

МРТ-ВИЗУАЛИЗАЦИЯ ПРОЦЕССОВ КОАРТИКУЛЯЦИИ В ПОСЛЕДОВАТЕЛЬНОСТЯХ VCCV (НА МАТЕРИАЛЕ РУССКОГО ЯЗЫКА)

Г.Е. КЕДРОВА, Н.В. АНИСИМОВ, Л.М. ЗАХАРОВ, Ю.А. ПИРОГОВ

Магнитно-резонансная томография (МРТ) в настоящее время является мощным методом медицинской диагностики благодаря высокой информативности и безопасности исследования [1]. Однако возможности современных томографов привлекают и исследователей из других, не связанных напрямую с медициной, областей. Примером нестандартного, не связанного с медицинской диагностикой, применения МРТ являются эксперименты по визуализации артикуляторных органов при произнесении звуков речи. Первые томографические исследования артикуляционных процессов при речепроизводстве датируются началом 1990-х гг. На этом этапе технические ограничения используемых томографов позволяли получать низко-контрастные изображения мягких тканей, которые можно было сформировать только за довольно продолжительные (сравнительно с временем протекания артикуляционных процессов) промежутки времени – от 20 секунд до 3 минут. Естественно, в таких условиях было возможно исследование только гласных звуков, продуцируемых к тому же в достаточно искусственных условиях (изолированное и растянутое произнесение). Результаты этих исследований можно считать классикой этого направления [2]. Однако лингвистов всегда интересовали не только изображения относительно статических артикуляторных процессов, но и возможность изучения конфигураций артикуляторных органов в динамике – при производстве слитной речи. Прежде всего, это касается такой языковой универсалии, как коартикуляция. Актуальность изучения коартикуляционных механизмов связана и с тем, что процессы коартикуляции могут быть по-разному реализованы на границе звуков внутри слога и на границе слогов, а также в рече-

вых единицах большего размера – в словах, на границе словосочетаний (фонетических слов) и т.д.

Для экспериментального изучения процессов коартикуляции было необходимо получить реальные изображения процесса артикуляции в динамике. Получение динамических изображений стало возможным при разработке так называемой функциональной магнитно-резонансной томографии (ФМРТ). В настоящее время с помощью ФМРТ анализируется работа сердца, других пульсирующих органов, исследуются суставы, выявляются зоны активности головного мозга при воздействии внешнего раздражителя. В общем виде задачи ФМРТ решаются за счет предельного сокращения времени сканирования, что достигается благодаря минимизации всех временных задержек в сканирующей импульсной последовательности. Это, в свою очередь, приводит к снижению пространственной разрешающей способности, уменьшению отношения сигнал/шум, упрощению картины контраста из-за снижения взвешенности, кроме того, повышается вероятность артефактных искажений сигналов из-за аппаратных несовершенств. Тем не менее, все издержки, связанные с ускорением процесса сканирования, считаются оправданными, если за счет повышения временного разрешения можно извлечь полезную информацию. Однако при изучении процессов речепроизводства при помощи функциональной магнитно-резонансной томографии исследователи столкнулись с достаточно серьезными проблемами. Прежде всего, это связано с высокой скоростью реализации артикуляционных жестов. Из акустической фонетики известно, что средняя длительность звуков в потоке речи существенно меньше секунды (напр., средняя длительность только собственно смычки у смычных согласных варьируется в диапазоне от 30 до 120 мсек., а длительность шумовой фазы обычно около 40-60 мсек.). Более длительными являются аффрицированные и аспирированные согласные – у них длительность шума может достигать до 100-120 мсек. Наиболее долгий период реализации характерен для шумных (фрикативных) согласных, однако и здесь он не превышает 200

мсек. (в зависимости от фонетического контекста он может быть 50 – 200 мсек.) [3]. В целом можно считать, что средняя длительность слога в потоке речи стремится к 200 мсек.

