерландского и английского языков затем было предложено полностью открытое задание (без готовых вариантов ответа, хотя часто в подобных исследованиях их указывают), и были получены аналогичные результаты. На анализ речи влияла как слуховая, так и визуальная информация. Результаты опровергли модель слухового доминирования (ADM), которая предполагает, что «вклад» визуальной составляющей речи зависит от того, насколько хорошо слышна аудиальная составляющая. Кроме того, была доказана ложность аддитивной модели восприятия (AMP), согласно которой слуховые и визуальные источники комбинируются линейно. Процесс хорошо описывает нечётко-логическая модель восприятия (FLMP), утверждающая, что множественные непрерывные источники информации оцениваются и интегрируются в восприятие речи. В целом эти результаты подтверждают предыдущие данные, полученные в ходе экспериментов с участием носителей английского, испанского и японского языков. Таким образом, анализ моделей не выявил статистически значимых различий в характере обработки информации в разных языковых группах. Небольшие различия в производительности у представителей разных языковых групп объясняются авторами как следствие различий в фонологических системах нидерландского и английского языков. Результаты этого исследования продемонстрировали, что основные механизмы восприятия речи одинаковы для разных языков[[47]](#footnote-47).

В классическом эксперименте Мак-Гурка и Макдональда, как и в подавляющем большинстве последующих, эффект проверялся на согласных. Однако существуют работы, доказывающие, что эффект обнаруживается и в случае с гласными. Это касается, в частности, исследования на базе шведского, где участникам предлагалось четыре слога (/gig/, /gyg/, /geg/ и /gøg/) с одними и теми же согласными, но разными гласными, произносимые четырьмя говорящими, в только аудиальном, только визуальном и в аудиовизуальном форматах (неконгруэнтными признаками были открытость и огубленность гласных). В случае с аудиовизуальными стимулами участники почти всегда воспринимали открытость на слух. Большинство участников воспринимали огубленность визуально, а не на слух. Это привело к слиянию, например, когда звук /geg/ в паре с артикуляцией /gyg/ преимущественно воспринимался как /gøg/. Кроме того, значительной оказалась разница между участниками женского и мужского пола: большая часть мужчин полагалась на зрение не так сильно, как женщины. Наличие огубленности (заметный признак) замечалось чаще, чем её отсутствие. Влияние оптической информации не было полностью объяснено на основе показателей успешности участников в чтении по губам по сравнению со слуховым восприятием[[48]](#footnote-48).

В другом эксперименте, уже на базе нидерландского языка, в качестве материала использовались только гласные, а именно звуки /i, y, e, Y/, которые различаются только по высоте и огубленности. В данном исследовании звук ухудшался путём добавления шума с расчётом на то, что по мере ухудшения слышимости участники будут всё больше полагаться на зрительную информацию. Как и в остальных исследованиях, посвящённых АВИ, конгруэнтные стимулы улучшали идентификацию, а неконгруэнтные приводили к слиянию и ухудшению восприятия[[49]](#footnote-49).

Кроме того, существует аналогичное исследование на материале тонов китайского языка. Поскольку аудиовизуальное восприятие речи было тщательно исследовано с точки зрения согласных и гласных, считалось, что визуальная информация от артикуляционных движений передаёт фонетическую информацию (например, место артикуляции), облегчая или меняя восприятие речи. Однако существуют данные о другом типе визуальной информации, передающем нефонетическую информацию (например, временную метку), влияющую на восприятие речи. Существование этих двух типов визуальной информации в процессе АВИ предполагает наличие двух уровней аудиовизуальной интеграции речи на разных этапах обработки. Исходя из этого, были проведены эксперименты, результаты которых показали, что визуальная информация влияет на слуховое восприятие лексического тона. Как и в большинстве работ, исследующих принципы работы АВИ, оказалось, что восприятие лексического тона улучшается при добавлении визуальной информации о соответствующем артикуляционном движении. Длительность распознавания тонов изменялась в случае, если зрительные стимулы были неконгруэнтными. Кроме того, исследования показали, что существует два типа визуальной информации, вовлеченных в процесс аудиовизуальной интеграции лексического тона китайского языка: временная (нефонетическая) реплика и тональная (фонетическая) реплика. Этот вывод также подтверждает, что аудиовизуальное восприятие речи включает нефонетический и фонетически специфический уровни обработки. Нефонетическая аудиовизуальная интеграция может начаться на ранней стадии, тогда как фонетическая аудиовизуальная интеграция может произойти на более позднем этапе обработки. Поскольку лексические тоны китайского редко берутся в основу исследований аудиовизуального восприятия речи, результаты этих экспериментов важны для дальнейших теоретических построений в области изучения АВИ[[50]](#footnote-50).

## 1.5. Другие факторы, влияющие на проявление эффекта Мак-Гурка

Традиционно в слоговых стимулах, которые применяют в качестве материала для наблюдения за проявлениями эффекта Мак-Гурка, используют гласный /a/. В работе С. Шигэно исследовалось, различается ли проявление эффекта Мак-Гурка в зависимости от гласных в слоге и при совпадении или несоответствии гласных в неконгруэнтных стимулах в японском языке. Были проведены два эксперимента для изучения процесса извлечения и интеграции аудиовизуальных фонетических признаков. Первый из них был разработан для сравнения степени проявления эффекта в случаях использования трёх разных гласных в стимулах. Результаты показали, что эффект был наибольшим в случаях с /i/, умеренным в случаях с /a/ и почти отсутствовал в случаях с /u/. Это говорит о том, что возникновение эффекта Мак-Гурка зависит от фонетических характеристик гласных и их артикуляции. Во втором эксперименте эффект Мак-Гурка наблюдался с конгруэнтными и неконгруэнтными стимулами, которые не совпадали ещё и по гласным. За исключением звука /u/, эффект проявлялся сильнее, когда гласные совпадали. Эти результаты тоже демонстрируют зависимость эффекта Мак-Гурка от гласных в слогах-стимулах[[51]](#footnote-51). Что касается согласных, то чаще всего в стимулах используются губные, переднеязычные и заднеязычные смычные согласные, т.е. такие, которые совпадают по способу, но различаются по месту артикуляции.

К. Мунхолл и др. провели ряд экспериментов, чтобы проверить наличие влияния различных временных соотношений на проявление эффекта Мак-Гурка. В первом эксперименте было показано, что для проявления эффекта Мак-Гурка не требуется строгой временной синхронности между слуховыми и зрительными речевыми стимулами. Визуальные стимулы сильно влияли на участников эксперимента только тогда, когда слуховые стимулы отставали от зрительных на 180 мс. Однако эффект Мак-Гурка проявлялся сильнее, если гласные в зрительном и звуковом стимуле совпадали. Во втором эксперименте объединялись аудиальные и визуальные стимулы, воспроизводимые в разных условиях речи (быстрая, нормальная, чёткая). Результаты показали, что манипуляции как со зрительными, так и со слуховыми условиями речи влияли на её восприятие независимо друг от друга. Кроме того, наблюдалась тенденция к тому, что стимулы, которые больше совпадали, вызывали большую интерференцию, чем те, которые не соответствовали друг другу по временным параметрам. В третьем эксперименте были объединены звуковые и зрительные стимулы в разных речевых условиях (быстрая, чёткая), и звуковые стимулы задерживались по сравнению со зрительными. Участники продемонстрировали те же результаты, что и во втором эксперименте. Результаты показывают, что участники исследования могут быть чувствительны к временной согласованности речи, но полное совпадение информации при этом не является необходимым[[52]](#footnote-52).

Существует теория о влиянии предпочтительной модальности восприятия на аудиовизуальную обработку данных и перцепцию в целом. Исходя из того, что люди могут в разной степени полагаются на ту или иную модальность восприятия, можно предположить, что перцепция напрямую зависит от канала, через который поступает информация. Это предположение подробно обсуждалось в рамках теории когнитивных стилей (стилей обучения). Однако на данный момент нет доказательств существования корреляции между ведущей перцептивной модальностью и эффективностью восприятия информации. Так, в ряде экспериментов, поставленных Л. Массой и Р. Майером, студенты колледжа (Эксперимент 1) и взрослые, не посещающие колледж (Эксперимент 2), изучали материал на уроке компьютерных технологий. Им предлагалась презентация из 31 кадра с текстом или иллюстрациями, а затем участники проходили тест. Кроме того, они участвовали в тесте на 14 когнитивных показателей, включая тесты на когнитивный стиль, предпочтения в обучении, пространственные способности и общие достижения. Результаты продемонстрировали, что так называемые «визуалы» и «аудиалы» не показали существенной разницы в тестах по результатам обучения. Таким образом, гипотеза о том, что ориентированность на ведущую перцептивную модальность может существенно улучшить показатели эффективности обучения для разных групп людей, не получила серьёзных доказательств[[53]](#footnote-53). Однако изучение эффекта Мак-Гурка могло бы предоставить новые доказательства роли когнитивных стилей и предпочтительных модальностей восприятия в аудиовизуальной обработке.

## 1.6. Аудиовизуальная интеграция у детей

Ещё одной важной областью исследований является сравнение процессов обработки информации разными группами реципиентов, в частности младенцами, детьми и подростками (в то числе в сопоставлении со взрослыми носителями языка). Некоторые исследования подтверждают, что младенцы чувствительны к конгруэнтности артикуляции и голоса в возрасте 2,5 месяцев[[54]](#footnote-54). Более того, уже в возрасте двух месяцев они могут сопоставлять слуховые и визуальные сигналы. Существуют исследования, в которых младенцам показывали видео без звука, на котором одновременно отображалось два одинаковых лица рядом друг с другом, и одно из них артикулировало один слог, а другое – другой слог. Позже младенцам показывали то же видео, но со звуком одного из слогов, и младенцы преимущественно смотрели туда, где зрительная артикуляция соответствовала звуку[[55]](#footnote-55),[[56]](#footnote-56). Дж. Холлич и др.[[57]](#footnote-57) выдвинули гипотезу о том, что помимо навыка сопоставления информации, получаемой из разных модальностей, визуальная информация облегчает слуховое восприятие речи младенцами. Также существуют данные о том, что у младенцев наблюдается эффект Мак-Гурка, как и у взрослых[[58]](#footnote-58),[[59]](#footnote-59). Д. Бернхэм и Б. Додд взяли в основу своего эксперимента парадигму привыкания (habituation-test paradigm), в которой экспериментальной группе давали неконгруэнтные стимулы, т.е. собственно стимулы Мак-Гурка (например, с артикуляцией /ga/ и звуком /bа/), а контрольной группе предлагались конгруэнтные стимулы (например, аудиовизуальный /ba/). Позже обеим группам были предъявлены только звуковые стимулы /bа/, /dа/ и /ðа/. Выяснилось, что младенцы из экспериментальной группы в тестовой фазе обрабатывали аудиосигналы для /dа/ или /ðа/ значительно дольше, чем слуховое /ba/. Автор исследования предполагает, что участники экспериментальной группы распознавали неконгруэнтные стимулы (звук /ba/ с артикуляцией /ga/) как звук /da/ или /ðа/, а не /ba/[[60]](#footnote-60).

