

В.В. ЕВДОКИМОВА, Д.А. КОЧАРОВ, П.А. СКРЕЛИН
**МЕТОД ПОСТРОЕНИЯ ФОРМАНТНЫХ КАРТИН ДЛЯ
ИССЛЕДОВАНИЯ ФОНЕТИЧЕСКИХ ХАРАКТЕРИСТИК
ГЛАСНЫХ**

Евдокимова В.В., Кочаров Д.А., Скрелин П.А. Метод построения формантных картин для исследования фонетических характеристик гласных.

Аннотация. Представлены результаты применения метода получения формантных составляющих гласных фонем. Проведен обзор существующих направлений разработки методов получения формантных характеристик гласных для разных языков с момента возникновения метода и до недавнего времени. Также проведен обзор степени использования методов оценки формантных картин в речевых технологиях и при обработке естественного языка. На массиве корпуса профессионального чтения CORPRES были получены данные по формантным составляющим для 351929 реализаций гласных фонем на материале 8 дикторов. Полученные данные сгруппированы в соответствии с обозначениями в реальной транскрипции, которую проводили сегментаторы-фонетисты в рамках работы по разметке корпуса. На формантных плоскостях представлено распределение ударных аллофонов гласных для всех дикторов. На примере одного диктора-мужчины представлена вариативность формантных характеристик в корпусе для предударных и заударных аллофонов. Также представлены результаты, свидетельствующие о различиях огубленных безударных /i/ и /a/, которые воспринимаются как /u/ не только наивными носителями языка, так и фонетистами-экспертами. В качестве экспериментального материала использовались записи чтения одним диктором мужского пола специально подобранных предложений, которые учитывали различные лингвистические факторы. Результаты анализа показали, что высока вероятность огубленности безударного неогубленного гласного, если в следующем слогое есть безударный огубленный гласный. Анализ данных формантных составляющих этих гласных показал, что значения первой форманты этих гласных близки к значениям у ударного гласного /u/ для этого диктора. Артикуляторно подъем этих гласных соответствует подъему /u/. Значения второй форманты у гласных [u], которые должны были реализовываться как [i] и [a], различаются. Они в большей степени продвинуты вперед по сравнению с ударным /u/.

Ключевые слова: фонетика, фонология, акустический анализ речевого сигнала, формантные характеристики гласных, характеристики огубленности.

1. Введение. Современные методы обработки речевых сигналов предполагают использование речевых технологий, обеспечивающих обработку материала автоматически без постоянного контроля процедуры оператором.

Однако для решения ряда фонетических задач, связанных с речепорождением в различных специфических условиях и восприятием такой речи, целесообразно обращаться к методам, которые были разработаны еще во второй половине XX века.

Одним из ведущих методов исследования фонетических характеристик гласных остается формантный анализ. Он позволяет делать выводы из наблюдаемых различий между значениями формант, обу-

словленных как разными гласными фонемами, так и общими фонемами в разных диалектах, стилях и типах речи.

2. Методы определения формантных характеристик. Основы формантного анализа были заложены еще в XIX веке в трудах немецкого физика Германа Гельмгольца [1]. Разработанный им метод основывался на сравнении речи с набором сигналов камертонов и резонаторов, причем каждому тембру соответствовал свой набор воспроизводимых частот. Восприятие тембра определялось только набором частот и не зависело от фазы воспроизводимых звуковых сигналов.

Объяснению механизма формирования речевого сигнала посвящены труды германского ученого Л. Германа (L. Hermann), опубликованные в 90-х годах. Им использован математический метод разложения полученной в эксперименте сложной кривой реализации звукового сигнала с помощью рядов Фурье. Конечным результатом, подвергавшимся дальнейшему анализу, стал набор синусоидальных составляющих речевого сигнала.

Проведенные исследования позволили Л. Герману сделать вывод, что тембр определяется сочетанием гармонических составляющих, степень присутствия которых в сигнале зависит от резонансных свойств надгортанных полостей. В надгортанные полости толчками поступает воздух из гортани. Эта воздушная волна модифицируется в соответствии с резонансными свойствами надгортанных полостей. Сформировавшиеся таким образом мощные псевдогармонические составляющие речевого сигнала Л. Герман первым назвал в 1894 году формантами, заменив действующий до этого термин «характерный тон» [1].

В конце XIX – первой половине XX века формантный анализ становится общепринятым, причем по-прежнему для определения формант используются как физические (камертоны), так и математические (метод Фурье) методы. Сложный самонастраивающийся речевой аппарат человека не позволяет исследователям получить устойчивые результаты по конкретным формантам, но дают возможность сформировать общую картину акустической природы гласных. Среди них можно отметить работы А. И. Томсона, который изучал ударные гласные с учетом их соседства с согласными. В. А. Богородицкий изучал формантный состав шепотных гласных при изолированном произношении [1]. Особенность работ Л. В. Щербы в том, что он сопоставлял слуховой анализ гласных с записями их реализации.

Результаты их работ прошли сравнение с современными компьютерными методами. Для гласных [i, e, o] выделенная учеными характерная частота полностью совпадает по диапазону с формантой F2, выделенной при компьютерном анализе. По другим гласным результаты незначительно отличаются. Для гласной [a] по данным Л. В. Щербы характерная ча-

стота 1120 – 1150 Гц лучше соответствует F2, чем в исследованиях А. И. Томсона, для [u] обратная картина – данные А. И. Томсона соответствуют лучше. Более сложная картина с гласным [i]. А. И. Томсон считает его дифтонгом с [i]-образным завершением, а Л. В. Щерба сосредотачивается только на его первой части. По результатам исследований В. А. Богородицкого, тон [u] соответствует частоте F1 [1].

Собранные исследователями акустические медико-биологические материалы позволили сделать ряд выводов о строении речевого тракта человека как сложной системы резонаторов, что дало возможность Г. Фанту представить его в форме акустической системы переменной структуры, имеющей ряд резонансных частот [2].

В составе этой системы Г. Фант выделил две основные составные части: голосовой источник и артикуляторный аппарат. Голосовой источник – нижний отдел речевого аппарата, основные части которого трахея, гортань, голосовые связки. Резонатором голосового источника является весь организм человека, а его костно-мышечная система служит для энергетического обеспечения воздушного потока из голосового источника. Им производится фонация – основная генерация звука речевым аппаратом.

Верхний отдел речевого тракта имеет в составе фарингальную, ротовую, носовую полости и набор активных и пассивных артикулирующих органов (язык, зубы, губы, мягкое небо). Его можно в процессе речи рассматривать как самонастраивающийся акустический фильтр. Именно он, перестраивая структуру по командам мозга, формирует речевой сигнал из акустической волны голосового источника.

Функциональное разделение речевого аппарата на два основных отдела позволило сосредоточить внимание на артикуляторном аппарате – наиболее сложной и важной составной части, которой в основном определяются фонетические особенности речи. В модели Г. Фанта это сочетание настраиваемых стандартных акустических элементов – труб, рупоров. Подобный подход позволил ввести стандартные передаточные функции этих элементов и обратиться к стандартным частотным методам исследования динамических систем различной природы. Сравнение с экспериментальными данными открыло возможность определения параметров, соответствующих резонансным частотам артикуляторного аппарата, то есть формантам.

Представим схему работы речевого аппарата человека. Первый этап – инициация. На нем костно-мышечная система формирует воздушный поток из легких в гортань. Частота формирования потока соответствует дыханию. При его прохождении через нелинейную систему гортань-голосовых связок в ней создается автоколебательный режим – связки вибрируют и при этом проходящий через них поток воз-

духа модулируется с частотой основного тона, создается, по терминологии Г. Фанта, глоттальная волна. Таким образом происходит фонация – формирование энергетической основы акустического сигнала. Глоттальная волна поступает в артикуляторный аппарат и фильтруется им в зависимости от расположения артикулирующих органов, положение которых задается мозгом. Процесс управления артикулирующими органами относится к разделу изучения деятельности мозга и обычно остается за рамками фонетических исследований.

Разработанная Г. Фантом описанная выше акустическая модель речи стала первым физически обоснованным описанием этого процесса и даже дала возможность определять формант расчетным путем. Описание функционирования каждого из отделов речевого тракта помогло проведению мысленных экспериментов – теоретическому описанию происходящих в нем при произнесении отдельных фонем процессов.

С развитием вычислительной техники и звукозаписи анализ речевого сигнала стал частным случаем обработки любого сигнала чисто математическими методами независимо от его физической природы. Машинная обработка открыла возможности решения ряда прикладных фонетических задач. Это и идентификация источника звука, и изучение супraseгментных характеристик, и машинный синтез речи. Важное место в их решении отводится методам определения и анализа вариаций формантных частот.

