УДК [811.111+811.161.1]'342.3(045)

Медведева Наталья Георгиевна
аспирантка кафедры фонетики
и практики англоязычной речи
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Natalia Medvedeva
PhD student of English Phonetics
and Speech Practice Department
Belarusian State University
of Foreign Languages
Minsk, Belarus
Enatalia7medvedeva@gmail.com

Яскевич Виталий Валерьевич
кандидат филологических наук,
заведующий кафедрой фонетики
и практики англоязычной речи
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Vitali Yaskevich
PhD in Philology,
Head of English Phonetics and Speech
Practice Department
Belarusian State University
of Foreign Languages
Minsk, Belarus
vitvalyas@gmail.com

ИСПОЛЬЗОВАНИЕ АЙТРЕКИНГА НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ИССЛЕДОВАНИИ ВОСПРИЯТИЯ АККОМОДАЦИОННОГО ВАРЬИРОВАНИЯ АНГЛИЙСКИХ И РУССКИХ СОГЛАСНЫХ

В исследовании используется окулографический метод для анализа восприятия аккомодационного варьирования согласных [р] и [п] в английском и русском языках. Эксперимент с носителями русского языка, владеющими английским, выявил, что артикуляторная близость гласных в дифонах с соответствующими согласными усиливает аллофоническую конкуренцию, увеличивая количество ошибок при распознавании и время идентификации, особенно в английском языке как иностранном.

Ключевые слова: айтрекинг; парадигма «Визуальный мир»; аккомодационное варьирование; перцептивная фонетика; иностранный язык; родной язык.

APPLICATION OF AI-POWERED EYE-TRACKING IN THE STUDY OF THE PERCEPTION OF ENGLISH AND RUSSIAN CONSONANTS ACCOMMODATIVE VARIATION

The study employs eye-tracking to investigate the perception of accommodative variation of the consonants [p] and [п] in English and Russian. Experiments with native Russian speakers proficient in English revealed that articulatory similarity of vowels intensifies allophonic competition, increasing errors and identification time, particularly in English as a foreign language.

Key words: eye-tracking; visual-world paradigm; accommodative variation; auditory phonetics; foreign language; native tongue.

Изучение восприятия речи представляет собой междисциплинарное направление, объединяющее лингвистику и нейронауки. Основой для развития этой области послужила модель Вернике–Гешвинда, предложенная более 150 лет назад немецким невропатологом Карлом Вернике и позднее усовершенствованная американским неврологом Норманом Гешвиндом [1, p. 408]. Эта модель стала отправной точкой для становления таких дисциплин, как нейро- и психолингвистика, заложив фундамент для дальнейших исследований восприятия речи. Развитие данных дисциплин неразрывно связано с появлением передовых цифровых технологий, позволяющих глубже понять нейронные и когнитивные механизмы, лежащие в основе процесса обработки речевой информации. Данные технологии включают в себя различные методы нейровизуализации, применение искусственного интеллекта для моделирования процесса восприятия речи [2], а также окулографию (айтрекинг) – технологию отслеживания движения глаз.

