

*А.В. Венцов**Санкт-Петербург*

## **Фонетическое аннотирование звучащего текста: проблема согласного /j/**

Одним из эффективных способов исследования сложных систем является функциональное моделирование, при котором исследуются и моделируются не сами физические процессы в системе, а правила (формулы, алгоритмы) преобразования информации в ней – так, чтобы реакция всей модели и составляющих ее модулей на входное воздействие совпадали с соответствующими реакциями самой системы.

Нет сомнений, что к числу таких систем относится и система восприятия речи человеком, особая сложность которой состоит в том, что ни один из ее условных «модулей» не доступен прямому наблюдению и гипотезы о структуре и функциях этих модулей строятся по косвенным данным, получаемым в экспериментах на животных, а также в психоакустических и психолингвистических экспериментах на человеке.

Акустический речевой сигнал обрабатывается прежде всего периферическими отделами слуховой системы [Физиология речи 1976; Чистович и др. 1986] и далее преобразуется, предположительно, в последовательность дискретных «символов», природа которых пока неизвестна, а при построении функциональной модели определяется теми требованиями ко входному сигналу, которые исходят от следующих по иерархии «модулей». Предполагается, что для модулей лексического поиска, грамматической и семантической обработки может быть достаточным описание сигнала набором символов, напоминающим символы фонетической транскрипции. При этом структура набора, количество и тип символов должны быть необходимыми и достаточными для успешной работы следующих модулей, т.е. не перегруженными слишком подробными деталями и отражающими всю акустическую инфор-

мацию, необходимую для последующей адекватной интерпретации обрабатываемого сигнала. Здесь необходимо особо подчеркнуть, что все отклонения от орфоэпических норм во входном речевом сигнале обязательно должны фиксироваться, поскольку именно в таких условиях и функционирует система восприятия речи носителя языка.

Если принять подобный подход, то без большого ущерба процесс создания полной функциональной модели системы восприятия речи можно разделить на две сравнительно независимые части: преобразование непрерывного акустического сигнала в цепочку дискретных символов и разработку алгоритмов дальнейшего «лингвистического» анализа полученного результата. Последняя часть неизбежно потребует исчерпывающих знаний о том, что именно представляет собой этот результат, т.е. понимания того, с каким именно «входным сигналом» предстоит работать «модулям» лингвистического анализа. Для получения таких знаний необходимо создание речевых корпусов, аннотированных по описанным выше правилам с обязательным сохранением всей информации об особенностях акустической реализации лексических единиц.

Опыт создания пробного корпуса спонтанной речи, размеченного в соответствии с изложенными выше принципами [Венцов и др. 2013], свидетельствует, что на основании оценок экспертов и результатов анализа динамических спектрограмм удастся без особых усилий безошибочно идентифицировать большинство сегментов в речевом сигнале, соответствующих согласным фонемам. Исключение составляет согласный «йот» (/j/): когда его «слышат» носители языка, спектрографический анализ далеко не всегда позволяет понять, какими именно измеряемыми инструментально акустическими характеристиками должен быть описан этот согласный, чтобы его «услышали» модули лингвистического анализа. Другими словами, возникает нетривиальный вопрос о том, всегда ли «услышанный» экспертом согласный должен отображаться символом «[j]» в строке акустико-фонетической транскрипции? Разумеется, речь здесь идет об описании речевого сигнала, предпо-

ложительно отражающем уровень его обработки, предшествующий принятию каких-либо лексических решений.

Чтобы понять, какие акустические события в речевом сигнале следует при транскрибировании описывать именно символом [j], стоит прежде всего проанализировать процессы, происходящие в речевом тракте говорящего при произнесении данного согласного.

«Из всех среднеязычных в русском языке имеется только один, а именно среднеязычный щелевой сонант [j]. Артикуляция его (в ударном слоге) очень похожа на артикуляцию гласного [и] (см. рис. 49 и 50): кончик языка лежит у нижних зубов, края языка упираются в боковые зубы, средняя часть спинки поднята к твердому нёбу (к его границе с мягким), задняя часть спинки и корень языка продвинуты вперед. По средней части языка проходит слабая воздушная струя, окрашенная голосом, дающая сонант. Отличается согласный [j] от гласного [и] несколько большей силой воздушной струи и большим напряжением в фокусе шумообразования» [Матусевич 1976: 156] (см. рис. 1 – А.В).

К этому нужно добавить, что щель, образующаяся в речевом тракте при произнесении [j] (рис. 1), подобна таковой у согласного [s], с учетом звонкости согласного [j] шумовая составляющая может быть сопоставима с таковой у звонкого щелевого [z] и на спектрограммах иногда может не просматриваться (рис. 2).

