

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК БЕЛАРУСИ
Объединённый институт проблем информатики

Б.М. Лобанов, Л.И. Цирульник

КОМПЬЮТЕРНЫЙ СИНТЕЗ И КЛОНИРОВАНИЕ РЕЧИ

**Минск
«Белорусская наука»
2008**

ОГЛАВЛЕНИЕ

Вместо предисловия. От голоса робота - к персональному клону голоса человека	7
Введение	18
Глава 1. Фонетико-акустические основы синтеза речи по тексту	22
1.1. Общие сведения о языке и речи	22
1.2. Система фонем русского языка	25
1.3. Оттенки фонем и система аллофонов русской речи.....	29
1.4. Особенности преобразования «буква-фонема» (ПБФ) для русского языка	34
1.5. Просодические характеристики русской речи.....	37
1.6. Акустические характеристики речи.....	43
1.7. Информационная и модуляционная структура речевого сигнала	48
Глава 2. Методы синтеза фонетико-акустических характеристик речи	54
2.1. Методы персонализированного синтеза речевого сигнала.....	54
2.1.1. Основные характеристики артикуляторного метода синтеза речи по тексту	55
2.1.2. Основные характеристики формантного метода синтеза речи по тексту	56
2.1.3. Основные характеристики компиляционного метода синтеза речи по тексту	58
2.1.4. Основные характеристики корпусного метода синтеза речи по тексту	62
2.1.5. Выбор метода синтеза речи по тексту для передачи индивидуальности голоса и манеры чтения.....	63
2.2. Методы моделирования просодических характеристик речи ..	63
2.2.1. Автосегментная просодическая модель	64
2.2.2. Просодическая модель IPO.....	65
2.2.3. Суперпозиционная просодическая модель	65
2.2.4. Непрерывная параметрическая просодическая модель.....	66
2.2.5. Просодическая модель портретов акцентных единиц.....	67
2.3. Методы выбора базовых речевых единиц для синтеза речи	69
2.4. Методы формирования речевых и текстовых корпусов для создания индивидуализированных речевых баз данных.....	73

2.5. Методы сегментации и маркировки естественного речевого сигнала	77
Глава 3. Экспериментальные исследования фонетико-акустических и просодических характеристик речи	80
3.1. Исследование внутрисловных и межсловных фонетико-акустических явлений в русской речи	81
3.1.1. Методика эксперимента	81
3.1.2. Гласные в сочетаниях «гласная-гласная»	82
3.1.3 Гласные в сочетаниях «согласная – гласная», «гласная - согласная»	84
3.1.4 Акустические характеристики согласных.....	86
3.2. Исследование фонетико-акустических особенностей сегментации последовательности слов на слоги.....	87
3.3. Исследование межъязыковых и междикторских различий просодических характеристик речи	92
3.3.1. Методика создания мелодических портретов акцентных единиц.....	92
3.3.2. Результаты исследования межъязыковых различий.....	98
3.3.3. Результаты исследования междикторских различий.....	100
3.4. Исследование персональных особенностей синтагматического членения речи	104
3.4.1. Методика эксперимента	104
3.4.2. Статистические характеристики синтагматического членения	106
3.4.4. Обсуждение результатов эксперимента	109
Глава 4 Компьютерная модель мультиволнового синтеза речи по тексту.....	111
4.1. Структура синтезатора речи по тексту	111
4.2. Текстовый процессор	112
4.2.1. Блок предварительной обработки.....	113
4.2.2. Блок морфологической и акцентной маркировки слов	120
4.3. Просодический процессор	125
4.3.1. Блок членения текста на предложения, пунктуационные и лексические синтагмы.....	127
4.3.2. Блок членения ПС и ЛС на синтаксические синтагмы и их разметка на акцентные единицы.....	129
4.3.3. Блок маркировки интонационного типа синтагм.....	140
4.4. Фонетический процессор	146
4.4.1. Блок преобразования слов – фонетических исключений.....	147

4.4.2. Блок преобразования буква-фонема	148
4.4.3. Блок преобразования фонема-аллофон	151
4.5. Акустический процессор.....	155
4.5.1. Блок синтеза акустико-просодических характеристик речевого сигнала.....	156
4.5.2. Блок синтеза акустико-фонетических характеристик речевого сигнала.....	160
4.5.3. Алгоритм просодической обработки речевого сигнала в акустическом процессоре	169
4.6. Программная реализация системы синтеза речи «Мультифон».....	180
4.6.1. Функциональная схема системы.....	180
4.6.2. Пользовательский интерфейс системы «Мультифон»	182
4.6.3. Оценка разборчивости синтезированной речи	185
Глава 5. Компьютерное клонирование индивидуальных характеристик речи.....	188
5.1. Факторы вариативности речевого сигнала.....	188
5.2. Основные этапы компьютерной технологии клонирования фонетико-акустических характеристик речи	191
5.2.1. Выбор и подготовка текстовых и речевых корпусов.....	191
5.2.2. Разметка речевого корпуса на фонетические и просодические сегменты	193
5.2.3. Автоматическая сегментация и маркировка речевого сигнала.....	195
5.2.4. Создание оптимального набора элементов компиляции.....	198
5.3. Алгоритмы автоматического создания БД фонетических элементов компиляции	200
5.3.1. Алгоритмы «отсекающего отбора» речевых сегментов.....	200
5.3.2. «Селекция» речевых сегментов.....	201
5.3.3. «Диагностика» и «коррекция» сегментов	202
5.3.4. «Размножение» элементов компиляции.....	204
5.4. Компьютерная система клонирования фонетико-акустических характеристик речи – «ФоноКлонатор».....	205
5.4.2. Пользовательский интерфейс системы «ФоноКлонатор».....	208
5.4.3. Результаты практического использования системы «ФоноКлонатор»	212
5.5. Компьютерная система клонирования просодических характеристик речи – «ИнтоКлонатор».....	215
5.5.1. Функциональная схема и алгоритмы работы системы.....	215

5.5.2. Пользовательский интерфейс системы «ИнтоКлонатор»	220
5.5.3. Результаты практического использования системы «ИнтоКлонатор»	222
5.6. Оценка правдоподобия синтезированного речевого клона	223
Приложение 1. Текстовый корпус для создания БД элементов компиляции	232
П 1.1. Текст для создания мини-набора аллофонов русской речи	232
П 1.2. Текст для создания макси-набора аллофонов, диаллофонов и аллослогов русской речи	233
П 1.3. Мини-текст для создания основного набора просодических портретов русской речи	243
Приложение 2. Таблицы генерации мини-набора аллофонов русской речи	253
П 2.1. Аллофоны гласных фонем	253
П 2.2. Аллофоны согласных фонем	255
Приложение 3. Статистика фонем, дифонем, фонослогов и их аллофонических вариантов	259
П 3.1. Фонемы, дифонемы, фонослоги	259
П 3.2. Позиционные диаллофоны и аллослоги	262
П 3.3. Позиционно-комбинаторные аллофоны, диаллофоны, аллослоги	265
Приложение 4. Спектральные портреты основных аллофонов русской речи	273
П 4.1. Основные аллофоны гласных фонем: А, Е, У, Y, I, О	273
П 4.2. Согласные фонемы	276
Приложение 5. Мелодические портреты основных интонационных типов вопросительной интонации русской речи	278
П 5.1. Мелодические портреты интонации частного и общего вопросов с различным количеством АЕ	278
П 5.2. Мелодические портреты интонации частного вопроса с вопросительным местоимением без частицы и с частицами бы, же, -нибудь, ну, а, и, да.	281
П 5.3. Мелодические портреты интонации частного вопроса с вопросительными частицами ли, разве, правда, никак, а, и.	283
Приложение 6. Списки служебных частей речи (предлогов, частиц, междометий, союзов), местоимений и слов- фонетических исключений	285

П 6.1. Предлоги.....	285
П 6.2. Частицы	287
П 6.3. Междометия	290
П 6.4. Союзы	293
П 6.5. Местоимения.....	295
П 6.6. Фонетические исключения	296
Литература.....	299
Сведения об авторах.....	316

Вместо предисловия

От голоса робота - к персональному клону голоса человека (К истории русскоговорящих машин)

“Мы впервые в истории вида начинаем широко пользоваться техническими говорящими орудиями – иначе говоря, не только изготавливать орудия (чем человек отличается от животных), но и обучать их нашему языку (чем мы начинаем отличаться от всех ранее живших людей)”.

Из книги академика РАН
Вячеслава Всеволодовича Иванова
“Лингвистика третьего тысячелетия”

Предыстория

Из всего живого только человека Создатель наградил даром речи, благодаря чему ему удалось столь значительно развить свои интеллектуальные способности и, по мнению многих философов, стать человеку человеком. Осмелимся предположить, что нечто подобное происходит на наших глазах и с компьютером, интенсивно овладевающим широким спектром речевых технологий от работы со звуковыми файлами до синтеза, распознавания и понимания речи [1]. В этой книге мы коснёмся лишь одного аспекта речевых технологий, а именно, синтеза речи, как наиболее близкого её авторам.

Первые попытки создания в России синтезаторов речи относятся к XVIII веку. Во времена правления Екатерины II Петербургская Академия Наук объявила конкурс на создание говорящей машины. Победителем конкурса стал профессор Петербургского университета Кратценштейн, который построил систему акустических резонаторов, возбуждаемых воздушным потоком и издававших гласные звуки русской речи.

Несколько позже Вольфганг фон Кемпелен разработал более сложную модель генерации связной речи (рис. 1). В ней в роли резонаторов речевого тракта выступала гибкая трубка из кожи, управляемая оператором. Имелись также отверстия для имитации носовых полостей и ручки управления свистками, создававшими фрикативные звуки.

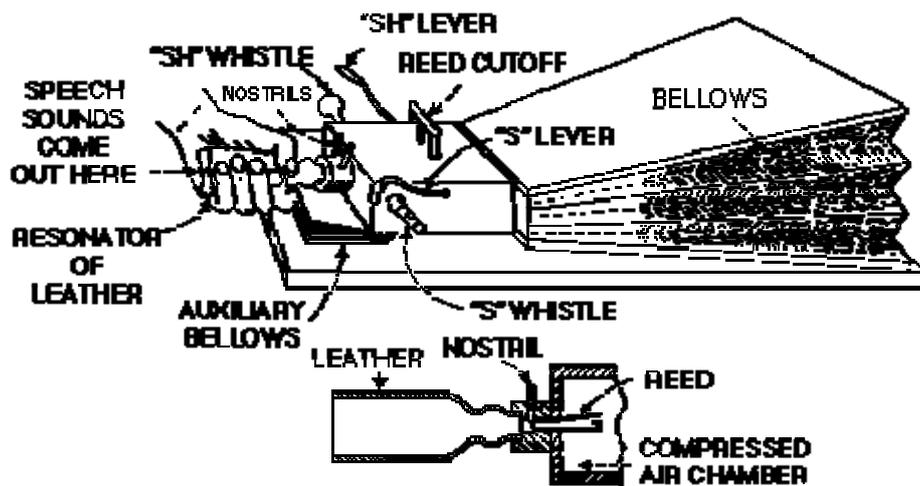


Рис.1. Синтезатор Кемпелена

Следующая заметная попытка синтеза русской речи уже относится к 30-м годам XX века, и была связана с развитием звукового кино и электронной музыки. В московской студии электронной музыки музея Скрябина инженер Е. А. Шолпо решил, что звуковую дорожку можно создать искусственно. Он рисовал в крупном масштабе рассчитанные им звуковые волны, фотографировал их кадр за кадром и проигрывал готовую пленку через кинопроектор. Хотя работа была очень трудоемкой и малопроизводительной, Шолпо озвучил этим способом несколько мультфильмов с помощью построенного им прибора - вариафона.

Хорошо знавший работы Шолпо другой сотрудник Студии – Мурзин, выбрал метод синтеза речи с помощью ряда Фурье – в виде суммы элементарных спектральных составляющих, в музыкальной акустике получивших название "чистые тона". Банк "чистых тонов" Мурзин сконструировал в виде стеклянного диска, очень похожего на современный компакт-диск. На его основе был создан синтезатор звуков под названием АНС (от инициалов композитора А.Н. Скрябина, которому посвятил свое изобретение автор). Первые модели говорящих устройств тех времен были очень похожи на музыкальные инструменты, а обучение операторов тоже напоминало обучение музыкантов и требовало немало времени и способностей.

История “средних” лет

Начало современной истории создания русскоговорящих машин датируется серединой 60-х годов 20 века и непосредственно связано с развитием электроники и вычислительной техники. Немаловажную роль в освоении мирового технологического уровня синтеза речи того времени сыграли научные стажировки в конце 60-х годов М.Ф. Деркача

в Лаборатории Фанта (Стокгольм) и одного их авторов этой книги в Лаборатории Лоренца (Эдинбург), где впервые были разработаны формантные синтезаторы речи (см. рис.2).



Рис. 2. Гуннар Фант с формантным синтезатором

С использованием формантных синтезаторов этих лабораторий были впервые получены образцы синтеза русской речи весьма высокого качества. В последующие годы наиболее интенсивные исследования и разработки синтезаторов речи в СССР проводились в Минске, Ленинграде, Москве, Таллине.

Первая, пока ещё довольно упрощённая модель синтезатора русской речи, разработанная в Минске, «ФОНЕМОФОН-1» (рис. 3) «заговорила» в начале 70-х гг. и успех в её создании был связан, прежде всего, с разработкой новых принципов формантного синтеза речевых сигналов.



Рис. 3. Синтезатор «Фонемофон-1»

Позже появилась усовершенствованная модель формантного синтеза речи, в которой были оптимизированы характеристики

формантных фильтров «*ФОНЕМОФОН-2*». В 1979 г. «*ФОНЕМОФОН-3*» демонстрировался на Всемирной выставке «Телеком-79» в Женеве (см. рис. 4). Артур Кларк, посетивший павильон СССР, записал в книгу отзывов по поводу синтезатора речи: «*Вы превзошли мои фантазии «Космической Одиссеи – 2001».*



Рис.4. Б.М. Лобанов и синтезатор «Фонемофон-3» на Всемирной выставке «Телеком-79» в Женеве

Важную роль в создании серии промышленных синтезаторов речи сыграла разработка цифрового синтезатора «*ФОНЕМОФОН-4*» (1984). Его серийный выпуск впервые в СССР был налажен в ПО «Кварц» г. Калининграда благодаря энтузиазму конструктора Валерия Афонасьева. К 1986 г., в сотрудничестве с профессором кафедры фонетики Минского лингвистического университета Еленой Карневской, была разработана англоязычная версия синтезатора, демонстрировавшаяся на Всемирном конгрессе фонетических наук. Вот факсимиле отзыва об этой демонстрации уже упоминавшегося основоположника формантного синтеза речи Гуннара Фанта (рис. 5).

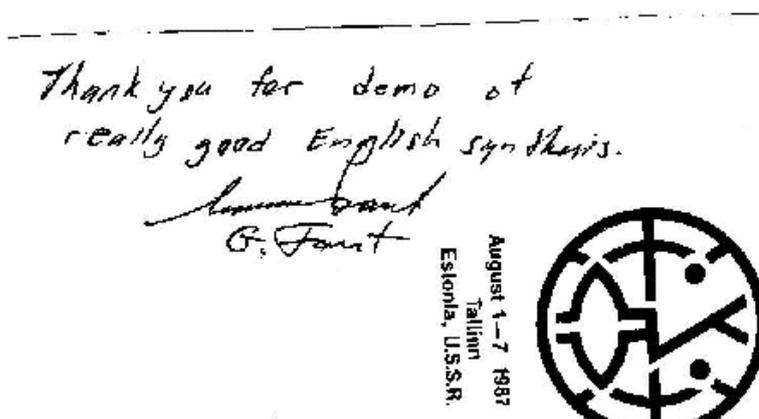


Рис.5. Отзыв Г. Фанта

Ещё долгое время формантный синтезатор играл ключевую роль в системах синтеза речи по тексту, пока в конце 80-х - начале 90-х годов не был предложен одним из авторов новый микроволновой (МВ) метод синтеза речевых сигналов, воплощённый в синтезаторе «ФОНЕМОФОН-5» инженером Александром Ивановым. Удивительная компактность созданного синтезатора (всего 64К байт) позволила оснастить синтезом речи первые персональные компьютеры класса ЕС-1840 и IBM-XT. До сих пор ещё он используется незрячими (более сотни комплектов программных продуктов для незрячих были созданы и распространены научным сотрудником Института технической кибернетики Георгием Лосиком в России, Украине и Белоруссии), а его вполне разборчивое звучание можно услышать в комплекте программ на CD ROM «Говорящая мышь», разработанных группой программистов из МГУ. На основе МВ-метода разработаны версии чешского и польского языков, а также автономный одноплатный модуль синтеза речи, украинско-язычная версия которого некоторое время работала на линиях киевского метро.

Новейшая история

К середине 90-х годов мощности персональных компьютеров так возросли, что можно было уже подумать не только о компактности программы и разборчивости речи, но и о естественности звучания синтезированного голоса. В этом направлении много сделано было на филфаке МГУ Ниной Зиновьевой и Ольгой Кривновой. В качестве элементарной единицы синтеза они предложили взять не микроволны (отдельные периоды сигнала), а целый звук – аллофон.

Следующий шаг в синтезе русской речи был сделан благодаря сотрудничеству Лаборатории экспериментальной фонетики Санкт-Петербургского университета с Национальным французским центром телекоммуникации (CNET). В течение 2-х лет (1995-96) сотрудники Лаборатории П. Скрелин и др. смогли успешно адаптировать их

дифонную технологию применительно к синтезу русской речи. Этот синтезатор стал коммерческим продуктом французской фирмы ELAN под названием DIGALO (см: www.digalo.com).

В конце 1999 г. в Минске в Институте технической кибернетики (сейчас – Объединённый институт проблем информатики НАН Беларуси) после почти 5-летнего перерыва вновь возобновились интенсивные работы по синтезу речи. Сравнительно небольшой коллектив способных молодых программистов сумел на современном уровне программно реализовать многолетний «речевой» опыт одного из авторов этой книги. К настоящему времени создана серия «движков», реализующих многоголосый синтез русской речи по тексту, а также синтез белорусской, польской и английской речи. Более того, положено начало и достигнуты обнадеживающие результаты в разработке принципиально новой технологии – технологии компьютерного «клонирования» персонального голоса и речи. Но об этом более подробно будет рассказано ниже.

О чём машине говорить? Зачем Она заговорила?

Говорящий компьютер – это принципиально новое средство человеко-машинного общения, преимущества и возможности использования которого до конца ещё не осознаны широкой общественностью. Здесь уместна аналогия с немым и звуковым кино. Дар речи "великий немой" – кинематограф – получил в конце 20-х годов, но ещё долгое время звуковые фильмы копировали приёмы режиссуры немого кино. Образно говоря, ситуация с современными компьютерами сейчас очень схожая. По-видимому, потребуется ещё немало времени, чтобы синтез речи стал органической частью компьютера и был широко востребован.

