

Юлия Олеговна Нигматулина

Ольга Васильевна Раева

Россия, Санкт-Петербургский государственный университет

Исследование русской спонтанной речи: новые методы – новые результаты¹

Ключевые слова: восприятие речи, методы исследования спонтанной речи, редуцированные формы слов, стяжение звуков

Key words: spoken word recognition, methods of spontaneous speech research, reduced word forms, sound contraction

Abstract

The study of spontaneous speech requires that it be transcribed and analyzed objectively. The method of continuous transcription minimizes the influence of lexical and grammatical information in the course of data processing. Asemantic (monosyllabic) speech fragments are analyzed aurally with simultaneous observation of oscillograms and spectrograms. New data concerning reduction and sound contraction in spontaneous Russian have been obtained. The results of the research allow formulating hypotheses concerning the recognition of spontaneous speech that should be verified in perceptual experiments.

Исследование возможных алгоритмов восприятия речи, осуществляемое сотрудниками Лаборатории моделирования речевой деятельности Санкт-Петербургского государственного университета, невозможно без изучения спонтанной речи, т. е. той акустической реальности, с которой носитель языка ежедневно сталкивается в процессе речевой коммуникации. В связи с этим требуется обращение к объективным методам анализа речевого материала.

Спонтанная речь «может характеризоваться наличием незаполненных и заполненных пауз хезитации, повторов, перебивов, недосказанных слов, ошибок и увеличением темпа речи» [Риехакайнен 2010: 87], о чем говорится в таких работах, как [Земская 2006 (1979); Бондарко и др. 1988 и др.]. Изменения, затрагивающие звуковой облик словоформ в русской речи, описаны, например,

¹ В статье представлены результаты исследований, поддержанных грантами Президента Российской Федерации № МК-3646.2013.6 и № НШ-1778.2014.6.

в [*Русская разговорная речь* 1973; Бондарко и др. 1988 и др.]. Тем не менее следует отметить, что сведения о качественной и количественной редукции единиц речи, а также о других характеристиках речевого материала были получены в результате слухового анализа, т. е. прослушивания и определения звукового состава речевого фрагмента методом экспертной оценки на слух. Некоторые данные получены на материале, фонетический анализ которого проводился выборочно [напр., Шерстинова и др. 2009 и проч.].

Существует также ряд исследований на материале других языков (например, голландского), описание спонтанной речи в которых осуществлялось с учетом результатов автоматического транскрибирования больших речевых фрагментов (процедура получения фонетической транскрипции для Корпуса голландской устной речи представлена, например, на сайте Корпуса в разделе *Broad Phonetic Transcription [Corpus Gesproken Nederlands 2004]*).

Однако субъективная оценка эксперта в процессе слухового анализа, а также несовершенство систем автоматического транскрибирования, результаты работы которых требуют последующей проверки и ручной правки, не позволяют сформировать исчерпывающее представление об акустических особенностях речевого материала.

Для получения более полной информации об акустической реализации естественной речи был проведен анализ, включающий в себя несколько этапов.

Прежде всего, была осуществлена орфографическая расшифровка записей с целью получить данные о семантике и синтаксической структуре исследуемых текстов.

С нашей точки зрения, важным представляется выявление акустических особенностей речевого сигнала, обрабатываемого начальными уровнями слуховой системы, что позволяет определить, как реализуются те или иные речевые единицы в потоке речи. Фиксация таких особенностей возможна в акустико-фонетической транскрипции.

Метод, используемый нами для получения акустико-фонетической транскрипции, заключается в последовательном прослушивании небольших (один слог) асемантичных фрагментов записей спонтанных текстов с одновременным анализом соответствующих им динамических спектрограмм. Такой метод позволяет принимать решение о качестве звука, в максимальной степени учитывающее акустические параметры сигнала, а также исключать, насколько это возможно, влияние лексико-грамматических знаний эксперта и таким образом фиксировать то, что на самом деле произносится в спонтанной речи. Использование динамических спектрограмм помогает в той или иной мере избегать субъективной оценки эксперта. При обработке нашего материала каждый фрагмент транскрибировался одним из 6 экспертов-фонетистов, затем затранскрибированные сегменты выборочно проверялись другими экспертами.