Для получения динамических изображений быстро протекающих процессов (каким и является процесс речепроизводства) наиболее перспективным представляется применение методов магнитно-резонансной томографии (МРТ) с использованием методик быстрого сканирования, позволяющих неинвазивно получать посрезовые изображения со скоростью порядка 10 кадров в секунду. В идеале это требует специального аппаратного и программного обеспечения, позволяющего реализовать наиболее быстрый на сегодняшний день метод МР-сканирования – эхо-планарный метод. Однако из-за невозможности реализовать данный метод на имеющемся в настоящее время в ЦМТС МГУ томографе Tomikon S50 основные усилия исследователей были направлены на разработку и реализацию новых технических методов оптимизации процесса получения томографического изображения и совершенствования методик так называемого «псевдодинамического» отображения артикуляционных движений, или методик стробоскопического сбора данных. В соответствии с «псевдодинамическим» принципом в процессе эксперимента формируется банк синхронизированных томографических изображений, которые получены последовательным традиционным способом съемки быстрых процессов, который производится с некоторым временным сдвигом. В ходе эксперимента диктор многократно повторяет короткие речевые фрагменты (слова/псевдослова, предложения, фразы), одновременно ведется последовательная съемка артикуляторного тракта с заданным временным сдвигом, на основе которой после расшифровки изображений проводится реконструкция процесса изменения конфигурации артикуляторных органов и их взаимного расположения в динамике.

В качестве примера успешно реализованного ФМРТ исследования, использующего эту методику, можно привести проведенное на филологи-

ческом факультете МГУ им. М.В. Ломоносова совместно с ЦМТС МГУ исследование по МР-визуализации артикуляторных органов в процессе порождения речи. Результаты одного из этапов этого исследования представлены ниже.

Экспериментальный речевой материал. Испытуемые (2 человека, мужчина и женщина) получили задание многократно произнести по возможности с ровной интонацией псевдослова со структурой VCC`V, где в качестве гласного звука выступал гласный [а], а в качестве согласных – взрывные звонкие [б], [д], и [г]. Серия повторов слов с каждым из согласных включала от 20 до 25 единиц, что позволило получить по 50 томографических срезов (изображений) на серию.

Технические параметры эксперимента.

В ходе эксперимента сканировалась зона размерами 20x13 см, охватывающая области рта и верхней части шеи, с пространственным разрешением 3x3 мм и толщиной среза 9 мм. Параметры импульсной последовательности следующие: TR=12 мс, TE=5.5 мс, FA=10⁰. Испытуемый располагался в обычном для томографического исследования положении лежа. Полученные МР-изображения представляют собой набор кадров, разделенных по времени на 0.5–0.8 сек. и отображающих одну и ту же зону сканирования.

Методики обработки полученных данных. В целом по всем трем согласным нами было получено 150 изображений, которые фиксировали состояние артикуляторных органов в разные моменты произнесения звуковой последовательности. Атрибуцией изображений занимались 2 профессиональных фонетиста, которые на первом этапе проводили эту атрибуцию независимо друг от друга. При сравнении полученных в ходе атрибуции результатов оказалось, что идентификация томографических снимков в отношении абсолютного большинства изучаемых согласных звуков у обоих испытуемых в значительной степени одинакова (совпадения составляют более 80% по звукам [б] и [г] и более 90% по звуку [д]).