Следует также ожидать, что смыкание языка с твердым небом при произнесении этого согласного сопровождается характерным для согласных повышением давления воздуха (внутриротового давления) за местом смычки и, соответственно, понижением частоты и амплитуды колебаний голосовых связок (вплоть до возможного оглушения согласного). Одновременно должна уменьшаться интенсивность акустического сигнала на интервале согласного (рис. 3, 1).

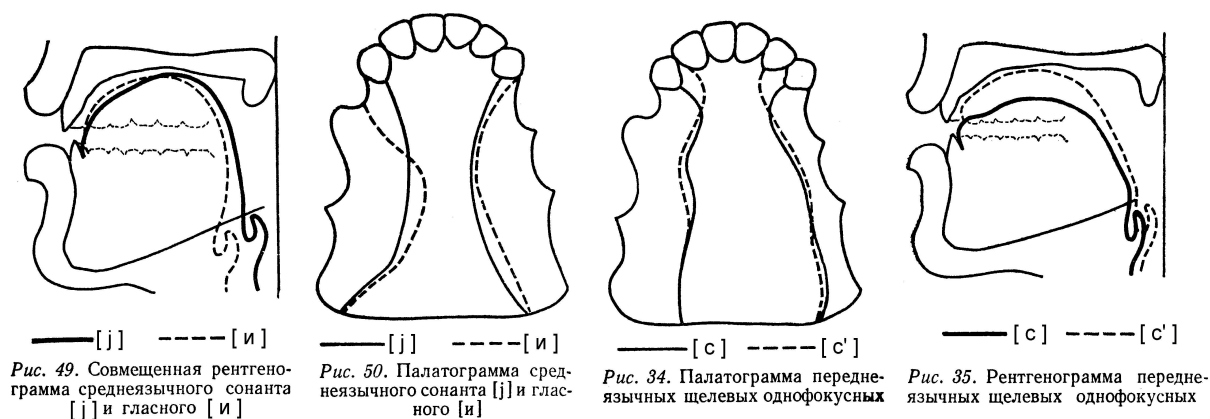


Рис. 1. Палатограммы и схематичные рентгенограммы речевого тракта при произнесении гласного [i], согласного [j] и твердого и мягкого согласного [s]. По: Матусевич 1976: 156, 136.

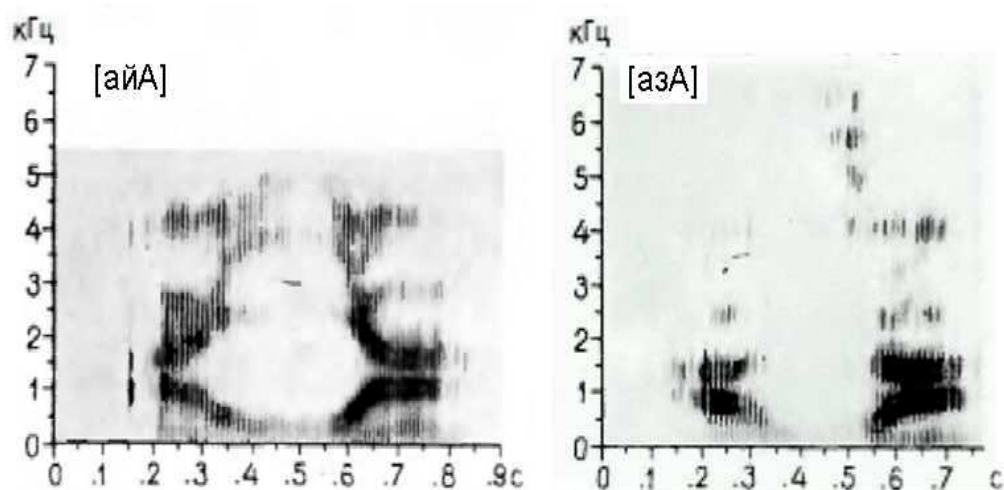


Рис. 2. Динамические спектрограммы двуслогов *ая* и *аза*. По: Деркач и др. 1983.

Таким образом, «образованный активным движением средней части тела языка к средней части твердого нёба, звук *й* является истинно палатальным звуком и его влияние на окружающие звуки неразрывно связано с палатализацией последних» [Деркач и др. 1983: 63]. В речевом сигнале это проявляется в виде так называемых *i*-образных формантных переходов в спектрах прилегающих гласных, что отчетливо проявляется на рис. 2 и 3.