Программы автоматического определения траекторий движения формант становятся в настоящее время обязательной составной частью алгоритма обработки речевого сигнала при решении ряда фундаментальных и прикладных фонетических задач, таких как идентификация звуков, изучение супraseгментных характеристик, настройка синтезаторов речи. Их разработка до сих пор остается крайне актуальной, и большое количество исследований посвящено поиску методов вычисления движения формантных частот.

Автоматическая посегментная обработка (обработка окном) реализаций речи дает богатый материал о структуре сигнала, но зачастую процесс выделения в нем именно формант делается по спектрограммам вручную и определяется опытом исследователя. Тем не менее при значительных объемах речевого материала (Big Data) необходимо отразить его в программе обработки и добиться полной автоматизации процесса.

Рассмотрим основные пути совершенствования методов автоматической обработки речевого сигнала.

Первая группа этих методов основывается на обобщенном математическом аппарате. В них отсутствует фонетический подход, не учтены физиологические особенности речевого тракта человека. Просто проводится обработка в общем случае нестационарного полича-

стотного сигнала. Применение этих методов открывает возможность построения передаточных функций речевого тракта, перехода от временного сигнала к его частотным характеристикам. Появляется возможность слежения за флюктуацией формантных частот во времени.

В [3] рассмотрено применение для построения формантной картины гласных с помощью вейвлет-преобразования. В отличие от часто используемых методов преобразования Фурье и LPC-преобразования при таком подходе сигнал раскладывается на частотные полосы, но при этом сохраняется информация о его изменении во временной области. Иначе говоря, сохранена информация о появлении отдельных частот в сигнале. Появляется возможность изучения вариаций формант во времени. Для необходимой коррекции результатов может быть использован ранее подготовленный банк формант гласных.

Другой метод изучения изменения формант во времени использован в статьях [4, 5]. В них предлагается построение вектора состояния априорно заданной динамической системы (в данном случае речевого тракта) на основе применения рекурсивного фильтра (фильтра Калмана). При таком подходе сглажены временные скачки формантных частот.

В метод LPC-преобразования в работе [6] дополнительно введена процедура «сглаживания» скачков формант с применением алгоритма быстрой сходимости Ньютона – Рафсона. Проводится итерация с оптимальным шагом. Метод требует априорного задания структуры передаточной функции речевого тракта.

Близкий подход использован в [7], где для заранее определенной структуры речевого тракта форманты определяются рядом различных методов. Это метод LPC-преобразования, метод стабилизированного взвешенного преобразования SWLP (расчет полюсов модели при учете действия остаточной энергии, варьируя величину окна обработки), метод минимизации дисперсии искажения реакции MVDR. Для каждого метода определены границы использования.

В [8] рассмотрены общие источники ошибок в методах извлечения гласных формант при помощи LPC-преобразования, рассматриваются причины того, как программное обеспечение влияет на полученные формантные значения. Для этого проводится моделирование измерений, где для каждого гласного звука делаются тысячи измерений с различными настройками.

В [9] описан алгоритм, который автоматически находит самые гладкие формантные траектории для коротких отрезков речи в программе PRAAT. Метод выбирает для каждого сегмента самый ровный из ряда альтернатив. Критерий гладкости основан на моделировании формантных треков с полиномиальными функциями и использует также дисперсии полиномиальных коэффициентов. Автором отмечены

достоинства рассматриваемого метода – полная автоматизация и возможность воспроизведения.

В [10] предлагается новый формантный трекер с использованием смешанных моделей t -плотностей (tMM) для вокала. Резонансные частоты речевого тракта оцениваются методом гибридного линейного прогнозирования (HLP). Гибридный целочисленный цикл, синхронизированный с частотой основного тона (ЧОТ), и метод линейного предсказания (LP) улучшает частотное разрешение на вокализованных сегментах и приводит к более близким формантным оценкам, чем те, которые предоставляются другими методами LP. В сочетании с HLP траектории формант, как утверждают авторы, более точно отслеживаются tMM, чем гауссовскими смешанными моделями (GMM).

Акустическая частотная модель речевого тракта человека находит применение в ряде исследований акустико-физиологического направления [11-15]. Сама акустическая модель в них рассматривается как инструмент воспроизведения голосового сигнала. Однако их результаты показывают, что даже усложнение структуры модели «мозг-речевой аппарат» не дает адекватного описания процесса речи. Они пригодны при синтезе речи, но мало полезны при ее анализе.

Правда, могут решаться частные задачи. Например, определение по формантным частотам особенностей строения тела диктора, в частности оценка длины речевого тракта [16]. Изучался диапазон длин 14...19 см. Параллельно собирались данные магнитно-резонансной томографии. Сделан вывод, что для поставленной цели в большей степени пригодны высшие форманты, как менее подверженные артикуляции.

Еще одна группа исследований основывается на LF-модели речевого тракта Г. Фанта, то есть на выделении голосового источника и артикуляторного аппарата [17, 18]. Формантная картина речевого сигнала рассматривается как результат фильтрации артикуляторным аппаратом глоттальной волны от голосового источника. Особенности глоттальной волны отражаются в структуре формантной картины [19, 20].

Существует ряд работ, построенных на методе обратной фильтрации речевого сигнала с определением формы глоттальной волны. Для решения прямой задачи определения формантной картины он пригоден, но только в сочетании с применением описанной выше двухзвенной модели речевого тракта [21]. В этом случае предполагается, что глоттальная волна имеет сплошной спектр с пиком на частоте основного тона и подвергается дальнейшему частотному преобразованию на артикуляторном аппарате.

3. Развитие метода определения формантных характеристик в прикладных исследованиях. Форманты, или резонансные частоты надгортанных полостей речевого тракта давно считаются одним из

самых компактных и информативных наборов параметров для описания гласных, с важными коррелятами в перцептивном и артикуляторном аспектах. Тем не менее, несмотря на важность этих параметров, их трудно измерить надежно с помощью автоматических методов. Поэтому проводятся работы по конструированию корпусов формантных характеристик, то есть упорядоченные наборы формантных частот и полос пропускания. Естественно, что эти корпуса записаны под конкретные исследования, что определило количество и тип дикторов, фонетическую сложность речевого материала и другие экспериментальные условия. В [22] мотивом для создания нового набора данных формант в первую очередь послужило стремление создать материал, который может помочь синтезировать речь с широким спектром легко выбираемых личных характеристик диктора.

В [23] предлагается метод прогнозирования формант из данных по пику и полосе пропускания, и этот метод применен к извлечению значений первой и второй формант из корпуса региональных диалектов Северной Америки, содержащего 134 000 единицы ручного измерения формант. Это предсказание формант, как показывают результаты, существенно увеличивает производительность расчетов по определению формант по сравнению со стандартными пакетами анализа речи. Авторы демонстрируют, что можно провести социолингвистический анализ на основе данных формант гласных, используя прогнозируемые значения. В дальнейшем социолингвисты могут применять эту методологию для того, чтобы иметь возможность эффективно анализировать большие объемы данных.

Расчет формантных характеристик как метод широко используется в современных исследованиях как для поиска связи с физическими характеристиками говорящего и их восприятием [16, 24, 25], так и для решения более сложных задач фонетики. Предпринимаются попытки исследовать разницу в значении формантных характеристик в зависимости от положения тела говорящего [26], связать их изменения с возрастом человека [27], исследовать разницу формантных характеристик у близнецов [28]. Метод расчета формантных характеристик используется для исследования такого сложного вида речи, как шепотная [29, 30].

Кроме чисто фонетических и физиологических направлений формантный анализ постепенно вошел и в разработку систем идентификации человека по голосу и речи. Если в предыдущем веке формантные характеристики использовались в основном в синтезаторах речи, то в настоящее время они помогают разработчикам создавать более сложные системы, сочетая формантный анализ с различными статистическими математическими методами обработки речи. Например, в [31] используется технология лингвистически ограниченных i-

векторов или векторов идентичности на основе формант для автоматического распознавания говорящего. В [32] применяется линейный дискриминантный анализ в задачах судебной идентификации для распознавания говорящего по формантным характеристикам.

В [33] предлагается новый метод выделения характеристик говорящего, основанный на формантах, вейвлет-энтропии и нейронных сетях и обозначенный как FWENN. На первом этапе пять формант и семь пакетов Шеннона для энтропийных волн извлекаются из сигналов динамиком в качестве вектора характеристик говорящего. На втором этапе эти 12 коэффициентов используются в качестве входных данных для нейронных сетей с прямой связью. Вероятностная нейронная сеть также предлагается для сравнения. В отличие от обычного метода распознавания говорящего, при котором признаки извлекаются из предложений (или слов), предлагаемый метод выделяет признаки из гласных. Использование формант гласных позволяет распознавать говорящих, когда доступны только частично записанные слова. Это может быть полезно для глухонемых людей или когда записи повреждены. Экспериментальные результаты показывают, что предложенный метод успешно справляется с задачами верификации и идентификации дикторов с высокой степенью классификации. Результат достигается с помощью минимального объема информации, с использованием только 12 эффективных признаков (то есть длины вектора) и только одного гласного сигнала. Результаты сравниваются с хорошо известными классическими алгоритмами распознавания говорящего и превосходят их.