Синтезатор речи - это одна из составных частей речевого интерфейса, без которой разговор с компьютером не может состояться. При этом имеется в виду прочтение вслух произвольной текстовой информации, а не проигрывание предварительно записанных звуковых файлов. Синтезатор речи обеспечивает выдачу в речевой форме заранее неизвестной информации непосредственно по орфографическому тексту, генерируемому компьютером.

С точки зрения пользователя, наиболее разумное решение технологии синтеза речи – это включение речевых функций (в перспективе – многоязычных, с возможностями перевода) в состав операционной системы. Точно так же, как мы пользуемся командой PRINT, можно будет применять команду TALK или SPEAK. Такие команды, по-видимому, появятся в ближайшем будущем в меню общеупотребительных компьютерных приложений и в языках

программирования. Важно отметить, что пользователь должен также иметь достаточные возможности по настройке голоса компьютера (индивидуальности звучания, тембра и темпа речи).

Фактически, благодаря синтезатору речи по тексту (имеющему в англоязычной литературе стандартную аббревиатуру TTS – Text-To-Speech), открывается еще один канал передачи данных от компьютера к человеку, аналогичный тому, который мы имеем благодаря монитору или принтеру. Конечно, малоэффективным было бы передавать рисунок голосом, но вот услышать электронную почту или результат поиска в базе данных в ряде случаев было бы весьма удобно, особенно если в это время глаза заняты чем-либо другим. Синтезатор речи совершенно незаменим, если вы хотите получить информацию, находясь далеко от компьютера или в движении. Воспользовавшись обычным или мобильным телефоном, можно будет связаться со своим компьютером и прослушать электронную почту или интересующую вас страничку интернет. В экстренных случаях компьютер сам сможет дозвониться до вас и, выполняя роль секретаря, сообщить необходимую информацию.

Вышеупомянутые функции синтезатора уже сейчас крайне необходимы для лиц, имеющих проблемы со зрением. Инвалидность по зрению имеет особо тяжелые социально-психологические последствия для человека. Как образно заметил д.п.н. Георгий Лосик, *«изобретая линзу, ученые не полагали, что она породит такое приспособление, как очки, которое сделает слабовидящих людей неотличимыми от зрячих. Точно так же, изобретая синтезатор речи, они не подозревали, что он совершит подобную революцию в жизни тотально незрячих людей, дела их равными со всеми в мире компьютерной информации»*.

Вообще, даже простое перечисление ситуаций, в которых будет полезен синтез речи - это материал для большого самостоятельного обзора.

Компьютерное "клонирование" персонального голоса и речи

Многолетние исследования, выполненные в XX веке, позволили создать синтезаторы, обеспечивающие качество и разборчивость речи, вполне пригодное для широкого спектра практических приложений. Однако, несмотря на все усилия, синтезированная речь оставалась ещё далёкой по качеству от естественной и обладала узнаваемым машинным акцентом. Причиной этому были не столько уровень наших знаний о процессах речеобразования и о фонетике, сколько нехватка вычислительных ресурсов компьютеров того времени. Сейчас мы можем не ограничивать себя ни объёмом оперативной и дисковой памяти, ни требуемым объёмом вычислений и приступить к созданию системы

синтеза речи по тексту с максимально возможным приближением по звучанию к голосу и манере чтения конкретного диктора.

Такая постановка задачи, хотя и отдалённо, напоминает широко известную биологическую проблему клонирования, когда делается попытка воспроизвести живую копию на основе генетического материала только одного родителя. В нашем случае, в отличие от классической задачи клонирования, ставится цель создания близкой копии, но не биологической, а компьютерной, и не всего существа в целом (в данном случае человека), а только одной из его интеллектуальных функций: чтение произвольного орфографического текста. При этом ставится задача максимально полного сохранения персональных акустических особенностей голоса, фонетических особенностей произношения и акцента, а также просодической индивидуальности речи (мелодика, ритмика, динамика). В принципе, в генетике рассматривается и такая возможность, как создание своеобразных "химер" из разнородного генетического материала. Применительно к технологии клонирования голоса и речи - это тот случай, когда в основу синтеза закладываются, например, акустика голоса одного диктора, фонетические особенности произношения - другого, а просодическая индивидуальность речи - третьего.

Клонирование акустических характеристик голоса

Персональные акустические характеристики голоса диктора обусловлены множеством факторов, таких как анатомические особенности строения и функционирования элементов речевого аппарата (гортань, голосовые связки, глотка, полость рта и др.), динамические особенности взаимодействия колебаний голосовых связок и резонаторов речевого аппарата ("coupling effect"), а также многое другое. Как известно, попытки имитации персональных характеристик голоса в системах «текст – речь» на основе моделирования физиологических и акустических процессов речеобразования из-за их чрезвычайной сложности до сих пор не привели к ощутимым результатам. В связи с этим наиболее разумным представляется использование отрезков натуральной речевой волны в качестве минимального "генетического материала" для клонирования голоса. В качестве таких отрезков целесообразно выбрать позиционно-комбинаторные варианты фонем – аллофоны или мультиаллофоны, ограниченный набор которых способен обеспечить порождение устной речи произвольного содержания. При этом звуковая волна содержит в себе все персональные особенности голосообразования, проявляющиеся в каждом конкретном аллофоне или мультиаллофоне.

Клонирование фонетических особенностей произношения

В отличие от персональных акустических характеристик голоса, обусловленных, в основном, статическими параметрами речевого аппарата, фонетические особенности произношения обусловлены главным образом динамикой артикуляторных движений, осуществляемых в процессе речеобразования. Присущие данному индивиду скорость артикуляторных движений, характерные запаздывание или опережение движений отдельных артикуляторов, индивидуальные особенности артикуляции того или иного звука (например /P/), региональный или иностранный акцент, дефекты произношения некоторых звуков обуславливают возникновение своеобразных позиционных и комбинаторных оттенков фонем и создают уникальный набор аллофонов. Таким образом, успех клонирования персональных фонетических особенностей произношения зависит главным образом от успеха имитации персональных особенностей фонемно-аллофонного преобразования.

Клонирование просодических характеристик речи

Комплекс просодических (интонационных) характеристик, включающий мелодику, ритмику и энергетику речи, задаётся закономерными изменениями во времени частоты основного тона – F_0 , длительности звуков – T и амплитуды звуковых сигналов – A . Характер этих изменений определяется не только конкретным текстом, но и персональной манерой его чтения. Решение задачи клонирования просодических характеристик речи конкретного диктора заключается в создании достаточно полного набора персональных «портретов просодем» его речи.

Технология клонирования

Для успешного клонирования персональных характеристик голоса и дикции необходимо создать достаточно полные наборы звуковых волн аллофонов и интонационных «портретов» речи. В случае, если клонируемый диктор физически доступен, для этой цели используется специально разработанный компактный звуковой массив слов и отрывков текста, начитываемый им в студии или в обычных условиях. Если же клонируемый диктор недоступен, то используются уже имеющиеся записи его голоса на радио, телевидении и др.

Первые результаты по клонированию (на примере персонального голоса и дикции одного из авторов этой книги) были получены в лаборатории распознавания и синтеза речи ОИПИ в 2000 году [2]. К настоящему времени набор клонов состоит уже из нескольких мужских и женских голосов, созданных на основе технологии компьютерного клонирования, разработанной авторами и достаточно детально

описанной в данной книге. Проведенные опыты по клонированию различных голосов показали, что с использованием специально подобранных массивов слов и отрывков текста достаточно хорошие результаты могут быть получены при длительности звуковой записи порядка 5 – 10 минут. В случае использования произвольных текстов минимально необходимая длительность звуковой записи составляет порядка 20 – 40 минут.

Перспективы компьютерного клонирования

Проводимая нами аналогия между биологической проблемой клонирования и лингво-акустической проблемой синтеза персонализированной речи по тексту может стать не только лишь красивой метафорой. Во-первых, она подчёркивает общенаучную значимость, современность и сложность поставленной задачи. Во-вторых, она выделяет эту задачу в отдельный самостоятельный класс в ряду других задач современных речевых технологий. И, наконец, в-третьих, она стимулирует создание новых специализированных методик, а также автоматических и полуавтоматических методов клонирования персонального голоса и речи в системах "Текст-Речь".

Отметим также некоторые возможные коммерческие аспекты компьютерного клонирования персонального голоса и речи. По нашему мнению, найдётся большое количество пользователей компьютера, желающих, чтобы их персональный компьютер заговорил его собственным голосом. И хотя это всего лишь компьютерный, а не биологический клон, однако обладатели такого "клона" всё же могут быть уверены, что хотя бы частица их сущности – их голос и манера чтения – останутся нетленными. Многим, наверное, было бы интересно, чтобы компьютер говорил голосом близкого ему человека или голосом любимого актёра. Интересным может быть также проект оживления давно ушедших от нас голосов великих людей по оставшимся от них грамофонным или студийным записям. Таким путём можно было бы услышать голос Есенина, читающего не читанные им ранее стихи, или голос знаменитого диктора Левитана, объявляющего новые указы Президента. В более практическом плане разработка эффективной технологии клонирования голоса и речи может значительно повысить привлекательность использования синтезаторов речи в разнообразных компьютерных системах управления благодаря высокому качеству и естественности речи, её персонализации и узнаваемости голоса.

Наряду с указанными положительными примерами применения технологии «клонирования» характеристик голоса и речи человека следует отметить также и определённую опасность её недобросовестного или криминального использования. Можно

представить себе, например, провокационные телефонные звонки компьютера, имитирующие голос знакомого человека, или же несанкционированное использование голоса известного актёра или общественного деятеля для целей озвучивания не вполне этичных рекламных роликов. Преступник может представиться по телефону, например, менеджером банка и вынудить вас разгласить данные персонального счета, или с определённой целью завести разговор от имени известного политического лица. Однако это уже выходит за рамки проблем клонирования речи и относится к самостоятельной области информационной безопасности, обладающей собственными мощными средствами противодействия.

В биологии есть понятие о двух основных классах экспериментов по клонированию – *in Vitro*” (т.е. в пробирке) и – *“in Vivo”* (т.е. в живом). Таким образом, можно сказать, что сегодня путём компьютерного воссоздания голоса человека закладываются основы нового класса экспериментов по клонированию – *“in Silico”* (т.е. в микросхемах). Это может стать увлекательной перспективой для многих направлений создания систем искусственного интеллекта, наделённых неповторимыми чертами личности конкретного человека, т.е. не только его голосом, но и определённой суммой знаний, его поведенческими особенностями в той или иной области деятельности.

Предлагаемая читателю монография является обобщением научных результатов, полученных авторами за последние годы. Она посвящена рассмотрению теоретических и экспериментальных основ, а также описанию конкретных практических решений и результатов компьютерного синтеза и клонирования речи.

Введение

Современный уровень развития вычислительной техники и ее повсеместное внедрение в человеко-машинных системах управления делают актуальной организацию общения человека и компьютера в одной из наиболее удобных для человека форм – в форме речевого диалога на естественном языке. Речевой способ общения в человеко-машинных системах имеют принципиальные преимущества, главными из которых являются следующие:

– удобство, естественность и простота общения, не требующая специальной подготовки, что существенно расширяет круг потенциальных пользователей вычислительных систем и повышает эффективность их использования;

– разгрузка зрительного канала при выводе информации и устранение ручных манипуляций при вводе, что увеличивает оперативность взаимодействия с компьютером и уменьшает число ошибок оператора;

– возможность использования в качестве терминалов телефонных аппаратов и существующей сети телефонной связи, что позволяет создавать системы массового обслуживания, в том числе с выходом в Интернет.

Создание систем речевого общения с ЭВМ требует решения двух основных проблем: проблемы автоматического синтеза и проблемы автоматического анализа и распознавания речи. Наглядным примером роста популярности систем обработки речи является следующий факт: с момента организации в 1989 году крупнейшей конференции по речевым технологиям EURO-SPEECH, а затем INTER-SPEECH, число ее участников к 2007 году выросло с 300 до 3000 человек [3, 4].

В данной книге освещаются вопросы, связанные с решением первой проблемы – проблемы синтеза речи по тексту – и её дальнейшим развитием – проблемой клонирования в процессе синтеза характеристик голоса и речи конкретной личности.

В полном объеме проблема синтеза речи решается в тех исследованиях и разработках, в которых ставится задача автоматического синтеза речи *неограниченного словарного состава* непосредственно по орфографическому тексту сообщения с любыми наперёд заданными характеристиками синтезированного голоса и речи. Эта задача удовлетворительно еще не решена ни для одного из языков, хотя на ее решение были затрачены усилия многих исследователей США, Японии, Франции и других странах. Исследователи сталкиваются здесь с принципиальными трудностями, связанными с созданием интеллектуальной системы, моделирующей речевое чтение человеком произвольных текстов. Следует отметить также, что далеко не все

результаты, полученные для одного языка, годятся при разработке синтезатора речи другого языка. Фонетические системы каждого конкретного языка уникальны. Необходима конкретно-языковая разработка правил ассимиляции, коартикуляции и редукции звуков, просодической (ритмико-мелодической) организации речи.

Исследования, проведенные для создания моделей синтеза речи по тексту, позволили вскрыть тонкие глубинные структуры звуковой организации речевого потока. Удовлетворительный по качеству синтез речи невозможен в принципе, если не созданы всеобъемлющая и детально проработанная акустическая модель полной системы фонем языка, модель их модификаций в речевом потоке под действием комбинаторных, позиционных и просодических факторов. По мере разработки все более качественной модели синтеза речи по тексту накапливается все более достоверная информация об эталонных образцах каждой фонемы и их модификаций в речевом потоке. Эта информация является исключительно ценной для смежной области – автоматического распознавания речи.

Несмотря на широкое развитие исследований в области синтеза речи по тексту, основным критерием качества до недавних пор оставалась степень разборчивости синтезированной речи. Недостаточное внимание, уделяемое естественности синтезированной речи, ставит барьер для широкого применения систем синтеза речи по тексту в составе интерфейсов компьютерных систем и устройств. Одним из путей повышения естественности синтезированной речи является разработка методов и средств анализа персональных особенностей голоса человека, его дикции и выразительности речи с последующей реализацией этих характеристик при синтезе речи по тексту. При этом решается не только проблема повышения естественности синтезированной речи, но и проблема воспроизведения произвольного текста с манерой чтения конкретного человека и его голосом.

Такая постановка задачи впервые сформулирована одним из авторов книги [5, 6] как проблема «компьютерного клонирования». Компьютерное клонирование – это дальнейший этап развития систем искусственного интеллекта, когда моделируется не только сама интеллектуальная функция, но и особенности её проявления у конкретного человека. Введение биологического термина «клонирование» не является самоцелью и возможно в будущем станет таким же продуктивным для компьютерных наук, как «нейронные сети» или «генетические алгоритмы».

В речевых технологиях целью компьютерного клонирования является создание системы синтеза речи по тексту с максимально возможным приближением по звучанию к голосу и манере чтения конкретного человека. При этом ставится задача максимально полного

сохранения персональных акустических особенностей голоса, фонетических особенностей произношения и акцента, а также просодической индивидуальности речи (мелодика, ритмика, динамика).

Таким образом, исследование проблемы синтеза речи по тексту актуально как в практическом плане создания универсальных систем речевого вывода информации из компьютера, так и в теоретическом плане создания интеллектуальных персонализированных моделей чтения текстов различных стилей и содержания человеком.

Книга содержит предисловие, введение, 6 глав и приложения.

В первой главе *«Фонетико-акустические основы синтеза речи по тексту»* излагаются основные сведения о фонетике русского языка. Рассмотрены артикуляторно-акустические особенности образования звуков речи и принципы просодического оформления речевого потока.

Во второй главе *«Методы синтеза фонетико-акустических характеристик речи»* рассматриваются методы синтеза речи по тексту и их возможности для передачи персональных характеристик речи, формулируется содержательная постановка задачи компьютерного клонирования персональных характеристик речи.

В третьей главе *«Экспериментальные исследования фонетико-акустических и просодических характеристик речи»* описаны результаты исследований фонетико-акустических и просодических характеристик речи дикторов, проведенных авторами в ходе реализации задач синтеза и клонирования.

В четвертой главе *«Компьютерная модель мультиволнового синтеза речи по тексту»* описаны структура системы синтеза речи по тексту и функционирование её основных блоков. Приведены особенности программной реализации и пользовательского интерфейса системы «МультиФон», а также результаты экспериментальных исследований разборчивости синтезированной речи.

В пятой главе *«Компьютерное клонирование индивидуальных характеристик речи»* формулируются основные этапы технологии компьютерного клонирования персональных характеристик речи, описаны методы реализации технологии компьютерного клонирования. Приведены особенности программной реализации и пользовательского интерфейса систем «ФоноКлонатор» и «ИнтоКлонатор», а также результаты экспериментальных исследований правдоподобия синтезированных с их помощью речевых клонов.

В приложениях приведен богатый фактический материал по описанию используемого текстового корпуса, звуковых, интонационных, лексических, морфологических и синтаксических характеристик устной речи, полезный разработчикам технологий и

компьютерных систем и распознавания и синтеза речи, а также исследователям в областях прикладной лингвистики, в частности, фонетики и просодики.

Глава 1

Фонетико-акустические основы синтеза речи по тексту

1.1. Общие сведения о языке и речи

Язык определяется как совокупность фонетических, лексических и грамматических средств для выражения мыслей и чувств человека [7]. Язык образует органическое единство с мышлением и служит важнейшим средством общения между людьми. Национальный язык представляет собой явление социальное и связан в своём возникновении и развитии с конкретным коллективом (народом).

Язык и речь тесно взаимосвязаны и образуют единство абстрактного и конкретного: систему средств общения и реализацию этой системы. Реализация языковых средств общения может осуществляться в формах письменной и устной речи. Письменная речь фиксируется в форме текста, устная речь реализуется в звуковой форме. Письменная форма речи подчиняется законам грамматики данного языка, правилам правописания и пунктуации. Устная речь – законам фонетики языка, правилам произношения звуков и интонирования (просодики) речи.

На физическом уровне устная речь реализуется в виде специфических звуков (акустических колебаний), генерируемых речевым аппаратом человека. По образному высказыванию Н.И. Жинкина [8], «речевой звук – это величайшее разнообразие, целый мир, сонм звуковых признаков. Здесь одни спектры сменяются мгновенно другими, модулируется частота, сила и длительность звука по многим ступеням».