Из этого следует, что собственно согласный [j] характеризуется комплексом акустических признаков: наличием *i*-образных формантных переходов на примыкающих гласных, уменьшением интенсивности шума и высоких формант (вплоть до полного их исчезновения), снижением частоты основно-

го тона (вплоть до оглушения) и общей интенсивности речевого сигнала на участке согласного. При наличии всех этих признаков соответствующий участок речевого сигнала следует маркировать символом [j].

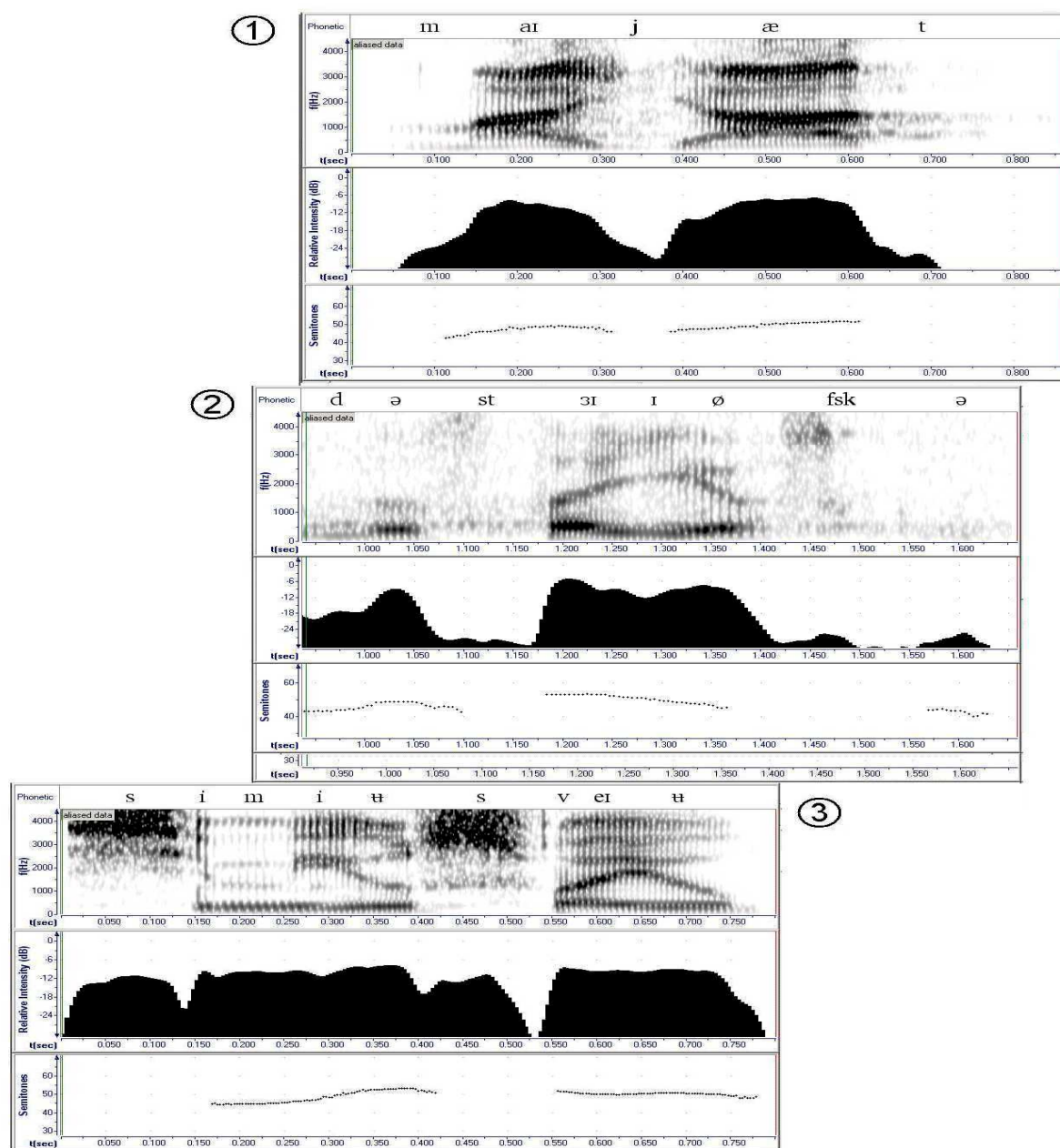


Рис. 3. Примеры акустической реализации согласного /j/ при разных степенях редукции. Сверху вниз: «маят», «Достоевского», «семью свою». Дикторы – мужчины.