Принципиально новым направлением исследований является изучение того, как представлены форманты в нейролингвистической структуре мозга человека [34]. В этом случае формантный анализ можно рассматривать как новый инструмент изучения психофизиологических явлений.

4. Предложения по методике построения формантных картин гласных. Воспользуемся для определения формантных картин гласных математическим аппаратом частотных методов исследования динамических систем. Именно амплитудные частотные характеристики (АЧХ) представляют собой аналог фонетической формантной картины [21]. Обработав записи конкретной реализации речевой посылки, найдем АЧХ каждого из отделов речевого тракта. Для этой цели проведем совместную обработку двух различных участков этой записи, причем используем участки записи различной длительности:

- первый сигнал соответствует по времени части сегмента данной фонемы;
- второй сигнал соответствует всей речевой посылке (не менее нескольких минут).

Выбранный подход позволяет обратиться при построении формантной картины гласной к математическому аппарату частотных методов, разработанных в теории автоматического регулирования и использовать частотные передаточные функции звеньев речевого аппарата как формальный аналог формантной картины и инструмент ее описания [21].

Применение этого аппарата дает возможность применять для количественной оценки динамических явлений в речевом тракте стандартные приемы описания динамических систем, удобные для последующего перехода к фонетическим характеристикам.

Пренебрегая обратными связями между отделами речевого тракта, можно упрощенно представить его в форме двухзвенной разомкнутой динамической системы, приведенной на рисунке 1 [21].

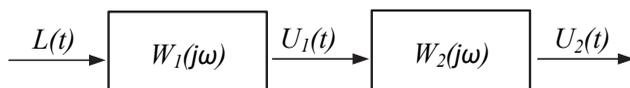


Рис. 1. Динамическая система речевого тракта человека, состоящая из двух динамических звеньев

Введены обозначения:

$L(t)$ — давление потока воздуха, формируемого органами дыхания, на входе голосового аппарата;

$W_1(j\omega)$ — частотная передаточная функция голосового источника;

$W_2(j\omega)$ — частотная передаточная функция артикуляторного аппарата;

$S(t)$ — внешнее воздействие;

$U_1(t)$ — выходной сигнал голосового источника;

$U_2(t)$ — выходной речевой сигнал.

Функция $S(t)$ описывает энергетическое воздействие при речи со стороны костно-мышечной системы и легких, иначе говоря, воздушный поток заданной интенсивности, действующий на голосовой источник. Можно считать, что этот поток пока не имеет частотной структуры, а все избирательные частотные свойства голосового источника отражены в частотной передаточной функции $W_1(j\omega)$, то есть собственно голосовой источник рассматривается по отношению к функции $S(t)$ как фильтрующее звено. Такому представлению хорошо соответствует биологическое описание работы нижнего отдела речевого тракта, включающего в себя

трахею, глотку и голосовые связки, как резонирующей нелинейной частотной системы, способной формировать автоколебательные процессы. Сформированный голосовым источником сигнал $U_1(t)$ поступает на вход второго фильтрующего звена артикуляторного аппарата, который физически представляет собой совокупность надгортанных полостей и артикулирующих органов. Таким образом, выходной сигнал с этого звена $U_2(t)$ определяется не только передаточной функцией $W_2(j\omega)$, но и структурой сигнала $U_1(t)$.

Предложенная двухзвенная модель речевого тракта открывает возможности поиска способов выделения сигналов $U_1(t)$ и $U_2(t)$. Один из этих способов построен на процедуре совместной обработки двух реализаций речевого сигнала, в которых априорно различна степень присутствия в сигнале $U_2(t)$ составляющих, обусловленных действием голосового источника и артикуляторного аппарата. В общем случае этот сигнал может рассматриваться как смесь из колебаний частоты основного тона и тембровых частот с одной стороны и колебаний с частотами, характерными для конкретных фонем — с другой. Степень присутствия каждой из этих групп составляющих может быть изменена при статистической обработке исходного аудиоматериала путем выбора участка и длительности обрабатываемого сегмента. Если сегментировать участок действия конкретной фонемы, то присутствие в спектрограмме составляющих ее собственных частот максимально.

С увеличением длительности обрабатываемого сегмента до фразы, нескольких фраз и далее до всего сообщения степень присутствия в спектрограмме частотных составляющих конкретной фонемы быстро падает из-за статистического усреднения. В то же время составляющие частот голосового источника присутствуют в речевой посылке почти постоянно (кроме глухих согласных и пауз). Их интенсивность почти не ослабевает. В результате для достаточно длинной речевой посылки (в расчетах бралось несколько минут) можно прийти к полностью сглаженной на высокочастотном участке спектральной плотности выходного сигнала $U_2(t)$, сохранившей в то же время пики на частотах основного тона и ее высших гармоник (если они были). В первом приближении такая спектральная плотность, в которой подавлено действие артикуляторного аппарата, может рассматриваться как спектральная плотность сигнала $U_1(t)$.

В отличие от частотной фильтрации описанный прием не вводит жесткой частотной границы фильтрации и позволяет сохранить в сигнале $U_1(t)$ как составляющую частоты основного тона, так и составляющие тембровых частот.

Вторым сравниваемым участком выбираем сегмент конкретной гласной фонемы. Частотная передаточная функция $W_2(j\omega)$ артикуляторного аппарата может быть описана соотношением (1) [21]:

$$|W_2(j\omega)|^2 = \frac{k_0^2 S_a(\omega)}{k_a^2 S_0(\omega)}, \quad (1)$$

где $S_a(\omega)$ – спектральная плотность сигнала $U_2(t)$, полученная при обработке сегмента фонемы; $S_0(\omega)$ — спектральная плотность сигнала $U_2(t)$, полученная при обработке всей речевой посылки; k_0 и k_a — коэффициенты усиления, полученные при обработке всей посылки и сегмента фонемы соответственно.

Сравнение полученных частотных характеристик с результатами определения формантных частот вручную по сонограммам подтверждают допустимость применения способа (рис. 2, 3).

Надо отметить, что основным искажением, которое сделанное допущение может внести в форму частотной характеристики, может быть неполная компенсация составляющей частоты основного тона. Причина в том, что при обработке речевой посылки в целом из-за просодических процессов пик, соответствующий частоте основного тона, может сформироваться более низкий и широкий, чем на графике спектральной плотности фонемы.

При необходимости, например для выделения сигнала $U_1(t)$, реально соответствующего сегменту фонемы, перед тем как провести операцию фильтрации экспериментально полученного сигнала $U_2(t)$ через звено с передаточной функцией обратной передаточной функции артикуляторного аппарата, надо провести структурирование этой передаточной функции, представив ее в виде набора элементарных динамических звеньев. Проведенные расчеты подтверждают применимость предлагаемого способа обработки экспериментального аудиоматериала для изучения частотных свойств голосового источника и артикуляторного аппарата.

Закономерно, что при математической обработке первого (короткого) участка на полученном графике АЧХ формантная картина исследуемой фонемы будет присутствовать в максимальной степени. Наоборот, на графике АЧХ длинной посылки эта формантная картина будет в максимальной степени «затерта», а наиболее рельефно будут присутствовать характеристики частоты основного тона, то есть по этой АЧХ можно оценить спектр глоттальной волны. Таким образом, сопоставление этих АЧХ методами обработки частотных характеристик позволяет выделить формантную картину исследуемой гласной, соответствующую данному короткому сегменту.