Естественный процесс речеобразования следует рассматривать в двух взаимно дополняющих аспектах: лингвистическом и акустическом. Лингвистическая теория рассматривает фонетические и просодические аспекты речи, акустическая теория описывает артикуляторно-акустические характеристики фонем и просодем. Лингвистический аспект речи является предметом исследования фонетики - науки о звуковой системе языка. Основы фонетики русского языка заложены в работах Бодуэна де Куртенэ, Богородицкого и Щербы еще в конце XIX - начале XX века. В состав звуковой системы языка входят основные звуки речи - фонемы и их модификации в речевом потоке - аллофоны, а также интонация (просодика), включающая мелодику, ритмику и энергетику речи. Интонация определяет коммуникативный тип высказывания, т.е. она помогает поддерживать диалог, коммуникацию между людьми, и, соответственно, поддерживает не только монологическую речь, но и диалогическую, где могут быть

всевозможные вопросительные и восклицательные конструкции, кроме обычных типов повествовательных конструкций.

Экспериментально-фонетические исследования артикуляторного и звукового аспектов речи были начаты фонетистами в 50 -х годах [8, 9]. Фонетистами получены рентгенограммы, палатограммы и спектрограммы основных звуков русской речи [10, 11]. Большую роль в понимании физиологических процессов речеобразования и восприятия речи сыграла работа Л.А.Чистович и др. [12]. Общее состояние электроакустических исследований звуков речи тех лет было отражено в известных монографиях М. А. Сапожкова [13], а затем Д.Фланагана [14]. Большое значение в исследовании артикуляторных и звуковых (спектральных) характеристик речи имела ставшая уже классической работа Г.Фанта [15]. В ней указанные аспекты речи фактически объединены в одну общую акустическую теорию речеобразования.

Большое внимание, начиная с 60-х годов XX века, было также уделено экспериментальному исследованию интонации русской речи [16-20]. В формировании речи интонация выполняет как грамматическую функцию, так и стилистическую. При помощи интонации передается законченность предложения и осуществляется его членение на синтаксические единицы, выражается эмоциональность речи, волевые побуждения, а также различные модальные оттенки значений. Очевидно, что если при синтезе речи исключить просодическую составляющую, то такая речь будет невыразительной и монотонной. Кроме того, смысл сказанного тоже может быть понят не всегда. Приведём известный пример такого рода: «Казнить, нельзя помиловать» или «Казнить нельзя, помиловать». Здесь конкретный вариант просодического членения, употреблённый говорящим, может сыграть очень серьезную роль в жизни данного человека.

Просодика выражается в физических параметрах речевого сигнала, таких как: частота основного тона – F_0 (мелодика речи), сила звука или амплитуда речевого сигнала – A (энергетика речи), а также длительность звуков в потоке речи – T (ритмика речи). Просодические модели для различных типов высказываний были предложены в работах [21-22]. Роль просодики для синтеза речи исследовалась в [23-24].

Рассматриваемая в данной монографии проблема создания компьютерных систем синтеза речи целиком укладывается в решение комплекса проблем автоматического преобразования письменной речи в устную, т.е. в решение задачи синтеза «текст - речь». При этом синтез устной речи должен осуществляться на основе лексико-грамматического анализа заданного текста путём моделирования процессов речеобразования с учётом правил произношения звуков и интонирования, свойственных данному языку.

Аналогично тому, как связный текст естественного языка членится на различные иерархические уровни, в потоке речи выделяются следующие основные иерархические составляющие:

- высказывание;
- фонетический период;
- фраза;
- синтагма;
- акцентная единица;
- фонетическое слово;
- слог;
- фонема;
- аллофон;
- звук.

В речевом потоке **высказывание** членится на **фонетические периоды** (или сверхфразовые единства), которым в тексте обычно соответствуют абзацы.

Фонетические периоды речи, в свою очередь, состоят из **фраз**, которым в тексте соответствуют повествовательные, вопросительные или восклицательные предложения.

Каждая фраза может состоять из одной или нескольких **синтагм**, которые являются относительно самостоятельными просодическими единицами речи. В письменном тексте нет точного эквивалента синтагмы. Тем не менее, определёнными индикаторами границы синтагмы могут являться знаки препинания внутри предложения. Синтагма может состоять как из одного слова, так и из нескольких слов – двух, трех и более, в зависимости от строения фразы и стиля речи говорящего.

Синтагмы могут состоять из одной или нескольких **акцентных единиц**. Под акцентной единицей понимаются одно или несколько слов, объединённых общим групповым (сильным) ударением.

Акцентная единица, в свою очередь, состоит из **фонетических слов**. Фонетическое слово включает знаменательное слово и присоединённые к нему служебные слова - предлоги или частицы.

Фонетическое слово может быть подразделено далее на **слоги**. С точки зрения физиологии образования слог представляет собой звук или несколько звуков, произносимых одним выдохательным толчком.

Слоги, в свою очередь, состоят из **фонем**, являющихся минимальными смыслоразличительными единицами речи.

Еще один, последний уровень составляющих, – **аллофоны**, или оттенки фонем, которые реализуются в речи в виде акустических колебаний – **звуков**.

Общелингвистические сведения, используемые в данном разделе, почерпнуты из наиболее известных работ отечественных и зарубежных

авторов, в которых отражен современный уровень развития фонетических исследований русского языка [16-24], а также из ряда собственных работ авторов [25-27].

1.2. Система фонем русского языка

Под фонемой в лингвистике понимается наименьшая смысловоразличительная единица речи [19]. Вопрос о существовании каждой отдельной фонемы решается положительно, если она участвует в смысловом противопоставлении слов языка. Другими словами, данный звук является фонемой, если найдутся слова, различающиеся только этим звуком. Например, фонемы /м/ и /в/ существуют, поскольку существуют слова *МОЛ* и *ВОЛ*.

В русском языке насчитывается 42 фонемы. Каждая фонема обладает определенными акустическими свойствами, которые обусловлены артикуляторными особенностями ее образования. Схематическое изображение артикуляторного аппарата человека дано на рис. 1.1. Различают активные и пассивные органы артикуляции.

К **активным** органам относятся:

- кончик, спинка, боковинки и тело языка,
- губы,
- нёбная занавеска,
- нижняя челюсть,
- голосовые связки.

К **пассивным** органам относятся:

- зубы,
- альвеолы,
- мягкое нёбо
- твёрдое нёбо,
- носовая полость,
- глотка и гортань.

Каждую фонему характеризует определенный "способ" и "место" ее образования. По способу образования русские фонемы делятся на две основные группы: гласные /у, о, а, э, ы, и/ и согласные (остальные 36 фонем). Образование гласных фонем характеризуется отсутствием преград в речевом тракте, в то время как при образовании согласных в ротовой полости обязательно присутствует полная или неполная смычка (щель), создаваемая языком или губами.

Согласные фонемы по способу образования делятся на группы сонорных, щелевых, взрывных и аффрикат.

Для группы сонорных согласных /м, м', н, н', л, л', р, р', й'/ характерно наличие относительно широкой щели. Эта щель образуется следующим образом:

- при опускании нёбной занавески у носовых сонорных /м, м', н, н'/,
- при опускании боковинки языка у боковых сонорных /л, л'/,
- между колеблющимся кончиком языка и альвеолами у дрожащих /р, р'/,
- между спинкой языка и твёрдым нёбом у плавной сонорной /й'/.

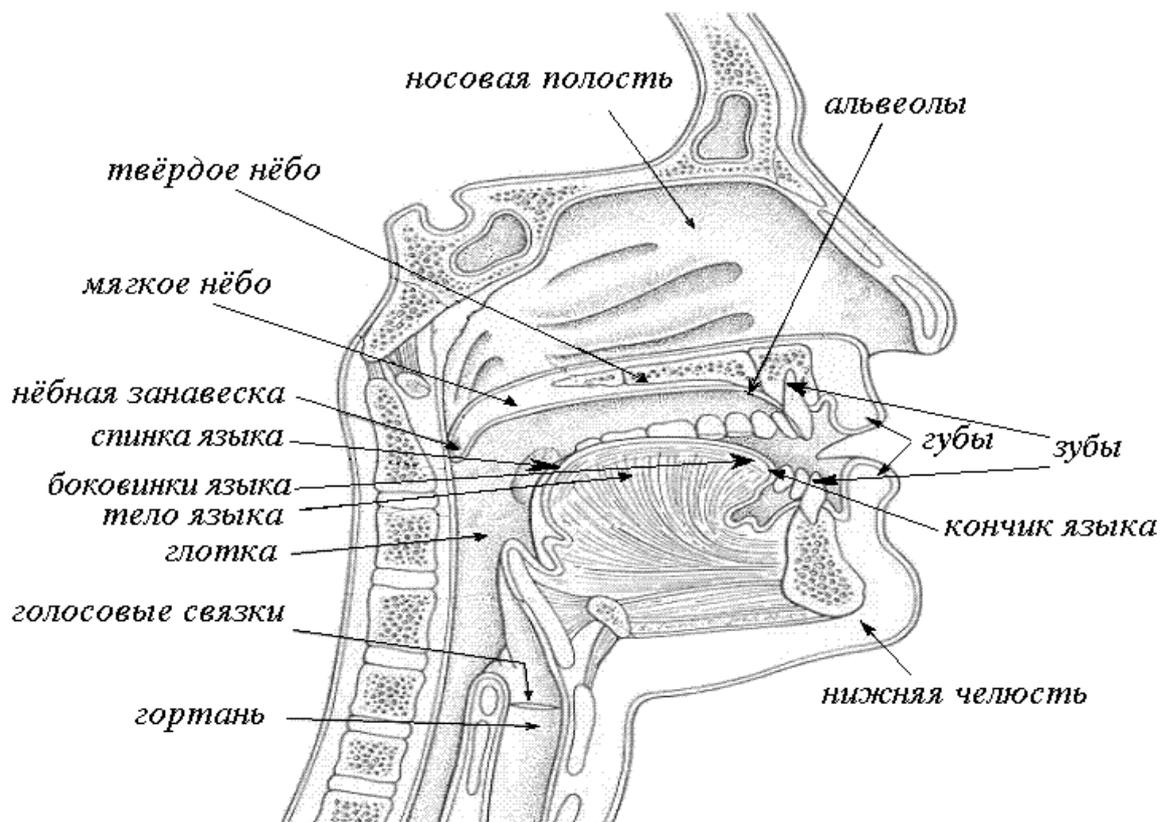


Рис. 1.1. Строение речеобразующего тракта

Группа щелевых согласных /в, в', з, з', ж, ф, ф', с, с', ш, ш', х, х'/ характеризуется наличием достаточно узкой шумообразующей щели, возникающей при неполной смычке артикуляторных органов. Щелевые согласные делятся далее на звонкие /в, в', з, з', ж/ и глухие /ф, ф', с, с', ш, ш', х, х'/ в зависимости от того, участвуют или не участвуют в их образовании голосовые связки.

Группа взрывных характеризуется наличием полной смычки в артикуляторном тракте с последующим ее резким открытием. Так же как и щелевые согласные, взрывные делятся на звонкие /б, б', д, д', г, г'/ и глухие /п, п', т, т', к, к'/.

И наконец, фонемы из группы глухих аффрикат /ц, ч'/ характеризуются тем, что в процессе их образования происходит смена фазы смычки на фазу шумообразующей щели.

Рассмотрим далее классификацию русских фонем по "месту" образования. Под "местом" образования в фонетике понимается положение

сужений в артикуляторном тракте, определяющее его конфигурацию и, в конечном итоге, его резонансные свойства.

Место образования гласных фонем определяется положением тела языка (подъем высокий/низкий; продвижение вперед/назад) и степенью сближения губ (огубление).

Место образования согласных фонем определяется положением в артикуляторном тракте смычки или щели, а также положением тела языка (мягкие или твердые согласные). Согласные фонемы по месту образования делятся на группы губных, зубных, альвеолярных и нёбных, каждая из которых может включать твёрдые /м, н, л, р, в, з, ж, ф, с, ш, х, б, д, г, п, т, к, ц/ или мягкие /м', н', л', р', й', в', з', ф', с', ш', х', б', д', г', п', т', к', ч'/ согласные.

К группе губных согласных относятся /м, м', в, в', ф, ф', п, п'/. При этом губному месту образования соответствует контакт нижней губы с верхними зубами или с верхней губой.

К группе зубных относятся /н, л, з, с, д, т, ц, н', л', з', с', д', т'/. При этом зубному месту образования соответствует контакт кончика языка с верхними зубами.

К группе альвеолярных относятся /р, ж, ш, р', ш', ч'/. Альвеолярному месту образования соответствует контакт кончика языка с альвеолами.

К группе нёбных относятся /х, г, к, х', г', к', й'/. Нёбному месту образования соответствует контакт спинки языка с твердым нёбом.

Мягкие (палатальные) согласные соответствующего места образования характеризуются еще дополнительным подъёмом спинки языка к мягкому нёбу.

В таблице 1.1 представлены фонемы русского языка (в русской и латинской транскрипциях), представленные в координатах «место» - «способ» образования, в соответствии с описанной выше их классификацией.

Для сравнения в таблице 1.2 в тех же координатах представлены фонемы русского языка (в верхних строках) и белорусского (в нижних строках), при этом для транскрипции используются буквы национальных алфавитов.

Отличительные особенности фонетических систем белорусского и русского языков заключаются в следующем.

В белорусском языке отсутствуют следующие фонемы:

- мягкие согласные *Т, Д, Ш, Ч, Р*;
- мягкая и твёрдая *Г*.

В белорусском языке имеется ряд специфических фонем, отсутствующих в русском:

- плавная *Ў*;
- мягкая *Ц* и твёрдая *Ч*;

- мягкая аффриката *Дз* и твёрдая *Дж*;
- мягкая и твёрдая щелевая *Гх*.

Вычисляя степень сходства фонетических систем русского и белорусского языка как отношение количества одинаковых для двух языков фонем к общему количеству фонем (см. табл. 2.2), получим, что фонетические системы этих языков совпадают на 71%.

Таблица 1.1

Система фонем русского языка в координатах «место образования» - «способ образования»

		Согласные									Гласные			
Способ образования		Шумные					Сонорные				Огубленные	Неогубленные	Способ образования	
		Глухие			Звонкие		Дрожание	Носовые	Боковые	Плавные				
		Взрывные	Аффрикаты	Щелевые	Взрывные	Щелевые								
Место образования		Взрывные	Аффрикаты	Щелевые	Взрывные	Щелевые	Дрожание	Носовые	Боковые	Плавные	Огубленные	Неогубленные	Место образования	
Небные (заднеязычные)	Мягкие	<i>к'</i> <i>К'</i>	~	<i>х'</i> <i>Н'</i>	<i>г'</i> <i>Г'</i>	~	~	~	~	<i>й'</i> <i>Ј'</i>	~	<i>и</i> <i>І</i>	Высокие	Передние
	Твёрдые	<i>к</i> <i>К</i>	~	<i>х</i> <i>Н</i>	<i>г</i> <i>Г</i>	~	~	~	~	~	~	<i>у</i> <i>У</i>		
Альвеолярные (среднеязычные)	Мягкие	~	<i>ч'</i> <i>Ch'</i>	<i>ш'</i> <i>Sh'</i>	~	~	<i>р'</i> <i>R'</i>	~	~	~	~	<i>э</i> <i>Е</i>	Низкие	
	Твёрдые	~	~	<i>ш</i> <i>Sh</i>	~	<i>ж</i> <i>Zh</i>	<i>р</i> <i>R</i>	~	~	~	~	<i>ы</i> <i>У</i>		
Зубные (Переднеязычные)	Мягкие	<i>т'</i> <i>T'</i>	~	<i>с'</i> <i>S'</i>	<i>д'</i> <i>D'</i>	<i>з'</i> <i>Z'</i>	~	<i>н'</i> <i>N'</i>	<i>л'</i> <i>L'</i>	~	<i>у</i> <i>У</i>	<i>ы</i> <i>У</i>	Высокие	Задние
	Твёрдые	<i>т</i> <i>T</i>	<i>ц</i> <i>С</i>	<i>с</i> <i>S</i>	<i>д</i> <i>D</i>	<i>з</i> <i>Z</i>	~	<i>н</i> <i>N</i>	<i>л</i> <i>L</i>	~	<i>о</i> <i>О</i>	<i>а</i> <i>А</i>		
Губные	Мягкие	<i>п'</i> <i>P'</i>	~	<i>ф'</i> <i>F'</i>	<i>б'</i> <i>B'</i>	<i>в'</i> <i>V'</i>	~	<i>м'</i> <i>M'</i>	~	~	<i>о</i> <i>О</i>	<i>а</i> <i>А</i>	Низкие	
	Твёрдые	<i>п</i> <i>P</i>	~	<i>ф</i> <i>F</i>	<i>б</i> <i>B</i>	<i>в</i> <i>V</i>	~	<i>м</i> <i>M</i>	~	~	<i>о</i> <i>О</i>	<i>а</i> <i>А</i>		

Таблица 1.2

Сравнительная таблица фонетических систем русского и белорусского языков

Согласные												Гласные			
Способ образования	Место образования	Шумные						Сонорные				Огубленные	Неогубленные	Способ образования	
		Глухие			Звонкие			Дрожание	Носовые	Боковые	Плавные			Место образования	
		Взрывные	Аффрикаты	Щелевые	Взрывные	Аффрикаты	Щелевые								
Нёбные (заднеязычные)	Мягкие	к', к'	~ ~	х', х'	г', ~	~ ~	~ гх'	~ ~	~ ~	~ ~	й', й'	~ ~	и і		Высокие
	Твёрдые	к к	~ ~	х х	г ~	~ ~	~ гх	~ ~	~ ~	~ ~	~ ~				
Альвеолярные (среднеязычные)	Мягкие	~ ~	ч', ~	ш', ~	~ ~	~ ~	~ ~	р', ~	~ ~	~ ~	~ ~	~ ~	э э	Низкие	
	Твёрдые	~ ~	~ ч	ш ш	~ ~	~ дж	~ ж ж	р р	~ ~	~ ~	~ ~				
Зубные (Переднеязычные)	Мягкие	т', ~	~ ц'	с', с'	д', ~	~ дз'	з', з'	~ ~	н', н'	л', л'	~ ~	у у	ы ы	Высокие	Задние
	Твёрдые	т т	ц ц	с с	д д	~ ~	з з	~ ~	н н	л л	~ ~				
Губные	Мягкие	п', п'	~ ~	ф', ф'	б', б'	~ ~	в', в'	~ ~	м', м'	~ ~	~ ~	о о	а а	Низкие	
	Твёрдые	п п	~ ~	ф ф	б б	~ ~	в в	~ ~	м м	~ ~	~ ў				

1.3. Оттенки фонем и система аллофонов русской речи

Описанные выше артикуляторные признаки места и способа образования характеризуют систему фонем русского языка в статическом виде. В потоке речи в зависимости от конкретных условий акустико-артикуляторные характеристики фонем изменяются, что

приводит к появлению оттенков фонем, или аллофонов. Аллофоны подразделяются на позиционные и комбинаторные [19].