Что касается более старших детей, большая часть исследований аудиовизуального восприятия речи говорит о том, что влияние визуальной информации у детей слабее по сравнению с подростками и взрослыми. Так, С. Дюпон и др.[[61]](#footnote-61) в 2005 году для своего эксперимента набрали восемь четырёх- и пятилетних детей и десять взрослых (все – носители французского языка, проживающие в Канаде) в качестве участников, которым предъявлялись конгруэнтные и неконгруэнтные стимулы в аудиальной, визуальной и аудиовизуальной модальностях. Взрослые записывали то, что они воспринимали, а детей просили повторять то, что, по их мнению, только что сказал говорящий. Результаты показали, что около 90% всех детских реакций на аудиовизуальные неконгруэнтные стимулы были идентичны их звуковой составляющей, а вот среди взрослых лишь около 40% ответов совпали с аудиальным стимулом. Это позволяет выдвинуть предположение о том, что взрослые гораздо больше, чем дети, ориентировались на визуальную информацию. Похожее исследование с более широким возрастным диапазоном участников (тоже франкоязычных канадцев) проводили К. Трэмблей и др. Они выделили три возрастные группы: от 5 до 9 лет (11 человек), от 10 до 14 лет (16 человек) и от 15 до 19 лет (11 человек). После предъявления стимулов участников просили повторять слоги, которые, по их мнению, они слышали. Это могли быть только звуковые стимулы /bа/ или /vа/, конгруэнтные аудиовизуальные стимулы /bа/, /va/ или неконгруэнтное сочетание звука /ba/ и артикуляции /va/. Было обнаружено, что в группе детей от 5 до 9 лет 60% ответов на неконгруэнтные стимулы были идентичны звуковой составляющей аудиовизуального стимула (стимул, сочетавший звук /ba/ и артикуляцию /va/, идентифицировали как /ba/). Этот результат резко контрастировал с показателями двух других групп (≈20%). При этом существенной разницы между группой от 10 до 14 лет и группой от 15 до 19 лет обнаружено не было. Результаты показали, что результаты аудиовизуального восприятия у детей старше 10 лет становятся похожими на показатели взрослых, а принципы аудиовизуальной обработки могут меняться в процессе развития, особенно в возрасте от 5 до 9 лет[[62]](#footnote-62). Поэтому сегодня в исследованиях эффекта Мак-Гурка стараются выделять более точные возрастные группы для определения конкретного возраста, когда дети проходят ту стадию развития, после которой аудиовизуальное восприятие уже практически не меняется.

К. Секийяма и Д. Бёрнхэм в экспериментах 2004[[63]](#footnote-63) и 2008[[64]](#footnote-64) годов набрали носителей японского и английского языков в качестве участников, которые были разделены на четыре возрастные группы, а именно 6, 8 и 11 лет и взрослые. Участников просили нажать одну из трёх кнопок для ответа (/ba/, /da/ или /ga/) сразу после того, как им предъявлялись аудиовизуальные стимулы (конгруэнтные/неконгруэнтные), только аудио- или только видеостимулы с различными соотношениями сигнал/шум. В соответствии с вышеупомянутыми исследованиями, авторы обнаружили, что у большинства детей до 6 лет в обеих языковых группах степень влияния визуальной информации была одинаково мала, а затем у носителей английского языка степень влияния визуальной составляющей на восприятие речи в целом с возрастом увеличивалась, причём это увеличение было особенно заметным между 6 и 8 годами, тогда как для японских детей степень влияния визуальной информации не менялась.

Итак, упомянутые нами работы демонстрируют некий паттерн, согласно которому визуальное влияние является более сильным в аудиовизуальном восприятии у младенцев и взрослых (за исключением носителей японского), однако у 6-летних детей при этом влияние зрительной информации слабое вне зависимости от их языковой принадлежности. Существует так называемая U-образная траектория развития аудиовизуального восприятия речи, предложенная в 2009 году С. Джергер и др.[[65]](#footnote-65) Согласно этой концепции, слуховое восприятие у младенцев, подростков и взрослых находится под сильным влиянием визуальной информации, в то время как у детей это влияние гораздо менее заметно. Однако такой паттерн в аудиовизуальном восприятии речи не является универсальным для всех языков, поскольку присутствует у носителей английского, но не работает, например, для носителей японского языка, что было продемонстрировано К. Секийямой и Д. Бёрнхэмом. Кроме того, лежащая в основе этой концепции закономерность остаётся неизученной. Так, неясно, почему именно возраст шести лет является критическим периодом для динамики развития аудиовизуального восприятия речи в случае с английским языком.

## 1.7. Исследования на материале русского языка

Существует множество исследований, которые описывают проявления эффекта Мак-Гурка у носителей разных языков, однако аналогичных экспериментов на базе русского языка до сих пор практически нет. В 2021 году было опубликовано исследование Чж. Янг[[66]](#footnote-66), в котором описана методика кросс-лингвистического исследования на материале русского языка, но лишь как одного из шести рассматриваемых в рамках эксперимента языков (автор также изучает американский английский, китайский, кантонский диалект китайского, нидерландский и японский). Все эти языки, за исключением русского, ранее уже изучались применительно к интересующему нас эффекту, однако никогда не становились частью одного исследования, хотя такое решение могло бы расширить знания о том, как эффект Мак-Гурка проявляется в случае с несколькими языками одновременно. Учитывая уже упоминавшиеся нами ранее исследования зависимости проявления эффекта Мак-Гурка от гласной в слоге-стимуле (эффект проявляется сильнее всего с /i/, средне с /a/ и хуже всего с /u/)[[67]](#footnote-67),[[68]](#footnote-68), автор предлагает гипотезу о том, что на визуальную разборчивость артикуляторных движений рта влияют гласные. Чем больше визуальная разборчивость языка, тем больше визуальной информации сможет получить тот, кто воспринимает информацию и, следовательно, тем сильнее в процессе аудиовизуального восприятия будет проявляться эффект Мак-Гурка. Шесть предлагаемых автором языков образуют три уровня визуальной разборчивости, среди них самый высокий уровень у английского и китайского, средний – у кантонского диалекта и нидерландского и самый низкий – у японского и русского. Русский язык был включен в исследование как раз из-за схожего с японским набора гласных. Предполагается, что это может служить причиной схожих проявлений эффекта Мак-Гурка для двух языков с одинаковой степенью визуальной разборчивости. Это исследование предполагает изучение восприятия, при котором реакция на стимул будет отлична и от визуального, и от аудиального сигнала (например, реакция /da/ на сочетание слухового стимула /ba/ и визуального /ga/), исключая при этом исследование комбинированных ответов (/gba/ на сочетание артикуляции /ba/ и звука /ga/). В исследовании 1993 года К. Сэкияма и Й. Тохкура[[69]](#footnote-69) обнаружили, что у носителей японского языка встречается меньше комбинированных ответов, и они предположили, что это связано с особенностями японской фонологической системы, не допускающей фонологических стечений согласных. Вероятно, это приводит к тому, что говорящие на японском языке более склонны к единой согласной артикуляции (например, только /ba/ или только /ga/), а не двойной артикуляции (например, /gba/) по сравнению с носителями других языков, которые допускают такие консонантные комплексы. Как и в японском, в мандаринском диалекте китайского языка группы согласных также не допускаются. Поэтому, чтобы избежать влияния наличия кластеров согласных, в исследовании предлагается исключить изучение комбинированных ответов на стимулы с применением эффекта Мак-Гурка.

Что касается слогов, автор предлагает включить в эксперимент только глухие смычные /pa/, /ta/ и /ka/, хотя традиционно в исследованиях используются звонкие /ba/, /da/ и /ga/. Однако в китайском языке нет звонких смычных, поэтому вполне возможно, что при предъявлении звонкого /b/ носители китайского языка будут произносить глухое /p/. Ещё одним доводом в пользу применения глухих смычных служит то, что звук, обозначаемый в нидерландском языке буквой «g», произносится носителями нидерландского языка как велярный фрикативный /x/ или /ɣ/, а не велярный смычный /g/, как в английском языке, тогда как все три предлагаемых автором слога не несут подобных сложностей ни в одном из шести языков.

Самой, вероятно, трудной для осуществления частью предлагаемого Чж. Янг исследования является не выбор стимулов, а их создание. В 2015 году Дж. Маньотти и др.[[70]](#footnote-70) подвергли критике стимулы, которые произносятся только двумя дикторами, поскольку разные говорящие могут вызывать разную степень эффекта Мак-Гурка, а фактор диктора должен быть исключён. Таким образом, Чж. Янг предлагает для создания стимулов набрать по 8 носителей для каждого языка с равномерным гендерным распределением (всего в таком случае получается 48 дикторов). При этом, по мнению автора, внешность говорящего должна соответствовать языку, на котором он/она говорит. Что касается участников исследования, их возрастной диапазон для каждой языковой группы должен составлять от 23 до 26 лет, и все они должны быть носителями, обрабатывая, таким образом, стимулы на родных им языках. Автор рассчитывает на 30 человек с равным гендерным распределением для каждого из шести языков. Все 180 предполагаемых участников эксперимента не должны иметь нарушений зрения или слуха или дефектов речи. То, каким образом это будет контролироваться, автор не поясняет.

И хотя методология этого исследования описана достаточно чётко и предусматривает множество нюансов, соблюдение всех вышеупомянутых условий на практике представляется трудноосуществимой и затратной задачей, хотя бы с точки зрения времени и усилий. Результаты данного исследования пока не представлены, но автор выдвигает предположения о том, какими они могут быть. Учитывая, что каждый участник будет даст по 80 ответов, а всего участников по 30 в каждой из шести групп, получается, что суммарные данные для последующей обработки составят 2400 ответов для каждого языка и 14400 ответов в общем. Все они будут разделены на три категории: fusion – ответы, которые не совпадают ни с одним из предложенных стимулов, auditory и visual, которые относятся соответственно к зрительной и звуковой составляющим стимула, идентифицированным участником. Автор предполагает, что для всех шести языков ответы, совпадающие с визуальными стимулами, будут встречаться реже всего, уступая ответам типа fusion и auditory, однако конкретные доли каждого типа ответов от их общего числа будут различаться в зависимости от визуальной разборчивости языка. Предполагается, что комбинированные ответы у носителей английского и китайского языков будут встречаться гораздо чаще, чем у носителей нидерландского и кантонского диалекта китайского. В японском и русском языках количество ответов смешанного типа должно быть значительно меньше, чем в случаях с нидерландским и кантонским. Кроме того, существенной разницы между английским и китайским, нидерландским и кантонским и между японским и русским языками быть не должно[[71]](#footnote-71).

На сегодняшний день это исследование является практически единственным, предлагающим учитывать данные русского языка. Однако в данном случае он был выбран скорее не сам по себе, а в качестве близкого аналога для японского. Это указывать на бесспорную необходимость наблюдения за проявлениями эффекта Мак-Гурка в русскоязычной среде. Представляется перспективным использование эффекта для изучения того, как происходит обработка информации, поступающей одновременно по слуховому и зрительному каналам восприятия. Учитывая, что для русского языка пока не существует экспериментальных данных о том, как проявляется этот эффект, в экспериментах, который описан в следующей главе данной работы, мы попытались выяснить, какие сочетания звуковых и зрительных стимулов вызывают наибольшее количество ошибок в восприятии слуховой информации носителями русского языка[[72]](#footnote-72).

## 1.8. Выводы по главе

В первой главе были рассмотрены теоретические вопросы, связанные с восприятием информации, интеграцией кросс-модальных данных, а также основные определения и положения, связанные с эффектом Мак-Гурка и факторами, влияющими на него.

Эффект Мак-Гурка является частным случаем аудиовизуальной интеграции, т.е. явления сочетания и взаимного влияния зрительной и звуковой информации, многие вопросы, связанные с которым, до сих пор не имеют чёткого ответа. В оригинальном эксперименте Мак-Гурка и Макдональда участникам предъявлялись данные, которые одновременно задействовали несколько модальностей (обычно для этого используют видеозапись и фонограмму). У некоторых участников происходил конфликт двух информационных потоков, именуемый интерференцией, и это противоречие на выходе рождало сигнал, отличный от исходных. Впоследствии выяснилось, что этот эффект крайне устойчив (проявляется даже у тех, кто о нём знает) и, как и большинство явлений, связанных с восприятием речи, является лингвоспецифичным. Помимо родного языка реципиента, существуют данные и о других факторах, влияющих на проявления эффекта. Например, некоторые исследования говорят в пользу гипотезы о зависимости проявления эффекта от возраста реципиента, хотя выделить один конкретный паттерн, который был бы верным для всех участников вне зависимости от их языковой принадлежности, пока не удалось. Также прослеживается зависимость интерференции при восприятии стимулов, направленных на возникновение эффекта Мак-Гурка, от их фонетической составляющей: некоторые гласные провоцируют большую интерференцию, чем другие. Кроме того, существует теория о влиянии предпочтительной модальности восприятия на проявление эффекта, но данная точка зрения пока не получила подтверждений.