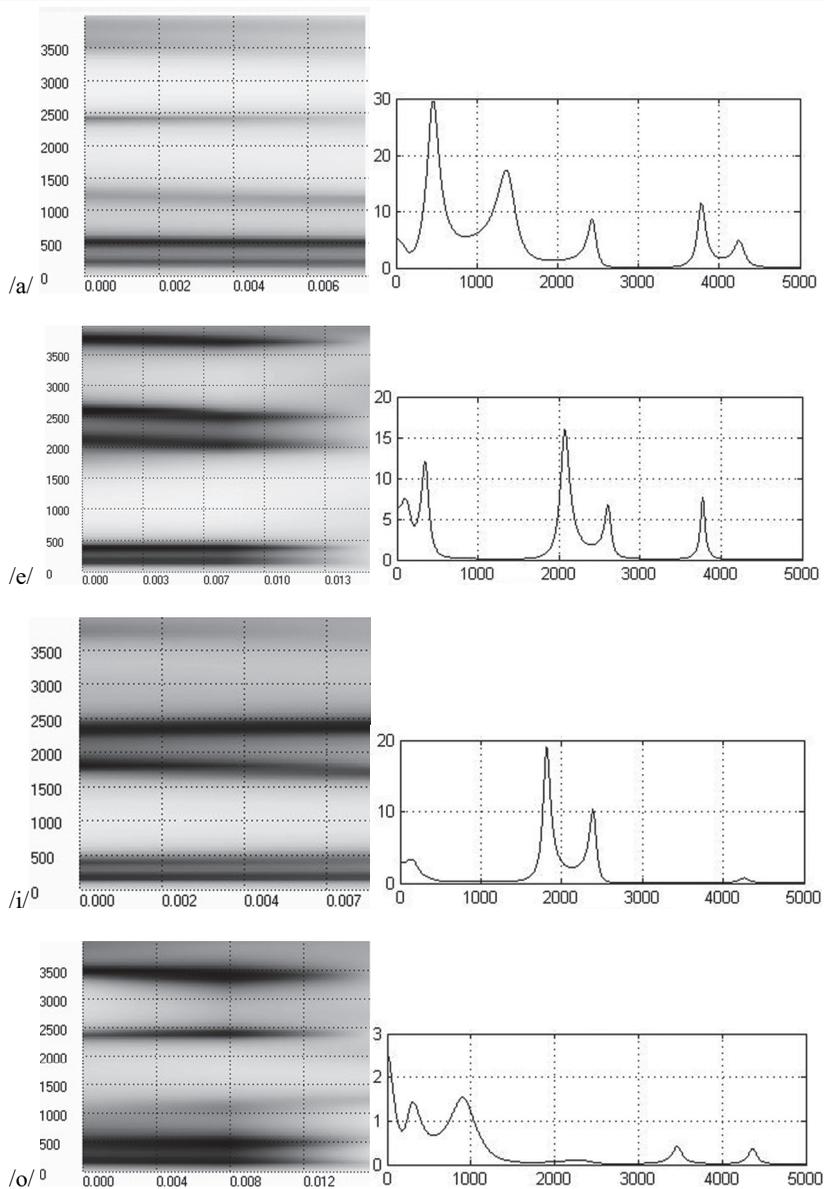


Рис. 2. Сонограммы (слева, ось X – время (с), Y – частота (Гц)) для сегментов гласных /a/, /e/, /i/, /o/; АЧХ артикуляторного аппарата того же сегмента гласной (справа, ось X – частота (Гц), Y – спектральная плотность мощности)

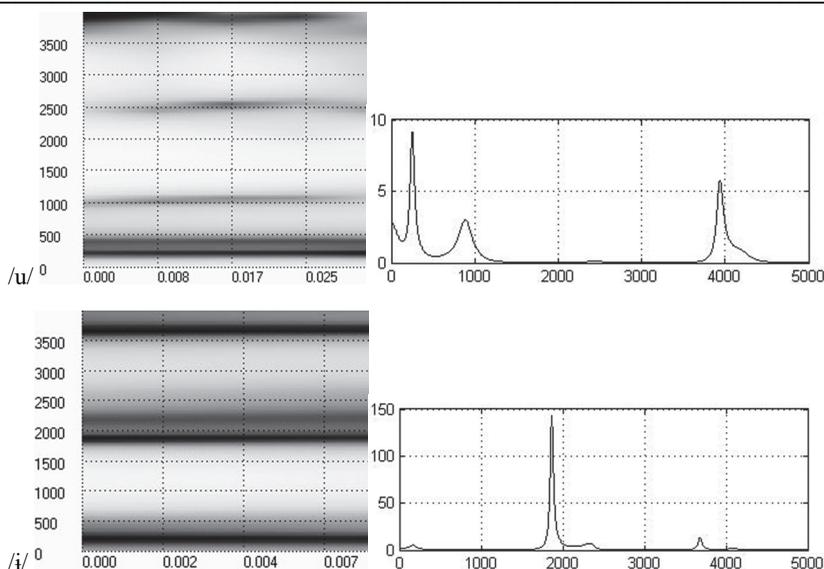


Рис. 3. Сонограммы (слева, ось X – время (с), Y – частота (Гц)) для сегментов гласных /u/, /i/; АЧХ артикуляторного аппарата того же сегмента гласной (справа, ось X – частота (Гц), Y – спектральная плотность мощности)

На рисунках 2, 3 приведены примеры рассчитанных АЧХ и соответствующие им сонограммы аудиосигнала для русских гласных. Сонограммы были сделаны в программе Wave Assistant. Амплитудно-частотные характеристики были получены с помощью оригинальных скриптов, написанных в программе Matlab. Соответствие частотных пиков подтверждает адекватность и применимость метода. На любом временном отрезке гласной может быть получена соответствующая формантная картина.

Основные особенности метода:

- описание в соответствии с LF-моделью Г. Фанта речевого тракта двумя последовательно соединенными динамическими звеньями, каждое из которых описывает динамику его соответствующего отдела;
- построение формантной картины гласной методом совместной обработки АЧХ отделов речевого тракта;
- степень детализации формантной картины гласной определяется выбором ширины окна обработки;
- форма АЧХ артикуляторного аппарата отражает энергетическое соотношение между отдельными формантами гласной;
- возможность автоматизированной обработки речевого материала по описанному методу.

Собранный при посегментной обработке реализации гласной массив АЧХ применим для решения фонетических задач, например для выявления соответствия между перцептивно отобранными свойствами речи и трансформациями формантных картин гласных (см. табл. 1). В таблице видно, что расчет формантных характеристик произведен таким образом, что значения каждой из формант на протяжении гласного участка не перемешиваются между собой. Таким образом, метод позволяет оценить значение резонансных частот надгортанных полостей речевого тракта в динамике и выделить каждую форманту отдельно.

Таблица 1. Трансформация формантных характеристик Гц во времени в одной реализации гласного /e/

F1	F2	F3
494	2730	3478
494	2776	3498
502	2846	3536
506	2876	3424
506	2746	3330
504	2546	3156
490	2336	3080

5. Расчет формантных характеристик дикторов на корпусе CORPRES. В работе проводился расчет формантных характеристик дикторов на корпусе CORPRES [35]. Это полностью аннотированный русский речевой корпус, разработанный на Кафедре фонетики и методики преподавания иностранных языков Санкт-Петербургского государственного университета в результате трехлетнего проекта по созданию корпуса для синтеза речи методом unit-selection. Целью корпуса было предоставить большой образец стандартной русской связной речи. Первоначально он был предназначен для использования в синтезе unit-selection TTS, однако, предполагалось, что он может быть пригоден для использования в более широком диапазоне фонетических исследований и разработок. Корпус включает образцы различных стилей речи, воспроизведенных 4 мужчинами и 4 женщинами. Шесть уровней аннотации охватывают всю фонетическую и просодическую информацию о записанных речевых данных, включая метки частоты основного тона, границы звуков и фонетических событий внутри звуков, узкую и широкую фонетические транскрипции, а также орфографическую расшифровку и просодическую аннотацию. Точная фонетическая транскрипция данных обеспечивает особенно ценный ресурс как для исследований, так и для целей разработки. Общий размер корпуса составляет 528 458 слов и содержит 60 часов речи, состоящих из 7,5 часов от каждого говорящего. 40% корпуса было сегментировано вручную и полностью анно-

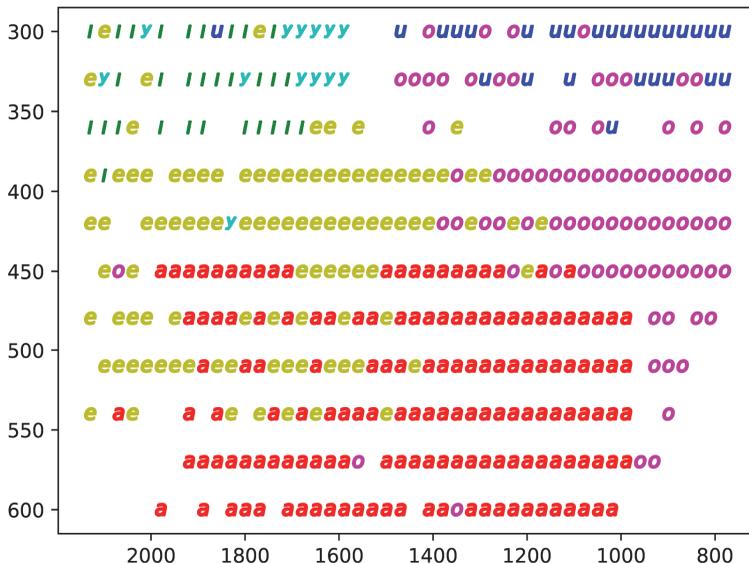


Рис. 5. Диктор 2. Мужской голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц)