Позиционные аллофоны определяются положением фонемы по отношению к ударному слогу в слове. Наличие сильноцентрированного словесного ударения (полного или частичного) является одной из главных фонологических особенностей русского языка [28]. Эта особенность поддерживается ярким фонетическим контрастом ударных и безударных гласных благодаря существенным различиям их количественных (длительность и амплитуда) и качественных (спектральная окраска) характеристик. Ударные слоги в качественном и количественном отношении реализуются наиболее полно, в то время как безударные слоги, а, в особенности, гласные в заударных слогах могут редуцироваться в очень сильной степени (эффект редукции). Редукция безударных гласных подчиняется строгим закономерностям, что позволило ещё в XIX веке вывести русскому филологу А.А. Потебня [29] знаменитую формулу ритмической организации русского слова:

$$2-1-0-2 \quad (1.1)$$

где номер обозначает степень редукции гласного; 0 – отсутствие редукции, 1 – первая и 2 – вторая степень редукции.

В расширенном виде, с учётом неприкрытого начального слога и открытого конечного слога в слове, эта формула может быть записана как:

$$G_1 - (CG_2 - \dots - CG_2) - CG_1 - CG_0 - (CG_2 - \dots - CG_2) - CG_1 \quad (1.2)$$

где G обозначает гласный звук, C – согласный; степень редукции гласного показана индексом при букве G .

Комбинаторные аллофоны определяются ближайшим контекстом фонемы и обусловлены наложением в потоке речи артикуляции одного звука на другой в результате действия эффектов коартикуляции и ассимиляции. В результате коартикуляции согласная фонема в значительной степени приобретает окраску последующей гласной фонемы (например, огубление согласной /x/ в слове «сухую - *сухуйу*» в сравнении с «сухая - *сухая*»). Гласная же фонема в значительной степени приобретает окраску предшествующей согласной (например, смягчение гласной /a/ в слове «тяпки - *т'апк'и*» в сравнении с «*тапки* - *тапк'и*»). Эффекты ассимиляции в наибольшей степени проявляются при стечении двух или более согласных (например, смягчение и потеря взрывного сегмента согласной /m/ в слове «отнять - *ат'н'ат'и*»).

Эффекты редукции, коартикуляции и ассимиляции проявляются в основном внутри слоговых комплексов, в меньшей степени на стыках слогов и в ещё меньшей степени, но всё ещё заметной, на стыках слов

внутри синтагмы. Описанные эффекты практически полностью прекращают своё действие только на стыках синтагм и фраз.

Описание фонем в потоке речи в виде последовательности аллофонов может быть осуществлено с различной степенью подробности. В общем случае невозможно дать точную оценку количества аллофонов, т.к. она напрямую зависит от степени детализации учёта влияния позиционных и комбинаторных факторов. При этом общее количество получаемых аллофонов может колебаться от нескольких сотен до нескольких десятков тысяч [30-31].

Рассмотрим два практически опробованных варианта аллофонных наборов, обеспечивающих достаточно подробное описание оттенков фонем в потоке речи, подтверждённое результатами тестирования синтезированной речи [32].

Схема генерации аллофонов гласных фонем по первому варианту представлена на рис. 1.2.

Фонемно-аллофонное преобразование обеспечивает генерацию следующих **позиционных** аллофонов гласных: ударный, частично ударный, безударный первой степени редукции и безударный второй степени редукции. Всего – 4 различных позиционных вариантов. С учётом ***левого контекста*** генерируются следующие **комбинаторные** аллофоны гласных: после паузы, после твёрдых губных, переднеязычных, заднеязычных, после /M/, после /N/, после /L/, после /R/, большинства мягких, после /M'/, после /N'/, после /L'/, после /R'/, после гласных /U/, /O/, /A/, /E/, /Y/, /I/. Всего - 19 левых контекстов. С учётом ***правого контекста*** генерируются следующие **комбинаторные** аллофоны гласных: перед паузой, перед твёрдыми губными, переднеязычными и заднеязычными, перед мягкими. Всего - 5 правых контекстов. Итого, для 5-ти гласных фонем /U, O, A, E, I(Y)/обеспечивается генерация $N_V = 5*4*19*5 = 1900$ аллофонов.

Аллофоны согласных также генерируются с учётом комбинаторных факторов левого и правого контекста (см. рис 1.3).

Левый контекст: после паузы, после глухих согласных, после звонких согласных, после гласных. ***Правый контекст***: перед паузой, перед глухими согласными, перед звонкими согласными, перед безударными гласными, перед ударными гласными. Итого, для 36-ти согласных фонем обеспечивается генерация $N_C = 36*4*5 = 720$ аллофонов.

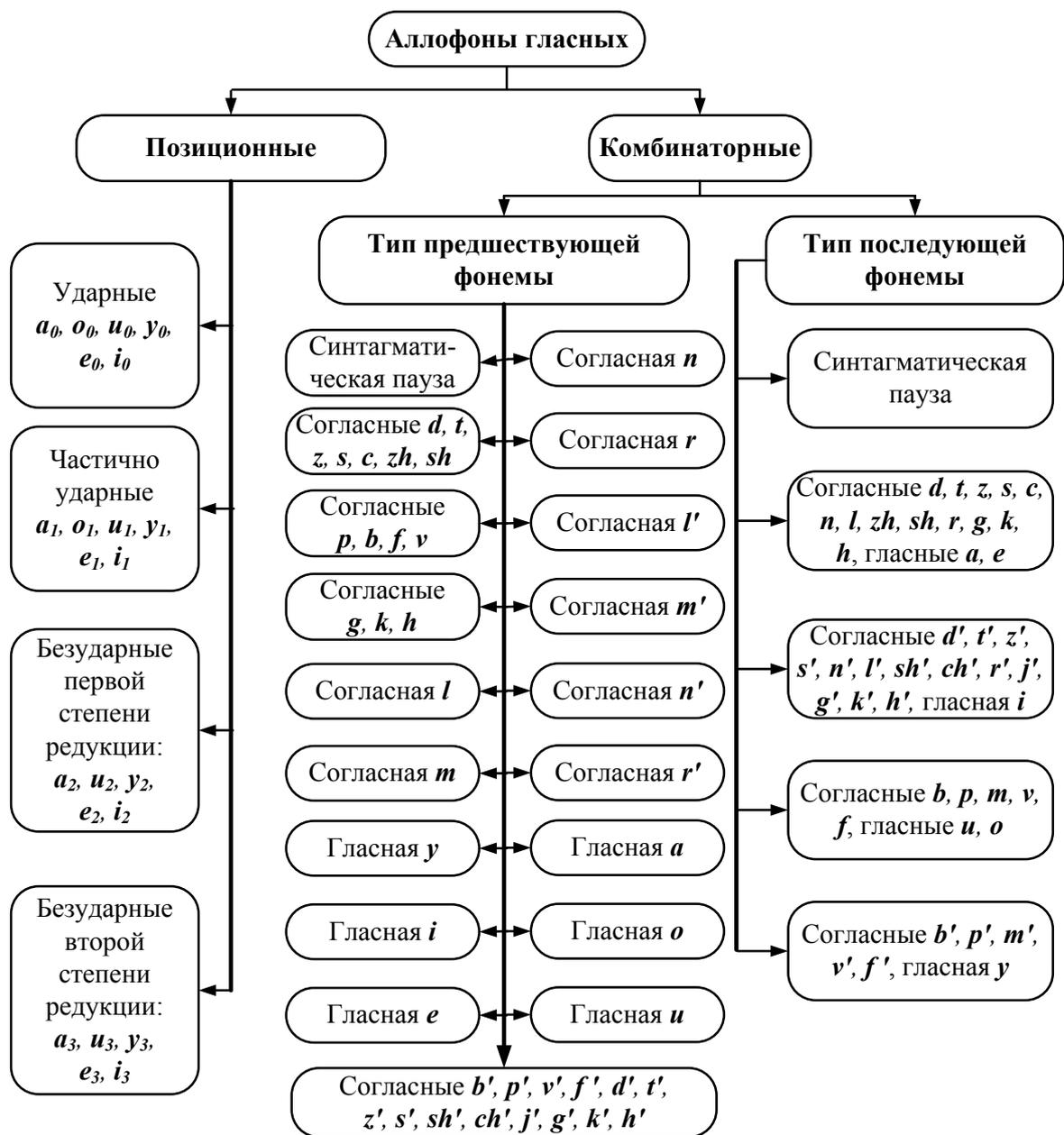


Рис.1.2. Схема образования аллофонов гласных фонем

Полученная оценка общего количества аллофонов $N_v + N_c = 1900 + 720 = 2620$ не претендует, конечно, на абсолютную точность, однако указывает на порядок этого числа и иллюстрирует степень вариативности фонем в речевом потоке. В работе [30] описанное количество аллофонов предложено называть **макси-набором**. Возможны и другие оценки количества аллофонов в зависимости от требуемой точности описания влияния позиционных и комбинаторных факторов при синтезе речи.

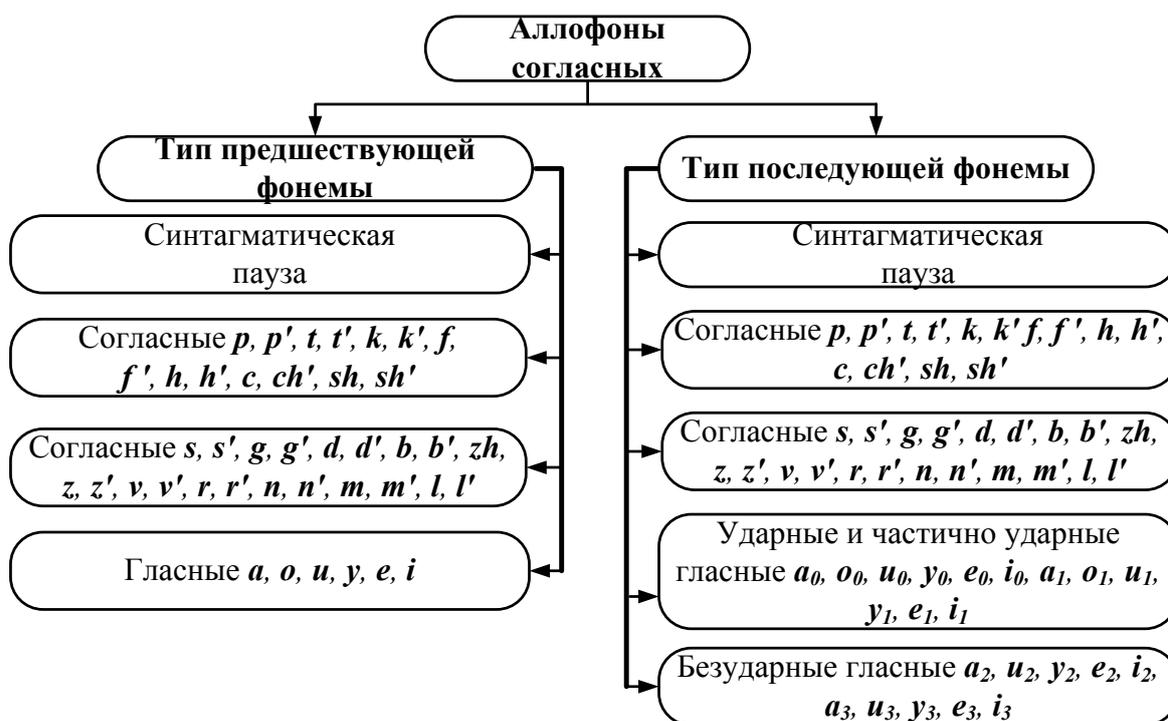


Рис. 1.3. Схема образования аллофонов согласных фонем

В работе [31] кроме *макси-набора* авторами предложен также *мини-набор* аллофонов. При использовании мини-набора для синтеза русской речи создаётся тот же набор позиционных аллофонов, однако набор комбинаторных аллофонов существенно сокращён. С учётом *левого контекста* создаются следующие комбинаторные аллофоны гласных: после синтагматической паузы, после твёрдых губных согласных, после твёрдых передне- и среднеязычных согласных, после твёрдых заднеязычных согласных и гласных, после мягких согласных. Всего 5 левых контекстов. С учётом *правого контекста* создаются следующие комбинаторные аллофоны гласных: перед паузой, перед переднеязычными и заднеязычными твёрдыми согласными и гласными /U/, /O/, /A/, /E/, /Y/, перед губными согласными, перед мягкими согласными и гласной /I/. Итого, для 5-ти типов гласных создаются $N_v = 5 \cdot 4 \cdot 5 \cdot 4 = 400$ аллофонов. Аллофоны согласных создаются только с учётом правого контекста: перед паузой, перед глухими согласными, перед звонкими согласными, перед безударными гласными, перед ударными гласными. Итого, для всех 36-ти согласных создаются $N_c = 36 \cdot 5 = 180$ аллофонов. Всего создаётся: $400 + 180 = 580$ аллофонов русской речи. Предлагаемый мини-набор аллофонов обеспечивает достаточно высокую разборчивость синтезированной речи при относительно невысокой в сравнении с макси-набором её естественности.

Каждый из генерируемых аллофонов обозначается именем фонемы и следующими за ним тремя целочисленными индексами: i, j, k . Индекс i

для гласной указывает её позицию по отношению к словесному ударению и может принимать одно из следующих значений: 0 – полноударный, 1 – частично-ударный, 2 – гласный первой степени редукции, 3 – гласный второй степени редукции. Индекс *i для согласной* указывает удвоение фонемы и может принимать одно из следующих значений: 0 – обычная фонема, 1 – удвоенная фонема. Для *гласных и согласных* индекс *j* указывает группу левого контекста, а индекс *k* – группу правого контекста.

1.4. Особенности преобразования «буква-фонема» (ПБФ) для русского языка

Орфографическое письмо по своей природе консервативно и отражает, как правило, произносительную форму языка в период становления письменности, в отличие от фонематической записи, которая должна отражать норму произношения, сложившуюся к настоящему времени. Поэтому при чтении орфографического текста человек наряду со знанием общих правил преобразования «буква – фонема» использует знания о целом ряде сложившихся исключений, таких как замена «ч» на «ш» в словах «что», «чтобы», устранение непроезжих согласных в словах «солнце», «лестный», замена окончаний прилагательных «-ого», «-его» на «ово», «ево» и др.

Характерно, что в русской орфографии в два раза больше гласных букв, чем гласных фонем, и в два раза меньше согласных букв, чем согласных фонем. Для целого ряда согласных фонем в русском языке нет специальных букв, и фонема определяется несколькими буквами или положением определённой буквы в слове. Например, мягкая согласная фонема может записываться соответствующей буквой и буквой «ь», стоящей за согласной. Мягкость согласной при написании может выражаться также стоящими за согласной гласными буквами «и», «е», «ё», «ю», «я». С последним способом выражения мягкости согласного связан тот факт русской орфографии, что фонемы гласных, кроме «ы» и «и», имеют по два буквенных эквивалента: «у» и «ю», «о» и «ё», «а» и «я», «э» и «е». Фонема «й'» имеет также несколько способов орфографической записи: с помощью специальной буквы «й», с помощью букв «е», «ё», «ю», «я», стоящих в начале слова, после гласного или «ъ», с помощью букв «о» и «и», стоящих после «ь». Для фонемы «ш'» имеется специальная буква «щ», но та же фонема может обозначаться также сочетанием «сч».

В русской орфографии имеются буквы «ь» и «ъ», выполняющие своеобразные функции. Буква «ь» указывает мягкость предшествующей согласной, парной по признаку твёрдость/мягкость. Кроме того, буква «ь» является разделительным знаком: после согласных перед «е», «ё»,

«ю», «я» она указывает на то, что эти буквы читаются как в начале слова. Буква «ъ» также несёт разделительную функцию.

Результатом ПБФ являются цепочки фонем, отображающих каждое слово входного текста. Используется общепринятый набор 42-х фонем русской речи, для обозначения которых приняты символы, приведенные в таблице 1.1.

Процедура ПБФ в целом опирается на известные регулярные правила внутрисловного преобразования буква-фонема [33-35], при этом для каждой текущей буквы текста учитывается её левое и правое окружение. Слова, не подчиняющиеся регулярным правилам преобразования, и их фонемная транскрипция содержатся в списке исключений.

В ходе применения процедуры ПБФ вначале проверяется каждое слово поступившей синтагмы на наличие его в списке исключений и, в случае нахождения, осуществляется его замена на соответствующую цепочку фонем. Если же слово в списке не найдено, к нему применяется регулярные правила ПБФ. Эти правила учитывают категорию букв, основанную на классификации фонем по месту и способу образования (см. табл. 1.1).

При обработке гласных букв (А, О, У, Э, Ы, И, Я, Ё, Ю, Е) используются следующие правила:

– *Если i-я буква гласная*, то проверяем не является ли (i-1)-й знак пробелом, Ъ, Ь или одной из гласных. Если ДА, то гласные Я, Ё, Ю, Е преобразуются, соответственно, в фонемы /J'A, J'O, J'U, J'E/, а гласные А, О, У, Э, Ы, И в фонемы /A, O, U, E, Y, I/.

Примеры: пьяный, подъём, дают, ель, атом, аорта.

– *Если (i-1)-й символ является любой согласной буквой*, то буквы А или Я переходят в фонему /A/, буквы О или Ё – в /O/, буквы У или Ю – в /U/, буквы Э или Е – в /E/, буква Ы – в /Y/ и буква И – в /I/ (кроме случаев, когда (i-1)-я согласная – буква Ц или Ш).

Примеры: пять, дать.

– *Если (i-1)-ый символ – буква Ц или Ш*, то буква И переходит в /Y/.

Примеры: цифра, шило.

– *Все ударные гласные* фонемы отмечаются знаком ударения.

– *Все безударные гласные* фонемы сохраняются, кроме /O/, которая в отсутствие ударения переходит в фонему /A/.

Примеры: молоко.

При обработке согласных, непарных по глухости - звонкости (Й, М, Н, Р, Л, Ц, Ч, Х, Щ) используются следующие правила ассимиляции звуков:

– *Непарные согласные* Й, Ц, Ч, Щ всегда переходят, соответственно, в фонемы /J', C, Ch', Sh'/.

Примеры: майка, цель, часто, щука.

– Если (i+1)-я буква - Ъ, Я, Ё, Ю, Е или И, то буквы М, Н, Р, Л, Х переходят в фонемы /M', N', R', L', H'/.

Примеры: мясо, небо, рюмка, люди, моль, хек.