В связи с большим количеством факторов, влияющих на этот эффект, он постоянно получает новые модификации в исследованиях на материале различных языков, однако экспериментов на материале русского языка, по всей видимости, пока не проводилось. Одним из немногих исключений можно назвать подробно описанный, но пока не воплощённый в жизнь эксперимент Чж. Янг, в котором автор берёт русский в числе шести языков, поставленных в парные параллели в зависимости от схожести их систем гласных (русский в этом случае примерно равен японскому). Автор предлагает гипотезу о том, что системы гласных влияют на восприятие визуальных стимулов, что, в конечном итоге, может стать причиной схожих проявлений эффекта Мак-Гурка для языков с одинаковой степенью визуальной разборчивости. Однако поставить такой эксперимент на практике пока представляется трудновыполнимым.

Таким образом, опираясь на теоретические предпосылки можно предположить, что и для русского языка существуют определённые закономерности восприятия, которые влияют на проявления эффекта Мак-Гурка. В следующей части работы мы попробуем проверить это на практике.

# Глава 2. Восприятие аудиовизуальных стимулов носителями русского языка

*В 1913 году за одним из столов*

*рулетки в Монте-Карло шарик*

*останавливался на чёрном 26 раз подряд.*

*И каждый раз не верящие своим глазам*

 *и верящие своей интуиции игроки*

*ставили на красное. И проигрывали.*

*И снова ставили на красное…*

## 2.1. Эксперимент 1[[73]](#footnote-73)

### 2.1.1. Цель

В первом эксперименте мы пытались выявить такие сочетания звуков и артикуляций в неконгруэнтных аудиовизуальных стимулах, для которых искажение воспринимаемой на слух информации было бы наиболее сильным для русскоговорящих слушателей. При этом мы попытались учесть фактор реципиента и проверить, будут ли отличаться результаты в зависимости от индивидуальных особенностей участников, а именно от ведущей перцептивной модальности восприятия и от возрастной группы.

### 2.1.2. Стимулы

Первая часть эксперимента включала в себя последовательность аудиовизуальных стимулов, которые были специально созданы для данного исследования. Необходимость в этом возникла в связи с отсутствием таких стимулов на русском языке (по крайней мере, в открытом доступе), тогда как, согласно теоретическим данным, описанным в предыдущей главе, фонологические особенности каждой языковой системы требуют, чтобы слоги произносились именно носителем языка. Диктором выступила женщина, носитель русского языка, 20 лет, без дефектов речи, занимающаяся лингвистикой, но не являющаяся профессиональным оратором. Поскольку этот эксперимент был пилотным, мы взяли лишь одного носителя языка в качестве диктора, хотя мы понимали, что такой выбор, вероятно, является самым большим методологическим ограничением этого эксперимента. Мы попросили диктора произнести шесть слогов русского языка, а именно /tа/, /dа/, /pа/ /bа/, /fа/ и /vа/. В отличие от классических экспериментов, направленных на изучение эффекта Мак-Гурка, в которых используются только смычные согласные, мы решили выбрать согласные, близкие по месту артикуляции и в то же время такие, артикуляция которых легко различима для слушателей. Поэтому в наш эксперимент мы включили губно-губные смычные согласные, губно-зубные щелевые согласные и переднеязычные смычные согласные. Гласной, которая использовалась во всех стимулах, была /а/, так как она традиционно используется в большинстве исследований эффекта Мак-Гурка. Каждый слог нужно было повторять по пять раз. Произнесение слогов было записано на видео. Затем из исходных видео были составлены стимулы (конгруэнтные и неконгруэнтные), часть из которых предположительно должна была спровоцировать у участников возникновение эффекта Мак-Гурка, а именно привести к неправильной интерпретации услышанного. Для этого звуковую дорожку одного слога совмещали с визуальным рядом другого. Процесс синхронизации выполнялся вручную одним из экспериментаторов, а затем проверялся другим. При этом записи с глухими согласными совмещались только с глухими согласными, а звонкие – только со звонкими. Всего получилось 18 комбинаций: шесть исходных, в которых аудио- и видеоряд совпадают (эти стимулы были контрольными), и ещё 12 стимулов с неконгруэнтными данными – по 6 для глухих и для звонких согласных.

### 2.1.3. Процедура эксперимента

Эксперимент проводился на платформе Google Forms. Он состоял из двух частей. Участники должны были прочитать инструкцию к эксперименту, предоставить согласие на участие в нём и некоторую персональную информацию (пол, возраст, год обучения (для школьников)). После этого участники должны были выбрать один из четырех вариантов опросника, которые отличались друг от друга только последовательностью стимулов (чтобы исключить влияние порядка предъявления стимулов на ответы участников, мы располагали стимулы в случайном порядке, для чего был использован интернет-сайт, выдававший случайные последовательности цифр; таким образом достигалась псевдорандомизация).

В первой части эксперимента участники должны были внимательно посмотреть и прослушать каждый из 18 стимулов и отметить, что произнёс говорящий, выбрав один из шести предложенных ответов или записав свой. Особенного внимания при этом требовал грамотный выбор формулировки задания. Важно было не акцентировать внимание респондента на какой-либо конкретной модальности восприятия. Напротив, через вопрос о том, что произнёс говорящий, была предпринята попытка уйти от прямых указаний, таких как «послушайте» или «посмотрите». Кроме того, было необходимо, чтобы участник эксперимента не просто смотрел или слушал, но совершал оба действия одновременно, поскольку в противном случае эффект Мак-Гурка, очевидно, не возник бы. Разумеется, у некоторых респондентов это провоцировало двоякое понимание задания, однако большинство из них, как показала практика, поняли задание верно.

В большинстве исследований эффекта Мак-Гурка используется парадигма вынужденного выбора (т.е. участнику предлагается выбрать один из предложенных вариантов)[[74]](#footnote-74), и мы также выбрали её, однако при этом оставили респондентам возможность написать свой ответ, если они считали, что ни один из предложенных вариантов не подходит. В качестве предлагаемых ответов на стимулы выступали 6 использованных нами в эксперименте слогов.

Вторая часть формы содержала анкету, направленную на определение предпочтительной модальности восприятия. Для этого использовалась анкета С. Ефремцева, состоящая из 48 вопросов, на которые необходимо ответить «да» или «нет». Считается, что этот опросник определяет, насколько сильно человек выразил предпочтения в отношении каждой из трёх модальностей восприятия: аудиальной, визуальной и кинестетической (в анкете по 16 вопросов для каждой модальности). Чем больше положительных ответов на вопросы из соответствующего блока даёт участник, тем большую роль для него в процессе восприятия играет именно этот перцептивный канал.

Эксперимент проводился в соответствии с Хельсинкской декларацией и действующими российскими и международными нормами, касающимися этики научных исследований. На прохождение эксперимента у участников уходило около 10 минут. С экспериментом можно ознакомиться по ссылке: https://forms.gle/riGhH1smjiPzchWdA. По окончании эксперимента участники могли оставить свой адрес электронной почты, чтобы получить результаты анкеты Ефремцева.

### 2.1.4. Участники

В эксперименте приняли участие две группы респондентов: школьники от 14 до 17 лет и взрослые от 18 до 50 лет. Изначально в каждой группе было по 70 человек. Мы решили проанализировать результаты школьников (подростков) отдельно, так как вопрос ведущей перцептивной модальности чаще всего поднимается в сфере школьного обучения. Кроме того, как уже было сказано ранее, существуют данные о том, что в процессе АВИ дети демонстрируют результаты, отличающиеся от взрослых.

### 2.1.5. Принципы обработки данных, аналитика

В рамках исследования «верными» считались те ответы, которые соответствовали звуковому сигналу в аудиовизуальном стимуле. В дальнейшем ответы, не соответствующие аудиосигналам, будут называться ошибочными. Из анализа исключались ответы тех участников, которые допустили хотя бы одну ошибку в ответах на контрольные (конгруэнтные) стимулы. Были также исключены ответы тех участников, которые после завершения эксперимента сообщили, что у них возникли проблемы с воспроизведением видео. В результате были проанализированы данные 60 школьников от 14 до 17 лет (Me=16,0; M=15,9; SD=0,9; 49 девушек) и 60 участников в возрасте от 18 до 50 лет (Me=20,5; M=23,5; SD=7,3; 46 женщин).

С помощью критерия χ2 мы сравнили количество правильных ответов на разные стимулы. С помощью критерия Манна – Уитни, критерия χ2 с поправкой на непрерывность и коэффициента ранговой корреляции Спирмена мы сравнили данные, полученные от школьников и взрослых. Корреляционный анализ применялся также для проверки гипотезы о зависимости количества правильных ответов от ведущей перцептивной модальности участника. Учитывались только ответы на целевые (неконгруэнтные) стимулы. Для статистической обработки использовалось бесплатное программное обеспечение JASP (<https://jasp-stats.org/>).

### 2.1.6. Результаты: различия по группам

Было обнаружено влияние фактора возрастной группы участника. Эффект Мак-Гурка у взрослых выражен сильнее, чем у школьников: школьники в среднем допустили меньше ошибок, хотя среднее количество правильных ответов различается незначительно (соответственно 10,5 и 9,8, W=1414; p=0,037; см. Рисунок 1).

Рисунок 1. Среднее количество верных ответов для обеих возрастных групп

Общее количество неверных ответов у школьников меньше, чем в группе взрослых (91 (12,6%) и 135 (18,8%) соответственно на 720 ответов в каждой из групп; χ²=9,704; р=0,002). Наименьшее количество правильных ответов, данных одним участником в группе школьников, составляет 2 из 12. Во взрослой группе двое участников не дали ни одного правильного ответа на целевые стимулы (при этом они правильно назвали все контрольные стимулы). Было решено не исключать их из выборки, поскольку нас интересовало, в том числе, влияние отдельных факторов (в частности, предпочтительной перцептивной модальности) на восприятие целевых стимулов.

Мы также проверили гипотезу о том, что количество правильных ответов уменьшается с возрастом. Обнаружена слабая отрицательная корреляция (Spearman's rho=-0,222; p=0,015), но следует отметить, что в обоих случаях, когда на целевые стимулы не было дано ни одного верного ответа, неверные ответы давали не самые старшие участники эксперимента, а две девушки, возраст которых составил 21 год.

Далее будут описаны результаты распознавания всех стимулов. Данные будут приведены в целом по всем 120 участникам, поскольку все тенденции, о которых пойдет речь ниже, были одинаковыми для обеих групп участников.

### 2.1.7. Результаты: анализ количества ошибок

Всего было получено 1214 правильных ответов и 226 неправильных ответов на 12 целевых стимулов. Статистически значимой разницы в количестве правильных ответов между глухими и звонкими звуками нет (χ²=0,635; p=0,426): участники допустили 119 ошибок в стимулах со звонкими согласными и 107 ошибок с глухими.

Было обнаружено влияние как звучащего (χ²=91,240; df=5; p=<0,001) и артикулируемого слогов (χ²=49,042; df=5; p=<0,001), так и их сочетаний (χ²=196,987; df=11; p=<0,001). Наибольшее количество ошибок при распознавании того, что звучало, провоцировали аудиостимулы /bа/ (76 ошибок) и /pа/ (60 ошибок) и артикуляции /vа/ (65 ошибок) и /fа/ (51 ошибка). Эти результаты свидетельствуют о том, что губно-губные смычные согласные интерпретируются хуже всего, если зрительная информация неконгруэнтна аудиальной (что подтверждено статистическим анализом: χ²=87,052; df=2; p=<0,001; см. Табл. 2).