Метод окулографии широко применяется в психолингвистических исследованиях благодаря разнообразию стационарного и мобильного оборудования, а также гибкому программному обеспечению, позволяющему адаптировать эксперименты под широкий спектр исследовательских задач и требований. Впервые применение айтрекера в психолингвистическом исследовании было осуществлено в рамках экспериментального подхода «парадигма визуального мира» (Visual-world paradigm), разработанного Р. М. Купером в 1974 году. Во время проведения эксперимента в данной парадигме, участникам предлагалось смотреть на изображения или слова на мониторе с параллельным прослушиванием речи, в то время как айтрекер фиксировал направление движения глаз, задержку взгляда на представленных объектах, синхронизируя данные наблюдения с воспринимаемыми участниками аудиосигналами [1, p. 458]. В основе данного подхода лежит гипотеза, согласно которой траектория движения глаз и фиксация взгляда на объектах отражают когнитивные процессы, происходящие у испытуемых во время восприятия звучащей речи, что обеспечивает возможность более глубокого изучения механизмов данного процесса [3, p. 455]. Так, в своем эксперименте Р. М. Купер обнаружил, что испытуемые направляют взгляд на изображение льва, когда слышат слово «лев» или его часть, а также фиксируют взгляд на изображениях льва, змеи и зебры, услышав слово «Африка» [4, с. 84]. Также было установлено, что активация движения взгляда в направлении изображения происходила в течение 200 мс с начала воспроизведения слова на аудиозаписи, иногда в пределах звучания первой фонемы слова [5, с. 152], что свидетельствовало о возможности испытуемых предвосхищать слова, которые должны прозвучать на аудиозаписи, основываясь на подсказках из контекста, синтаксической и грамматической структурах фраз, а также акустических характеристиках начальных звуков слов.

Несмотря на значительный потенциал айтрекинга для исследования процессов восприятия речи, данный метод получил широкое признание в психолингвистическом сообществе лишь спустя два десятилетия, когда в 1995 году М. К. Таненхаус провел исследование, посвященное изучению процесса обработки синтаксических структур с использованием окулографии. Таненхаус также поспособствовал началу применения данного метода в фонетических исследованиях. Одним из первых наиболее влиятельных исследований в этой сфере стала совместная работа П. Д. Аллопенны, Дж. С. Магнусона и М. К. Таненхауса, проведённая в 1998 году. В ходе окулографического эксперимента испытуемым предлагалось перемещать изображения на мониторе, следуя инструкциям на параллельно прослушиваемой аудиозаписи, например, «Pick up the beaker; now put it below the diamond» («Возьмите мензурку; теперь поместите ее ниже ромба»), при этом на экране были представлены четыре изображения: *beaker* ‘мензурка’ – так называемый «целевой объект» эксперимента, *beetle* ‘жук’ – «когортный конкурент», имеющий идентичные начальные фонемы с целевым объектом, *speaker* ‘динамик’ – «конкурент по рифме», имеющий идентичную целевому слову финальную часть, и *carriage* ‘коляска’ – «отвлекающий стимул», не связанный фонологически с целевым *beaker*. Полученные данных показали, что фиксация взгляда на целевом объекте *beaker* и когортном конкуренте *beetle* начиналась с момента произнесения целевого слова (поскольку оба объекта начинались на идентичные звуки), но фиксации на конкуренте *beetle* уменьшались по мере произнесения целевого слова. При этом фиксации на изображениях начинались уже спустя 200–300 мс после начала воспроизведения слова, что свидетельствует о быстрой интеграции речевой и визуальной информации. Важно, что конкурент по рифме, несмотря на несовпадение звуков в начале слова, имел более длительные фиксации по сравнению с *carriage*. Полученные результаты свидетельствуют о том, что в процессе восприятия речевого сигнала слушатели одновременно оценивают все слова (изображения), частично соответствующие входной информации до тех пор, пока одно из них не будет иметь достаточно признаков для окончательной идентификации. Авторы исследования также отмечают, что «лексическая конкуренция», подразумевающая одновременную активацию конкурирующих слов вместе с целевым, сохраняется даже после того, как целевой объект становится фонетически различимым, что свидетельствует о вероятностной природе процесса распознавания устной речи. Исследование, однако, не рассматривало влияние акустических и артикуляционных характеристик звуков на лексическую активацию, но, согласно авторам, айтрекинг может быть особенно эффективен в изучении влияния акустической информации на распознавание слов [6, p. 422, 437–438].