Так, например, последовательный слуховой анализ каждого слога речевого фрагмента, соответствующего словоформе *когда*, с опорой на спектрограмму

(рис. 1) позволяет получить следующую акустико-фонетическую транскрипцию: [kada].

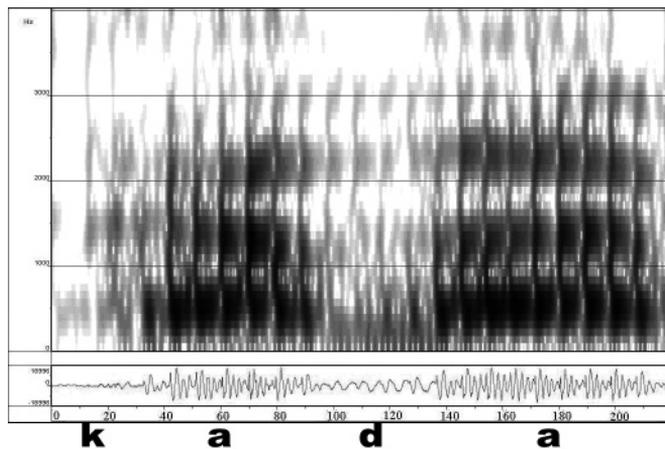


Рис. 1. Спектрограмма для словоформы *когда* из речевого фрагмента *когда меня пригласили*

В связи с предполагаемой в дальнейшем компьютерной обработкой результатов подобной разметки текстов, для их описания не используются символы, требующие специальных наборов шрифтов. Вместо этого принята система, аналогичная набору SAMPA, особенности которой представлены в таблице 1.

Также, при разметке использовались маркеры для глухих пауз (pause), вдохов (inh), для неразборчиво произнесенных фрагментов (неразб), для одновременной речи нескольких дикторов и пр. Ударение обозначалось знаком «+» после соответствующего гласного.

Результаты транскрибирования речевого материала представленным методом доступны путем создания конкорданса по текстам, размещенным на сайте Корпуса русского литературного языка (<http://www.nagusco.ru/>), в разделе *Наши ресурсы*.

Все проанализированные записи разбиты на межпаузальные интервалы и снабжены орфографической расшифровкой. Сплошная акустико-фонетическая транскрипция на данный момент осуществлена для фрагментов записей спонтанной речи, общей продолжительностью звучания около 90 минут (10 488 словоупотреблений).

Кроме того, по записям корпуса с уже имеющейся фонетической транскрипцией был создан орфографически-транскрипционный словарь произносительных вариантов русской речи [Венцов и др. 2013]. Суть такого словаря состоит в том, что в процессе поиска словоформы в орфографической записи пользователь получает все варианты ее акустической реализации в пределах описанного материала в виде фонетической транскрипции с учетом частоты встречаемости (см. таблицу 2).

Гласные	Символы (латиница)		
	Между твердыми	После или между мягкого(ими)	Перед мягким
а	а	а:	aI
о	о	о:	oI
э	е	е:	eI
и	і	і	i
ы	Q	—	Q
у	u	u:	uI
шва	@	@	@I
Согласные	Символы (латиница)		
	Глухие	Звонкие и озвонченные	
ж	—	Z	
ц	C	D	
х	X	H	
ч	C	G	
ш	S	—	
щ	\$	S':	

Табл. 1. Символы, используемые для описания результатов акустико-фонетической разметки спонтанных текстов

В качестве описываемых единиц словаря были взяты графические слова, то есть такие фрагменты речи, которые на письме разделены пробелами. В процессе транскрибирования оказалось, что между словами иногда наблюдается стирание фонетической границы и разделить такое сочетание слов в транскрипции невозможно. Тогда такие сочетания признаются цельной самостоятельной единицей, и для их обозначения в орфографической расшивке используется знак «_». Прежде всего, это относится к так называемым составным словам [см. Венцов и др. 2004]. Свойства таких слов на грамматическом, семантическом, а также фонетическом уровнях полностью соответствуют свойствам неделимого слова. Для таких слов, как правило, наблюдается слитность произнесения: например, *то_есть* [teIs'], [tQs'] и др.