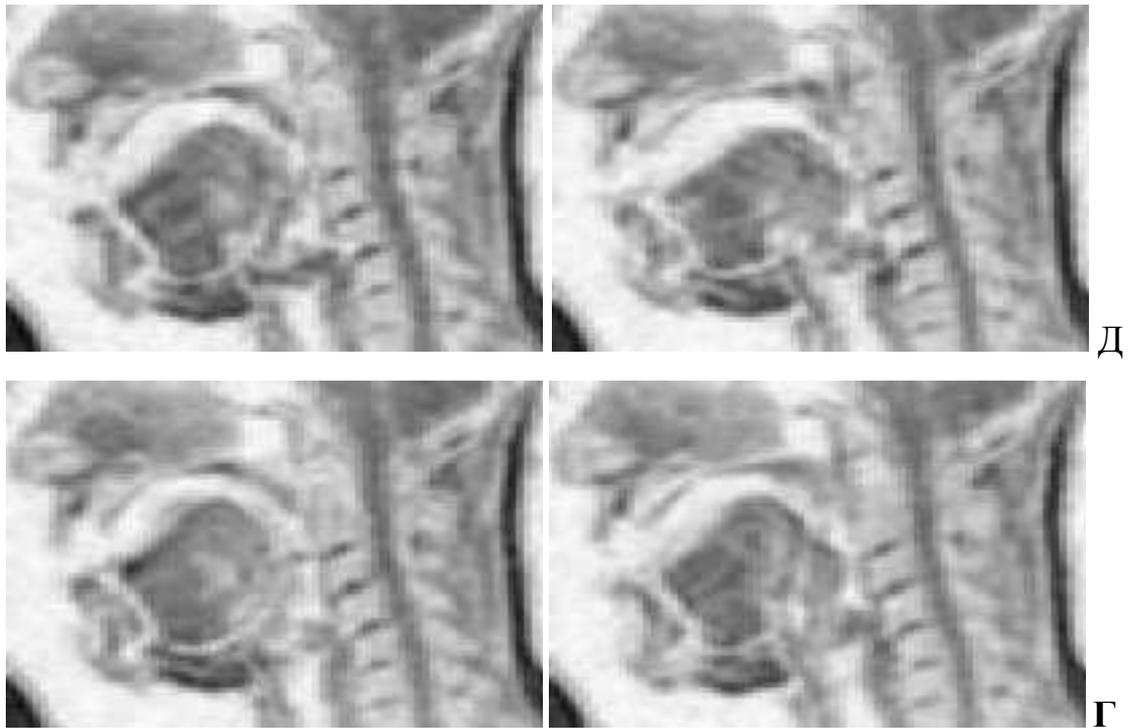
Для оценки степени информативности изображений, которые предположительно соответствовали гласным фрагментам звуковой последовательности, был проведен специальный перцептивный эксперимент. В рамках этого эксперимента оценивалась возможность кластеризации МР-изображений на основе воспринимаемых глазом различий в особенностях конфигурации артикуляторных органов, а также степень совпадения полученных кластеров с такими факторами, как основные различительные признаки согласного и положение гласного в слове (начальное или конечное). В результате удалось доказать постоянство артикуляторных «поз», характерных для каждого из звуков (гласного и согласных) в произнесении каждого диктора; показать высокую информативность полученных изображений в отношении коартикуляторного воздействия на гласный такого признака согласного, как место образования, а также с высокой степенью достоверности представить некоторые коартикуляционные различия в зависимости от положения гласного звука в слове по отношению к согласному.

На рис. 1 приведены МР-изображения гласного [а] в произнесении диктора К (женский голос); показаны наиболее типичные изображения гласного, предшествующего и следующего за согласными звуками (для наглядности томографические изображения даны в инверсном формате, который позволяет более отчетливо показать основные конфигурации, в первую очередь, мягких тканей – губ, языка и увулы).

Рис. 1. МР-изображения предшествующего (левый ряд) и следующего (правый ряд) гласного звука [а] в сочетании с согласными [б], [д], [г].



Б



Для сравнения приведем МР-изображение тянутого гласного [a] в изолированной позиции в произнесении того же диктора (диктор К., женский голос), которое было получено нами на предыдущем этапе исследования [5].