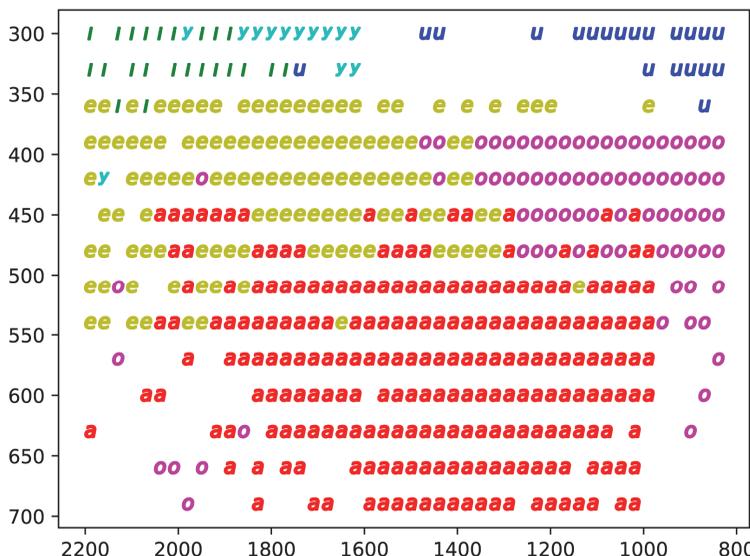


Рис. 6. Диктор 3. Мужской голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц)



Рис. 7. Диктор 4. Мужской голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц)

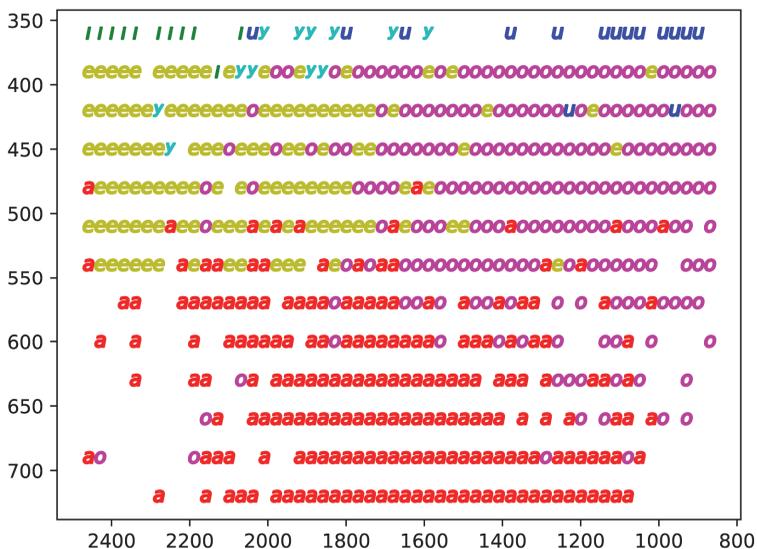


Рис. 8. Диктор 5. Женский голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц)

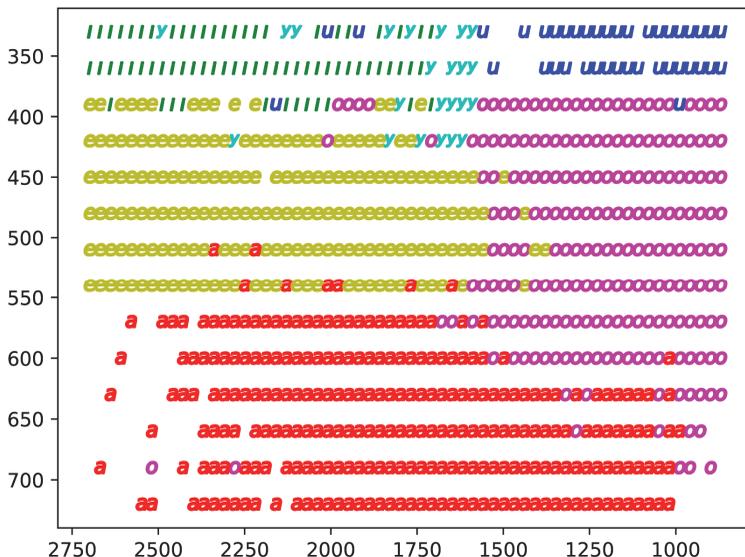


Рис. 9. Диктор 6. Женский голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц)

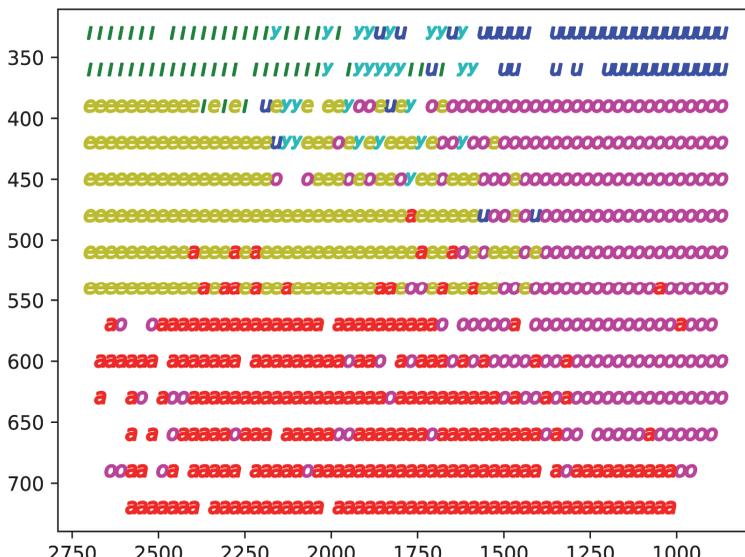


Рис. 10. Диктор 7. Женский голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц)

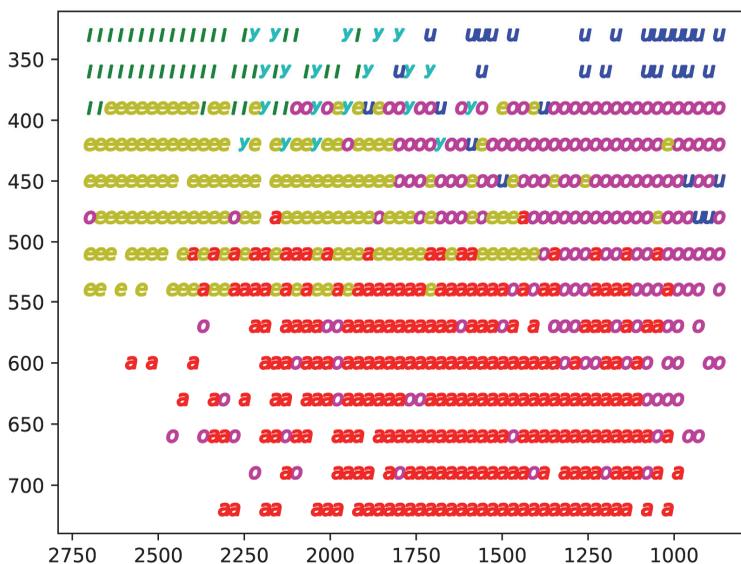


Рис. 11. Диктор 8. Женский голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц)

Данные по каждому диктору были рассчитаны на массиве примерно 43000 реализаций гласных. Из них для данных формантных плоскостей были взяты только ударные реализации. На рисунках видно, что ударные гласные у дикторов в целом лежат в тех местах, где обычно и находятся реализации основных аллофонов гласных. Также на отображение данных повлияла собственная частотность некоторых гласных фонем в русском языке. Например, /a/ является самой частотной фонемой, а /i/ встречается реже. Полученные данные также позволяют оценить формантные характеристики безударных гласных в соответствии с тем, как они были обозначены на уровне реальной транскрипции аннотаторами корпуса. Для фонемы /a/ в корпусе обозначались для безударных звуков первый предударный, второй предударный и заударные аллофоны, для других фонем – только предударные и заударные аллофоны. Применение данного метода дает также возможность оценить расположение и безударных аллофонов гласных.