Кроме составных слов, «стирание» границ между словами может затрагивать и любые другие словоформы, если на их стыке оказываются два гласных или два согласных звука, реализующихся как один. Например,

Словоформа	Вариант реализации	Частота реализации	Словоформа	Вариант реализации	Частота реализации
людей	[l'u:d'i#+]	5	будет	[buI+d'it]	7
	[l'u:d'i+]	1		[buId'i+tx]	1
	[l'@Id'i#+]	1		[bu+Qt]	1
	[v'id'e:#+]	1		[buI+d'i(t)]	1
	[l'u:d'i+j]	1		[buI+d'i]	1
	[l'id'e:]	1		[bu+@t]	1
	[l'u:d'e:i]	1		[buI+d]	1
	[lQd'i+]	1		[buI+it]	1

Табл. 2. Пример результата поиска по словарю произносительных вариантов для словоформ *людей* и *будет*

должны опираться [doZne+p'eras], *дома или* [do+m@l'] и т. д. Поскольку проведение границы между такими сочетаниями в фонетической расшифровке не представляется возможным, в корпусе эти словоформы были также соединены в орфографической записи с помощью знака «_» и вошли в словарь произносительных вариантов как самостоятельные единицы.

Рассмотренный метод анализа позволяет получить описание звукового состава речи, максимально приближенное к тому, что было произнесено в реальности. Результаты обработки речевого материала, во-первых, подтверждают имеющиеся сведения об особенностях спонтанной речи, во-вторых, позволяют дополнить и расширить данные других исследований, а в-третьих, могут быть использованы в качестве входного «сигнала» в предполагаемой функциональной модели восприятия речи.

Так, существующие описания устной разговорной речи свидетельствуют о том, что «звуки могут менять свои качества, ослабляться или совсем исчезать» [*Русская разговорная речь* 1973: 40]. Однако кроме изменений, допускаемых нормами литературного языка, в предшествующих исследованиях было выявлено, например, качественное изменение ударных гласных в неподготовленной речи [Бондарко и др. 1988].

Полученные нами данные об акустических характеристиках спонтанной речи подтверждают данный факт. Так, ударные гласные [e] и [o], которые являются более закрытыми, могут заменяться на ударные [i] и [u] соответственно (например, [d'il'i+] из словосочетания *такое определение* [tako+e ap'r'id'il'i+n'i]). Более того, были отмечены случаи полного выпадения ударных слогов (например, [n'imo+ZQt @tkrt@gra+f'iu:] *не может открыть фотографию*) [Риехакайнен 2010; Венцов и др. 2011].

Результаты количественного анализа указывают на то, что в спонтанной речи «чаще всего встречается выпадение одного звука (57,6%)», полная количественная редукция затрагивает два звука в 31,5% случаев, выпадение трех и более элементов характерно для 10,9% словоформ [Риехакайнен 2010: 93]. Эллиптированными в основном оказываются компоненты заударной части слова [Риехакайнен 2010: 93]. Сопоставление количества редуцированных элементов заударной и преударной частей слова показало, что выпадение компонентов заударной части настолько масштабно, что дальнейшая классификация внутри данного явления представляется трудноосуществимой [Горлова, Слепокурова 2012].

Выпадение звуков в преударной части словоформы характерно только для 7,4% случаев (332 словоформы) [Горлова, Слепокурова 2012: 15]. Так, эллипсис гласных наблюдается в 39% случаев (129 словоформ; например, *сегодня* [s'voI+n'i]), полная количественная редукция одновременно нескольких звуков характерна для 20% случаев (66 словоформ; например, *говорит* [geIgr'i+t]). Выпадение согласных является наиболее частотным и отмечается в 41% случаев (137 словоформ); в 83% случаев согласный выпадает из группы согласных (например, *систему* [s'is'e:+mo]) [Горлова, Слепокурова 2012: 18]. Также отмечается, что «чаще всего [...] утрачиваются смычные согласные [t], [d] и шелевые [f], [v], сонорные согласные выпадают реже, чем шумные. Кроме того, выявлено, что в интервокальной позиции чаще выпадают мягкие согласные» [Горлова, Слепокурова 2012: 18–19].