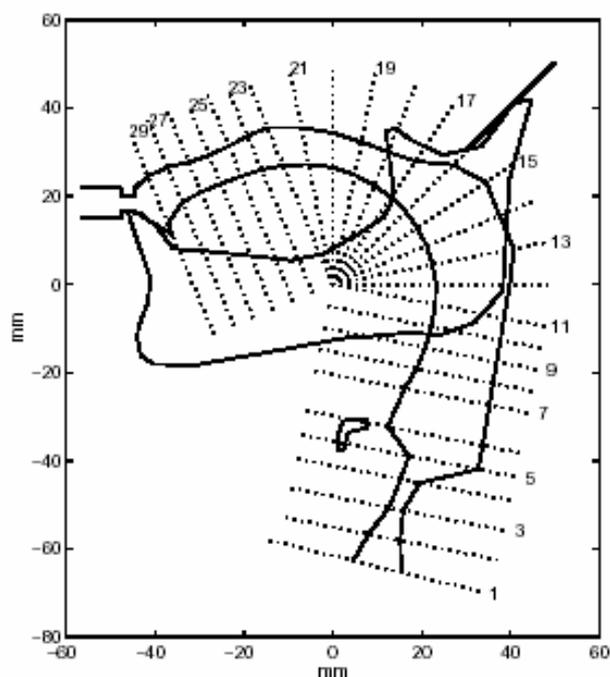
Рис. 3. МР-изображение гласного [a] в изолированном произнесении (диктор К).



Для большей наглядности и сопоставимости результатов по всем полученным изображениям были проведены измерения язычной артикуляции в соответствии с масштабной сеткой, разработанной в работе [6]. Эта координатная система позволяет очень подробно описать и зафиксировать практически все релевантные параметры состояния артикуляторного тракта; она включает в себя не только постоянные точки измерений, но и переменные реперы, которые позволяют произвести масштабирование измере-

ний без искажения конфигурации для дикторов с разными размерами речевого аппарата (рис. 4).

Рис. 4. Масштабная сетка для измерения артикуляционных конфигураций, состоящая из векторов, местоположение которых фиксировано (линии 7-22), и векторов, которые привязаны к местоположению определенных точек – кончика языка (линии 23-29) и гортани (линии 1-6).



Для нашего исследования использовались векторы с 12 по 28. По всем томографическим изображениям, которые соответствовали разным этапам произнесения звука [а] в сочетании с согласными звуками [б], [д], [г] (как в начальной позиции, так и в конце псевдослова), были подсчитаны величины каждого из векторов; для сравнения таким же образом были обработаны изображения артикуляторных органов в процессе произнесения согласного звука и в период паузы между словами. Значения каждого из векторов для наиболее типичных артикуляторных конфигураций согласных и гласных звуков, а также паузального компонента представлены на схеме 1.