– Если (i+1)-я буква или знак не являются Ъ, Я, Ё, Ю, Е или И, то буквы М, Н, Р, Л, Х переходят в фонемы /M, N, R, L, H/.

Примеры: мать, нос, рот, код, мох.

– Если i-я буква - Н, (i+1)-я буква - Н, Д, Т, С, а (i+2)-я буква - Ъ, Я, Ё, Ю, Е или И, то Н переходит в фонему /N'/.

Примеры: антихрист, бандюга.

– Если i-я буква и (i+1)-я буква одинаковы, а (i+2)-я буква - Ъ, Я, Ё, Ю, Е или И, то обе буквы переходят в соответствующие мягкие согласные.

Примеры: интеллигент, комиссия.

При обработке согласных, парных по глухости - звонкости (Б,П), (Д,Т), (Г,К), (В,Ф), (З,С), (Ж,Ш) используются следующие правила ассимиляции звуков:

– Если i-я парная согласная буква стоит перед пробелом (), или перед любым знаком препинания, или перед (i +1)-й глухой согласной (П, Т, К, Ф, Ц, Ш, Щ, Ч, Х, Щ), то буквы Б, П переходят в фонему /P/; Д,Т – в /T/; В,Ф – в /F/; З,С – в /S/; Ж, Ш – в /Sh/; Г,К – в /K/), (кроме случая, когда Г стоит перед К).

Примеры: резцы, покровь.

– Если буква Г стоит перед К, то она переходит в фонему /H/.

Примеры: легко, легкий .

– Если (i+1)-я буква – Б, Д, Г, З или Ж, то буквы П, Б переходят в фонему /B/; Д,Т – в /D/; В,Ф – в /V/; З,С, – в /Z/; Ж,Ш – в /Zh/; Г,К – в /G/.

Примеры: отдых, с горы.

– Если (i+1)-я буква – Ъ, А, О, У, Э или Ы, то буквы П, Б, Д, Т, Г, К, В, Ф, З, С, Ж, Ш переходят, соответственно в фонемы /B, P, D, T, G, K, V, F, Z, S, Zh, Sh/.

Примеры: база, садик.

– Если (i+1)-я буква - Я, Ё, Ю, Е или И, то буквы П, Б, Д, Т, Г, К, В, Ф, З, С, Ж, Ш переходят, соответственно в фонемы /B', P', D', T', G', K', V', F', Z', S', Zh, Sh/.

Примеры: день, пятый.

– Если (i+1)-я буква – Б и (i+2)-я буква глухая согласная (П, Т, К, Ф, Ц, Ш, Щ, Ч, Х, Щ) или конец слова, то буквы Б,П переходят в фонему /P'/; Д,Т – в /T'/; Г,К – в /K'/; В,Ф – в /F'/; З,С – в /S'/; Ж,Ш – в /Sh'/.

Примеры: кровь, медь.

– Если (i+1)-я буква – Ъ и (i+2)-я буква Б, Д, Г, З или Ж, то буквы П, Б переходят в фонему /B'/; Д,Т – в /D'/; Г,К – в /G'/; В,Ф – в /V'/; З,С – в /Z'/; Ж,Ш – в /Zh/.

Примеры: *косьба.*

– Если (i+1)-й символ – буква Ъ и (i+2)-й символ – одна из букв Й, М, Н, Р, или Л, то буквы П, Б, Д, Т, Г, К, В, Ф, З, С, Ж, Ш переходят, соответственно в фонемы /P', B', D', T', G', K', V', F', Z', S', Zh, Sh/.

Примеры: *Кузьма, письмо.*

– Если i-я буква – С, (i+1)-я буква – Н, Т или С, а (i+2)-я буква – Ъ, Я, Ё, Ю, Е или И, то С переходит в фонему /S'/.

Примеры: *кость, усни, пустяк, массив.*

– Если i-я буква и (i+1)-я буква одинаковы, а (i+2)-я буква – Ъ, Я, Ё, Ю, Е или И, то обе буквы переходят в соответствующие мягкие согласные.

Примеры: *оттяжка, ассистент.*

1.5. Просодические характеристики русской речи

Просодика играет важную роль при восприятии речи человеком. По интонации определяется коммуникативная направленность высказывания, логический смысл, выделение главного и общего (рема и тема), осуществляется вычленение семантически связанных отрезков речи и объединение речевых элементов внутри этих отрезков. Зачастую в зависимости от интонационного оформления логически идентичные высказывания могут иметь различное семантическое значение.

Основными компонентами просодики являются интонация и ударение. Кроме словесных ударений, в просодике речи рассматриваются также другие виды ударений, которые делают естественную речь более выразительной и, в частности, помогают передавать ее интонацию. Физически интонация и ударение реализуются совокупностью акустических средств (просодических характеристик речи), к числу которых относятся:

– **мелодика** – движение частоты основного тона (F_0);

– **ритмика** – текущее изменение длительности звуков и пауз (T);

– **энергетика** – текущее изменение силы (амплитуды) звука (A).

Синтез речи по тексту предполагает наличие автоматической процедуры формирования текущих контуров мелодии, силы звука, фонемной длительности и длительности пауз на основе анализа определенных свойств входного текста. Это значит, что этап преобразования орфографического текста в последовательность фонем должен сопровождаться выделением информации, необходимой для задания просодических характеристик синтезируемых речевых сигналов.

В фонетике и физиологии речи [19, 12] считается установленным, что минимальной самостоятельной интонационной единицей речи

является синтагма. Самостоятельность её выражается в артикуляторной целостности, семантико-синтаксической связанности ее элементов и в наличии объективных характеристик её вычленения из речевого потока. В печатном тексте несомненной границей синтагмы является любой знак препинания. Однако число синтагм может существенно превосходить количество знаков препинания в тексте. Как правило, синтагма может включать в себя от 1 до 5 слов, причем наиболее частотной является трехсловная синтагма. Конкретное место границы синтагмы выбирается, исходя из оптимального удовлетворения требований семантико-синтаксического, фонетического и физиологического характера. Первое из названных требований предписывает объединение внутри синтагмы семантически связанных слов и запрещает их разъединение на две синтагмы. Второе требование выражает стремление фонетической системы языка к определенным ритмическим конструкциям (например, по 2-3 слова), объединяемым в одну синтагму. И, наконец, третье предписывает формирование синтагмы с числом слов не более того количества, которое можно успеть произнести за один такт выдоха.

Таким образом, с учетом вышеизложенных требований, осуществляется первичный шаг синтагматического чтения – выделение первой, затем второй и т.д. синтагм для каждого предложения текста.

Следующим шагом является определение интонационного типа синтагмы. Основными интонационными типами синтагм принято считать следующие: завершенность, незавершенность, вопрос и восклицание. Количество интонационных подтипов для основных интонационных типов может достигать нескольких десятков [19-24]. Задача определения интонационного подтипа решается, исходя из анализа двух факторов: положения синтагмы в тексте и ее семантического наполнения. При анализе первого фактора учитываются особенности расположения синтагмы относительно ближайших знаков пунктуации, а именно:

- находится ли синтагма непосредственно в начале текста, абзаца, фразы или следует за знаком раздела синтагмы внутри фразы;
- предшествует ли синтагма концу текста, абзаца, фразы или находится внутри фразы перед другой синтагмой;
- за каким конкретным знаком пунктуации следует, и какому знаку предшествует анализируемая синтагма.

При анализе второго фактора учитывается смысловая направленность синтагмы. Определяется её логически выделенный интонационный центр. Выясняется, не является ли синтагма, даже при отсутствии в тексте специальных знаков пунктуации, выражением побуждения, пояснения, переспроса и др.

Окончательное решение об интонационном типе синтагмы выносится с учетом обоих факторов.

После того, как определены границы синтагмы и ее интонационный тип, синтагма должна быть разбита на составляющие ее фонетические слова и акцентные единицы (группы) с одновременным указанием мест словесного и группового ударений.

Фонетическое слово состоит из знаменательной части речи и примыкающих к ней односложных служебных частей речи: предлогов и (или) частиц. Фонетическое слово характеризуется наличием единого словесного ударения, которое, как правило, несет знаменательное слово. Входящее в состав фонетического слова служебное слово может находиться либо перед знаменательным словом – такое безударное слово называется "проклитикой", либо после – такое безударное слово называется "энклитикой".

Положение словесного ударения в знаменательном слове для русского языка свободное (в отличие, например, от французского, где ударным является, как правило, последний слог, или от чешского, где, как правило, ударный слог – первый). Поэтому слова в текстах, предназначенных для синтеза речи, должны маркироваться знаками словесного ударения. Постановка словесных ударений может быть алгоритмизирована. Такие алгоритмы основываются на знании положения ударения в лексемах используемого словаря и на знании правил переноса ударения при переходе к различным словоформам данной лексемы.

Структурно на более высоком уровне, чем фонетическое слово и словесное ударение, находятся понятия акцентной единицы и её группового ударения.

Под акцентной единицей (АЕ) понимается одно или несколько полнозначных слов синтагмы, объединенных общим просодическим (мелодическим, ритмическим и энергетическим) контуром, привязанным к единому групповому ударению. Количество акцентных единиц чаще всего совпадает с числом фонетических слов в синтагме, а сама акцентная единица – с фонетическим словом. Однако синтагма, состоящая из нескольких слов, может включать и только одну акцентную единицу, и две, и три, в зависимости от передаваемого смыслового значения. Например, синтагма-фраза "Саша варил кашу" будет состоять из одной АЕ, если она является ответом на вопрос "*Кто варил кашу?*"; из двух АЕ, если она является ответом на вопрос "*Что делал Саша?*", и, наконец, из трёх АЕ, если она является ответом на вопрос "*Что варил Саша?*".

В первом случае групповое ударение стоит на первом слове, а акцентная единица включает все три слова. Во втором случае - на первом и втором слове, причем вторая АЕ включает два слова. В

третьем случае все три слова несут групповое ударение, а акцентные единицы включают по одному слову.

Выделение слова в акцентной единице делает его словесное ударение сильным (или полным). Другие слова в акцентной единице получают слабое (или частичное) ударение. Чаще всего полное ударение получает первое слово акцентной единицы. Это позволяет сформулировать простое правило установления границ акцентной единицы: левая граница совпадает с началом слова, содержащего полное ударение, правая – с началом слова следующей акцентной единицы.

Акцентные единицы ранжируются относительно границ синтагмы на конечные, начальные и срединные. Эти категории акцентных единиц вносят различный по значимости вклад в формирование просодического контура синтагмы. Основное разнообразие мелодических контуров реализуется на конечной акцентной единице, чаще всего несущей и логическое ударение, существенно меньшее разнообразие – на начальной и срединной. Мелодические, ритмические и энергетические характеристики акцентных единиц являются теми минимальными элементами, из которых складывается интонация синтагмы, фразы и текста в целом.

В основе алгоритмов мелодического, ритмического и энергетического оформления речи, рассматриваемых в книге, лежит предложенная одним из авторов книги модель интонационных портретов акцентных единиц (ПАЕ-модель) [36, 37], более подробно рассматриваемая в последующих главах.

Как уже указывалось, акцентная единица является той минимальной единицей, на которой задаются интонационные характеристики и из совокупности которых складывается интонация синтагмы в целом. В литературе описано большое количество разнообразных интонационных типов синтагмы [17, 19, 21, 23, 24]. Практически совершенно необходимым для синтеза речи по тексту является изучение тех интонационных типов, которые связаны с грамматической (синтаксической) функцией интонации. Среди них, как уже было сказано выше, наиболее важными являются интонация завершенности, незавершенности, вопроса и восклицания (побуждения).

Интонация синтагмы складывается последовательным соединением начальной, срединной и конечной акцентных единиц, на каждой из которых в соответствии с требуемым интонационным типом формируются необходимые контуры мелодики, ритмики и энергетика. При наличии в синтагме более 3-х АЕ добавляется необходимое количество срединных единиц. При наличии менее 3-х АЕ исключается вначале срединная, а затем – начальная акцентные единицы.

Мелодика, ритмика и энергетика каждой акцентной единицы задаются нормированными значениями частоты, длительности и

интенсивности на трех ее участках: ядре, предьядре и заядре. Ядром акцентной единицы является ударный слог, отмеченный знаком группового ударения. Предьядром и заядром – соответственно предшествующие ему и следующие за ним фонемы акцентной единицы.

Обобщённые мелодические портреты одноакцентной синтагмы для вопросительного, восклицательного, незавершённого и завершённого типов интонации представлены на рис. 1.4. Кривые для каждой АЕ изображены в нормированных координатах «время-частота». Интервал нормированного времени $[0, \frac{1}{3}]$ соответствует предьядру, $[\frac{1}{3}, \frac{2}{3}]$ – ядру, и интервал $[\frac{2}{3}, 1]$ – заядру. Интервал нормированной частоты основного тона $[0, \frac{1}{3}]$ соответствует её низкому уровню, $[\frac{1}{3}, \frac{2}{3}]$ – среднему, и $[\frac{2}{3}, 1]$ – высокому.

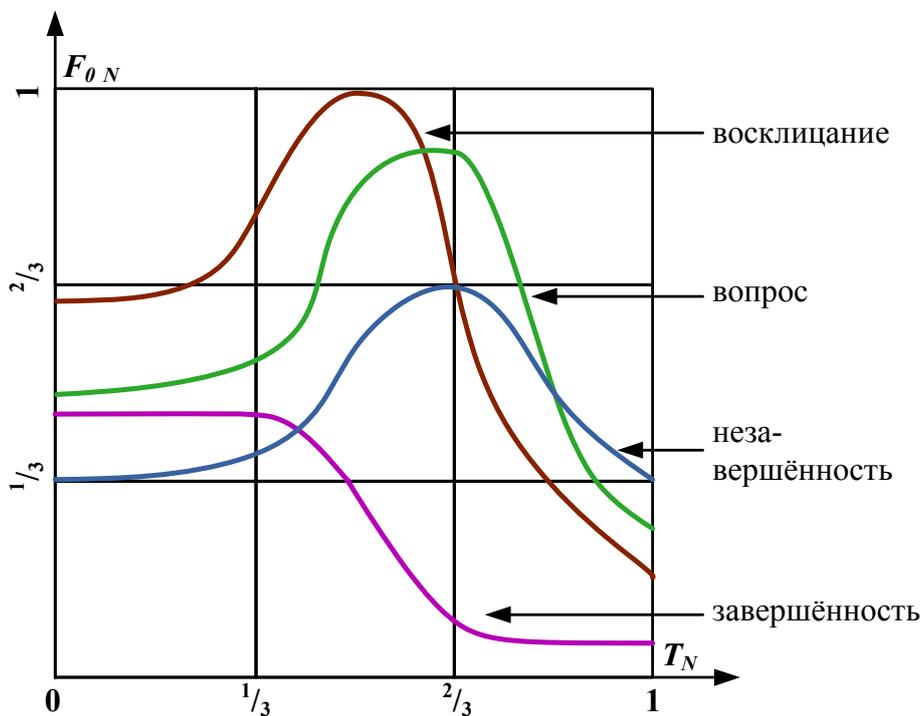


Рис. 1.4. Обобщённые мелодические портреты синтагм четырёх интонационных типов

На рис. 1.5 представлен набор обобщённый мелодический портрет синтагмы вопросительного типа, состоящей из трёх АЕ: начальной (рис. 1.5а), серединной (рис. 1.5б) и конечной (рис. 1.5в).

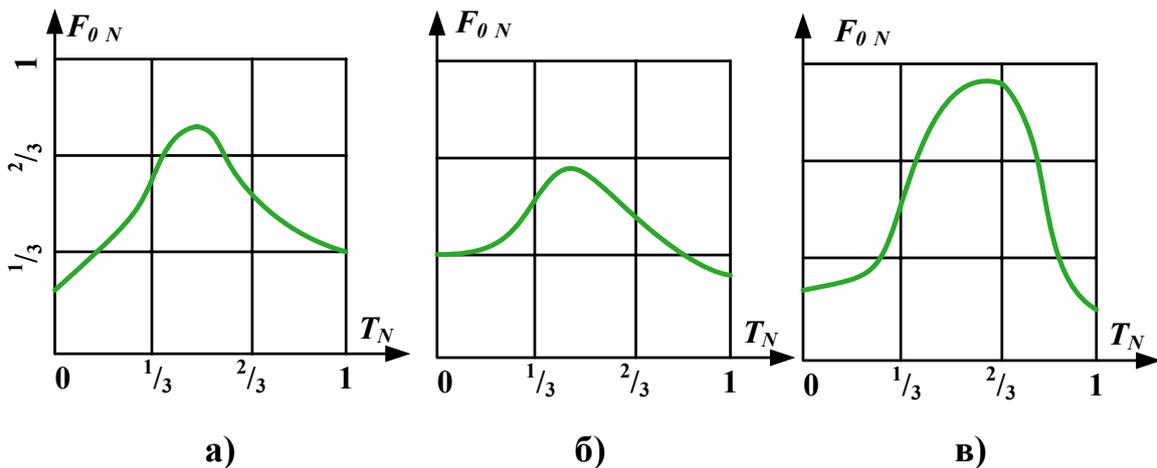


Рис. 1.5. Мелодический портрет синтагмы вопросительного типа: а) начальная АЕ, б) срединная АЕ, в) конечная АЕ

Обобщённый вид ритмического и энергетического портретов конечной АЕ синтагмы вопросительного типа приведены на рис. 1.6 а, б. Эти кривые показывают, что ударные гласные (ядра) выделяются большей длительностью и силой звука. Предударные звуки по мере удаления от ударного гласного постепенно укорачиваются и ослабляются, а затем несколько удлиняются и усиливаются. Заударные звуки постепенно ослабляются, а их длительность по мере удаления от ядра АЕ уменьшается, а затем увеличивается.