Работы, посвященные фонетическим изменениям в звучащей речи, в основном описывают явления, происходящие в рамках словоформы или сочетания словоформы с предлогом. Однако нередко в слитной речи модификация возникает на месте стыка двух слов, выходя таким образом за пределы словоформы. Как говорилось выше, в ходе транскрибирования записей с применением описанного метода наблюдалось стирание границ между словами, если на стыке двух любых слов оказываются два гласных или два согласных звука (например, *как-то_использовать* [ka+ktQspo+l'z@v@t']). Слияние двух смежных звуков, в результате которого возникает один звук на месте двух прежних, в фонетике принято называть стяжением [Розенталь 2008 и др.]. Результаты проведенного эксперимента по восприятию гласных звуков в месте стяжения на стыке словоформ показали, что большинство «стяженных» гласных в составе бессмысленных фрагментов естественного речевого сигнала распознаются испытуемыми как один звук [Риехакайнен и др. 2011].

Для того чтобы оценить масштаб этого явления в звучащей речи, на нашем материале было проведено сравнение орфографической записи всех мест потенциальных стяжений на стыке словоформ с транскрипцией данных фрагментов. В качестве таких случаев были рассмотрены все пограничные соседства гласных, а также сочетания согласных, одинаковых или различающихся по признаку глухости–звонкости и/или мягкости–твердости. Составные

слова при этом не учитывались. Общий объем записей, использованный для поиска стяжений, составил 161 минуту: 115 минут звучания спонтанной речи, а также – для сопоставления данного явления в разных типах речи – 46 минут звучания прочитанных текстов.

Решение о наличии или отсутствии стяжения на месте звукового сочетания принималось с учетом данных инструментального анализа (см. рис. 2). Стяжением признавались те случаи, когда на месте стыка словоформ на спектрограмме и осциллограмме ничто не указывало на присутствие двух звуков, а именно: 1) не происходило заметных изменений или прерываний; 2) для гласных не наблюдалось существенных изменений формант, которые свидетельствовали бы о переходе от одного гласного к другому.

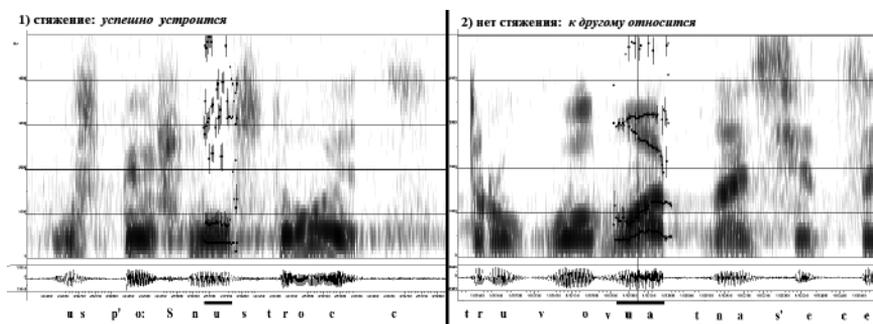


Рис. 2. Стяжение гласных на стыке в сочетании *успешно устроится* и отсутствие стяжения в сочетании *к другому относится*

Относительно общего количества слов, содержащихся в записях, была подсчитана регулярность появления стяжений и регулярность появления сочетаний, которые могли бы привести к стяжению. Значения вычислялись по формулам:

$$R_{\text{стяжений}} = (1000 * N_{\text{стяжений}}) / N_{\text{слов}},$$

$$R_{\text{сочетаний}} = (1000 * N_{\text{сочетаний}}) / N_{\text{слов}},$$

где R – регулярность; N – общее количество (сочетаний; стяжений).

Количество стяжений на стыке словоформ, а также количество всех сочетаний, которые могли бы привести к образованию стяжения, в расчете на 1000 слов представлены в таблицах 3 и 4.

Сочетания согласных на стыке слов возникают реже сочетаний гласных и при этом гораздо чаще реализуются в виде стяжения.

Как показали полученные результаты, в русском языке на стыке слов регулярно появляются сочетания звуков, которые могли бы привести к возникновению стяжения, явлению, свойственному русской звучащей речи – как спонтанной, т. е. неподготовленной, так и прочитанной, т. е. наиболее подготовленной.