Схема 1

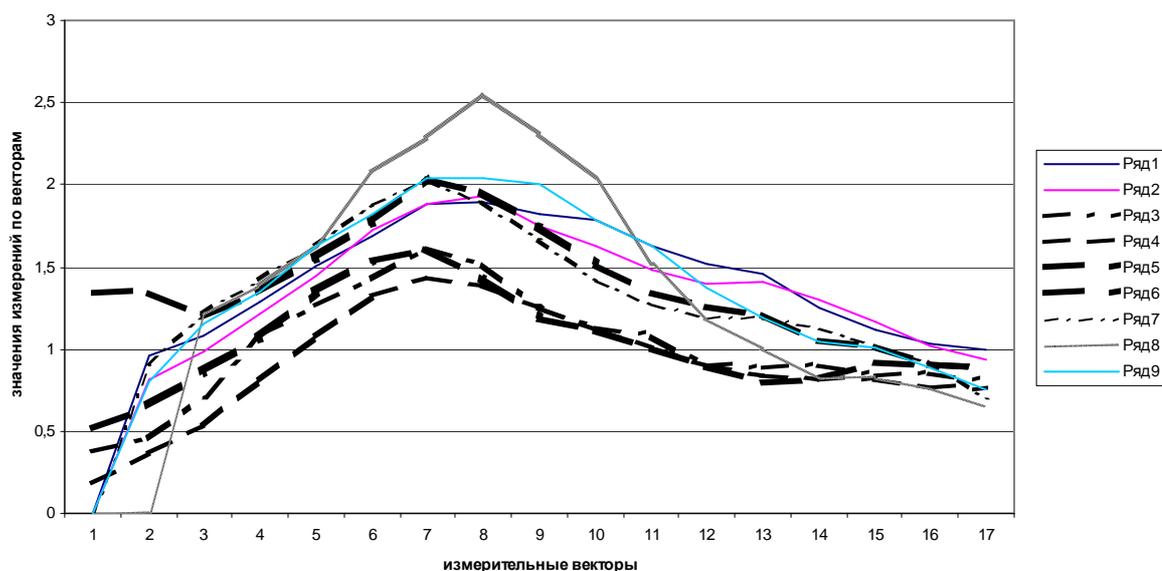


Схема 1. Значения измерений по векторам 28-12 (на оси X обозначены цифрами от 1 до 17) для сегментов: [а]д (ряд 1), [а]б (ряд 2), б-пауза (ряд-3), [б] (ряд 4), д-пауза (ряд 5), [д] (ряд 6), г-пауза (ряд 7), [г] (ряд 8), [а]г (ряд 9).

Обсуждение результатов и основные выводы исследования. Полученные данные демонстрируют, что в слитной речи именно коартикуляторное воздействие согласного звука становится определяющими для способа реализации гласных звуков. Это утверждение верно не только для последующего за согласным гласного звука, но также в определенной степени и для предшествующего ему гласного, равно как и для периодов паузы между однотипными повторяющимися звуковыми последовательностями. Наибольшая степень коартикуляции отличает гласный звук, которые соседствует с заднеязычным согласным, наименьшая отмечена нами для губного взрывного согласного. Можно также отметить, что переднеязычный взрывной ([д]) в большей степени влияет на артикуляцию последующего гласного звука, чем на артикуляционные позы предыдущего. В целом, можно утверждать, что коартикуляция гласного по отношению к согласному существенно возрастает при уменьшении степеней свободы язычных конфигураций.

Таким образом, нам представляется достаточно перспективным использовать полученные по разработанной методике МРТ-изображения для построения анимационных последовательностей, отображающих процесс произнесения сложных звуковых конструкций – слов, фраз и т.п. Это, в свою очередь, создает возможности для решения ряда фундаментальных задач теоретической фонетики и фонологии (прежде всего совершенствования теории речепорождения), а также множества прикладных областей речеведения (постановки произношения, автоматического распознавания и синтеза речи, компьютерных речевых баз данных и т.д.). Возможно, что МРТ-изображения окажутся полезными как для исследования и коррекции речевых патологий – например, заикания, так и для совершенствования методов восстановительной медицины (в частности, послеоперационного тренинга компенсаторных артикуляций).

Литература

1. *Анисимов Н.В., Пирогов Ю.А., Губский Л.В., Гладун В.В.* Управление контрастом и информационные технологии в магнитно-резонансной томографии // Под ред. Ю.А. Пирогова. – М.: Московский государственный университет, 2005.
2. *Baer T., Gore J.C., Gracco L.C., Nye P.W.* Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels // *Journal of the Acoustical Society of America*, 1991, 90 (2), pp. 799-828.
3. *Кодзасов С.В., Кривнова О.Ф.* Общая фонетика: Учебник. – М.: РГГУ, 2001.
4. *Кедрова Г.Е., Захаров Л.М., Пирогов Ю.А., Анисимов Н.В.* Исследование артикуляторной базы русского языка методами магнитно-резонансной томографии // XIII сессия Российского акустического общества (25-29 августа, 2003). Сб. трудов. Т. 3. Акустика речи. Медицинская и биологическая акустика. – М., ГУОС, 2003.
5. *Bailly G., Badin P., Vilain A.* Contribution de la mâchoire à la géométrie de la langue dans les modèles articulatoires statistiques // XXIIèmes Journées d'Études sur la Parole. – Martigny, Switzerland, 1998. pp. 287-290.