Таблица 2. Количество верных и неверных ответов в зависимости от типа согласного в звучащем слоге

|  |  |  |
| --- | --- | --- |
|  | **Тип звучащего согласного** |  |
| **Ответ** | **Переднеязычный** | **Губной** | **Губно-зубной** | **Всего** |
| Неверный | 47 | 136 | 43 | 226 |
| Верный | 433 | 344 | 437 | 1214 |

И наоборот: наибольшее количество ошибок при восприятии аудиальных сигналов провоцировала артикуляция губно-зубных щелевых согласных в видео (χ²=44,381; df=2; р=<0,001; см. Табл. 3).

Таблица 3. Количество верных и неверных ответов в зависимости от типа согласного в артикулируемом слоге

|  |  |  |
| --- | --- | --- |
|  | **Тип артикулируемого согласного** |  |
| **Ответ** | **Переднеязычный** | **Губной** | **Губно-зубной** | **Всего** |
| Неверный | 42 | 68 | 116 | 226 |
| Верный | 438 | 412 | 364 | 1214 |

С точки зрения сочетаний аудиальных и визуальных сигналов, чаще всего ошибки встречались в парах /bа/ в аудиальной модальности и /vа/ в визуальной (57 ошибок (47%)). Кроме того, много ошибок встречалось в интерпретациях слога /pа/ в комбинации с артикуляцией /fа/ (48 ошибок (40%)) (см. Табл. 4; первые две строчные буквы в названиях стимулов соответствуют звуковому сигналу (аудиальной составляющей), а две большие буквы означают слог, который артикулировался (т.е. визуальную составляющую)). Меньше всего ошибок встречалось при интерпретации звучащего /tа/ в сочетании с визуальным /fа/ (3 ошибки, в обратном сочетании 6 ошибок), а также звучащего /vа/ в сочетании с визуальным /dа/ (5 ошибок, в обратном сочетании 8 ошибок).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Всего** | **Визуальное доминирование** | **Аудиовизуальная последовательность (оба согласных из стимулов)** | **Ответы со слогами, не совпавшими ни с одним из предложенных в стимуле** | **Ответы с заменами на парную для аудиального сигнала по признаку звонкости /****глухости согласную** |
| baDA | 19 | 4 | 4 | 11 | 0 |
| baVA | 57 | 50 | 7 | 0 | 0 |
| daBA | 14 | 6 | 8 | 0 | 0 |
| daVA | 8 | 3 | 4 | 1 | 0 |
| faPA | 16 | 13 | 2 | 0 | 1 |
| faTA | 6 | 1 | 4 | 1 | 0 |
| paFA | 48 | 43 | 5 | 0 | 0 |
| paTA | 12 | 2 | 3 | 7 | 0 |
| taFA | 3 | 3 | 0 | 0 | 0 |
| taPA | 22 | 6 | 15 | 0 | 1 |
| vaBA | 16 | 10 | 6 | 0 | 0 |
| vaDA | 5 | 1 | 0 | 3 | 1 |

Таблица 4. Неверные интерпретации неконгруэнтных стимулов

### 2.1.8. Выводы по ошибкам

Все ошибки, допущенные участниками эксперимента, можно разделить на четыре группы:

1) с визуальным доминированием: ответы с согласными, артикуляция которых была на видео (или с их парами по признаку звонкости/глухости);

2) аудиовизуальная последовательность: ответы, в которых встречаются два согласных: тот, который звучал, и тот, который артикулировался на видео (при этом возможны разные варианты: например, ба-да-ба-да-ба или бда-бда-бда-бда-бда для аудиовизуального сигнала baDА), либо сочетание согласного, встречавшегося в одном из стимулов, с каким-либо другим согласным (например, ба-ва-ба-ва-ба для сигнала bаDА);

3) ответы, содержащие только один согласный звук, который не совпадал ни с тем, который звучал, ни с тем, что был в видео (например, ва-ва-ва-ва-ва для bаDА);

4) ответы, которые содержали звонкий/глухой парный для звучавшего в стимуле согласного (например, ва-ва-ва-ва-ва для стимула fаPА).

Распределение ответов по каждому стимулу представлено в Таблице 4. В случаях со стимулами, которые провоцировали наибольшее количество неправильных ответов (bаVА и pаFА), наблюдалось в основном визуальное доминирование (замены звуками, артикулировавшимися в видео). Интересно, что случаи аудиовизуальной последовательности (при наличии в ответе двух и более согласных) наблюдались практически для всех стимулов, кроме vаDА и tаFА (ошибки в которых вообще возникали реже всего). Наибольшее количество ответов, содержащих согласный звук, не совпадающий ни с тем, который звучал, ни с тем, что был в видео, встречалось для пар bаDА и pаTА. Это были согласные /v/ и /f/ соответственно, артикуляция которых находится между местом артикуляции звука, который звучал, и звука, который артикулировался.

### 2.1.9. Результаты: влияние предпочтительной модальности восприятия

Корреляционный анализ с помощью теста Спирмена не выявил связи между количеством правильных ответов на целевые (неконгруэнтные) стимулы и предпочтением какой-либо из модальностей восприятия (выявленной с помощью опросника Ефремцева) ни у всех участников в целом, ни у школьников или взрослых по отдельности (см. Табл. 5).

Таблица 5. Корреляция между количеством правильных ответов на стимулы и количеством ответов «да» на каждую из трёх модальностей восприятия в тесте С. Ефремцева

|  |  |  |  |
| --- | --- | --- | --- |
| **Модальность** | **Все участники** | **Школьники** | **Взрослые** |
| Аудиальная | rho=-0,028p=0,759 | rho=0,012p=0,929 | rho=-0,133p=0,311 |
| Визуальная | rho=0,012p=0,893 | rho=0,014p=0,916 | rho=0,077p=0,558 |
| Кинестетическая | rho=-0,051p=0,583 | rho=-0,019p=0,888 | rho=-0,060p=0,649 |

### 2.1.10. Общие выводы по Эксперименту 1

Эксперимент показал, что наибольшее количество неверных ответов встречается в парах губного смычного согласного в звуковой части стимула и губно-зубного щелевого в визуальной части стимула (т.е. bаVА и pаFА), и при этом наблюдалось зрительное доминирование. Строго говоря, такие случаи не свидетельствуют о проявлениях эффекта Мак-Гурка, но показывают, что губно-зубная артикуляция русских согласных достаточно чёткая и даже может привести к неправильной интерпретации произнесенного. Губные смычные согласные наиболее уязвимы с точки зрения слухового восприятия: респонденты чаще всего заменяли их на другие звуки в своих ответах.

Примерами проявления эффекта Мак-Гурка можно считать ответы с сочетанием нескольких согласных (прежде всего тех, что были представлены в стимуле, как утверждает в одной из своих работ К. Секийяма[[75]](#footnote-75)), а также те случаи, когда артикуляция звука в ответе находится между тем согласным, который озвучивается, и тем, который артикулируется (это примеры с ответами на стимулы bаDА и pаТА из пункта 2.1.8).

Отсутствие разницы в обработке глухих и звонких согласных в эксперименте показывает, что в дальнейших аналогичных исследованиях с участием носителей русского языка могут использоваться стимулы как со звонкими, так и с глухими согласными, и это не должно влиять на результаты.

Было выявлено влияние фактора группы участников: школьники дали больше правильных ответов, чем те, кто уже закончил школу. При этом мы понимаем, что в данном исследовании граница между двумя группами участников во многом условна: в группе взрослых встречались студенты первого курса вуза, т. е. те, кому 18 лет, а общий средний возраст взрослой группы – 23,5 года. Поэтому вопрос о влиянии возраста участников на результаты исследования требует дальнейшего изучения. Возможно, большее количество правильных ответов в группе школьников связано с тем, что они более внимательно и ответственно подходили к выполнению инструкций в эксперименте в силу привычки выполнять школьные задания.

Тот факт, что влияния предпочтительной модальности восприятия на успешность идентификации неконгруэнтных стимулов обнаружено не было, можно объяснить по-разному. Например, это может подтверждать ранее сделанные Л. Массой и Р. Майером[[76]](#footnote-76), а также Дж. Куэвасом и Б. Доусоном[[77]](#footnote-77) выводы о том, что людей сложно однозначно разделять на «аудиалов», «визуалов» и «кинестетов», а когнитивные стили не играют такой большой роли при аудиовизуальной обработке речи. Кроме того, такой результат может свидетельствовать о несовершенстве анкеты, выбранной для выявления этих групп, хотя именно она чаще всего используется для определения предпочтительной перцептивной модальности в исследованиях с русскоязычными информантами. Так или иначе, этот аспект исследования требует дальнейшего развития. Возможно, в будущем имеет смысл проверить гипотезу о влиянии предпочтительной модальности восприятия на мультимодальную обработку только на тех стимулах, которые вызывают наибольшую интерференцию у носителей русского языка, однако с привлечением большего числа дикторов. Последнее необходимо также для того, чтобы убедиться, что полученные результаты не обусловлены индивидуальными особенностями артикуляции диктора.

В данный эксперимент вошли только шесть согласных, артикуляция которых показалась нам наиболее заметной для русскоговорящих слушателей. Следующий шаг предполагает включение в эксперимент и других согласных (в частности, заднеязычных смычных, как предлагается в большинстве экспериментов, направленных на получение эффекта Мак-Гурка).

## 2.2. Эксперимент 2

### 2.2.1. Цель

В отличие от первого эксперимента, в котором предлагалось учитывать фактор ведущей перцептивной модальности реципиента, во втором и третьем экспериментах мы сосредоточились на том, как сочетания разных слогов в аудиовизуальных стимулах влияют на их интерпретацию. Кроме того, мы попытались создать такие условия, чтобы у участника не было возможности отклоняться от инструкции или пропускать определённые этапы во время прохождения эксперимента.

### 2.2.2. Стимулы

Для этого исследования был создан ещё один набор аудиовизуальных стимулов, которые в этот раз записывались на профессиональном оборудовании в студии. Диктором выступила женщина 25 лет, носитель русского языка без дефектов речи. В этот раз диктор произносил восемь слогов русского языка: /bа/, /vа/, /gа/, /dа/, /kа/, /pа/, /tа/ и /fа/, то есть и щелевые (губно-зубные), и смычные (губно-губные, переднеязычные и заднеязычные). Каждый слог записывался по 7 раз, а затем два первых и последний вырезались (за первые два повторения носитель обычно привыкает к ритму, а последний часто отличается от предыдущих, вероятно потому, что концентрация внимания к этому времени падает). В итоге получалось по четыре повторения для каждого из 8 слогов. Затем составлялись стимулы: как и в прошлом эксперименте, звуковую дорожку одного слога совмещали с визуальным рядом другого. Процесс синхронизации вновь выполнялся вручную одним из экспериментаторов, а затем проверялся другим. Хотя первый эксперимент не выявил зависимости от звонкости/глухости согласного в стимуле, в этот раз было принято решение вновь совмещать глухие согласные с глухими, а звонкие – только со звонкими. Всего получилось 32 комбинации: восемь исходных конгруэнтных стимулов, которые стали контрольными, и ещё 24 стимула с неконгруэнтными данными – по 12 для глухих и для звонких согласных.