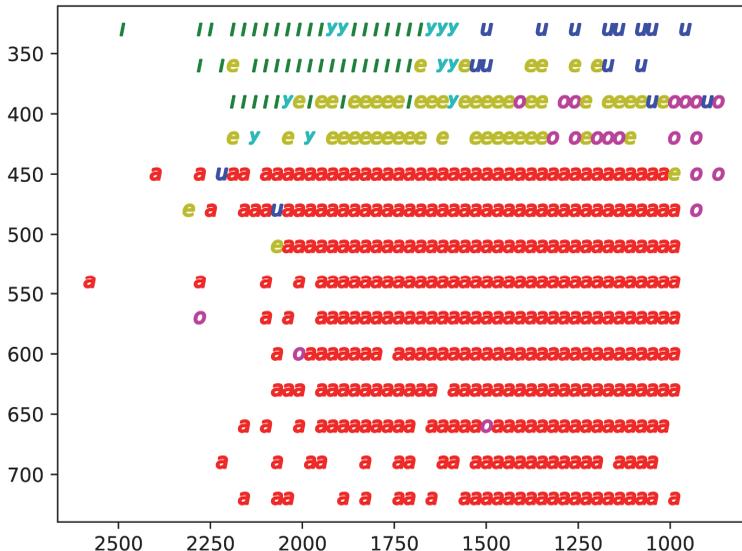


Рис. 12. Диктор 4. Мужской голос. Предударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц). Оси повернуты в соответствии с тем, как лежат гласные на артикуляторно-акустическом треугольнике

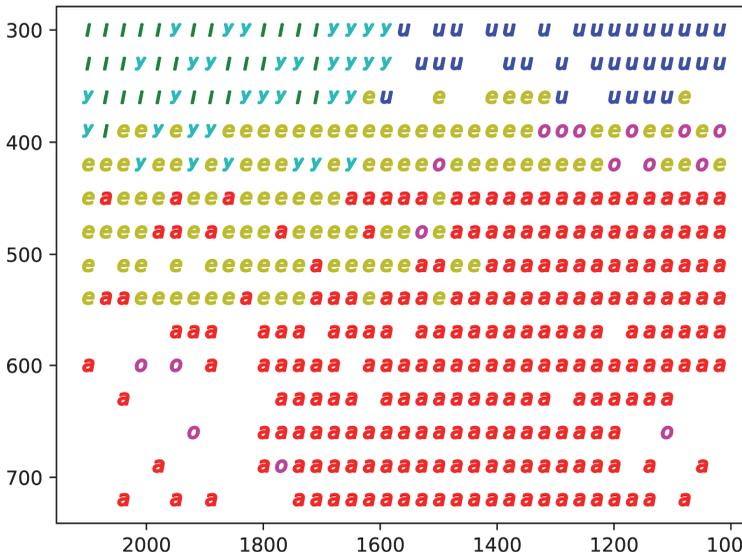


Рис. 13. Диктор 4. Мужской голос. Заударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц). Оси повернуты в соответствии с тем, как лежат гласные на артикуляторно-акустическом треугольнике

6. Использование метода расчета формант гласных для исследования характеристик безударных гласных в потоке речи. Предлагаемый метод определения формант использовался при исследовании акустических характеристик безударных гласных, которые, будучи изначально неогубленными, становятся огубленными под влиянием огубленных гласных, находящихся в следующем слоге, например произношение слова «голубому» как [gulubomu] вместо [gəlubomu] или слова «факультет» как [fukul^ltʲet] вместо [fəkul^ltʲet].

В качестве экспериментального материала использовались записи чтения одним диктором мужского пола «ОМ1» специально подобранных предложений, которые учитывали различные лингвистические факторы, такие как: (1) находится ли перед гласным губной согласный или нет («попугай», «голубой»); (2) на каком расстоянии (в слогах) находится огубленный гласный, который оказывает влияние: в следующем слоге («голубой») или через один («годовую»); (3) находятся ли эти два гласных в одном слове или в соседних словах («попугай», «папа купил»); (4) в какой позиции относительно ударения находится гласный: 2-й предударный, 1-й предударный, ударный, заударный («голубой», «голубка», «каску», «нескольку»).

Результаты анализа показали, что высока вероятность огубленности безударного неогубленного гласного, если в следующем слоге есть безударный огубленный гласный. Подробный анализ влияния различных факторов на огубленность безударных гласных описан в [37].

Вслед за анализом факторов был проведен акустический анализ тех гласных, которые становятся огубленными. В рамках анализа исследовались первые форманты, значения которых были получены при помощи описанного выше метода. На рисунке 14 представлены значения формант ударных гласных в речи диктора ОМ1, это акустическое пространство гласных использовалось в качестве референтного при анализе акустических реализаций безударных гласных, которые представлялись интерес. На рисунке 15 представлены значения формант безударных гласных /i/ и /a/ в речи диктора ОМ1, которые стали огубленными под влиянием безударных /u/, находящихся в следующем слоге.

Анализ данных формантных составляющих гласных показал, что значения F1 этих гласных почти такие же, как у ударного гласного /u/ и отличаются незначительно. Артикуляторно подъем этих гласных соответствует подъему /u/. Значения F2 у гласных [u], которые должны были реализовываться как [i] и [a] и обозначены как U_a и U_i , соответственно, различаются. И те, и другие гласные более продвинуты вперед по сравнению с ударным /u/, однако U_i гораздо более продвинуты вперед, чем U_a . U_i оказывается в акустической области,

соответствующей ударному /i/ (на рисунке он обозначен как «у»), в то время как U_a можно обозначить как гласный заднего ряда, хоть и более передний, чем ударный /o/.

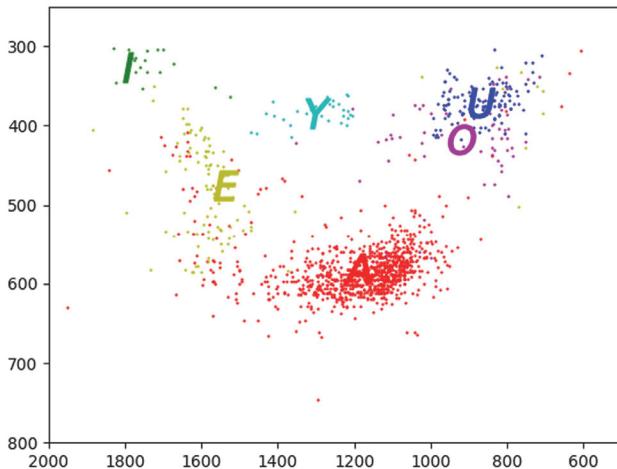


Рис. 14. Диктор ОМ1. Мужской голос. Ударные гласные. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц). Оси повернуты в соответствии с тем, как лежат гласные на артикуляторно-акустическом треугольнике

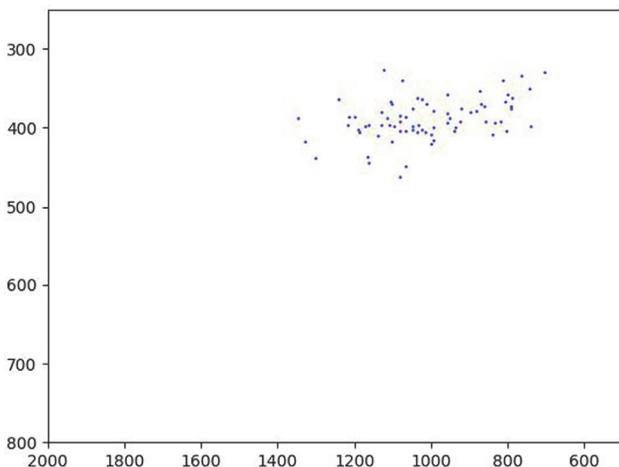


Рис. 15. Диктор ОМ1. Мужской голос. Безударные гласные огубленные под влиянием соседней гласной. Ось X – вторая форманта (Гц), ось Y – первая форманта (Гц). Оси повернуты в соответствии с тем, как лежат гласные на артикуляторно-акустическом треугольнике

7. Заключение. Разработана методика оценки вариативности формантной картины на материале корпуса профессионального чтения русского языка. Выполнен отбор акустического аудиоматериала. Сделана поаллофонная группировка реализаций с учетом позиции гласной в слове. Построены картины девиации основной форманты по сегменту для всех отобранных реализаций. Для каждой отобранной реализации построена картина девиации доминирующей форманты на протяжении сегмента. Сравнительным анализом проверено соответствие структурных и частотных отличий при трансформации формантной картины для разных дикторов и различных позиций гласного в слове. Подтверждено, что в режиме чтения позиция ударных аллофонов на плоскости, образованной первыми двумя формантами, подобна позиции основных аллофонов гласных фонем в русском языке.

Разработанная методика позволяет проводить оценку формантных частот гласных не только в зависимости от ударности гласного, но и от его акустического окружения и других фонетических факторов. При помощи предложенного метода были получены результаты об акустической реализации огубленности неогубленных безударных гласных. Анализ данных показал, что их ряд такой же, как у безударных аллофонов, а подъем такой же как у /u/.