Однако полностью этот комплекс акустических параметров согласного проявляется только при так называемом полном стиле произношения, когда все артикуляторные команды выполняются «до конца» и соответствующие

органы занимают теоретически предписанное им положение в речевом тракте. В обычных же условиях, а особенно при быстром темпе и в спонтанной речи, эти условия как правило не выполняются, что приводит к значительной редукации согласного /j/.

В фонетической литературе отмечается, что фонема /j/ в речевом сигнале может быть реализована как собственно согласный [j] (рис. 3,1) или неслоговой гласный (и, э, ы) (рис. 3,2) [Аванесов 1984; Касаткин 2003; Касаткин, Касаткина 2004; Касаткин 2009; Матусевич 1976; Панов 1979]. Наконец, в тех же работах указывается, что в некоторых положениях эта фонема «реализуется» нулем звука.

Поскольку фонетика, по определению, занимается изучением и описанием акустической картины речи, подобное наблюдение можно считать исчерпывающим. Правда, стоило бы уточнить, что понимается под «нулем звука»: возникает ли эта ситуация, когда в акустическом речевом сигнале (на осциллограмме или на динамической спектрограмме) не удастся (или удастся лишь с большими оговорками) идентифицировать участок, который можно было бы принять за согласный (рис. 3,3), или это значит, что согласный не может быть услышан, т.е. в сигнале полностью отсутствуют признаки, необходимые для его идентификации.

При этом надо учитывать, что при «акустическом» аннотировании звучащего на родном языке текста, фонетист-исследователь всегда знает позицию в сигнале, где должен был бы реализоваться данный согласный, т.е. сегментная структура известна и надо только описать ее акустическое выражение. Отсюда, очевидно, исходит возможность оперировать термином «неслоговой» применительно к гласным, образующимся на месте редуцированного согласного.

Например, акустическая реализация слов *поиска* и *войска* в определенных условиях может быть описана фонетистами как [pɔiskə] и [vɔiskə]. К сожалению, спектральные характеристики не дают возможности различить слоговой и неслоговой гласный «и». Как следствие, акустическая аннотация

этих слов не будет содержать метки «неслоговой» и система восприятия речи вынуждена будет интерпретировать эту пару слов как *поиска-войска* или *поиска-воиска*. Т.е. в очередной раз дело сводится к принятию решения уже на уровне лексического поиска в условиях, когда в результате редукции сегментная информация оказывается неполной или искаженной.

Возможно, идентифицировать «неслоговой» гласный в качестве «заместителя» согласного /j/ проще в комбинации «гласный+неслоговой гласный+гласный», поскольку в русском языке не встречаются последовательности из трех гласных подряд (они возможны, но маловероятны лишь на стыках между словоформами: см. слово *радиоинтервью* ниже, в подписи к рис. 5).

Если термин «нуль звука» понимать буквально, т.е. как невозможность услышать согласный /j/, то при «выпадении» согласного в процессе редукции подобная ситуация возникает только в трех комбинациях (в орфографии): «согласный + ь+я/е/ё/ю/и»(рис.4,1), «согласный + ъ+я/е/ё/ю»(рис.4,2) и «гласный + й + мягкий согласный» (рис.5).