Целью нашего пилотного исследования было изучение особенностей восприятия аккомодационного варьирования, а также установление характеристик гласных звуков, повышающих успех распознавания согласных аллофонов в комбинациях согласный+гласный (СГ) в начальном положении в слове. Дополнительно мы также стремились выявить темпоральные особенности процесса перцептивной идентификации, определяя степень когнитивной нагрузки данного процесса. Для реализации поставленных целей был организован окулографический эксперимент, в котором приняли участие десять носителей русского языка, владеющих английским на продвинутом уровне. Участникам предлагалось прослушать изолированные аллофонические реализации английского смычно-взрывного согласного [p], а также русского аллофона [п] в начальной позиции в слове, включающего переходной участок последующего гласного звука, и определить соответствие воспринимаемых звуков объектам на параллельно предъявляемых на экране изображениях, идентифицируя слова, начинающиеся с данных аллофонов. Всего каждую реализацию можно было прослушать 3 раза. Выбор данного согласного звука был обусловлен его сложной акустической структурой, позволяющей проводить разносторонние наблюдения. Материал на русском языке был включен для возможности проведения сопоставительного анализа, так как полагается, что перцептивная идентификация звуков родного языка в данном эксперименте будет проходить намного успешнее, чем в иностранном. Изображения были отобраны с учетом высокой частотности употребления соответствующих им слов в речи. Применение изображений вместо напечатанных слов способствовало упрощению процесса соотнесения акустической и визуальной информации, минимизации когнитивной нагрузки и повышению точности измерений, что обеспечило более надежные экспериментальные результаты. Одновременно с воспроизведением аудиостимулов испытуемым предъявлялись несколько изображений. Соответствующие изображениям лексические единицы были представлены пятью английскими и шестью русскими словами, обладающими как дистантными, так и близкими артикуляционными характеристиками ударных гласных, например в словах: *paw* ‘лапа’, *pearl* ‘жемчужина’, *pig* ‘свинья’, *puppy* ‘щенок’, *pool* ‘бассейн’ и *пыль, Пэн, повар, пиво, пальма, пуговица*.

Для получения окулографических данных использовалась программа «Beam Eye Tracker» версии 1.0 с использованием вебкамеры с частотой 30 кадров в секунду. Данная программа использует алгоритмы машинного обучения (AI-powered eye-tracking), при помощи которых искусственный интеллект распознаёт положение глаз, зрачков и головы, рассчитывая фиксации взгляда, саккады (быстрые движения глаз) и повороты головы. Во время эксперимента проводилась запись экрана с отображением направления движения взгляда в виде круга, который не был виден испытуемым (см. рисунок

ниже). Запись экрана производилась при помощи программы OBS Studio версии 25.0, с последующей ручной покадровой обработкой данных айтрекинга в программе Adobe Premiere Pro версии 24.5.0.