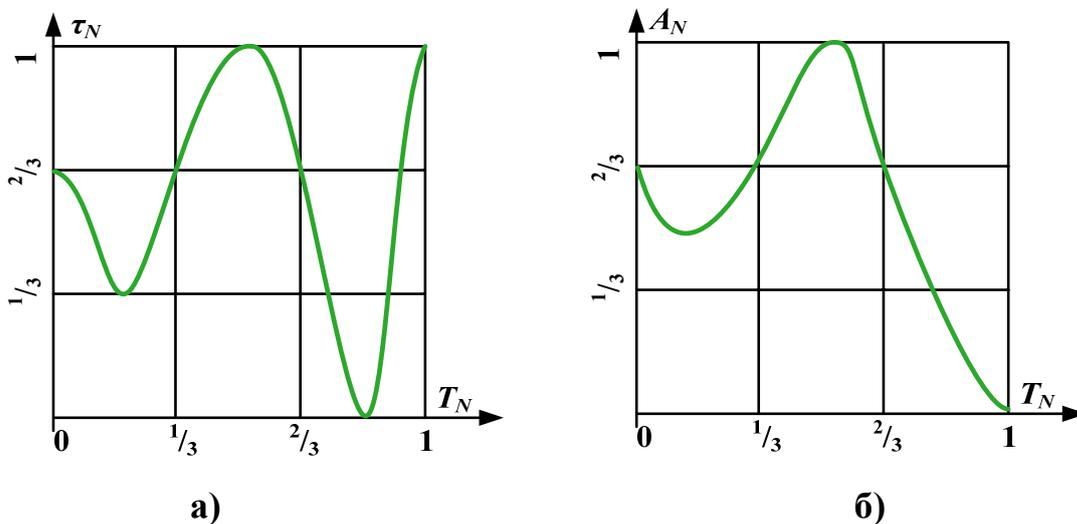


Рис 1.6. Обобщённые просодические портреты конечной АЕ синтагмы вопросительного типа: а) ритмический портрет, б) энергетический портрет

Поясним в заключение смысл введенного понятия нормированного времени. Это понятие (см. рис. 1.4) отражает при восприятии интонации субъективное равенство длительности предъядра, ядра и заядра, хотя в состав ядра всегда входит лишь одна фонема – ударная гласная, а

предъядро и заядро чаще всего включает несколько фонем. На акустическом уровне это означает, что при последующем переходе к реальному времени кривые на этих участках подвергаются модификации в зависимости от конкретных значений длительности фонем, входящих в каждую из зон, т.е. мелодические кривые предъядерной и заядерной зонах должны быть растянуты или сжаты в зависимости от числа фонем в этих зонах.

1.6. Акустические характеристики речи

Основы акустической теории речеобразования достаточно подробно изложены в монографиях Г. Фанта [15] и Д. Фланагана [14]. Схематическое изображение генерации речи человеком приведено на рис. 1.7.

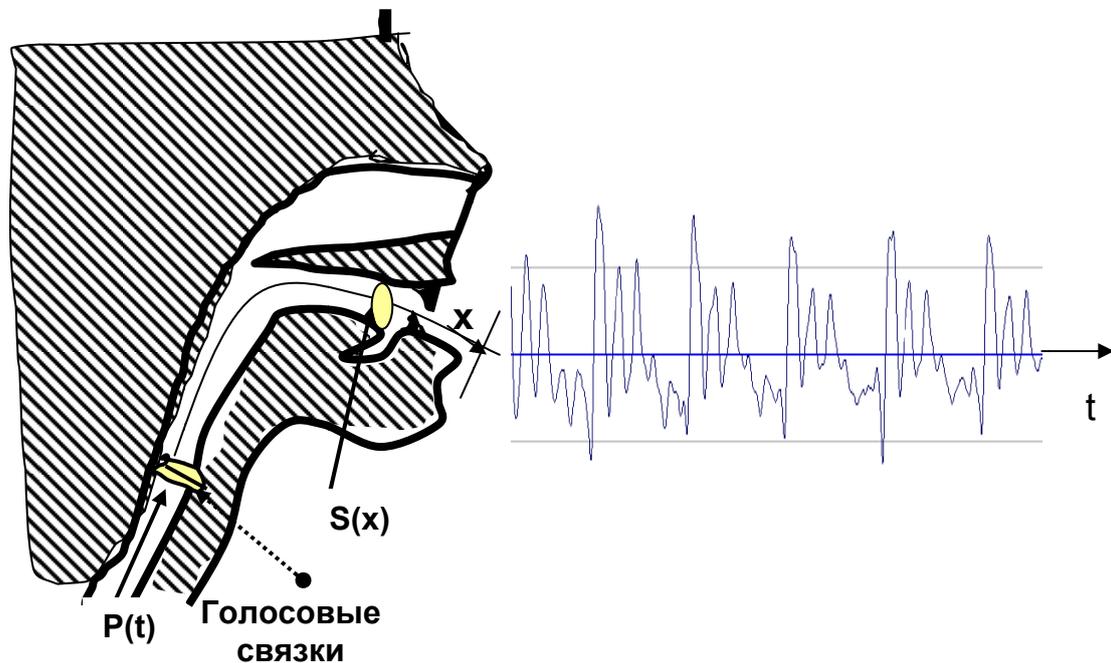


Рис. 1.7. Схематическое изображение генерации речи

Два параллельных канала (ротовой и носовой) образуют единую акустическую систему, возбуждаемую колебаниями голосовых связок, находящихся у основания глотки, либо турбулентным шумом, возникающим в месте сужения ротового канала. Энергия возбуждения создается за счет легочного усилия, затрачиваемого на создание избыточного давления в речевом аппарате. В процессе речеобразования под действием управляющих команд состояние активных артикуляторных органов непрерывно меняется, изменяя конфигурацию всей акустической системы. Как следствие, изменяются её резонансные свойства, т.е. частотная характеристика речевого (голосового) тракта.

Импульсы возбуждения, создаваемые колебаниями голосовых связок, проходя через речевой тракт, трансформируются в речевой сигнал.

Распространение акустических волн в такой системе не удается описать точными математическими методами. Однако при малых потерях и при условии, что длина возбуждаемых волн велика по сравнению с поперечными размерами тракта, а также в случае отсутствия резких изгибов в профиле речевого тракта, распространение акустических волн может быть описано уравнением Вебстера:

$$\frac{1}{S(x)} * \frac{\partial}{\partial x} \left(S(x) \frac{\partial p}{\partial x} \right) = \frac{1}{C^2} * \frac{\partial^2 p}{\partial t^2} \quad (1.3)$$

В этом уравнении, имеющем в качестве аргумента x – расстояние до голосовых связок, $S(x)$ является текущей площадью сечения тракта, нормального к направлению распространения волны, $p(t)$ – переменное воздушное давление, создаваемое голосовыми связками, C - скорость распространения звука в воздухе и t – время. Уравнение Вебстера можно проинтегрировать численным методом и рассчитать передаточную функцию речевого тракта. Однако практически непрерывное сечение речевого тракта выгоднее представить последовательностью коротких цилиндрических труб постоянного сечения.

Для однотрубной модели (что подходит для описания нейтрального звука /Э/) максимумы передаточной функции появляются на частотах

$$\omega_n = \pm(2n + 1) \frac{\pi c}{2l}, n = 0,1,2, \dots \quad (1.4)$$

Эти максимумы называются формантами. Если принять $l = 17$ см (длина голосового тракта взрослого мужчины), то из (1.4) получим, что частоты формант равны: $F_1 = 500$ Гц, $F_2 = 1500$ Гц, $F_3 = 2500$ Гц и т.д. Эти значения формант близки к значениям, полученным экспериментально для реального гласного.

При возбуждении акустической трубы переменным воздушным давлением, создаваемым голосовыми связками, на её выходе появляется речевой сигнал. На осциллограмме речевого сигнала (рис. 1.8) обнаруживаются два типа колебаний: колебания с частотой основного тона – F_0 – и дополнительные формантные колебания с частотами F_i .

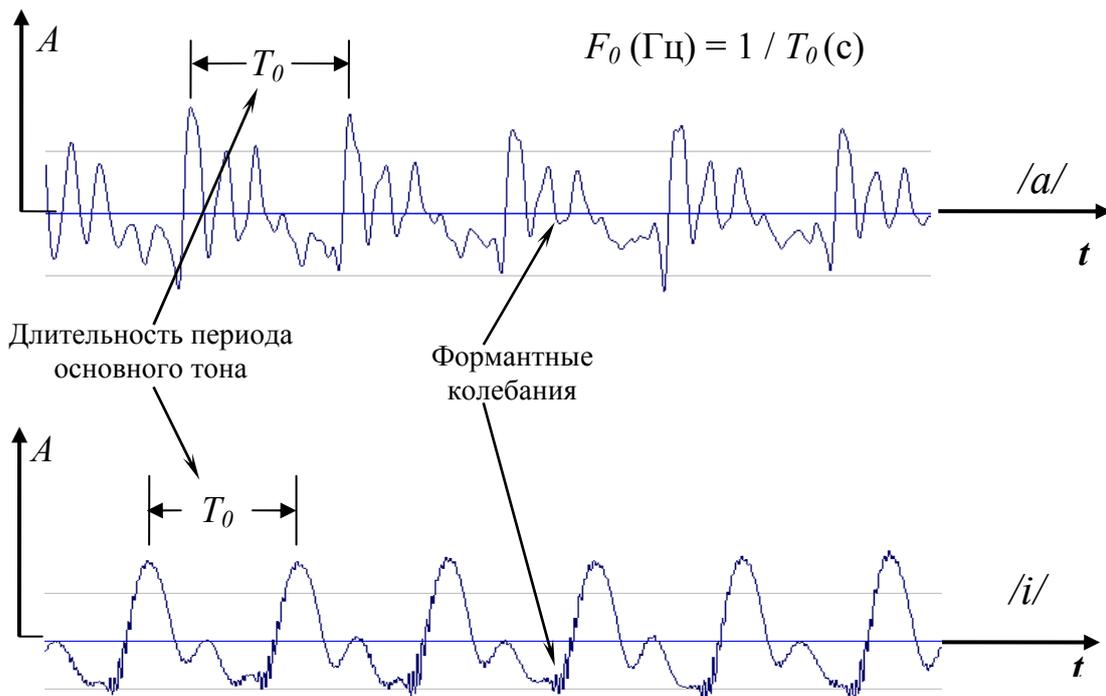


Рис. 1.8. Осциллограмма речевого сигнала для звуков /а/, /и/.

Речевой сигнал $s(t)$ можно представить как свертку возбуждающего сигнала $e(t)$ и импульсной характеристики голосового тракта $v(t)$. Данная модель формирования речевого сигнала во временной области, с примером для звонкого звука, представлена на рисунке 1.9.

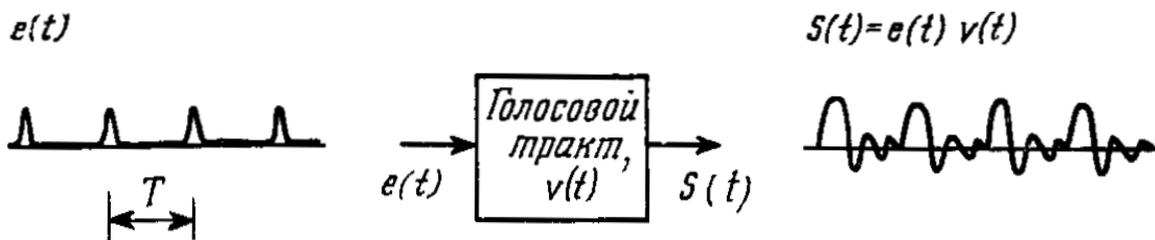


Рис. 1.9. Модель формирования речевого сигнала во временной области

Модель формирования речевого сигнала в частотной области представлена на рисунке 1.10. Спектр речевого сигнала – это произведение спектра Фурье возбуждающей функции (импульсов основного тона) и комплексной частотной характеристики голосового тракта. Последовательности импульсов с периодом T соответствует линейчатый спектр с интервалом между соседними линиями $F_0 = 1/T$.

Частотная характеристика речевого тракта является гладкой функцией частоты, что обусловлено физической структурой голосового тракта, обладающего акустическими резонансами, называемыми формантами.

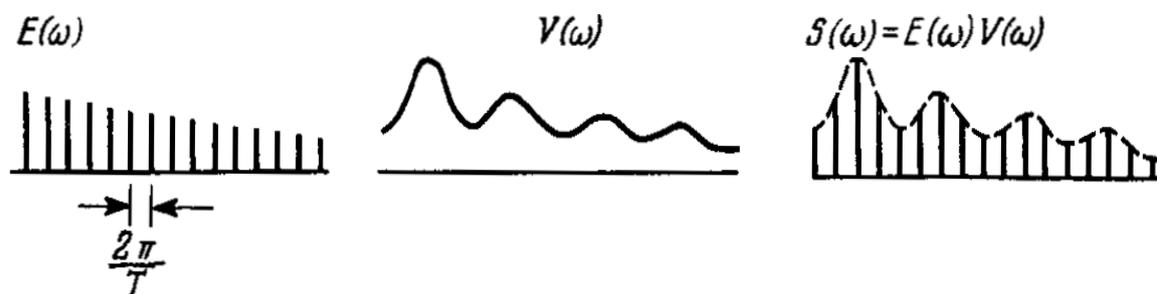


Рис. 1.10. Модель формирования речи в частотной области

Результирующий спектр речевого сигнала является произведением линейчатого спектра возбуждающего сигнала и частотной характеристики голосового тракта, в результате чего спектр речевого сигнала так же является линейчатым, с огибающей, характеризующей передаточную функцию голосового тракта.

На рис. 1.11 представлены двумерные (в координатах «частота – амплитуда») спектрограммы для звуков /a/, /i/. На спектрограммах показаны положения частоты основного тона – F_0 – и частоты формант – F_1, F_2, F_3 .

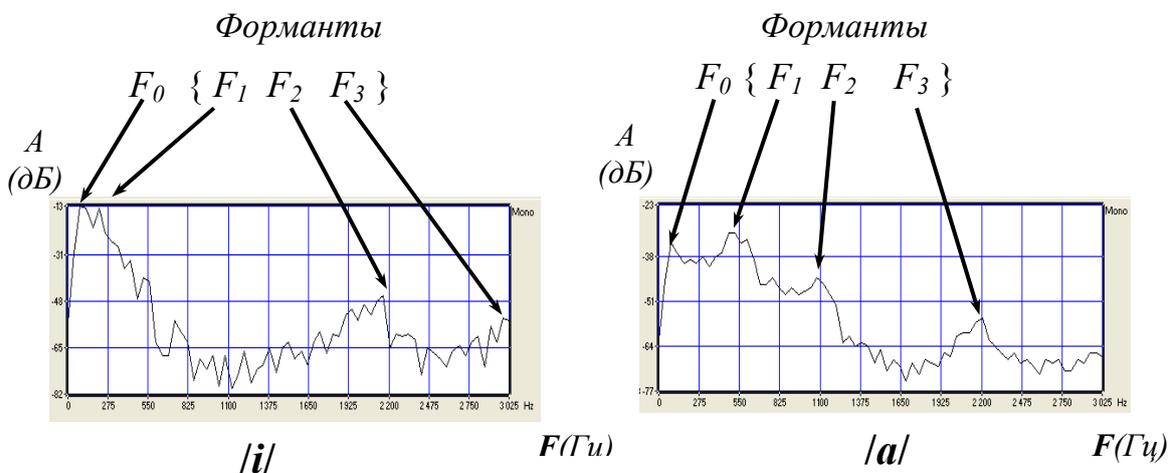
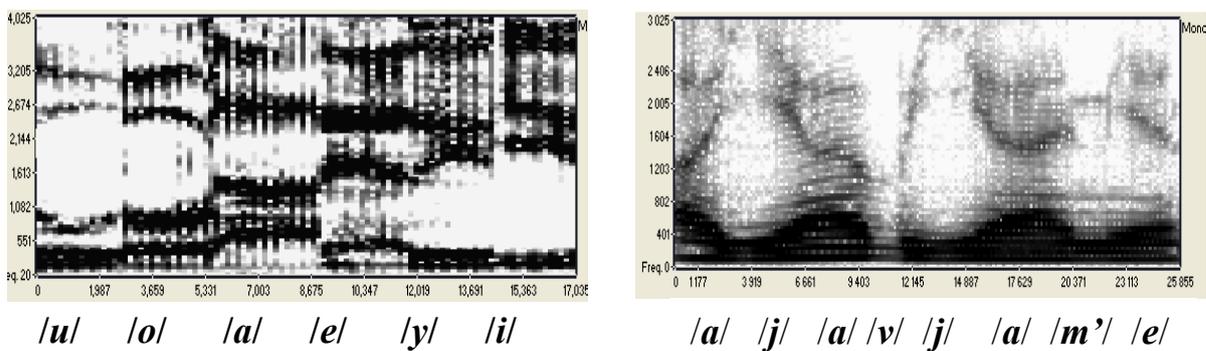


Рис. 1.11. Двумерные спектрограммы для звуков /a/, /i/

Для отображения временной динамики изменения спектральных характеристик используется трёхмерное отображение (т.н. сонограмма), на которой по оси абсцисс отображается время – t , по оси ординат частота – f , а амплитуды спектральных составляющих отображаются степенью почернения. Примеры сонограмм приведены на рис. 1.12. На сонограммах хорошо прослеживаются движения первых 3-х формант.



а) б)
Рис. 1.12. Примеры сонограмм: а) слитно произнесённая последовательность гласных, б) фразы «А я в яме».

Различают широкополосную и узкополосную сонограммы. Широкополосная сонограмма имеет лучшее, чем узкополосная сонограмма, разрешение по времени, что позволяет наблюдать периоды возбуждения в речевом сигнале. В то же время узкополосная сонограмма позволяет наблюдать частотные гармоники возбуждающего сигнала, которые не видны на широкополосной сонограмме. Осциллограмма фразы “Катя уехала”, а так же ее узкополосная и широкополосная сонограммы представлены на рисунке 1.13.

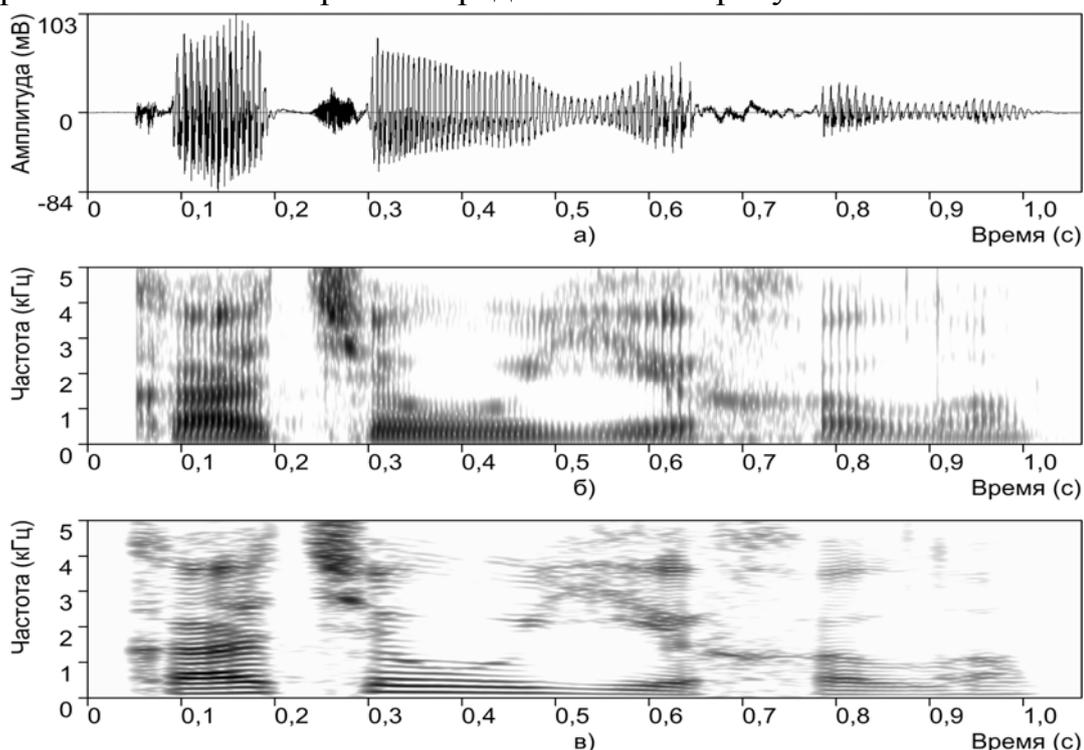


Рис. 1.13. Спектрограммы фразы “Катя уехала”:
а) осциллограмма; б) широкополосная спектрограмма;
в) узкополосная спектрограмма

1.7. Информационная и модуляционная структура речевого сигнала

Общепризнано, что речевой сигнал является по своей природе полиинформативным, что проявляется в многообразии типов информации, передаваемой с помощью речи. Информация, заключенная в речевом сигнале, может быть разделена на два вида: основную, заключающуюся в передаче смыслового содержания речи, и дополнительную, заключающуюся в передаче индивидуальных особенностей говорящего, его физического и эмоционального состояния, а также характеристик передающей среды. В результате вид спектральных характеристик речи зависит не только от произносимого звука, как это видно из рис. 1.12, 1.13, но и от ряда индивидуальных факторов. Это наглядно представлено на рис. 1.14, где показаны сонограммы одного и того же звука /А/, произнесённого 7-ю дикторами.