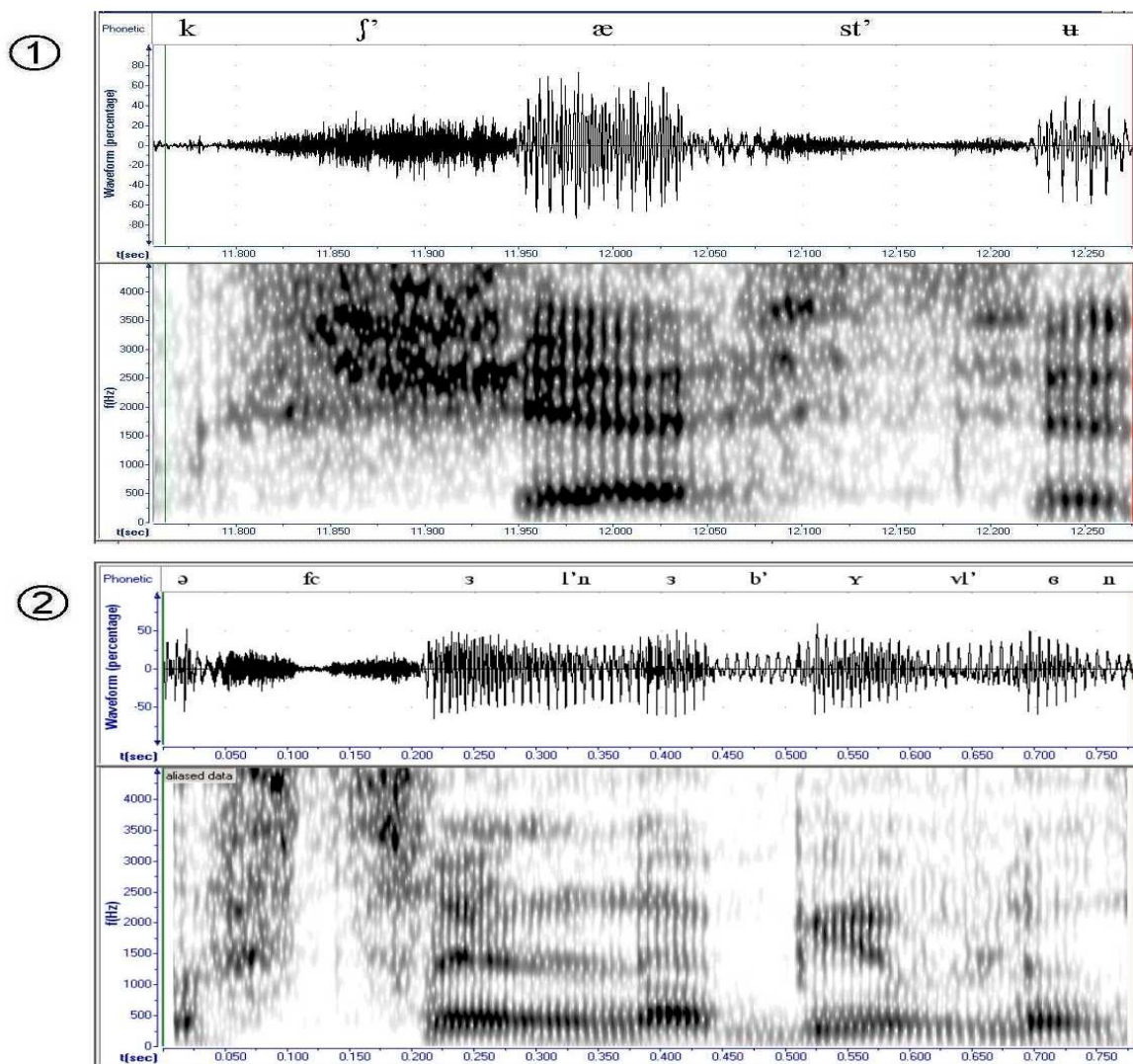


Рис. 4. Осциллограммы и динамические спектрограммы, иллюстрирующие ситуацию «нуля звука» в последовательности «согласный+/j/+гласный».

1: фонетическое слово «к\_счастью», диктор мужчина (файл academ\_zaliz\_slovo\_021.wav из мультимедийного подкорпуса Национального корпуса русского языка); 2: словосочетание «официально\_объявлен», диктор мужчина (файл academ\_zorin\_erop\_dvor\_170.wav оттуда же).