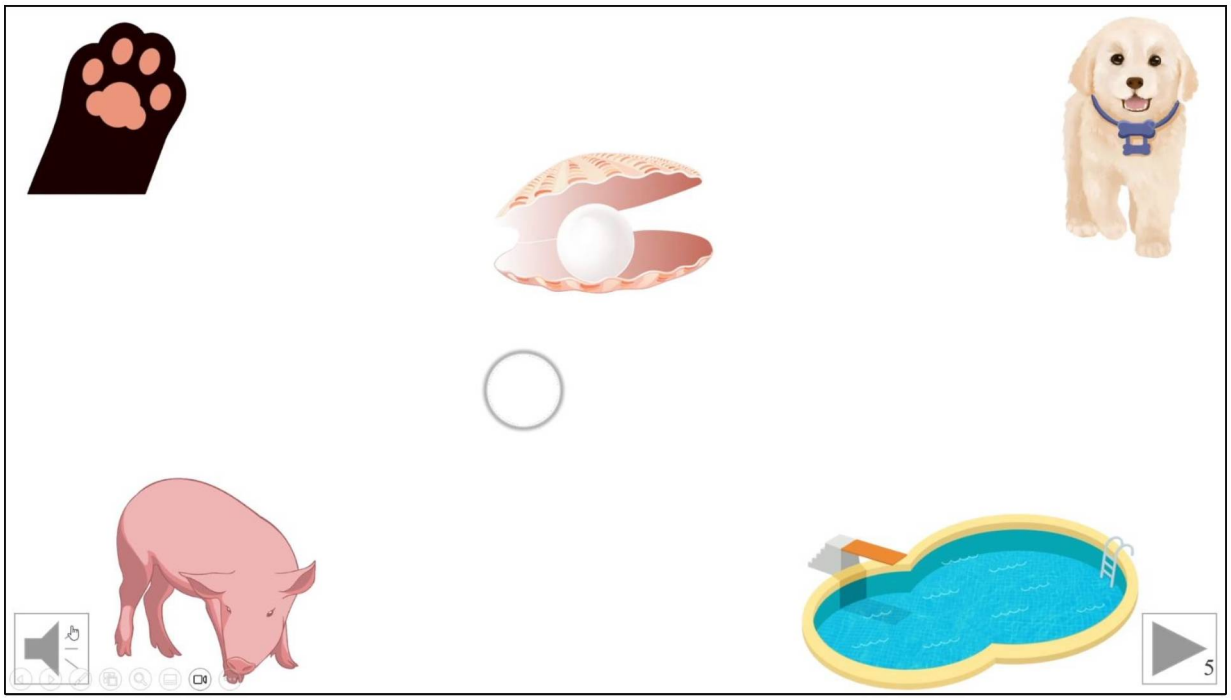

Рис. 1. Пример предъявления визуальных стимулов
и отображения движения взгляда

В результате проведённого эксперимента были определены: средняя длительность фиксации взгляда на каждом визуальном стимуле; общее время, затраченное на принятие решения по каждому слайду с визуальными стимулами (в миллисекундах, от начала первого прослушивания до выбора ответа); количество прослушиваний и ответы, выбранные участниками. Результаты представлены в таблицах 1 и 2 ниже.

Таблица 1

Английский язык, процент распознаваний
и конкурирующие аллофоны

Стимул	Процент распознаваний	Конкурирующие аллофоны
pu:	80 %	pɔ:; pɪ
pɜ:	40 %	pʌ; pɔ:
pʌ	80 %	pɜ:; pɔ:
pɔ:	80 %	pu:

Таблица 2

Русский язык, процент распознаваний
и конкурирующие аллофоны

Стимул	Процент распознаваний	Конкурирующие аллофоны
пы	90 %	по
пу	100 %	–
пэ	70 %	пы; па
па	100 %	–
по	80 %	пу
пи	100 %	–

Среднее количество прослушиваний реализаций английского языка составило 2,8 раза, русского – 1,9 раза. Среднее время принятия решения для английских аллофонов достигло 6447 мс, для русских – 3423 мс, с более быстрыми реакциями с примерами из русского языка, как и предполагалось ранее.

Наибольшие трудности в распознавании английских аллофонов вызвали целевые аллофоны слова *pearl* (ошибочно идентифицировано как *puppy* в 50 % случаев, *pearl* – в 40%, *paw* – в 10%) и *pool* (*pool* – 30 %, *pig* – 30 %, *pearl* – 20 %, *paw* – 10 %, *puppy* – 10 %). Ошибочная идентификация *pearl* как *puppy* в 50 % случаев объясняется артикуляторной близостью гласных /ɜ:/ и /ʌ/, которые относятся к среднему подъему и смешанному ряду, различаясь лишь по степени подъема (узкий и широкий соответственно). Неверная идентификация *pool* как *paw* в 10 % случаев, а *paw* как *pool* в 20 % случаев может объясняться близостью гласных звуков по признаку огубленности в рассматриваемых дифонах, а также закрытой артикуляцией звука /ɔ:/ в современной орфоэпической норме. Более продвинутая вперед артикуляция звука /u:/ в современном английском также объясняет его смешение с кратким гласным /ɪ/ в 10 % случаев при распознавании дифона pu: [7, с.14].