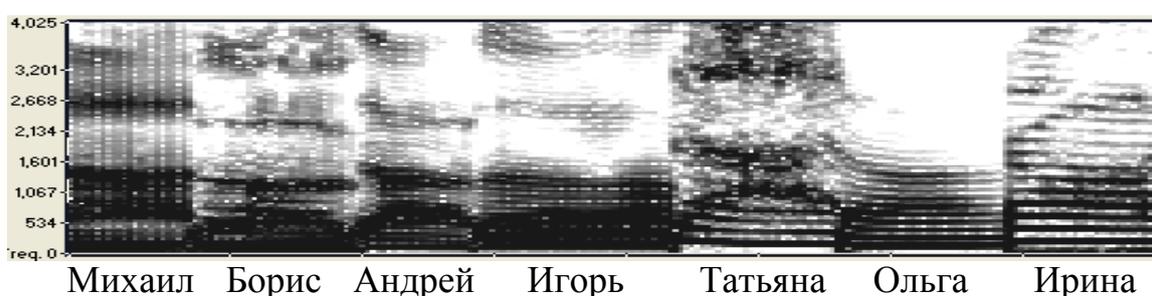


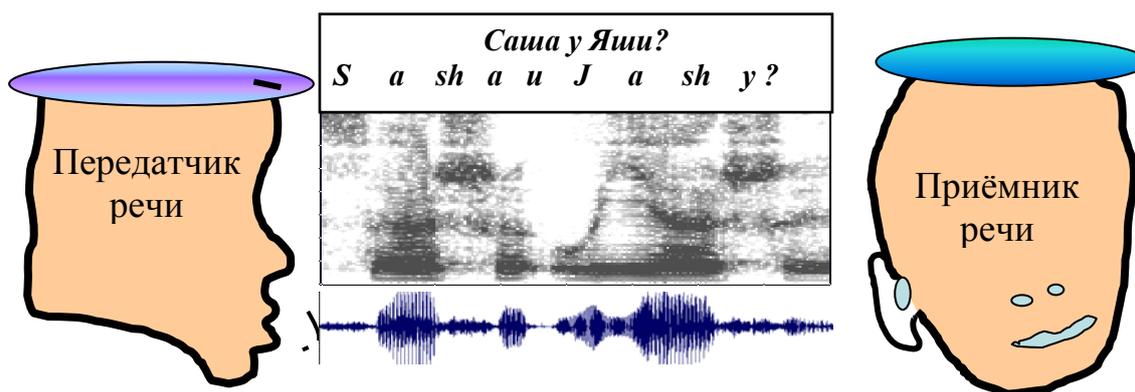
Рис. 1.14. Сонограммы звука /А/, произнесённого семью дикторами

Во временной области речевой сигнал имеет многоуровневую структуру, состоящую, как это показано на рис. 1.15, из непрерывно следующих друг за другом сегментов различного уровня: фразы, слова, фонемы, пичи (периоды основного тона).



**Рис. 1.15. Пример многоуровневой сегментации речевого сигнала:
а) фраз на слова, б) слова на фонемы, в) вокализованные участки на
ПИТЧИ**

Сложную информационную и модуляционную структуру речевого сигнала можно отчётливо выявить, если рассматривать как сигнал в специфической биосистеме речевой связи «человек – человек» (рис. 1.16). При таком подходе речевой сигнал рассматривается не в отрыве от всего процесса общения, а как следствие речевой связи между двумя и более говорящими.



**Рис. 1.16. Схематическое изображение биосистемы речевой связи
«человек – человек»**

Как известно из теории связи, для передачи любых сообщений требуется вполне определенная структура системы связи, которая

предполагает наличие источников информации, модуляторов и переносчиков информации. Основной особенностью речевого сигнала, вытекающей из такого рассмотрения, является его полиинформативность и полимодуляционность.

Рассмотрим более подробно структуру передатчика в речевой системе связи. В передатчике определим схему формирования информационной структуры речевого сигнала (рис. 1.17).

Сначала выделим и уточним самые важные виды информации, которые передаются по каналу связи:

- смысловая информация (в процессе общения человек пытается передать информацию, содержащую некоторый смысл);
- индивидуальность голоса и речи (слушая речь, мы получаем информацию не только о том, что говорится, но и том, кто говорит);
- эмоциональное и физическое состояние человека (по речи мы можем определить, человек взволнован или спокоен, охрип или с заложенным носом);
- информация о характеристиках среды и каналах передачи (на слух мы можем определить, используется ли, например, высококачественное вещание или это передача по узкополосному телефонному каналу).

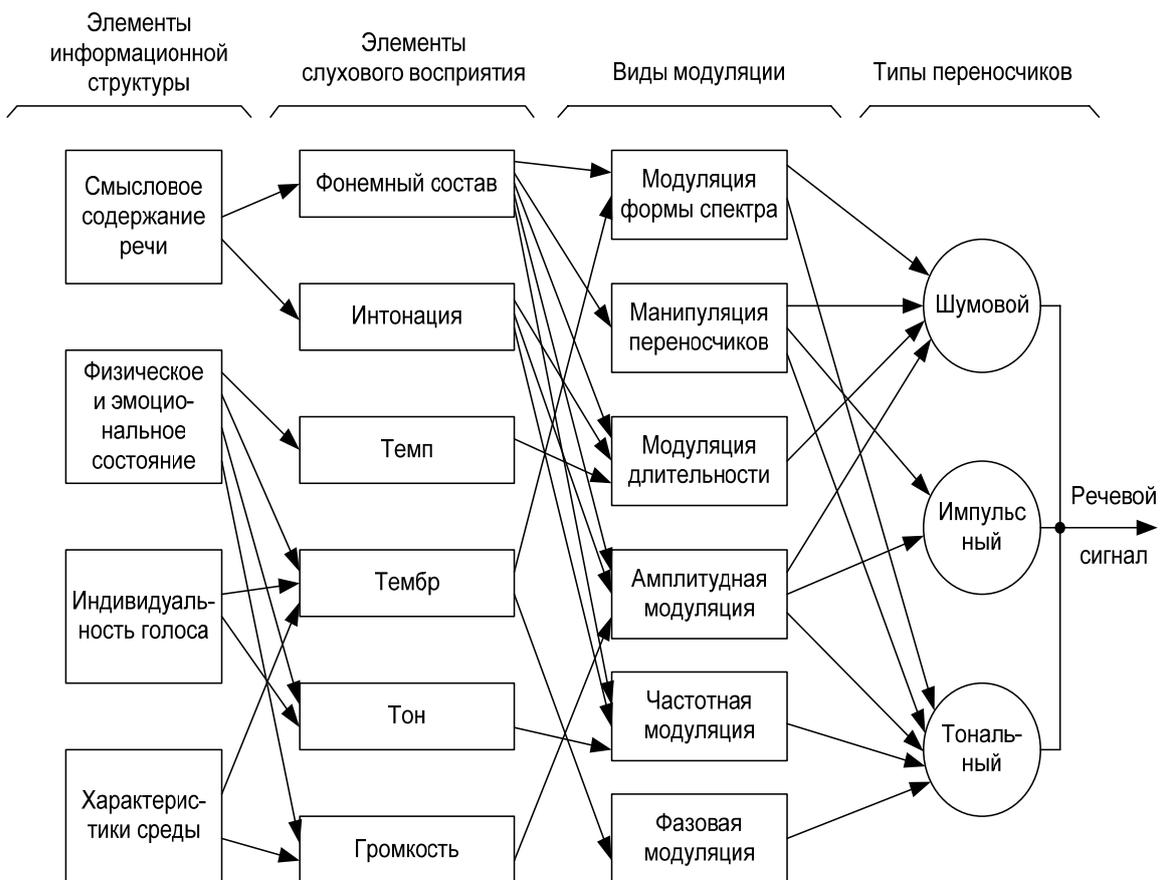


Рис. 1.17. Структура передатчика в речевой системе связи

Указанные виды информации связаны со слуховым восприятием. Смысл речевого высказывания передается с помощью фонетики и, в меньшей степени, просодики. Канал связи меняет тембр звука, вносит искажения и помехи.

Определим элементы слухового восприятия:

- фонетический состав;
- просодика (интонация);
- тембр звука;
- громкость звука;
- высота голоса;
- темп речи;
- искажения и помехи.

В «речевой» системе связи имеется, по крайней мере, 3 типа переносчика информации:

- тональный, с помощью которого передается тон (колебания голосовых связок) – это периодический переносчик;
- шумовой, переносящий шум в речи;
- импульсный, с помощью которого передаются импульсы (взрывы) в речи.

Рассмотрим теперь физику работы описанной схемы формирования информационной структуры речи.

Генерирование речевого сигнала связано с функционированием трех источников звука – переносчиков: тонального, шумового и импульсного (рис. 1.17). Передача информации осуществляется путем модуляции некоторых параметров α , β , γ , ... каждого из указанных переносчиков.

Тональный переносчик в отсутствие модуляции может быть описан как периодическое колебание $y(t)$, создаваемое движением голосовых связок, со спектром:

$$y(t) = A \sum_{k=1}^n a_k \cos(k\omega_1 t + \varphi_k) \quad (1.5)$$

где A – среднеквадратическое значение амплитуд спектральных составляющих; a_k – нормированные амплитуды k -х гармоник; ω_1 – частота первой гармоники; φ_k – фазовые сдвиги k -х гармоник; n – число гармоник.

В процессе речеобразования модулируемыми параметрами могут быть амплитуда A – амплитудная модуляция, $\{a_k\}$ – модуляция формы спектра и фаза $\{\varphi_k\}$ – фазовая модуляция спектра.

Шумовой переносчик в отсутствие модуляции может быть описан как гауссовский стационарный процесс, исчерпывающей

характеристикой которого является его функция корреляции или спектр мощности:

$$R(\tau) = P \sum_{k=1}^n d_k \cos k\Omega\tau \quad (1.6)$$

где $\Omega = \pi/T$; T – интервал корреляции; d_k – коэффициенты разложения нормированной функции корреляции в ряд Фурье; P – мощность случайного процесса.

В процессе речеобразования основными модулируемыми параметрами являются P – амплитудная модуляция и $\{d_k\}$ – модуляция формы спектра.

Импульсный переносчик характеризуется резким нарастанием скорости воздушного потока с дальнейшим экспоненциальным его спадом при раскрытии смычки у взрывных звуков:

$$I(t) = \begin{cases} U \exp[\alpha(t - \tau_0)] & \text{при } t > \tau_0 \\ U & \text{при } t \leq \tau_0 \end{cases} \quad (1.7)$$

Основным модулирующим параметром является U – амплитудная модуляция.

Кроме указанных видов модуляции целесообразно выделить в отдельный класс модуляцию длительности звуков и манипуляцию переносчиков как предельный случай амплитудной модуляции, характеризуемой дискретным процессом их независимого включения или выключения.

Как видно из рис. 1.17, наибольшее число видов модуляции, а, следовательно, и наибольшую информационную емкость, имеет тональный переносчик, а наименьшую – импульсный. Переносчики включаются или выключаются в связи с фонетическим составом сообщений. Каждый из переносчиков может менять свои параметры, т.е. может происходить:

– модуляция формы спектра или спектральная модуляция – несет наибольшую нагрузку и моделирует все источники (коэффициенты a_k и b_k), она модулирует и тональный переносчик.

– манипуляция переносчиков (включение / выключение переносчиков, связана с фонетическим составом сообщений);

– модуляция длительности звуков – это просодика, фонетика, темп.

– амплитудная модуляция – ей подвергаются все типы переносчиков. Амплитуда играет важную роль при передаче фонетической и просодической информации;

– частотная модуляция (модуляция частоты основного тона) – участвует в передаче просодической информации: передает

эмоциональное состояние человека через высоту его голоса, воздействует только на тональный переносчик.

Ни один из видов модуляции не передаётся одним источником. Почти все типы модуляции используются для передачи нескольких источников информации.

Роль каждого вида модуляции в передаче элементов информационной структуры речевого сигнала поясним на примере передачи информации о фонемном составе и интонации речи.

При передаче информации о фонемном составе речи осуществляется непрерывный процесс перестройки речевого тракта. Это приводит к непрерывному изменению формы мгновенных спектров речевого сигнала – модуляция формы спектров – и его средней мощности – амплитудная модуляция. При перестройке речевого тракта осуществляется также фазовая модуляция, хотя полной ясности ее роли в восприятии речевых звуков нет. Кроме того, процесс передачи информации о фонемном составе связан с постоянной сменой комбинаций включения переносчиков – манипуляция переносчиков – и с изменением частоты основного тона на смычках звонких взрывных звуков – частотная модуляция. Информация об интонационной окраске речи, а также о физическом и эмоциональном состоянии, об индивидуальных особенностях голоса и характеристике электроакустической среды передается в основном путем модуляций частоты основного тона – частотная модуляция – и общего уровня звуков – амплитудная модуляция.

Рассмотренная схема передающего устройства речевой системы связи не претендует, конечно, на полноту охвата всех возможных видов передаваемой и получаемой информации, а также всего разнообразия видов модуляции. На схеме рис. 1.17 не указана такая важная операция, непосредственно предшествующая модуляции переносчиков, как кодирование передаваемых сообщений. Однако в целом из рассмотрения схемы на рис. 1.17 можно сделать несколько важных выводов:

- в процессе речевого общения осуществляется параллельная передача различных видов информации;

- для передачи информации каждого вида одновременно используются несколько видов модуляции;

- ни один из видов модуляции (кроме манипуляции переносчиков) не используется для передачи только одного вида информации.

Глава 2

Методы синтеза фонетико-акустических характеристик речи

2.1. Методы персонализированного синтеза речевого сигнала

Различные подходы к синтезу речи по тексту достаточно подробно описаны в [38-40]. В данной главе рассматриваются принципиальные различия в методах обработки текста и генерации речевого сигнала, важные с точки зрения передачи индивидуальности голоса и речи.

В общем виде структурная схема синтезатора речи по тексту представлена на рис. 2.1.

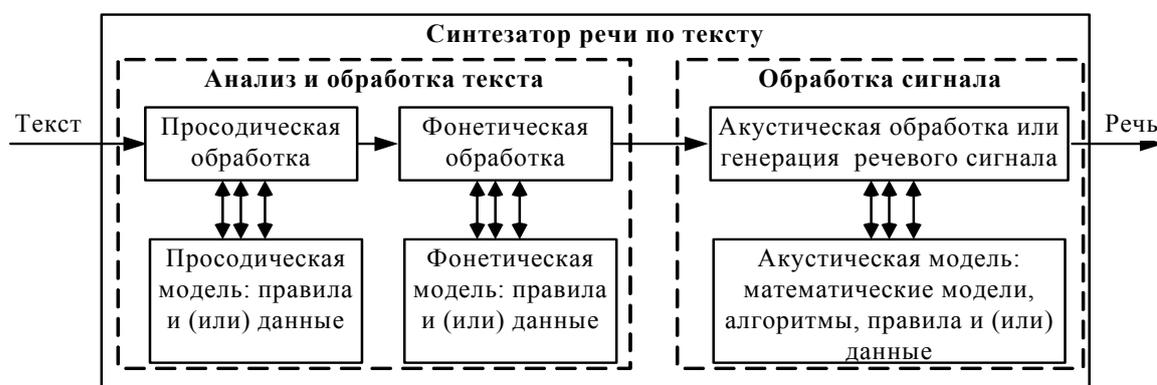


Рис 2.1. Общая структура синтезатора речи по тексту

Анализ и обработка текста включает просодические и фонетические преобразования, в результате которых генерируется последовательность фонетических элементов с маркерами значений просодических параметров: частоты основного тона F_0 , амплитуды A и длительности звуков T .

Полученная последовательность поступает на вход блока обработки сигнала, который на её основе генерирует речевой сигнал, обладающий заданными фонетическими и просодическими характеристиками.

Как качество синтезируемой речи, так и точность передачи индивидуальных характеристик голоса напрямую зависят от акустической модели, используемой блоком генерации сигнала.

Существует несколько подходов к генерации речевого сигнала в системах синтеза речи по тексту (СРТ-системах), использующих следующие методы [39]:

- артикуляторный, или бионический, метод;
- формантный метод;
- компиляционный, или конкатенативный, метод;
- корпусный метод.

2.1.1. Основные характеристики артикуляторного метода синтеза речи по тексту

Исторически первым получил развитие артикуляторный метод [40, 41], использующий физическую модель, основанную на детальном описании физиологии речеобразования. При этом речевой аппарат представляется как акустическая система, состоящая из последовательности коротких цилиндрических труб переменного сечения, возбуждаемая голосовым и (или) шумовым источниками. При генерации речевого сигнала характеристики речевого аппарата представляются электро-акустическими аналогами [14]. Общая структурная схема артикуляторного синтезатора речи по тексту представлена на рис. 2.2.