Литература

1. *Гордина М.В.* История фонетических исследований: (от античности до возникновения фонологической теории) // Филологический факультет Санкт-Петербургского государственного университета. 2006. С. 34–37.
2. *Фант Г.* Акустическая теория речеобразования // М.: Наука. 1964. 284 с.
3. *Chaari S., Ouni K., Ellouze N.* Wavelet ridge track interpretation in terms of formants // Ninth International Conference on Spoken Language Processing. 2006. pp. 1017–1020.
4. *Özbek I.Y., Demirekler M.* Tracking of visible vocal tract resonances (VVTR) based on kalman filtering // Ninth International Conference on Spoken Language Processing. 2006. 4 p.
5. *Mellahi T., Hamdi R.* LPC-based formant enhancement method in Kalman filtering for speech enhancement // AEU-International Journal of Electronics and Communications. 2015. vol. 69. no. 2. pp. 545–554.
6. *Weruaga L., Al-Khayat A.* All-pole model estimation of vocal tract on the frequency domain // Ninth International Conference on Spoken Language Processing. 2006. pp. 1001–1004.
7. *Magi C., Bäckström T., Alku P.* Stabilised weighted linear prediction—a robust all-pole method for speech processing // Eight Annual Conference of the International Speech Communication Association. 2007. pp. 522–525.
8. *Kendall T., Vaughn C.* Measurement variability in vowel formant estimation: a simulation experiment // Proceedings of The Scottish Consortium for ICPHS 2015. 2015. 5 p.
9. *Weenink D.* Improved formant frequency measurements of shortsegments // The Scottish Consortium for ICPHS 2015. 2015. 4 p.

10. *Ramirez M.A.* Hybrid Autoregressive Resonance Estimation and Density Mixture Formant Tracking Model // IEEE Access. 2018. vol. 6. pp. 30217–30224.
11. *Arai T.* Sliding Vocal-tract Model and its Application for Vowel Production // Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 72–75.
12. *Ghosh P.K. et al.* Estimation of articulatory gesture patterns from speech acoustics // Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 2803–2806.
13. *Fang Q., Nishikido A., Dang J.* Feedforward Control of A 3D Physiological Articulatory Model for Vowel Production // Tsinghua Science and Technology. 2009. vol. 14. no. 5. pp. 617–622.
14. *Arai T.* Simple Physical Models of the Vocal Tract for Education in Speech Science // Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 756–759.
15. *Lu X.B., Thorpe W., Foster K., Hunter P.* From experiments to articulatory motion – A three dimensional talking head model // Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 64–67.
16. *Lammert A.C., Narayanan S.S.* On Short-Time Estimation of Vocal Tract Length from Formant Frequencies // PloS one. 2015. vol. 10(7). pp. e0132193.
17. *Fant G., Liljencrants J., Lin Q.* A four-parameter model of Glottal Flow // STL-QPSR 1985. vol. 4. no. 1985. pp. 1–13.
18. *Fant G.* The voice source in the connected speech // Speech Communication. 1997. vol. 22. no. 2-3. pp. 125–139.
19. *Murphy P.J.* Relationship between Harmonic Amplitudes and Spectral Zeros and Glottal Open Quotient // International Conference on Phonetic Science (ICPhS). 2007. pp. 889–892.
20. *Uezu Y., Kaburagi T.* Analysis of voice register transition focused on the relationship between pitch and formant frequency // The Scottish Consortium for ICPhS. 2015. 5 p.
21. *Евдокимова В.В.* Системный подход к определению параметров речевого тракта // Вестник Санкт-Петербургского университета. Язык и литература. 2007. № 2-II. pp. 144–148.
22. *Mokhtari P., Tanaka K.A.* Corpus of Japanese Vowel Formant Patterns // Bulletin of The Electrotechnical Laboratory (ETL). 2000. vol. 64. pp. 57–66.
23. *Evanini K., Isard S., Liberman M.* Automatic formant extraction for sociolinguistic analysis of large corpora // Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 1655–1658.
24. *Barreda S.* Investigating the use of formant frequencies in listener judgments of speaker size // Journal of Phonetics. 2016. vol. 55. pp. 1–18.
25. *Macari A.T. et al.* Correlation Between the Position of the Hyoid Bone on Lateral Cephalographs and Formant Frequencies // Journal of voice. 2016. vol. 30. no. 6. pp. 757.
26. *Hoedl P.* Defying gravity: formant frequencies of English vowels produced in upright and supine body position // The Scottish Consortium for ICPhS 2015. 2015. 5 p.
27. *Eichhorn J.T., Kent R.D., Austin D., Vorperian H.K.* Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women // Journal of Voice. 2018. vol. 32. no. 5. pp. 644. e1-644. e9.
28. *Zuo D., Mok P.P.K.* Formant dynamics of bilingual identical twins // Journal of Phonetics. 2015. vol. 52. pp. 1–12
29. *Heeren W.F.L.* Can formant shifts and effort cues enhance boundary tone perception in whispered speech? // The Scottish Consortium for ICPhS 2015. 2015. 5 p.

30. *Zhao Y., Lin W.* Study of the formant and duration in Chinese whispered vowel speech // *Applied Acoustics*. 2016. vol. 114. pp. 240–243.
31. *Franco-Pedroso J., Gonzalez-Rodriguez J.* Linguistically-constrained formant-based i-vectors for automatic speaker recognition // *Speech Communication*. 2016. vol. 76. pp. 61–81.
32. *Skarnitzl R., Vaňková J.* Speaker discrimination using formant trajectories from case-work recordings: can LDA do it? // *The Scottish Consortium for ICPhS* 2015. 2015. 5 p.
33. *Daqrouq K., Tutunji T.A.* Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers // *Applied Soft Computing*. 2015. vol. 27. pp. 231–239.
34. *Fisher J.M., Dick F.K., Levy D.F., Wilson S.M.* Neural representation of vowel formants in tonotopic auditory cortex // *NeuroImage*. 2018. vol. 178. pp. 574–582.
35. *Skrelin P.A. et al.* A Fully Annotated Corpus of Russian Speech // *Proceedings of the International Conference on Language Resources and Evaluation*. 2010. pp. 109–112.
36. *Евдокимова В.В.* Вариативность формантной структуры гласного в разных видах речи // *Первый междисциплинарный семинар «Анализ разговорной речи» (АРЗ-2007)*. 2007. С. 49–54.
37. *Kocharov D., Evdokimova V., Evgrafova K., Morskovatykh M.* Labialization of unstressed vowels in Russian: phonetic and perceptual evidence // *International Conference on Speech and Computer*. 2018. pp. 301–310.

Евдокимова Вера Вячеславовна — канд. филол. наук, доцент, кафедра фонетики и методики преподавания иностранных языков, Санкт-Петербургский государственный университет (СПбГУ). Область научных интересов: языкознание, фонетика, фонология, речевые технологии, акустический анализ речевого сигнала, идентификация человека по голосу и речи, синтез и распознавание речевого сигнала, структура речевого поведения. Число научных публикаций — 41. postmaster@phonetics.spb.ru; Университетская наб., 11, 199034, Санкт-Петербург, Россия; р.т.: +7(812)328-9565.

Кочаров Даниил Александрович — канд. филол. наук, доцент, кафедра фонетики и методики преподавания иностранных языков, Санкт-Петербургский государственный университет (СПбГУ). Область научных интересов: автоматическая обработка речи и текста, математическая лингвистика, речевые технологии, фонетика, фонология. Число научных публикаций — 42. kocharov@phonetics.spb.ru; Университетская наб., 11, 199034, Санкт-Петербург, Россия; р.т.: +79219940104.

Скрелин Павел Анатольевич — д-р филол. наук, профессор, заведующий кафедрой, кафедра фонетики и методики преподавания иностранных языков, Санкт-Петербургский государственный университет (СПбГУ). Область научных интересов: языкознание, фонетика, фонология, речевые технологии, акустический анализ речевого сигнала, идентификация человека по голосу и речи, синтез и распознавание речевого сигнала, структура речевого поведения, автоматическая обработка речи и текста. Число научных публикаций — 100. skrelin@phonetics.spb.ru; Университетская наб., 11, 199034, Санкт-Петербург, Россия; р.т.: +7(812)328-9565.

Поддержка исследований. Исследование безударных гласных было выполнено при финансовой поддержке РФФИ (проект № 18-012-00767).

V. EVDOKIMOVA, D. KOCHAROV, P. SKRELIN
**METHOD FOR CONSTRUCTING FORMANTS FOR STUDYING
PHONETIC CHARACTERISTICS OF VOWELS**

V. Evdokimova, D. Kocharov, P. Skrelin. Method for Constructing Formants for Studying Phonetic Characteristics of Vowels.

Abstract. This article presents the results of applying method for obtaining formant components of vowel phonemes for the corpus of professional reading in Russian. In this paper, a review of existing areas of development of methods for obtaining formant characteristics of vowels for different languages was made. A review was also made of the extent to which formant picture patterns are used in speech technologies and natural language processing. On the corpus of professional reading CORPRES, data was obtained on formant components for 351929 realizations of vowel phonemes on the material of 8 speakers. The data obtained are grouped in accordance with the symbols in the real transcription, which was performed by phoneticians within the framework of segmenting the corpus. The formant planes represent the distribution of allophones of vowels for all speakers according to the two first formants. The variability of formant characteristics in the corpus for pre-tonic and post-tonic allophones are presented for one male speaker. The article also presents the results testifying the difference between the rounded unstressed /i/ and /a/, which are perceived by both naive speakers and expert phoneticians as /u/. As an experimental material, the recordings of reading by one male announcer of specially selected sentences, which took into account various linguistic factors, were used. Analysis of the data of the formant components of these vowels showed that the values of the first formant of these vowels are close to the values of the stressed vowel /u/ for this speaker. The closure of these vowels corresponds to the closure of /u/. The second formant values in the vowels [u], which were to be realized as [i] and [a] are different. They are more advanced in comparison with /u/.