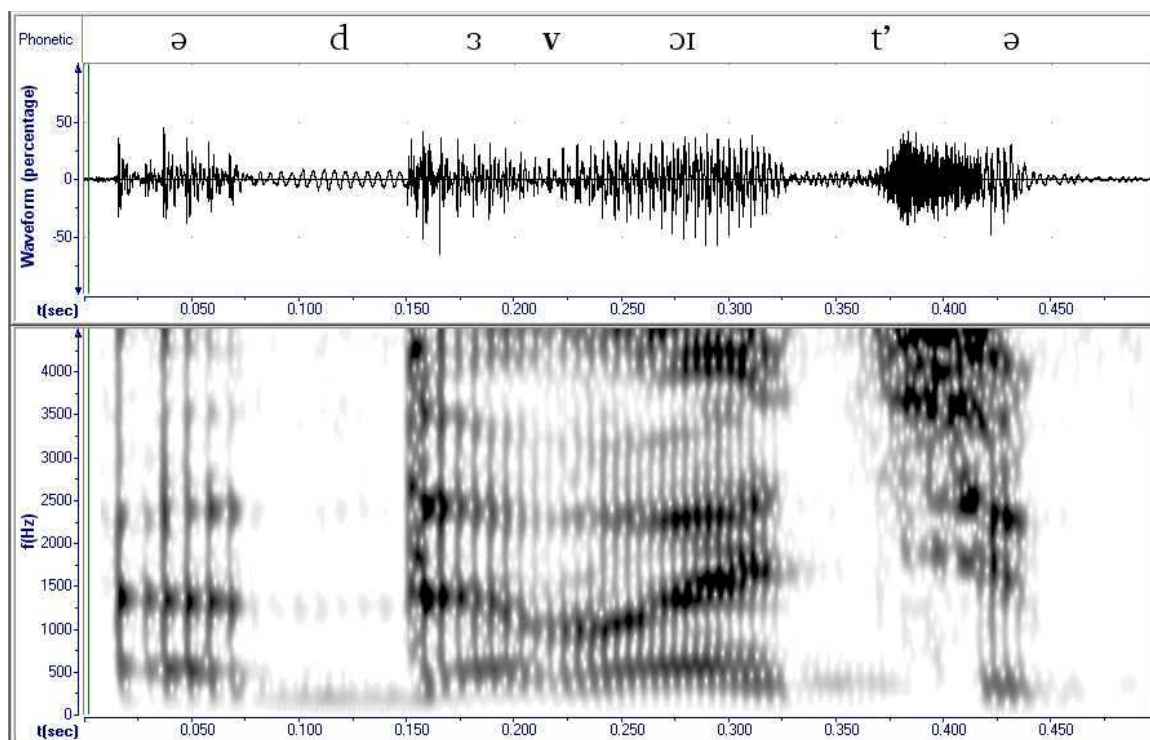


Рис. 5. Осциллограмма и динамическая спектрограмма, иллюстрирующие ситуацию «ноля звука» в последовательности «гласный+/j/+мягкий согласный».

Фонетическое слово «а давайте» из фразы «а давайте тогда начнем нашу беседу» (диктор – мужчина, радиointервью).

Как следует из приведенных рисунков, при выпадении согласного /j/ акустические характеристики сигнала в точности воспроизводят типичные изменения формантной картины гласных в соседстве с палатализованными согласными. В результате, опираясь только на акустический сигнал, слушатель не в состоянии «услышать» выпавший согласный, его когнитивная система должна была бы идентифицировать представленные на рисунках сигналы, как «к счастью, объявлен, давате», и лишь знание языка позволяет слушателю восстановить утраченное.

Совсем иную акустическую картину можно наблюдать в комбинациях «гласный+/j/+гласный»(рис. 3,3), «гласный+/j/+твердый согласный»(рис. 6) и

«гласный+/j/+пауза», в которых при выпадении согласного /j/ фонетисты фиксируют «нуль звука» [Касаткин 2009].

При выпадении согласного /j/ в последовательности «гласный+/j/+твердый согласный» (в орфографии) спектральные характеристики гласного, предшествующего твердому согласному, соответствуют параметрам гласного перед палатализованным согласным (рис. 6). Этого оказывается достаточно, чтобы когнитивная система «восстановила» редуцированный согласный, а носитель языка «услышал» его.