В части эксперимента, включающего согласные аллофоны русского языка, умеренные трудности возникли при идентификации аллофонов в словах *Пэн* и *повар*. Аккомодационное сочетание /пэ/ в слове *Пэн* было верно распознано в 70 % случаев, тогда как в 20 % и 10 % случаев его ошибочно идентифицировали как /пы/ (*пыль*) и /па/ (*пальма*) соответственно. Эти ошибки обусловлены относительной артикуляторной близостью гласных в данных комбинациях по вертикальному положению языка.

При идентификации аллофонов участники допустили меньше ошибок в русском языке (6) по сравнению с английским (12), что указывает на боль-

шую сложность распознавания иноязычных аллофонов, вероятно, из-за меньшей автоматизации процессов восприятия в иностранном языке. Анализ ошибочных распознаваний в английском языке показал следующее распределение: 50 % ошибок связаны с комбинациями согласный+гласный, где гласные близки по ряду и подъему; 25 % – с гласными, близкими по ряду; 25 % – с гласными, близкими по подъему. Таким образом, близость гласных по ряду и подъему является ключевым фактором ошибок. В русском языке ошибки распределились следующим образом: 83 % связаны с гласными, близкими по ряду, и 17 % – с гласными, близкими по ряду и подъему. Отсутствие ошибок, связанных только с подъемом, указывает на большую надежность этого признака при идентификации аллофонов в родной речи.

Ошибки возникали среди аккомодационных сочетаний с гласными, максимально близкими по своим артикуляционным характеристикам. Это подтверждает, что близость гласных в комбинациях СГ усложняет идентификацию аллофонов независимо от языка, что согласуется с данными Аллопенны и др. о вероятностной активации конкурирующих лексических единиц на основе частичного фонетического совпадения. Подобно их наблюдениям, где фиксации взгляда отражали одновременную активацию когортных и рифмованных конкурентов, наши результаты показывают, что артикуляторная близость гласных в дифонах усиливает конкуренцию между лексическими кандидатами, затрудняя точное распознавание слов.

Участники эксперимента демонстрировали склонность к более продолжительной фиксации взгляда на объектах, которые они затем выбирали. Анализ данных айтрекинга выявил тенденцию к вторичной по длительности фиксации на объектах, названия которых содержали аллофон, близкий по акустическим характеристикам к предъявленному аудиостимулу. Результаты эксперимента приведены в таблицах 3 и 4, где длительность фиксации на объектах с соответствующими аллофонами указана в порядке убывания, а выбранный участниками вариант выделен жирным шрифтом.

Таблица 3

Средняя длительность фокуса на объектах
при предъявлении аудиостимула (в мс)

Русский вариант		
пы (0,865)	пэ (0,237)	по (0,237)
пу (0,751)	по (0,324)	пэ (0,269)
пэ (1,295)	па (0,631)	по (0,240)
па (0,691)	по (0,236)	пэ (0,100)
по (0,877)	пи (0,120)	пэ (0,111)
пи (0,776)	пэ (0,328)	пы (0,133)

Таблица 4

Средняя длительность фокуса на объектах
при предъявлении аудиостимула (в мс)

Английский вариант		
pu: (1,357)	pɜ: (0,696)	pɔ: (0,550)
pɜ: (0,916)	**pʌ (0,639)**	pu: (0,404)
pʌ (1,578)	pɜ: (1,115)	pɔ: (0,508)
pɔ: (0,669)	pɜ: (0,335)	pu: (0,308)

При воспроизведении аудиостимула [pu:] участники дольше фиксировали взгляд на объекте с начальным [pɜ:], гласный которого имеет близость к /u:/ по признаку ряда, тогда как следующая по длительности фиксация на [pɔ:] определялась сходством с /ɜ:/ по подъему. Эти гласные также характеризуются относительной близостью по ряду. При предъявлении [pʌ], ошибочно воспринятого большинством испытуемых как [pɜ:], фиксации распределялись между указанными сочетаниями с преимуществом в 277 мс для [pɜ:], что отражает повышенную когнитивную сложность при распознавании аллофона [р] в сочетании с гласными смешанного ряда и среднего подъема. Данные гласные, обладая наибольшей акустической схожестью среди всех стимулов исследования, вызывали частые затруднения при идентификации. Повторное прослушивание [pʌ] и [pɔ:] подтверждало устойчивую вторичную фиксацию на [pɜ:], указывая на определяющую роль признака ряда в идентификации английского аллофона [р] для данной группы слов.