Keywords: phonetics, phonology, formants, acoustic analysis, digital signal processing, vowels, Russian.

Evdokimova Vera — Ph.D., Associate Professor, Department of Phonetics, Saint Petersburg State University (SPSU). Research interests: linguistics, phonetics, phonology, speech technologies, acoustic analysis of a speech signal, identification of a person by voice and speech, synthesis and recognition of a speech signal, structure of speech behavior. The number of publications — 41. postmaster@phonetics.spb.ru; 11, Universitetskaya emb., 199034, Saint-Petersburg, Russia; office phone: +7(812)328-9565.

Kocharov Daniil — Ph.D., Associate professor, Department of Phonetics, Saint Petersburg State University (SPSU). Research interests: automatic speech signal and text processing, computational linguistics, speech technologies, phonetics, phonology. The number of publications — 42. kocharov@phonetics.spb.ru; 11, Universitetskaya emb., 199034, Saint-Petersburg, Russia; office phone: +79219940104.

Skrelin Pavel — Ph.D., Dr.Sci., Professor, Head of Department, Department of Phonetics, Saint Petersburg State University (SPSU). Research interests: automatic speech signal and text processing, linguistics, phonetics, phonology, speech technologies, acoustic analysis of a speech signal, identification of a person by voice and speech, synthesis and recognition of a speech signal. The number of publications — 100. skrelin@phonetics.spb.ru; 11, Universitetskaya emb., 199034, Saint-Petersburg, Russia; office phone: +7(812)328-9565.

Acknowledgements. The research of unstressed vowels was supported by RFBR (grant 18-012-00767).

References

1. Gordina M.V. *Istorija foneticheskikh issledovanij (ot antichnosti do vozniknovenija fonologicheskoi teorii)* [History of phonetic studies (from the ancient times till modern times)]. Filologicheskij fakul'tet Sankt-Peterburgskogo gosudarstvennogo universiteta. 2006. 538 p. (In Russ.).
2. Fant G. *Akusticheskaja teorija recheobrazovanija* [Acoustic theory of speech production]. M.: Nauka. 1964. 284 p. (In Russ.).
3. Chaari S., Ouni K., Ellouze N. Wavelet ridge track interpretation in terms of formants. Ninth International Conference on Spoken Language Processing. 2006. pp. 1017–1020.
4. Özbek I.Y., Demirekler M. Tracking of visible vocal tract resonances (VVTR) based on kalman filtering. Ninth International Conference on Spoken Language Processing. 2006. 4 p.
5. Mellahi T., Hamdi R. LPC-based formant enhancement method in Kalman filtering for speech enhancement. *AEU-International Journal of Electronics and Communications*. 2015. vol. 69. no. 2. pp. 545–554.
6. Weruaga L., Al-Khayat A. All-pole model estimation of vocal tract on the frequency domain. Ninth International Conference on Spoken Language Processing. 2006. pp. 1001–1004.
7. Magi C., Bäckström T., Alku P. Stabilised weighted linear prediction—a robust all-pole method for speech processing. Eight Annual Conference of the International Speech Communication Association. 2007. pp. 522–525.
8. Kendall T., Vaughn C. Measurement variability in vowel formant estimation: a simulation experiment. Proceedings of the Scottish Consortium for ICPhS 2015. 2015. 5 p.
9. Weenink D. Improved formant frequency measurements of shortsegments. The Scottish Consortium for ICPhS 2015. 2015. 4 p.
10. Ramirez M.A. Hybrid Autoregressive Resonance Estimation and Density Mixture Formant Tracking Mode. *IEEE Access*. 2018. vol. 6. pp. 30217–30224.
11. Arai T. Sliding Vocal-tract Model and its Application for Vowel Production. Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 72–75.
12. Ghosh P.K. et al. Estimation of articulatory gesture patterns from speech acoustics. Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 2803–2806.
13. Fang Q., Nishikido A., Dang J. Feedforward Control of A 3D Physiological Articulatory Model for Vowel Production. *Tsinghua Science and Technology*. 2009. vol. 14. no. 5. pp. 617–622.
14. Arai T. Simple Physical Models of the Vocal Tract for Education in Speech Science. Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 756–759.
15. Lu X.B., Thorpe W., Foster K., Hunter P. From experiments to articulatory motion – A three dimensional talking head model. Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 64–67.
16. Lammert A.C., Narayanan S.S. On Short-Time Estimation of Vocal Tract Length from Formant Frequencies. *PLoS one*. 2015. vol. 10(7). pp. e0132193.
17. Fant G., Liljencrants J., Lin Q. A four-parameter model of Glottal Flow. *STL-QPSR*. 1985. vol. 4. no. 1985. pp. 1–13.
18. Fant G. The voice source in the connected speech. *Speech Communication*. 1997. vol. 22. no. 2-3. pp. 125–139.

19. Murphy P.J. Relationship between Harmonic Amplitudes and Spectral Zeros and Glottal Open Quotient. International Conference on Phonetic Science (ICPhS). 2007. pp. 889–892.
20. Uezu Y., Kaburagi T. Analysis of voice register transition focused on the relationship between pitch and formant frequency. The Scottish Consortium for ICPhS. 2015. 5 p.
21. Evdokimova V.V. [A systematic approach to determining the parameters of the vocal tract]. *Vestnik Sankt-Peterburgskogo universiteta. Yazyk i literatura – Vestnik of Saint Petersburg University. Language and Literature*. 2007. vol. 2-II. pp. 144–148.
22. Mokhtari P., Tanaka K.A. Corpus of Japanese Vowel Formant Patterns. *Bulletin of the Electrotechnical Laboratory (ETL)*. 2000. vol. 64. pp. 57–66.
23. Evanini K., Isard S., Liberman M. Automatic formant extraction for sociolinguistic analysis of large corpora. Tenth Annual Conference of the International Speech Communication Association. 2009. pp. 1655–1658.
24. Barreda S. Investigating the use of formant frequencies in listener judgments of speaker size. *Journal of Phonetics*. 2016. vol. 55. pp. 1–18.
25. Macari A.T. et al. Correlation Between the Position of the Hyoid Bone on Lateral Cephalographs and Formant Frequencies. *Journal of voice*. 2016. vol. 30. no. 6. pp. 757.
26. Hoedl P. Defying gravity: formant frequencies of English vowels produced in upright and supine body position. The Scottish Consortium for ICPhS 2015. 2015. 5 p.
27. Eichhorn J.T., Kent R.D., Austin D., Vorperian H.K. Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women. *Journal of Voice*. 2018. vol. 32. no. 5. pp. 644. e1-644. e9.
28. Zuo D., Mok P.P.K. Formant dynamics of bilingual identical twins. *Journal of Phonetics*. 2015. vol. 52. pp. 1–12
29. Heeren W.F.L. Can formant shifts and effort cues enhance boundary tone perception in whispered speech? The Scottish Consortium for ICPhS 2015. 2015. 5 p.
30. Zhao Y., Lin W. Study of the formant and duration in Chinese whispered vowel speech. *Applied Acoustics*. 2016. vol. 114. pp. 240–243.
31. Franco-Pedroso J., Gonzalez-Rodriguez J. Linguistically-constrained formant-based i-vectors for automatic speaker recognition. *Speech Communication*. 2016. vol. 76. pp. 61–81.
32. Skarnitzl R., Vaňková J. Speaker discrimination using formant trajectories from case-work recordings: can LDA do it? The Scottish Consortium for ICPhS 2015. 2015. 5 p.
33. Daqrouq K., Tutunji T.A. Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers. *Applied Soft Computing*. 2015. vol. 27. pp. 231–239.
34. Fisher J.M., Dick F.K., Levy D.F., Wilson S.M. Neural representation of vowel formants in tonotopic auditory cortex. *NeuroImage*. 2018. vol. 178. pp. 574–582.
35. Skrelin P.A. et al. A Fully Annotated Corpus of Russian Speech. Proceedings of the International Conference on Language Resources and Evaluation. 2010. pp. 109–112.
36. Evdokimova V.V. [Variability of the formant structure of a vowel in different types of speech]. *Pervyj mezhdisciplinarnyj seminar "Analiz razgovornoj rechi"* [First Interdisciplinary Seminar "Analysis of Conversational Speech"]. 2007. pp. 49–54. (In Russ.).
37. Kocharov D., Evdokimova V., Evgrafova K., Morskovatykh M. Labialization of unstressed vowels in Russian: phonetic and perceptual evidence. International Conference on Speech and Computer. 2018. pp. 301–310.