### 2.2.3. Процедура эксперимента

В предыдущем эксперименте мы не могли контролировать выполнение инструкций или удерживать внимание участников, поскольку форма была размещена онлайн, что помогает набирать больше участников, но делает результаты эксперимента не вполне объективными. Поэтому второй эксперимент был запрограммирован с помощью бесплатного программного обеспечения PsychoPy[[78]](#footnote-78) таким образом, чтобы у участников не было возможности нарушить инструкцию (например, не дослушать один стимул, сразу перейдя к следующему). В начале каждой сессии экспериментатор вручную вносил данные об участнике (имя, пол, возраст, родной язык и наличие проблем со слухом). Затем респонденты должны были прочитать инструкцию к эксперименту и дать своё согласие на участие в нём. Далее они сами проходили его на ноутбуке, пользуясь наушниками с шумоподавлением (что также должно было уменьшить влияние фактора шума во время восприятия сигнала), а экспериментатор в это время следил за тем, чтобы респондент не отвлекался и выполнял инструкцию. Рандомизация в этот раз осуществлялась для каждого участника по отдельности самой программой. Таким образом, у каждого респондента был свой порядок предъявления стимулов, исключающий влияние фактора их последовательности. Формулировка задания осталась неизменной: участников после каждого стимула просили ответить на вопрос о том, что произнёс говорящий.

В этот раз парадигма вынужденного выбора была заменена на задание открытого типа, в котором участник должен был самостоятельно ввести свой ответ буквами русского алфавита с помощью клавиатуры, причём участникам не было дано инструкций по поводу того, сколько раз нужно было записывать повторявшийся слог. Прохождение эксперимента у большинства участников занимало около 10 минут.

### 2.2.4. Участники

В эксперименте участвовало 40 респондентов в возрасте от 18 до 31 года. Участники подбирались так, чтобы среди них не было филологов или лингвистов. Таким образом мы постарались снизить вероятность того, что респондент знает об эффекте Мак-Гурка (несмотря на то что эффект считается устойчивым, так что знание о нём теоретически не должно влиять на результат).

### 2.2.5. Принципы обработки данных, аналитика

Ответы, соответствовавшие аудиосигналу стимула, по-прежнему идентифицировались как «верные». После исключения из анализа ответов тех участников, которые допустили хотя бы одну ошибку в контрольных стимулах, были проанализированы данные 32 человек в возрасте от 18 лет до 31 года.

С помощью критерия χ2 мы сравнили количество правильных ответов на разные стимулы. Учитывались только ответы на целевые (неконгруэнтные) стимулы. Для статистической обработки использовалось бесплатное программное обеспечение JASP (<https://jasp-stats.org/>).

### 2.2.6. Результаты

Всего было получено 768 ответов, из которых 693 (90,2% от общего числа) соответствовали аудиальным составляющим стимулов и, соответственно, считались верными в рамках исследования. Остальные 75 условно неверных ответов можно разделить на две группы:

1) ответы, соответствовавшие визуальным составляющим стимулов (артикулируемым слогам); их было 26 (3,4% от общего числа);

2) ответы, которые не совпадали ни с одной из составляющих стимулов; таких оказалось 49 (6,4% от общего числа).

Таким образом, среди неверных ответов чаще встречались те, которые не совпадали ни с одной из составляющих стимулов, в то время как визуальную составляющую аудиовизуального стимула люди идентифицировали реже всего.

Что касается ответов отдельных участников, у большинства из них общее число ошибок колеблется в диапазоне от 0 до 3, что является очень низким показателем, хотя четверо реципиентов дали аномально высокое число неверных ответов (от 7 до 23 ошибок из 24 возможных). Однако, поскольку в контрольных стимулах они не ошибались, было принято решение не исключать их из дальнейшего анализа.

Анализ зависимости результатов от звонкости/глухости согласных в слогах-стимулах подтвердил уже полученные ранее данные о том, что этот фактор не влияет на проявление эффекта Мак-Гурка (χ2=0,946; df=1; p=0,331). Кроме того, анализ зависимости результата от фонологической характеристики согласного в слоге также не показал статистически значимой разницы ни для видео- (χ2=0,884; df=3; p=0,829), ни для аудиостимулов (χ2=3,472; df=3; p=0,324).

Среди неконгруэнтных пар слогов наибольшую интерференцию провоцировало сочетание звука /bа/ и артикуляции /vа/ (7 ошибок, это 21,9% от общего числа ответов на этот стимул). Это согласуется с результатами предыдущего эксперимента, в котором сочетание губно-зубного щелевого и губно-губного смычного тоже провоцировало наибольшее количество ошибочных ответов. Однако, в отличие от того, что показали полученные ранее результаты, остальные слоги, тогда провоцировавшие интерференцию, в данном эксперименте дали иные результаты. Так, сочетание артикуляции /fа/ (губно-зубной щелевой) со звучащим /pа/ (губно-губной смычный), в случае с которым участники предыдущего эксперимента часто ошибались, в этот раз дало прямо противоположный результат, оказавшись среди наиболее надежно распознаваемых стимулов (всего 2 ошибки; если же из статистики исключить людей, которые неверно интерпретировали большинство стимулов (18 и 23 из 24), то данное сочетание вообще не вызывало ошибок). А вот на стимулы, которые состояли из визуального /bа/ и аудиального /gа/, а также визуального /gа/ и аудиального /vа/, ошибки возникали чаще (на каждый из случаев приходится по 6 ошибок).

Выделить слоги, которые провоцируют наименьшее количество ошибок, трудно, поскольку в данном случае сразу на пять разных сочетаний было получено лишь по одной ошибке (что являлось наименьшим числом ошибок для различных пар слогов в этом эксперименте): /dа/ (а – артикуляция) + /gа/ (з – звук); /gа/ (а) + /dа/ (з); /kа/ (а) + /tа/ (з); /tа/ (а) + /kа/ (з) и /vа/ (а) + /gа/ (з).

Следующим этапом анализа была проверка влияния пола участников на результат. Учитывались статистические данные 22 женщин и 10 мужчин. Женщины совершали ошибки значимо чаще, чем мужчины (χ2=30,150; df=1; p<0,001): в 528 ответах женщины совершили 73 ошибки, тогда как мужчины ошиблись лишь дважды в 240 ответах. Мы предположили, что такое маленькое количество ошибок у мужчин может быть связано с тем, что те мужчины, кто ошибался чаще, были исключены из-за неверной интерпретации контрольных стимулов. Однако анализ показал примерно схожую статистику для мужчин и женщин и в том случае, когда ошибившиеся в контрольных стимулах были включены в анализ.

### 2.2.7. Выводы по Эксперименту 2

В этом эксперименте подтвердилось, что признак звонкости/глухости согласного в стимуле не влияет на его интерпретацию. Статистически значимой разницы для согласных, различающихся по месту образования, также выявлено не было. Наибольшую интерференцию по-прежнему демонстрирует сочетание видеосигнала /vа/ и аудиосигнала /bа/, однако остальные показатели для разных пар слогов в двух экспериментах различаются. Была также выявлена зависимость интерпретации от пола участника: мужчины совершали значительно меньше ошибок.

Стоит отметить также, что в первом эксперименте общее число ошибок, допущенных при интерпретации стимулов, было гораздо больше, чем во втором. Более того, то количество неверно интерпретированных стимулов, которое демонстрирует статистика второго эксперимента, во многом связано с респондентами, которые в целом допустили большое количество ошибок. В остальном же этот эксперимент продемонстрировал практически полное отсутствие эффекта Мак-Гурка у участников, и причина этому ясна не до конца. Мы предполагаем, что такая значительная разница в результатах первого и второго экспериментов может быть связана с применением наушников во втором эксперименте, что могло стать причиной рассинхронизации аудио- и видеосоставляющих стимула. В таком случае то, что участники Эксперимента 2 не были лингвистами, могло усугубить этот эффект, смещая фокус внимания респондентов с артикуляции диктора. Вероятно, те, кто изучает язык, будут внимательнее к тому, как говорящий артикулирует слоги на видео, а потому могут показать другие результаты. В следующем эксперименте мы постарались проверить это предположение.

## 2.3. Эксперимент 3

### 2.3.1. Цель

Прошлый эксперимент показал неоднозначные результаты. Мы предполагаем, что это может быть связано с тем, что звук в наушниках, работающих по Bluetooth, может отставать от видеосигнала и, таким образом, не только сильно снижать проявления эффекта Мак-Гурка, но и влиять на всю статистику исследования. В Эксперименте 3 было принято решение отказаться от использования наушников и проверить, как может проявляться эффект в случае, если демонстрировать стимулы нескольким участникам одновременно. Кроме того, в этот раз было принято решение включить в группу участников лингвистов для проверки предположения о том, что артикуляция в видео может влиять на них больше, чем на других людей.

### 2.3.2. Процедура

Для этого эксперимента были взяты те же стимулы, которые были использованы в предыдущем эксперименте, однако процедура отличалась: теперь вместо работы в PsychoPy мы приняли решение соединить все стимулы в две последовательности (псевдорандомизация вновь осуществлялась с помощью сайта – генератора случайных последовательностей чисел). Без сомнения, такой подход снизил возможность контроля за исполнением инструкций в эксперименте, однако упростил саму процедуру и позволил собирать данные от нескольких участников одновременно.

В этом эксперименте респондентам выдавали листы бумаги, на которых они писали свой пол, возраст и родной язык. Затем включалась одна из двух записей, различавшихся только порядком следования стимулов, и в десятисекундных перерывах между сигналами участники записывали, что, как им показалось, произносил диктор. Как и в предыдущем случае, это было задание открытого типа, в котором участник должен был самостоятельно написать свой ответ.

### 2.3.3. Участники

В эксперименте приняли участие 56 респондентов в возрасте от 17 до 29 лет. Они были разделены на группы по 8–15 человек, которым показывалась та или иная последовательность стимулов: один вариант записи был предъявлен 35 респондентам, тогда как остальным 21 достался второй. Опираясь на результаты прошлого эксперимента, мы приняли решение набирать для участия филологов и лингвистов (студентов 1 курса Филологического факультета СПбГУ), чтобы посмотреть, будут ли отличаться результаты двух экспериментов не только в зависимости от внешних условий восприятия и внимательности человека к артикуляции говорящего, но и от возможного знания о существовании изучаемого нами эффекта, которое, как уже было сказано ранее, не должно влиять на успешность прохождения эксперимента.

### 2.3.4. Принципы обработки данных, аналитика

Верными считались ответы, соответствовавшие аудиосигналу в предлагаемом участнику стимуле. После исключения из анализа тех, кто допустил хотя бы одну ошибку в контрольных стимулах, и носителей других языков были проанализированы данные 40 человек в возрасте от 18 до 22 лет. С помощью критерия χ2 мы сравнили количество правильных ответов на разные стимулы. Учитывались только ответы на целевые (неконгруэнтные) стимулы. Для статистической обработки использовалось бесплатное программное обеспечение JASP (<https://jasp-stats.org/>).

### 2.3.5. Результаты

В начале мы проверили наличие зависимости результата от последовательности стимулов, т.е. от предъявляемого участникам варианта записи. Статистически значимой разницы обнаружено не было (χ2=0,015; df=1; p=0,903), более того, процент верных ответов в двух группах почти совпадает (89,7% и 89,2%). Следовательно, далее результаты для двух вариантов можно будет анализировать вместе.

Всего участниками было дано 960 ответов, 859 (89,5%) из которых были верными, т.е. совпадали с аудиальным сигналом предлагаемого стимула. Остальные ответы были либо идентичными видеосигналам (26 ответов, 2,7% от общего числа ответов), либо не совпадали ни с одной составляющей стимула (75 ответов, 7,8% от общего числа ответов). Из этого соотношения видно, что, как и в предыдущем эксперименте, ошибаясь в интерпретации аудиосигнала, люди гораздо чаще выбирают что-то отличное от поступающих сигналов, а не видеосигнал.

Диапазон ошибок у отдельных участников в этом эксперименте получился следующим: большая часть участников совершала от 0 до 4 ошибок. Есть те, кто совершил 5 или 6 ошибок, и три аномальных результата: 9, 13 и 19 ошибок.