Определенной закономерности появления стяжения звуков выявить не удалось. При возникновении комбинации гласных на стыке словоформ язык

	Спонтанная	Прочитанная
Гласные	20,5	56,7
Одинаковые согласные	7,0	11,4
t/t'/d/d'_s/s'/z/z'	4,1	9,0

Табл. 3. Частота появления стяжений: количество стяжений на 1000 слов

	Спонтанная	Прочитанная
Гласные	37,0	99,9
Одинаковые согласные	9,5	11,7
t/t'/d/d'_s/s'/z/z'	6,2	10,2

Табл. 4. Частота появления сочетаний, которые могли бы привести к стяжению: количество сочетаний на 1000 слов

в равной степени допускает два варианта звуковой реализации: зияние, т. е. сохранение нескольких звуков, и стяжение, т. е. сокращение количества звуков. Однако были отмечены условия, которые в некоторой степени располагают к появлению стяжений: 1) сочетание одинаковых или близких по качеству звуков; 2) отсутствие ударения на гласных сочетания. Стяжения появляются чаще на стыке недлинных (но длиннее одного звука), частотных слов, среди которых оказываются местоимения, предлоги, союзы, частицы и т. д.

Кроме стяжений звуков можно наблюдать и более глубокое взаимное наложение (или слияние) словоформ, когда две лексические единицы делят не только один общий звук, но, например, слог или группу согласных звуков: *можно_найти* [mo+ZnaIt'i+] (см. рис. 3), *из уст_взрослых* [Qz u+zro+slQx].

Данное явление представляет большой интерес с точки зрения описания механизмов обработки речевого сигнала, поскольку размывает границы между словами и, соответственно, затрудняет сегментацию в процедурах автоматического распознавания. При этом реализация стяжения, как правило, не зависит от тесноты связи между словоформами, на стыке которых оно происходит. Стяжение может возникать как между синтаксически тесно связанными словами (например, *не_обсуждаю* [n'e:+psuZdajo:]), так и между словами, принадлежащими разным предложениям (например, между простыми предложениями в составе сложного: *что_есть_семьи_от_которых_хочется_сбежать_куда_угодно, даже_в_школу* [StQs':i+m'a:t kado+rox#oI+Cc@_zb'e:ZeI+t' kuda ugo+dn@_da+ZfSko+lu]). Искажение сигнала, приводящее к качественной и количественной редукции и затрагивающее одновременно два слова, при инструментальном анализе затрудняет принятие однозначного решения о месте границы между словоформами, а в некоторых случаях и между предложениями.

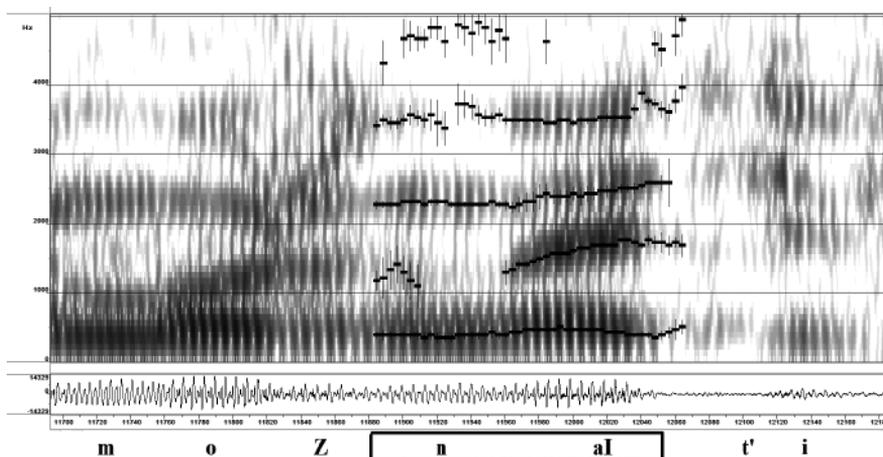


Рис. 3. Спектрограмма сочетания *можно_найти* со взаимным наложением слогов

Заключение

В статье представлен метод сплошного анализа спонтанной речи, который 1) применяется для исследования свободно порождаемой непрочитанной речи; 2) представляет собой сплошной, а не выборочный анализ речевого материала; 3) совмещает в себе экспертную оценку и инструментальный анализ; 4) сопровождается использованием программ с хорошим качеством спектрограмм и возможностью изменения параметров настройки (таких, как Speech Analyzer, WinSnoori и др.).