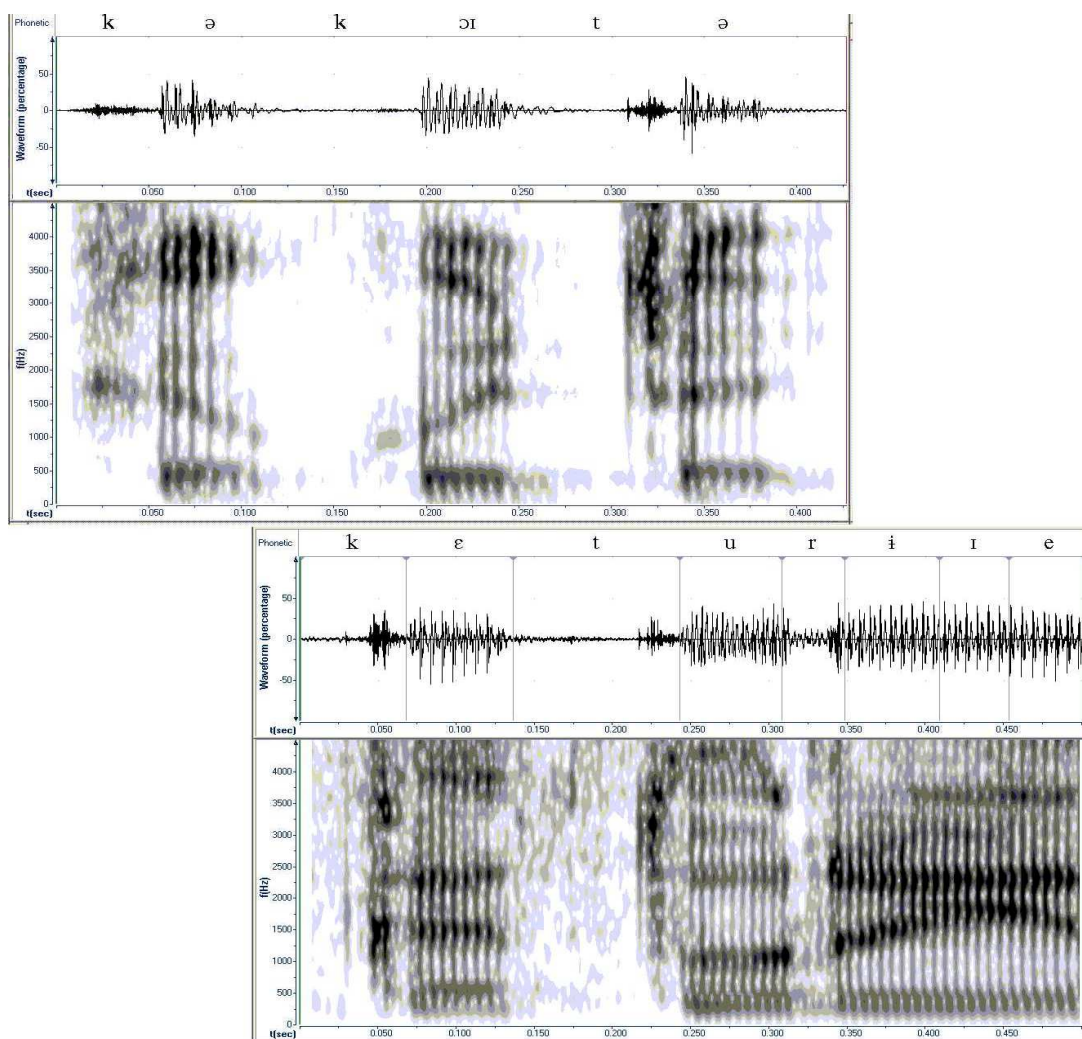


Рис. 6. Осциллограммы и спектрограммы словоформ «какой-то» (вверху) и «которого» (внизу): сравнение формантной структуры гласных перед твердым согласным.

Способность носителей русского языка слышать в подобных (конфликтных) ситуациях согласный /j/ убедительно продемонстрировали результаты экспериментов с короткими асемантическими стимулами, проведенные М. Блохиной [Блохина 2013].

Один из стимулов в ее эксперименте представлял собой выделенные из слова «Достоевского» (рис. 3,2) первый предударный и ударный слоги с удаленным неслоговым гласным (стимул 5, рис. 7). Другой – двуслог «ава» из произнесенного изолированно слова «говорит» (стимул 3, рис. 8).

В опытах участвовало 40 аудиторов, и их результаты свидетельствуют (рис. 7 и 8), что испытуемые воспринимали в обоих стимулах согласный /j/, а поскольку опора на лексику была невозможна, следует признать, что они действительно слышали этот согласный.

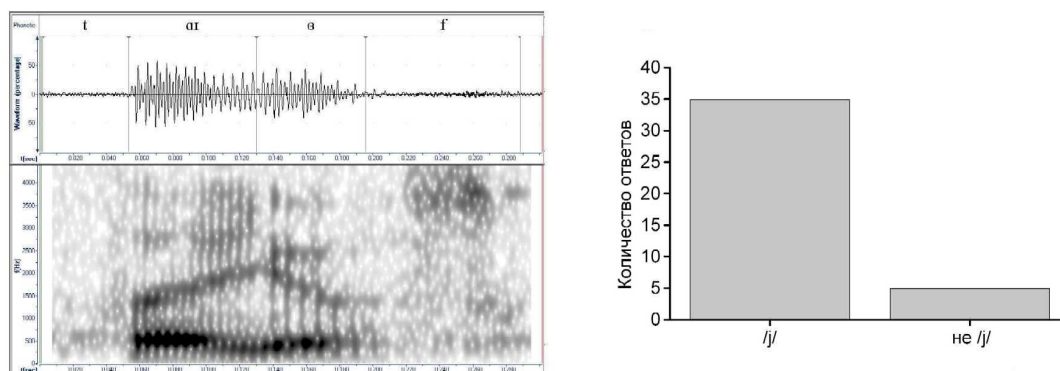


Рис. 7. Акустические параметры стимула 5 (слева) и результаты оценки его испытуемыми (справа).