Анализ зависимости результата от звонкости/глухости согласного в слоге-стимуле по-прежнему не выявляет влияния данного фактора на интерпретацию стимулов (χ2=1,107; df=1; p=0,293). Не было обнаружено влияния фактора места образования согласного в видеостимулах (χ2=1,804; df=3; p=0,614), однако он был выявлен для звуковых стимулов: люди гораздо чаще ошибались в идентификации губно-зубных и губно-губных согласных (χ2=20,836; df=3; p<0,001).

Участники чаще всего ошибались в идентификации сочетаний звучащего слога /bа/ и артикуляции /vа/ и наоборот (в обоих вариантах было допущено по 9 ошибок, что составило по 22,5% от общего числа ответов на этот стимул). Это отчасти согласуется с предыдущими результатами, хотя сочетание артикуляции /bа/ и звука /vа/ ранее не выделялось как провоцирующее наибольшее количество ошибок. Также реципиенты часто ошибались в сочетаниях артикуляции /gа/ и звука /vа/ (8 ошибок, 20% от общего числа ответов), артикуляции /dа/ и звука /vа/ и артикуляции /tа/ и звука /pа/ (по 7 ошибок, что составляет по 17,5% от общего числа ответов на этот стимул). Стимулы, сочетавшие артикуляцию /gа/ и звук /dа/, артикуляцию /kа/ и звук /tа/ и наоборот всегда идентифицировались верно. Это говорит о том, что сочетания заднеязычных и переднеязычных смычных идентифицировались участниками лучше всего.

Что касается влияния пола реципиента на результаты, то в данном эксперименте, как и в прошлом, было установлено, что женщины совершают больше ошибок при распознавании неконгруэнтных стимулов, однако статистически значимой разницы обнаружено не было (χ2=3,499; df=1; p=0,061). Интересно, что при добавлении в анализ участников, ранее исключённых в связи с ошибками в контрольных стимулах, фактор пола становится статистически значимым (χ2=6,657; df=1; p=0,010).

### 2.3.6. Выводы

В этом эксперименте, как и в предыдущем, участники хуже всего воспринимали визуальную информацию: чаще всего в неконгруэнтных стимулах реципиенты либо воспринимали аудиальную составляющую стимула, либо давали ответ, который не совпадал ни с одним из сигналов аудиовизуального стимула.

Оба эксперимента подтвердили, что фактор звонкости/глухости согласного в слоге-стимуле не влияет на его восприятие. Аудиовизуальный стимул, сочетавший звук /bа/ и артикуляцию /vа/, интерпретировался хуже всего во всех трёх экспериментах, а среди всех стимульных согласных чаще всех интерференцию провоцировал губно-зубной щелевой /v/.

Хотя и во втором, и в третьем экспериментах женщины делали больше ошибок, чем мужчины, фактор пола участника сыграл статистически значимую роль только во втором эксперименте. Однако эта тема ещё не изучена на носителях русского языка, и для более однозначных выводов необходимо большее количество участников. Кроме того, представляется целесообразным набирать стимульный материал от нескольких носителей для того, чтобы исключить фактор влияния персональных особенностей речи диктора на результат.

Несмотря на то что в обоих экспериментах диапазон количества ошибок оказался довольно широким, для каждого эксперимента это объясняется наличием участников, которые верно интерпретировали контрольные стимулы, но совершали много ошибок в целевых. Среднее количество ошибок во втором эксперименте составляет 2 (75 ошибок на 32 носителя), а в третьем – 3 (101 ошибка на 40 носителей), что свидетельствует о сравнительно низкой интерференции при восприятии неконгруэнтных стимулов в проведенных эксприментах.

# Заключение

Аудиовизуальная интеграция является частным случаем объединения многомодальных данных, при котором задействуется одновременно аудиальная и визуальная информация. Существуют исследования, которые доказывают, что восприятие устной речи напрямую зависит не только от того, что человек слышит, но и от того, что он видит. Эффект Мак-Гурка – это феномен, при котором тот, кто воспринимает речь, не может правильно определить, что слышит, если движения губ говорящего не совпадают со звуковым сигналом. Учёные по всему миру изучают проявления этого эффекта. Существуют данные о корреляции между интерференцией, возникающей при восприятии неконгруэнтных стимулов, и возрастом и полом слушающего, а также о влиянии конкретных сочетаний слогов, при которых аудиовизуальные стимулы чаще воспринимаются неверно.

Однако восприятие речи лингвоспецифично, и данные для каждого языка уникальны. Подобных исследований на базе русского практически нет. Опираясь на опыт иностранных коллег, мы предположили, что существует ряд факторов, которые могут влиять на то, как эффект выражается у русскоговорящих слушателей. Чтобы проверить это, был проведён ряд экспериментов, направленных на выявление того, от чего зависят проявления эффекта Мак-Гурка. Мы также хотели узнать, повлияет ли так называемая ведущая перцептивная модальность на успешность прохождения экспериментов и можно ли изучаемый в рамках исследования эффект использовать для диагностики предпочтительного канала восприятия.

Для всех трёх экспериментов были созданы специальные стимулы. Для этого носители русского языка проговаривали на камеру определённые наборы слогов. Затем видеодорожка одного стимула накладывалась на звуковую дорожку другого, и таким образом получались неконгруэнтные стимулы, т.е. такие, визуальная и аудиальная информация которых не совпадала. Поскольку традиционно для подобного рода экспериментов используются слоги с гласным /a/, мы также выбрали слоги только с этим гласным. При создании стимулов звонкие согласные объединялись только со звонкими, а глухие – с глухими. Это было сделано для того, чтобы проверить, влияет ли фактор звонкости/глухости согласного в слоге-стимуле на результат его интерпретации. Верными во всех случаях признавались такие ответы, которые совпадали со звуковой составляющей аудиовизуального стимула.

В первом эксперименте 60 школьникам и 60 взрослым, проходившим эксперимент онлайн, предлагались различные комбинации слогов /ba/, /va/, /da/, /pa/, /ta/ и /fa/ (из них было составлено 18 стимулов). Результаты эксперимента показали, что у школьников эффект был выражен не так сильно, как у взрослых, однако различия между группами были относительно небольшими и достаточно условными. Для выявления ведущего канала восприятия был использован опросник С. Ефремцева, который состоит из 48 вопросов, предполагающих ответы «да»/«нет». Данные эксперимента показали, что ведущая перцептивная модальность не влияла на то, как участники интерпретировали стимулы Мак-Гурка. Таким образом, гипотеза о значимой роли когнитивных стилей в процессе аудиовизуальной обработки речи не подтвердилась.

Во втором и третьем экспериментах использовались одни и те же стимулы, записанные на профессиональном оборудовании. Было выбрано восемь слогов: /ba/, /va/, /ga/, /da/, /ka/, /pa/, /ta/ и /fa/, из которых были составлены 32 сочетания.

Поскольку контролировать следование участником инструкций в прошлый раз было невозможно, было принято решение запрограммировать второй эксперимент с помощью ПО PsychoPy и собирать ответы для каждого участника по отдельности и в присутствии экспериментатора. В третьем эксперименте участники для прохождения эксперимента объединялись в группы по 8–15 человек. Каждой группе предлагалось одновременно посмотреть видео с последовательностью стимулов и записать свой ответ на бумаге. Целью обоих экспериментов была проверка того, как изменение условий восприятия (например, использование более качественных видео и наушников или прохождение эксперимента в составе группы) повлияет на проявление эффекта Мак-Гурка.

Данные 32 человек в возрасте от 18 до 31 года во втором эксперименте и 40 человек в возрасте от 18 до 22 лет в третьем эксперименте показали, во-первых, существенно меньшее количество ошибок среди участников (около 90% всех ответов в обоих экспериментах были верными). Что касается неверных ответов, то здесь интересно сравнить результаты всех трёх экспериментов: в первом из них большая часть ошибочных ответов совпадала с визуальной составляющей стимула. Анализ данных второго и третьего экспериментов показал иной результат: даже ошибаясь, визуальную составляющую участники идентифицировали реже всего.

Все три эксперимента показали, что фактор звонкости/глухости согласного в слогах-стимулах не влияет на количество условно «верных» ответов на неконгруэнтные стимулы. Стимулом, спровоцировавшим наибольшее количество ошибок во всех трёх экспериментах, оказался vaBA, где первый слог обозначает артикуляцию, а второй – звук. Это означает, что наибольшее количество неверных ответов встречается при произнесении губного смычного согласного и артикуляции губно-зубного щелевого.

Хотя и во втором, и в третьем эксперименте женщины давали больше неверных ответов, чем мужчины, только во втором эксперименте этот фактор оказался статистически значимым, что говорит о перспективах дальнейшего изучения влияния гендерного фактора на количество правильных ответов у участников.

Таким образом, нами были выявлены некоторые закономерности в проявлении эффекта Мак-Гурка у носителей русского языка. Однако существует множество пока не изученных вопросов, которые требуют дальнейшего рассмотрения. Ведь, помимо теоретического вклада в науку, подобные работы могут найти практическое применение. Например, в сфере дубляжа, где чёткое совпадение артикуляции и звука могло бы привести к лучшему пониманию и большему погружению в фильм или сериал. К тому же эффект Мак-Гурка активно изучается как один из важных феноменов, связанных с распознаванием речи компьютером. В будущем, опираясь на эти данные, люди смогут улучшить и синтез, и распознавание речи и «научить» машину лучше понимать человека.

# Список литературы

1. Аллахвердов В.М., Аллахвердов М.В. Феномен Струпа: интерференция как логический парадокс // Вестник СПбГУ. Серия 16: Психология. Педагогика. 2014. №4. С. 90–102.

2. Басов О.О., Карпов А.А. Анализ стратегий и методов объединения многомодальной информации // Информационно-управляющие системы. 2015. №2 (75). С. 7–14.

3. Горбунова Т.С., Фахрутдинова А.В. Применение аудиовизуальных средств обучения иностранному языку в неязыковом вузе // Ученые записки КГАВМ им. Н.Э. Баумана. 2014. С. 64–69.

4. Дружинин В.Н. Психология общих способностей: учебное пособие для бакалавриата, специалитета и магистратуры. М.: Издательство Юрайт, 2019. 349 с.

5. Затевалова Е.С., Зубов В.И., Риехакайнен Е.И., Прокаева В.О. Смотреть или слушать: о методах изучения аудиовизуальной интеграции. Психология познания / отв.ред. И.Ю. Владимиров, С.Ю. Коровкин. Ярославль: Филигрань, 2023. С. 109–113.

6. Иванько Д.В., Кипяткова И.С., Карпов А.А. Анализ методов многомодального объединения информации для аудиовизуального распознавания речи // Научно-технический вестник информационных технологий, механики и оптики. 2016. С. 387–417.

7. Система аудиовизуального синтеза русской речи / Лобанов Б.М. [и др.] // Информатика. 2008. С. 67–78.

8. Animated virtual characters to explore audio-visual speech in controlled and naturalistic environments / Thézé R. [et al.] // Scientific Reports. 2020. №10(1). P. 1–12.

9. Atrey P.K., Hossain M.A., Kankanhalli M.S. Multimodal Fusion for Multimedia Analysis: a survey // Multimedia Systems. 2010. №16(6). P. 345–389.

10. Audiovisual events in sensory memory / Besle J. [et al.] // Journal of Psychophysiology. 2007. №21. P. 231–240.

11. Audiovisual perception of congruent and incongruent Dutch front vowels / Valkenier B. [et al.] // Journal of Speech, Language, and Hearing Research. 2012. №55(6). P. 1788–1801.

12. Audiovisual semantic interference and attention: evidence from the attentional blink paradigm / Van der Burg E. [et al.] // Acta Psychologica. 2010. №134(2). P. 198–205.

13. Baddeley A.D., Hitch G.J. Working memory // Recent advances in learning and motivation. New York: Academic Press. 1974. №8. P. 47–115.