В случае с русскими аллофоническими сочетаниями в последовательности [пы] – [пэ] – [по] наблюдалась близость гласных по ряду, а также сближение [пэ] и [по] по подъему. В порядке фиксаций [пу] – [по] – [пэ] ключевыми факторами выступили ряд и огубленность. При прослушивании [пэ] вторичные фиксации приходились на [па] и [по], которые схожи преимущественно по подъему и в меньшей степени по ряду. Аллофоны [па] – [по] – [пэ] также демонстрировали большую близость по подъему. В последовательности [пи] – [пэ] – [пы] гласные были относительно равноудалены, при этом /и/ и /ы/ ближе по подъему, а /э/ – по ряду. В случае русских аллофонов основными критериями идентификации аллофонического варьирования выступали ряд и подъем, тогда как огубленность оказывала влияние в редких случаях.

Полученные результаты демонстрируют способность айтрекинга выявить перцептивную чувствительность испытуемых к тонким фонетическим

различиям, таким как аккомодационное варьирование звуков, что делает его эффективным инструментом для изучения процессов восприятия речи, как отмечали П. Д. Аллопенна, Дж. С. Магнусон и М. К. Таненхаус. Кроме того, выявленные закономерности указывают на необходимость учета артикуляторной близости гласных при разработке моделей распознавания речи и методик обучения иностранным языкам. В частности, повышенная когнитивная нагрузка при восприятии иноязычных аллофонов (на 88 % больше времени в английском варианте) подчеркивает важность тренировки фонетической чувствительности для улучшения перцептивной идентификации в иностранном языке. Дальнейшие исследования могут быть направлены на изучение влияния других фонетических факторов на аллофоническую конкуренцию в момент слухового восприятия, а также на расширение выборки для повышения статистической надежности данных.

ЛИТЕРАТУРА

1. Knight R. A., Setter J. The Cambridge Handbook of Phonetics. Cambridge : Cambridge Univ. Press, 2022. 720 p.
2. Дмитриев В. Я., Игнатьева Т. А., Пилявский В. П. Развитие искусственного интеллекта и перспективы его применения // Экономика и управление. 2021. Т. 27. No 2. С. 132–138.
3. An Eye-tracking Technique– Visual-World Paradigm / G. Guan, A. Hu, L. Guo, H. Yu // Proceedings of the 2019 3rd International Conference on Economic Development and Education Management (ICEDEM 2019) : Proceedings of the 2019 3rd International Conference on Economic Development and Education Management (ICEDEM 2019). 2019. P. 455–457. DOI: 10.2991/icedem-19.2019.108
4. Cooper R. M. The control of eye fixation by the meaning of spoken language // Cognitive Psychology. 1974. Vol. 6, № 1. P. 84–107.
5. Huettig F., Rommers J., Meyer A. S. Using the visual world paradigm to study language processing: A review and critical evaluation // Acta Psychologica. 2011. Vol. 137, № 2. P. 151–171. DOI: 10.1016/j.actpsy.2010.11.003
6. Allopenna P. D., Magnuson J. S., Tanenhaus M. K. Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models // Journal of Memory and Language. 1998. Vol. 38, № 4. P. 419–439.
7. Ефимова Е. В. Вариативность английских гласных в акцентной структуре фразы (на материале спонтанной речи) :дис. ... канд. филол. наук : 10.02.04. Минск, 2017. 232 л.