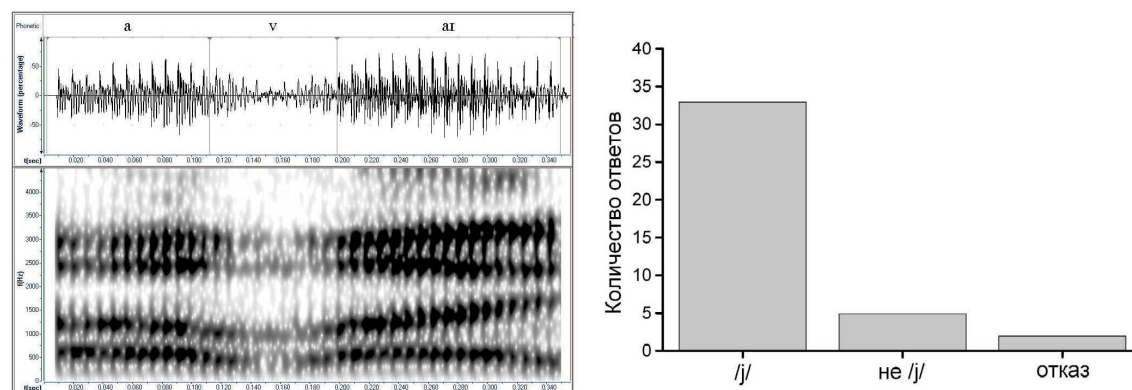


Рис. 8. Акустические параметры стимула 3 (слева) и результаты оценки его испытуемыми (справа).

Таким образом, оказывается, что наличие характерных *i*-образных формантных переходов в спектре предшествующих (последующих) гласных может обеспечить восприятие согласного /j/ даже в тех случаях, когда остальные признаки этого согласного утрачены в результате редукции.

Соответственно, при акустическом аннотировании речевых корпусов с целью последующего использования результатов такой разметки в качестве «входного сигнала» для следующих модулей функциональной модели восприятия речи наличие таких формантных переходов необходимо отражать в символах транскрипции. Иначе возникает риск, что дальнейшая обработка будет приводить к ошибочным результатам. Именно это обстоятельство учтено в транскрипции звучащих текстов, результаты которой доступны на сайте [www.russpeech.spbu.ru](http://www.russpeech.spbu.ru).

Приведенные выше данные позволяют уточнить значение термина «реализован» в применении к особенностям «звучания» согласного /j/ в естественной речи, отмечаемых в работах фонетистов.

Это определение может быть понято двояко: либо как констатация особенностей проявления данного согласного в результирующем акустическом сигнале, либо как указание на структуру артикуляторных команд, обеспечивающих порождение речевого сигнала.

Результаты анализа спектрально-временной структуры акустического речевого сигнала позволяют утверждать, что основные артикуляторные команды, характеризующие согласный /j/, выполняются даже в условиях сильной редукции, свойственной спонтанной речи.

#### Список литературы

1. Аванесов Р.И. Русское литературное произношение / 6-е изд. перераб. и доп.— М.: Просвещение, 1984.— 384 с.
2. Блохина М.Э. Особенности реализации согласного /j/ в русской спонтанной речи.— Выпускная квалификационная работа.— СПбГУ, 2013.

3. Венцов А.В., Нигматулина Ю.О., Раева О.В., Риехакайнен Е.И., Слепокурова Н.А. Корпус русских спонтанных текстов: структура и единицы // Труды международной конференции «Корпусная лингвистика–2013».– СПб.: С.-Петербургский гос. университет, Филологический факультет, 2013.– С. 223–231.
4. Деркач М.Ф., Гумецкий Р.Я., Гура Б.М., Чабан М.Е. Динамические спектры речевых сигналов.– Львов: Вища школа, 1983.– 168 с.
5. Касаткин Л.Л. Фонетика современного русского литературного языка.– М.: Изд-во МГУ, 2003.– 224 с.
6. Касаткин Л.Л. Орфоэпические правила реализации фонемы /j/ в современном русском литературном языке // Russian Linguistics.– 2009.– Vol. 33.– P. 177–194
7. Касаткин Л.Л., Касаткина Р.Ф. Особенности реализации фонемы /j/ в позиции после согласного перед гласным в русском литературном языке // Russian Linguistics.– 2004.– Vol. 28.– P. 227–236.
8. Матусевич М.И. Современный русский язык. Фонетика.– М.: Просвещение, 1976.– 288 с.
9. Панов М.В. Современный русский язык. Фонетика.– М.: Высшая школа, 1979.– 256 с.
10. Физиология речи. Восприятие речи человеком / Л.А. Чистович [и др.].– Л.: Наука, 1976.– 388 с.
11. Чистович Л.А., Венцов А.В., Люблинская В.В., Столярова Э.И., Чистович И.А. Слуховые уровни восприятия речи. Функциональное моделирование // Акустика речи и слуха: Сборник научных работ / Ред. Л.А. Чистович.– Л.: Наука, 1986.– С. 97-127.