14. Brown V.A., Strand J.F. “Paying” attention to audiovisual speech: Do incongruent stimuli incur greater costs? // Attention, Perception & Psychophysics. 2019. №81. P. 1800–1856.

15. Burnham D., Dodd B. Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect // Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology. 2004. №45(4). P. 198–224.

16. Changes in the McGurk effect across phonetic contexts / Hampson M. [et al.]. Technical Report. Boston University. 2003.

17. Cuevas J., Dawson B.L. A test of two alternative cognitive processing models: Learning styles and dual coding // Theory and Research in Education.2018. №16(1). P. 40–64.

18. Designing and deploying an interaction modality for articulatory-based audiovisual speech synthesis / Almeida N. [et al.] // SPECOM 2021. 2021. P. 36–49.

19. Desjardins R.N., Werker, J.F. 4-month-old female infants influenced by visible speech // Infant Behavior and Development. 1996. №19. P. 421–437.

20. Developmental shifts in children’s sensitivity to visual speech: A new multimodal picture–word task / Jerger S. [et al.] // Journal of Experimental Child Psychology. 2009. №102(1). P. 40–59.

21. Dodd B. Lip reading in infants: Attention to speech presented in-and out-of synchrony // Cognitive Psychology. 1979. №11(4). P. 478–484.

22. Dupont S., Aubin J., Ménard L. A study of the McGurk effect in 4 and 5-year-old French Canadian children // ZAS Papers in Linguistics. 2005. №40. P. 1–17.

23. Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker / Van Engen K.J. [et al.] // Journal of Speech, Language, and Hearing Research: JSLHR. 2014. №57(5). P. 1908–1921.

24. Erber N. P. Interaction of audition and vision in the recognition of oral speech stimuli // /Journal of Speech & Hearing Research. 1969. №12(2). P. 423–425.

25. Gottfried J.A., Dolan, R.J. The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. Neuron. 2003. P. 375–386.

26. Grant K.W., Walden B.E., Seitz P.F. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration // The Journal of the Acoustical Society of America, 1998. №103(5). P. 2677–2690.

27. Green K. P., Kuhl P. K. Integral processing of visual place and auditory voicing information during phonetic perception // Journal of Experimental Psychology: Human Perception and Performance. 1991. №17. P. 278–288.

28. Green K.P., Gerdeman A. Cross-Modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels // Journal of Experiment Psychology: Human Perception and Performance. 1995. №21(6), P. 1409–1426.

29. Hollich G., Newman R.S., Jusczyk P.W. Infants’ use of synchronized visual information to separate streams of speech // Child Development. 2005. №76(3). P. 598–613.

30. Inter-language differences in the McGurk effects for Dutch and Cantonese listeners / de Gelder B. [et al.] // Eurospeech 1995: Proceedings of the Fourth European Conference on Speech Communication and Technology. 1995. №18(21). P. 1699–1702.

31. Karpov A.A. An Automatic Multimodal Speech Recognition System with Audio and Video Information // Automatic Remote Control. 2014. №75(12). P. 2190–2217.

32. Kelly S.D., Kravitz C., Hopkins M. Neural correlates of bimodal speech and gesture comprehension // Brain Lang. 2004. №89(1). P. 253–260.

33. Kuhl P. K., Meltzoff A.N. The bimodal perception of speech in infancy // Science. 1982. №218. P. 1138–1141.

34. Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration / Tye-Murray N. [et al.] // Psychology and Aging. 2016. №31(4). P. 380–389.

35. Massa L.J., Mayer R.E. Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? // Learning and Individual Differences. 2006. №16(4). P. 321–335.

36. Massaro D.W., Cohen M.M., Smeele P.M. Cross-linguistic comparisons in the integration of visual and auditory speech // Memory & Cognition. 1995. №23(1). P. 113–131.

37. Mayer R.E. Cognitive Theory of Multimedia Learning // Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press. 2010. P. 31–48.

38. McGurk H., MacDonald J. Hearing lips and seeing voices // Nature. 1976. P. 746–748.

39. Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory / Colin C. [et al.] // Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology. 2002. №113. P. 495–506.

40. Oldakowski T. A multimodal assignment that enriches literacy learning: the problem // Insight A Journal of Scholarly Teaching. 2014. №9. P. 70–77.

41. Petrova T. Text presentation and information processing in Russian // 12th International Conference of Experimental Linguistics. International Society of Experimental Linguistics. 2021. P. 164–167.

42. Riekhakaynen E., Skorobagatko L. Written, not spoken or too much to read: How to present information more effectively? // Neurobiology of Speech and Language. Proceedings of the 5th International Conference on Neurobiology of Speech and Language. 2021. P. 15–16.

43. Riekhakaynen E., Zatevalova E. Should we believe our eyes or our ears? Processing incongruent audiovisual stimuli by Russian listeners // Speech and Computer. SPECOM 2022. Lecture Notes in Computer Science, 13721. Springer, Cham, 2022. P. 604–615.

44. Rosenblum L.D., Schmuckler M.A., Johnson J.A. The McGurk effect in infants // Perception & Psychophysics. 1997. №59(3). P. 347–357.

45. Sekiyama K. Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects // Perception & Psychophysics. 1997. №59(1). P. 73–89.

46. Sekiyama K., Burnham D. Impact of language on development of auditory-visual speech perception // Developmental Science. 2008. №11(2). P. 306–320.

47. Sekiyama K., Burnham D. Issues in the development of auditory-visual speech perception: Adults, infants, and children // Eighth International Conference on Spoken Language Processing INTERSPEECH-2004. 2004. P. 1137–1140.

48. Sekiyama K., Tohkura Y.I. Inter-language differences in the influence of visual cues in speech perception // Journal of Phonetics. 1993. №21(4). P. 427–444.

49. Shigeno S. Influence of vowel context on the audio‐visual speech perception of voiced stop consonants // Japanese Psychological Research. 2000. №42(3). P. 155–167.

50. Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers / Magnotti J.F. [et al.] // Experimental Brain Research. 2015. №233(9). P. 2581–2586.

51. Soto-Faraco S., Navarra J., Alsius A. Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task // Cognition. 2004. №92(3). P. 13–23.

52. Speaker-Dependent Visual Command Recognition in Vehicle Cabin: Methodology and Evaluation / Ivanko D. [et al.] // SPECOM 2021. 2021. P. 291–302.

53. Speech and non-speech audio-visual illusions: a developmental study / Tremblay C. [et al.] // PLOS ONE. 2007. №2(8). P. 742.

54. Summerfield Q. Some preliminaries to a comprehensive account of audiovisual speech perception // Hearing by eye: Psychology of lipreading Hillsdale. 1987. P. 3–51.

55. Svärdemo Åberg E., Åkerfeldt A. Design and recognition of multimodal texts: Selection of digital tools and modes on the basis of social and material premises? // Journal of Computers in Education. 2017. №4(3). P. 283–306.

56. Temporal constraints on the McGurk effect / Munhall K.G. [et al] // Perception & psychophysics. 1996. P. 351–362.

57. The influence of color and label information on flavor perception / Shankar M.U. [et al.] // Chemosensory Perception. 2009. P. 53–58.

58. The intermodal representation of speech in newborns / Aldridge M.A. [et al.] // Developmental Science. 1999. №2(1). P. 42–46.

59. Traunmüller H., Öhrström N. Audiovisual perception of openness and lip rounding in front vowels // Journal of Phonetics. 2007. №35(2). P. 244–258.

60. Wang R. Audiovisual perception of Mandarin lexical tones. Doctoral dissertation, Bournemouth University. 2018. 210 p.

61. Wu J. Speech perception and the McGurk effect: A cross cultural study using event-related potentials. 2009. 102 p.

62. Yang Z. A cross-linguistic examination on the McGurk effect in different developmental states. Master’s Thesis in Linguistics. Utrecht: Utrecht University, 2021. 73 p.

63. Zampini M., Spence, C. The role of auditory cues in modulating the perceived crispness and staleness of potato chips // Journal of Sensory Studies. 2004. P. 347–363.

1. Дружинин В.Н. Психология общих способностей: учебное пособие для бакалавриата, специалитета и магистратуры. М.: Издательство Юрайт. 2019. С. 114. [↑](#footnote-ref-1)
2. Басов О.О., Карпов А.А. Анализ стратегий и методов объединения многомодальной информации // Информационно-управляющие системы. 2015. №2 (75). С. 7-14. [↑](#footnote-ref-2)
3. Иванько Д.В., Кипяткова И.С., Карпов А.А. Анализ методов многомодального объединения информации для аудиовизуального распознавания речи // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 3. С. 387–401. [↑](#footnote-ref-3)
4. Gottfried J. A., Dolan, R. J. The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. Neuron. 2003. P. 375–386. [↑](#footnote-ref-4)
5. The influence of color and label information on flavor perception / Shankar M.U. [et al.] // Chemosensory Perception. 2009. P. 53–58. [↑](#footnote-ref-5)
6. Zampini M., Spence, C. The role of auditory cues in modulating the perceived crispness and staleness of potato chips // Journal of Sensory Studies. 2004. P. 347–363. [↑](#footnote-ref-6)
7. Karpov A.A. An Automatic Multimodal Speech Recognition System with Audio and Video Information