Результаты анализа с использованием такого метода подтверждают полученные ранее данные о вариативности речевого материала, с которым слушающий сталкивается в рамках повседневного общения. В спонтанной речи можно выделить единицы с разной степенью редукиции, но предсказать, будет ли редуцирована единица в каждом конкретном случае, не представляется возможным. Кроме фонетических изменений внутри словоформ, зафиксированы изменения, происходящие на стыке слов. Звук, образованный в результате стяжения, становится общим одновременно для двух словоформ, что делает словесную границу при сегментации речи неочевидной.

Все полученные данные ставят вопрос о возможных алгоритмах обработки речевого сигнала человеком, а также о том, как ребенок усваивает речь, единицы которой имеют различные варианты произнесения и в ряде случаев не могут быть разграничены.

Литература

- Бондарко Л., Вербицкая Л., Гейльман Н. и др., 1988, *Фонетика спонтанной речи*, Н. Светозарова (ред.), Ленинград: Издательство Ленинградского университета.
- Венцов А., Касевич В., Ягунова Е., 2004, Идиома, слово, фонетическое слово [в:] Г. Кедрова, В. Потапов (ред.), *Язык и речь: проблемы и решения: Сборник научных трудов к юбилею проф. Л. В. Златоустовой*, Москва: Макс Пресс, с. 357–363.
- Венцов А., Нигматулина Ю., Раева О., Риехакайнен Е., Слепокурова Н., 2013, Корпус русских спонтанных текстов: структура и единицы (Corpus of spontaneous Russian texts: structure and items) [в:] *Труды международной конференции «Корпусная лингвистика – 2013»*, Санкт-Петербург: Издательство СПбГУ, с. 223–230.
- Венцов А., Слепокурова Н., Риехакайнен Е., Апушкина И., Корешкова Е., 2011, Из опыта работы с русской спонтанной речью: создание фонетически транскрибированных текстов [в:] Р. Смулаковская (отв. ред.), *X выездная школа-семинар «Проблемы порождения и восприятия речи»: Материалы*, Череповец: ГОУ ВПО «Череповецкий государственный университет», с. 169–179.
- Горлова А., Слепокурова Н., 2012, Редукция предупредных компонентов словоформ в спонтанной речи [в:] *Анализ разговорной русской речи (АРЗ–2012): Труды шестого междисциплинарного семинара*, Санкт-Петербург: ГУАП, с. 15–20.
- Земская Е., 2006 (1979), *Русская разговорная речь: Лингвистический анализ и проблемы обучения*, Москва: Флинта – Наука.
- Корпус русского литературного языка*, 2009, <http://www.narusco.ru/> (дата обращения: 10.10.2014).
- Риехакайнен Е., 2010, *Взаимодействие контекстной предсказуемости и частотности в процессе восприятия спонтанной речи (на материале русского языка): Диссертация на соискание ученой степени кандидата филологических наук*, Санкт-Петербург: СПбГУ.
- Риехакайнен Е., Нигматулина Ю., 2011, Сегментация спонтанной речи: восприятие стяжений гласных на стыке словоформ [в:] Т. И. Ерофеева (отв. ред.), *Проблемы социо- и психолингвистики. Сборник статей*, вып. 15: *Пермская социопсихолингвистическая школа: идеи трех поколений: к 70-летию Аллы Соломоновны Штерн*, Пермь: Пермский университет, с. 31–38.
- Розенталь Д., 2008, *Справочник по русскому языку. Словарь лингвистических терминов*, Москва: Оникс 21 век.
- Русская разговорная речь*, 1973, Е. Земская (ред.), Москва: Наука.
- Шерстинова Т. Ю., Рыко А. И., Степанова С. Б., 2009, Система аннотирования в звуковом корпусе русского языка «Один речевой день» [в:] *Формальные методы анализа речи: Материалы XXXVIII Международной филологической конференции*, Санкт-Петербург: Факультет филологии и искусств СПбГУ, с. 66–75.
- Corpus Gesproken Nederlands*, 2004, http://lands.let.ru.nl/cgn/doc_English/topics/version_1.0/annot/phonetics/info.htm#werkwijze (дата обращения: 10.10.2014).