// Automatic Remote Control. 2014. №75(12). P. 2190–2200. [↑](#footnote-ref-7)
8. Atrey P.K., Hossain M.A., Kankanhalli M.S. Multimodal Fusion for Multimedia Analysis: a survey // Multimedia Systems. 2010. №16(6). P. 345–379. [↑](#footnote-ref-8)
9. Riekhakaynen E., Zatevalova E. Should we believe our eyes or our ears? Processing incongruent audiovisual stimuli by Russian listeners // Speech and Computer. SPECOM 2022. Lecture Notes in Computer Science, 13721. Springer, Cham, 2022. P. 604–615. [↑](#footnote-ref-9)
10. Oldakowski T. (2014). A multimodal assignment that enriches literacy learning: the problem // Insight A Journal of Scholarly Teaching. 2014. №9. P. 70–77. [↑](#footnote-ref-10)
11. Svärdemo Åberg E., Åkerfeldt A. Design and recognition of multimodal texts: Selection of digital tools and modes on the basis of social and material premises? // Journal of Computers in Education. 2017. №4(3). P. 283–306. [↑](#footnote-ref-11)
12. Горбунова Т.С., Фахрутдинова А.В. Применение аудиовизуальных средств обучения иностранному языку в неязыковом вузе // Ученые записки КГАВМ им. Н.Э. Баумана. 2014. [↑](#footnote-ref-12)
13. Baddeley A.D., Hitch G.J. Working memory // Recent advances in learning and motivation. New York: Academic Press. 1974. №8. P. 47–90. [↑](#footnote-ref-13)
14. Richard E. Mayer. Cognitive Theory of Multimedia Learning // Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press. 2010. P. 31–48. [↑](#footnote-ref-14)
15. Riekhakaynen E., Skorobagatko L. Written, not spoken or too much to read: How to present information more effectively? // Neurobiology of Speech and Language. Proceedings of the 5th International Conference on Neurobiology of Speech and Language. 2021. P. 15-16. [↑](#footnote-ref-15)
16. Petrova T. Text presentation and information processing in Russian // 12th International Conference of Experimental Linguistics. International Society of Experimental Linguistics. 2021. P. 164–167. [↑](#footnote-ref-16)
17. Иванько Д.В., Кипяткова И.С., Карпов А.А. Анализ методов многомодального объединения информации для аудиовизуального распознавания речи // Научно-технический вестник информационных технологий, механики и оптики. 2016. С. 387-399. [↑](#footnote-ref-17)
18. Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker / Van Engen K.J. [et al.] // Journal of Speech, Language, and Hearing Research: JSLHR. 2014. №57(5). P. 1908–1918. [↑](#footnote-ref-18)
19. Erber N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli // Journal of Speech & Hearing Research. 1969. №12(2). P. 423-425. [↑](#footnote-ref-19)
20. Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration / Tye-Murray N. [et al.] // Psychology and Aging. 2016. №31(4). P. 380–389. [↑](#footnote-ref-20)
21. Grant K.W., Walden B.E., Seitz P.F. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration // The Journal of the Acoustical Society of America, 1998. №103(5). P. 2677–2690. [↑](#footnote-ref-21)
22. Kelly S.D., Kravitz C., Hopkins M. Neural correlates of bimodal speech and gesture comprehension // Brain Lang. 2004. №89(1). P. 253–260. [↑](#footnote-ref-22)
23. Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration / Tye-Murray N. [et al.] // Psychology and Aging. 2016. №31(4). P. 380–389. [↑](#footnote-ref-23)
24. Brown V.A., Strand J.F. “Paying” attention to audiovisual speech: Do incongruent stimuli incur greater costs? // Attention, Perception & Psychophysics. 2019. №81. P. 1743–1756. [↑](#footnote-ref-24)
25. Затевалова Е.С., Зубов В.И., Риехакайнен Е.И., Прокаева В.О. Смотреть или слушать: о методах изучения аудиовизуальной интеграции. Психология познания / отв.ред. И.Ю. Владимиров, С.Ю. Коровкин. Ярославль: Филигрань, 2023. С. 109. [↑](#footnote-ref-25)
26. Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory / Colin C. [et al.] // Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology. 2002. №113. P. 495–506. [↑](#footnote-ref-26)
27. Soto-Faraco S., Navarra J., Alsius A. Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task // Cognition. 2004. №92(3). P. 13–23. [↑](#footnote-ref-27)
28. Audiovisual semantic interference and attention: evidence from the attentional blink paradigm / Van der Burg E. [et al.] // Acta Psychologica. 2010. №134(2). P. 198–205. [↑](#footnote-ref-28)
29. Brown V.A., Strand J.F. “Paying” attention to audiovisual speech: Do incongruent stimuli incur greater costs? // Attention, Perception & Psychophysics. 2019. №81. P. 1743–1756. [↑](#footnote-ref-29)
30. Система аудиовизуального синтеза русской речи / Лобанов Б.М. [и др.] // Информатика. 2008. С.67-78. [↑](#footnote-ref-30)
31. Animated virtual characters to explore audio-visual speech in controlled and naturalistic environments / Thézé R. [et al.] // Scientific Reports. 2020. №10(1). P. 1–12. [↑](#footnote-ref-31)
32. Designing and deploying an interaction modality for articulatory-based audiovisual speech synthesis / Almeida N. [et al.] // SPECOM 2021. 2021. P. 36–49. [↑](#footnote-ref-32)
33. Speaker-Dependent Visual Command Recognition in Vehicle Cabin: Methodology and Evaluation / Ivanko D. [et al.] // SPECOM 2021. 2021. P. 291–302. [↑](#footnote-ref-33)
34. McGurk H., MacDonald J. Hearing lips and seeing voices // Nature. 1976. P. 746–748. [↑](#footnote-ref-34)
35. Green K.P., Gerdeman A. Cross-Modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels // Journal of Experiment Psychology: Human Perception and Performance. 1995. №21(6), P. 1409–1426. [↑](#footnote-ref-35)
36. Green K. P., Kuhl P. K. Integral processing of visual place and auditory voicing information during phonetic perception // Journal of Experimental Psychology: Human Perception and Performance. 1991. №17. P. 278–288. [↑](#footnote-ref-36)
37. Аллахвердов В.М., Аллахвердов М.В. Феномен Струпа: интерференция как логический парадокс // Вестник СПбГУ. Серия 16: Психология. Педагогика. 2014. С. 90–101. [↑](#footnote-ref-37)
38. Аллахвердов В.М., Аллахвердов М.В. Феномен Струпа: интерференция как логический парадокс // Вестник СПбГУ. Серия 16: Психология. Педагогика. 2014. С. 90–101. [↑](#footnote-ref-38)
39. Audiovisual events in sensory memory / Besle J. [et al.] // Journal of Psychophysiology. 2007. №21. P. 231–238. [↑](#footnote-ref-39)
40. Inter-language differences in the McGurk effects for Dutch and Cantonese listeners / de Gelder B. [et al.] // Eurospeech 1995: Proceedings of the Fourth European Conference on Speech Communication and Technology. 1995. №18(21). P. 1699–1702. [↑](#footnote-ref-40)
41. Summerfield Q. Some preliminaries to a comprehensive account of audiovisual speech perception // Hearing by eye: Psychology of lipreading Hillsdale. 1987. P. 3–51. [↑](#footnote-ref-41)
42. Sekiyama K., Tohkura Y.I. Inter-language differences in the influence of visual cues in speech perception // Journal of Phonetics. 1993. №21(4). P. 427–444. [↑](#footnote-ref-42)
43. Sekiyama K. Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects // Perception & Psychophysics. 1997. №59(1). P. 73–80. [↑](#footnote-ref-43)
44. Wu J. Speech perception and the McGurk effect: A cross cultural study using event-related potentials. 2009. [↑](#footnote-ref-44)
45. Sekiyama K., Burnham D. Impact of language on development of auditory-visual speech perception // Developmental Science. 2008. №11(2). P. 306–320. [↑](#footnote-ref-45)
46. Inter-language differences in the McGurk effects for Dutch and Cantonese listeners / de Gelder B. [et al.] // Eurospeech 1995: Proceedings of the Fourth European Conference on Speech Communication and Technology. 1995. №18(21). P. 1699–1702. [↑](#footnote-ref-46)
47. Massaro D.W., Cohen M.M., Smeele P.M. Cross-linguistic comparisons in the integration of visual and auditory speech // Memory & Cognition. 1995. №23(1). P. 113–131. [↑](#footnote-ref-47)
48. Traunmüller H., Öhrström N. Audiovisual perception of openness and lip rounding in front vowels // Journal of Phonetics. 2007. №35(2). P. 244–258. [↑](#footnote-ref-48)
49. Audiovisual perception of congruent and incongruent Dutch front vowels / Valkenier B. [et al.] // Journal of Speech, Language, and Hearing Research. 2012. №55(6). P. 1788–1801. [↑](#footnote-ref-49)
50. Wang R. Audiovisual perception of Mandarin lexical tones. Doctoral dissertation, Bournemouth University. 2018. [↑](#footnote-ref-50)
51. Shigeno S. Influence of vowel context on the audio‐visual speech perception of voiced stop consonants // Japanese Psychological Research. 2000. №42(3). P. 155–167. [↑](#footnote-ref-51)
52. Temporal constraints on the McGurk effect / Munhall K.G. [et al.] // Perception & psychophysics. 1996. P. 351-362. [↑](#footnote-ref-52)
53. Massa L.J., Mayer R.E. Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? // Learning and Individual Differences. 2006. №16(4). P. 321–335. [↑](#footnote-ref-53)
54. Dodd B. Lip reading in infants: Attention to speech presented in-and out-of synchrony // Cognitive Psychology. 1979. №11(4). P. 478-484. [↑](#footnote-ref-54)
55. The intermodal representation of speech in newborns / Aldridge M.A. [et al.] // Developmental Science. 1999. №2(1). P. 42-46. [↑](#footnote-ref-55)
56. Kuhl P. K., Meltzoff A.N. The bimodal perception of speech in infancy // Science. 1982. №218. P. 1138-1141. [↑](#footnote-ref-56)
57. Hollich G., Newman R.S., Jusczyk P.W. Infants' use of synchronized visual information to separate streams of speech // Child Development. 2005. №76(3). P. 598-613. [↑](#footnote-ref-57)
58. Desjardins R.N., Werker, J.F. 4-month-old female infants influenced by visible speech // Infant Behavior and Development. 1996. №19. P. 421. [↑](#footnote-ref-58)
59. Rosenblum L.D., Schmuckler M.A., Johnson J.A. The McGurk effect in infants // Perception & Psychophysics. 1997. №59(3). P. 347-357. [↑](#footnote-ref-59)
60. Burnham D., Dodd B. Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect // Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology. 2004. №45(4). P. 204-220. [↑](#footnote-ref-60)
61. Dupont S., Aubin J., Ménard L. (2005). A study of the McGurk effect in 4 and 5-year-old French Canadian children // ZAS Papers in Linguistics. 2005. №40. P. 1-17. [↑](#footnote-ref-61)
62. Speech and non-speech audio-visual illusions: a developmental study / Tremblay C. [et al.] // PLOS ONE. 2007. №2(8). P. 742. [↑](#footnote-ref-62)
63. Sekiyama K., Burnham D. Issues in the development of auditory-visual speech perception: Adults, infants, and children // Eighth International Conference on Spoken Language Processing INTERSPEECH-2004. 2004. P. 1137-1140. [↑](#footnote-ref-63)
64. Sekiyama K., Burnham D. Impact of language on development of auditory-visual speech perception // Developmental Science. 2008. №11(2). P. 306–320. [↑](#footnote-ref-64)
65. Developmental shifts in children’s sensitivity to visual speech: A new multimodal picture–word task / Jerger S. [et al] // Journal of Experimental Child Psychology. 2009. №102(1). P. 40-59. [↑](#footnote-ref-65)
66. Yang Z. A cross-linguistic examination on the McGurk effect in different developmental states. Research Master’s Thesis in Linguistics, Utrecht University. 2021. [↑](#footnote-ref-66)
67. Shigeno S. Influence of vowel context on the audio‐visual speech perception of voiced stop consonants // Japanese Psychological Research. 2000. №42(3). P. 155–167. [↑](#footnote-ref-67)
68. Changes in the McGurk effect across phonetic contexts / Hampson M. [et al.]. Technical Report. Boston University. 2003. [↑](#footnote-ref-68)
69. Sekiyama K., Tohkura Y.I. Inter-language differences in the influence of visual cues in speech perception // Journal of Phonetics. 1993. №21(4). P. 427-444. [↑](#footnote-ref-69)
70. Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers / Magnotti J.F. [et al.] // Experimental Brain Research. 2015. №233(9). P. 2581-2586. [↑](#footnote-ref-70)
71. Yang Z. A cross-linguistic examination on the McGurk effect in different developmental states. MA Thesis in Linguistics, Utrecht University. 2021. [↑](#footnote-ref-71)
72. Затевалова Е.С. и др. Смотреть или слушать: о методах изучения аудиовизуальной интеграции. Психология познания / отв.ред. И.Ю. Владимиров, С.Ю. Коровкин. Ярославль: Филигрань, 2023. С. 109. [↑](#footnote-ref-72)
73. Исследование, описанное в этом разделе работы, описано в статье: Riekhakaynen E., Zatevalova E. Should We Believe Our Eyes or Our Ears? Processing Incongruent Audiovisual Stimuli by Russian Listeners // Prasanna S.R.M., Karpov A., Samudravijaya K., Agrawal S.S. (eds.) Speech and Computer. SPECOM 2022. Lecture Notes in Computer Science. Vol. 13721. Cham: Springer, 2022. P. 604–615. [↑](#footnote-ref-73)
74. Yang Z. A cross-linguistic examination on the McGurk effect in different developmental states. MA Thesis in Linguistics, Utrecht: Utrecht University. 2021. [↑](#footnote-ref-74)
75. Sekiyama K. Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects // Perception & Psychophysics. 1997. №59(1). P. 73–80. [↑](#footnote-ref-75)
76. Massa L.J., Mayer R.E. Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? // Learning and Individual Differences. 2006. №16(4). P. 321–335. [↑](#footnote-ref-76)
77. Cuevas J., Dawson B.L. A test of two alternative cognitive processing models: Learning styles and dual coding // Theory and Research in Education.2018. №16(1). P. 40–64. [↑](#footnote-ref-77)
78. <https://www.psychopy.org/> [↑](#footnote-ref-78)