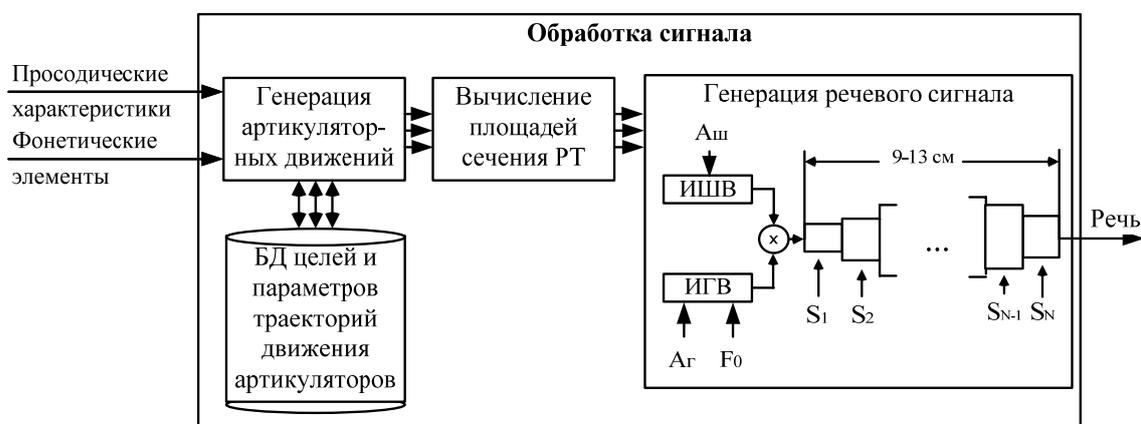


Рис 2.2. Схема артикуляторного синтезатора речи по тексту

Последовательность фонетических элементов с заданными просодическими параметрами подаётся в блок генерации артикуляторных движений, который извлекает из БД соответствующие значения целей и параметров траекторий движения артикуляторов. Сгенерированные значения функций движения артикуляторов передаются в следующий блок, который, используя геометрическую модель речевого тракта (РТ), вычисляет текущие значения площадей сечений секций РТ - S_i , где $1 \leq i \leq N$ [42].

Количество секций N равно отношению длины моделируемого речевого тракта к длине одной секции, которая может варьироваться от 0,2 до 1 см. Средняя длина речевого тракта у женщин составляет в 9-11 см, у мужчин – 11-13 см) Последовательность значений S_i , а также параметры сигналов источников голосового (A_g , F_0) и шумового ($A_{ш}$) возбуждения (ИГВ и ИШВ) передаются в блок генерации речевого сигнала. Каждая секция РТ может быть представлена эквивалентным Т-образным электрическим звеном. Аналоговая реализация речевого тракта [15] содержит 45 таких звеньев. Для представления речевого сигнала в терминах геометрических размеров речевого тракта цифровым

методом используются так называемые коэффициенты отражения - k_i [43]. Эти величины можно интерпретировать как коэффициенты частичной корреляции между отсчётами сигнала x_n и x_{n+1} . В работе [43] показано, что в случае акустической трубы с одинаковыми по длине звеньями площади сечения звеньев связаны с коэффициентами отражения формулой

$$\frac{s_i}{s_{i+1}} = \frac{1 + k_i}{1 - k_i} \quad (2.1)$$

В этой же работе предложен метод реализации синтезатора речевых сигналов в виде цифрового фильтра лестничной формы, описываемого коэффициентами отражения $\{k_i\}$.

Персонализированная артикуляторная модель речеобразования может быть задана базой данных целей и параметров траекторий движений артикуляторов. Для создания такой БД исследуются и моделируются нейрофизиологические [44], механические [45], аэродинамические [46] и акустические [47] аспекты речеобразования. Однако к настоящему времени достигнутые результаты исследования механики движения артикуляторных органов и аэродинамики процессов всё ещё не готовы для практического использования из-за чрезвычайной сложности и нерешённости базовых теоретических проблем моделирования процессов речеобразования.

2.1.2. Основные характеристики формантного метода синтеза речи по тексту

При формантном методе, в отличие от артикуляторного, моделируются не физиологические процессы образования речи, а результат этих процессов: акустические характеристики речевой волны [15, 48, 49]. При этом достигается компактность описания речевого сигнала [50] при достаточно высокой разборчивости синтезируемой речи. Фундаментальное понятие акустической теории речеобразования - форманта, может быть, с одной стороны, определена расчетным путем для каждой фонемы по конфигурации речевого тракта, а с другой - измерена экспериментально по спектру звука. Главным выводом акустической теории является то, что различия в конфигурациях речевого тракта, обусловленные местом образования фонем, отражаются на акустическом уровне в положении на частотной оси полюсов (резонансов) акустической системы. Эти полюса проявляются на амплитудно-частотном спектре в виде определенных максимумов, называемых формантами.

Общая структура формантного синтезатора речи [51] показана на рис. 2.3.

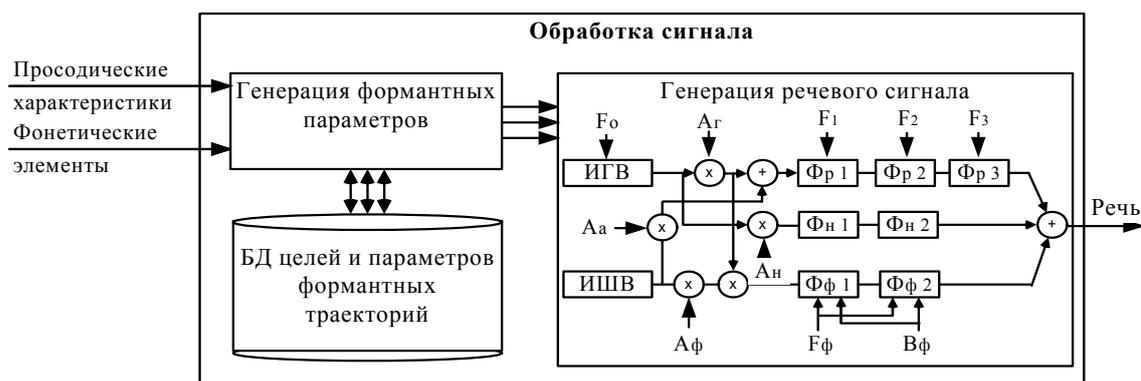


Рис 2.3. Схема формантного синтезатора речи по тексту

Последовательность фонетических элементов, размеченных значениями просодических параметров, поступает в блок генерации формантных параметров речи, использующий данные БД целей и параметров траекторий перестройки формант. Базовая акустическая модель, используемая данным блоком, состоит из источников возбуждения и набора формантных фильтров. Каждый из фильтров описывает характеристики формант и, вместе с источником возбуждения, моделирует речевой спектр, который отражает динамику движения артикуляторных органов.

Формантные параметры, генерируемые данным блоком, различны в разных системах синтеза, и их количество может достигать до 60 [52]. Основными параметрами модели являются следующие [51]: частота основного тона - F_0 , амплитуды голосового - A_g и аспиративного - A_a возбуждения ротовых формант, амплитуда шумового возбуждения - A_ϕ фрикативных формант, амплитуда голосового возбуждения - A_n носовых формант, частоты - F_1, F_2, F_3 ротовых формант, частота - F_ϕ и полоса пропускания - B_ϕ фрикативных формант. Временная последовательность наборов полученных параметров поступает в блок генерации речевого сигнала, управляя источниками шумового и голосового возбуждения (ИШВ, ИГВ) и характеристиками формантных фильтров.

В качестве источника шумового возбуждения используется генератор широкополосного шума [53, 54]. Импульсы голосового возбуждения могут аппроксимироваться треугольной функцией [55], отрезками синусоидальной функции [56], комбинацией синусоидальной и экспоненциальной функций [56], либо представляться с помощью аэродинамической модели голосообразования [51]. Формантные фильтры могут быть представлены цифровыми фильтрами второго порядка, описываемыми уравнением [58]

$$y_n = x_n + (2y_{n-1} - x_{n-1})k_1 - y_{n-2}k_2 \quad (2.2)$$

где n – текущий отсчёт времени,

$n-1, n-2$ – предыдущие отсчеты времени,
 x_n, x_{n-1} – входные сигналы фильтра,
 y_n, y_{n-1}, y_{n-2} – выходные сигналы,
 k_1, k_2 – коэффициенты, отражающие частоту и полосу пропускания фильтра, и определяемые следующими формулами:

$$k_1 = \left(1 - \pi \frac{\beta_i}{f_t}\right) \cos 2\pi \frac{F_i}{f_t} \quad (2.3)$$

$$k_2 = \left(1 - \pi \frac{\beta_i}{f_t}\right)^2 \quad (2.4)$$

где B_i – полоса пропускания i -той форманты,
 F_i – частота i -той форманты,
 f_t – частота дискретизации сигнала.

Для передачи индивидуальности голоса формантным синтезатором речи необходимо создать персонализированные БД целей и параметров траекторий перестройки формант, а также создать адекватную модель голосообразования. Однако автоматическое выделение значений формантных параметров из речевого сигнала, как правило, сопровождается большим количеством ошибок [59, 60], которые в конечном итоге влияют и на качество синтезируемой речи, и на точность передачи индивидуальных характеристик голоса.

Кроме того, существенным недостатком формантного метода является представление процессов речеобразования линейными моделями, не учитывающими взаимовлияние колебаний голосовых связок и резонаторов речевого тракта (так называемый «coupling effect» [61, 62]). Как следствие, формантный метод синтеза не позволяет с достаточной степенью правдоподобия передать индивидуальные характеристики голоса диктора. Эксперименты по узнаваемости голоса, синтезированного с использованием формантного метода [63], показывают, что степень опознавания (субъективная оценка) не превышает 80%.

2.1.3. Основные характеристики компиляционного метода синтеза речи по тексту

Основным отличием компиляционного, или иначе конкатенативного метода, от двух описанных выше является использование при синтезе речи элементарных отрезков естественной речевой волны [64]. При таком подходе исключается необходимость моделирования сложных акустических процессов речеобразования. В процессе синтеза скомпилированный из сегментов естественной речи сигнал подвергается дополнительной модификации: сглаживаются

переходы между соединёнными участками и изменяются просодические параметры в соответствии со значениями, полученными на этапе просодической обработки.

Общая структурная схема компиляционного синтезатора речи [65] показана на рис. 2.4.

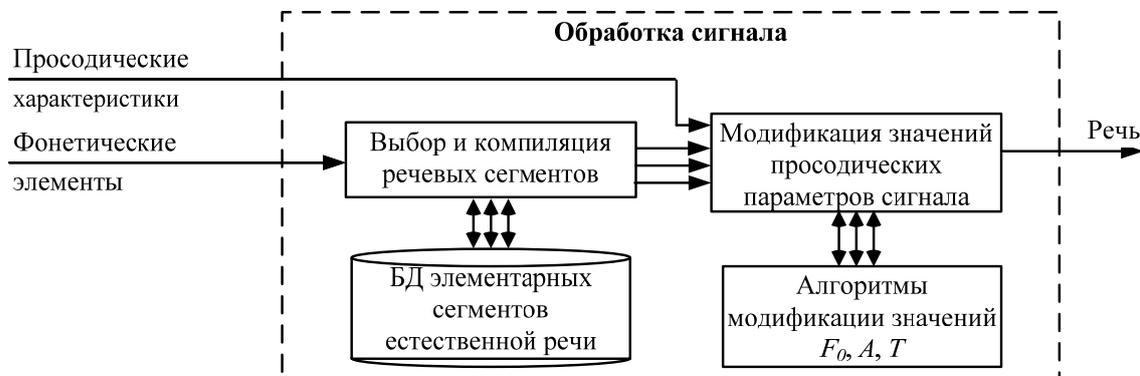


Рис 2.4. Схема компиляционного синтезатора речи по тексту

Последовательность фонетических элементов подаётся в блок обработки сигнала, который выбирает из БД сегментов естественной речи соответствующие звуковые реализации элементов и соединяет их в непрерывный речевой сигнал. Сформированный сигнал подаётся в блок акустической обработки, выполняющий модификацию значений F_0 , A , T речевой волны в соответствии с входными значениями просодических параметров. При этом используются различные алгоритмы модификации сигнала: TD-PSOLA, алгоритм плавной сшивки, модель «гармоники плюс шум».

Наиболее распространённым алгоритмом модификации речевой волны является TD-PSOLA (Time-Domain Pitch-Synchronous OverLap-Add) [66 – 68], изменяющий непрерывный речевой сигнал синхронно с периодом основного тона. При этом граница периода должна быть установлена в максимуме амплитуды сигнала.

Для модификации речевой волны вычисляется последовательность фреймов $s_i(n)$, получаемых синхронно с периодом основного тона:

$$s_i(n) = s(n)\omega(n - iT_0) \quad (2.5)$$

где n – номер отсчёта дискретного сигнала,

$s(n)$ – значение исходного сигнала,

$w(n-iT_0)$ – оконная функция взвешивания, длина которой кратна текущему периоду основного тона,

T_0 – период основного тона исходного сигнала.

Модифицированный сигнал $\tilde{s}(n)$ вычисляется в соответствии с формулой

$$\tilde{s}(n) = \sum_{i=-\infty}^{+\infty} s_i k_i(n - i(T - T_0)) \quad (2.6)$$

где T – результирующий период основного тона,
 k – коэффициент, компенсирующий изменение энергии, вызванное изменением периода основного тона.

Бесконечная сумма в формуле (2.6) ограничивается в практических приложениях, как правило, четырьмя слагаемыми.

Алгоритм TD-PSOLA обеспечивает достаточно хорошее качество модификации просодических параметров, сохраняя частоты и полосы пропускания формант исходного сигнала, если степень изменения периода основного тона, выражаемая как отношение результирующего периода основного тона T к исходному периоду основного тона T_0 , находится в диапазоне $[0,5 \dots 2]$.

Алгоритм плавной сшивки, предложенный одним из авторов этой книги [69], также изменяет речевой сигнал синхронно с периодом основного тона. В отличие от TD-PSOLA, для данного алгоритма необходима разметка на периоды в начале той его части, которая соответствует моменту закрытия голосовых связок, а также привязка процедуры модификации речевой волны к участкам, в точности соответствующим периоду основного тона.

Модификация речевой волны при изменении частоты основного тона осуществляется в соответствии с формулой

$$\tilde{s}(n) = s(n)L_1(n) + s(n + T_0 - T)L_2(n) \quad (2.7)$$

где T_0 – период основного тона исходного сигнала,
 T – результирующий период основного тона,
 $L_1(n)$, $L_2(n)$ – линейные функции, задаваемые, соответственно, формулами (2.8) и (2.9)

$$L_1(n) = \begin{cases} 1, & n \leq (T - N) \\ 1 - \frac{n}{T - N}, & (T - N) < n < N \end{cases} \quad (2.8)$$

$$L_2(n) = \begin{cases} 0, & n \leq (T - N) \\ 1 - \frac{n}{T - N}, & (T - N) < n < N \end{cases} \quad (2.9)$$

где N – коэффициент сшивки, зависящий от результирующего значения T .

В практических приложениях N принимается равным $0,5 * T$.

Алгоритм плавной сшивки, так же, как и TD-PSOLA, применим для модификации F_0 , если степень изменения периода основного тона находится в диапазоне $[0,5 \dots 2]$. Преимуществом алгоритма плавной

сшивки в сравнении с TD-PSOLA является сохранение неизменным исходного речевого сигнала в начальной части периода, соответствующей отрезку времени, когда голосовые связки закрыты. Благодаря этому обеспечивается максимально полное сохранение индивидуальных акустических характеристик голоса.

При использовании модели «гармоники плюс шум» [70 – 72] исходный речевой сигнал представляется суммой периодической и стохастической функций:

$$\tilde{s}(n) = \tilde{s}_p(n) + \tilde{s}_r(n) \quad (2.10)$$

где $\tilde{s}_p(n)$ – периодическая составляющая сигнала,
 $\tilde{s}_r(n)$ – стохастическая составляющая сигнала.

Стохастическая функция $\tilde{s}_r(n)$ обычно задаётся с Гауссовым или равномерным распределением.

Периодическая функция представляется как сумма гармоник

$$\tilde{s}_p(n) = \sum_i a_i(n) \cos \varphi_i(n) \quad (2.11)$$

где i – индекс гармоники,
 $a_i(n)$ – амплитуда i -той гармоники,
 φ_i – фаза i -той гармоники, выражаемая формулой

$$\varphi_i(n) = \varphi_i(0) + \sum_{j=1}^i [i\omega_0(j)]T \quad (2.12)$$

где $\omega_0(j)$ – мгновенная частота сигнала,
 T – интервал временной выборки.

Модификация сигнала в модели «гармоники плюс шум» осуществляется путём вычисления соответствия временных значений t' синтезированного сигнала временным значениям t исходного сигнала, для чего используется функция соответствия $t(t')$.

Применение модели «гармоники плюс шум» имеет некоторые ограничения. В частности, при соединении сегментов естественной речи в точках соединения может возникнуть фазовая рассогласованность. Распространение изменения фазы на соседние фреймы влечёт искажения вокализованных щелевых звуков. Определённым недостатком метода является высокая вычислительная сложность его программной реализации.

Для передачи индивидуальности голоса и дикции личности методом компиляционного синтеза речи по тексту необходимо создать БД элементарных сегментов естественной речи, причём в ней должен содержаться, по крайней мере, один экземпляр речевого сегмента для каждого возможного фонетического элемента, получаемого на этапе фонетической обработки. Определённым недостатком компиляционного

метода в сравнении с двумя описанными выше являются повышенные требования на объем памяти, используемой для хранения БД элементов компиляции.

2.1.4. Основные характеристики корпусного метода синтеза речи по тексту

Последний из рассматриваемых подходов – корпусный – так же, как и компиляционный, использует БД естественной речи. Однако используемая БД состоит не из отдельных специально отобранных элементов компиляции, а представляет собой корпус фонограмм естественной речи, размеченной на элементы фонемной размерности с маркерами их просодических характеристик. Важным отличием корпусного подхода является также возможность использования нескольких сегментов с одинаковыми фонетическими, но различными просодическими характеристиками [73 – 75], благодаря чему естественный речевой сигнал в процессе синтеза во многих случаях не подвергается дополнительной просодической модификации. Правда, такая возможность не исключается и при использовании компиляционного метода в случае, если в БД синтезатора будут дублироваться одинаковые фонетические сегменты, но с различными просодическими характеристиками.

Схема корпусного метода синтеза речи по тексту [76] показана на рис. 2.5. Последовательность фонетических элементов и соответствующие целевые просодические параметры подаются в блок выбора речевых сегментов, который обращается к БД, содержащей речевой корпус, предварительно размеченный на акустико-фонетические элементы с маркерами просодических параметров. Из БД выбираются сегменты, наиболее близкие к требуемым как по фонетическим, так и по просодическим параметрам, которые затем, зачастую без какой-либо дополнительной модификации, соединяются в непрерывный речевой сигнал.

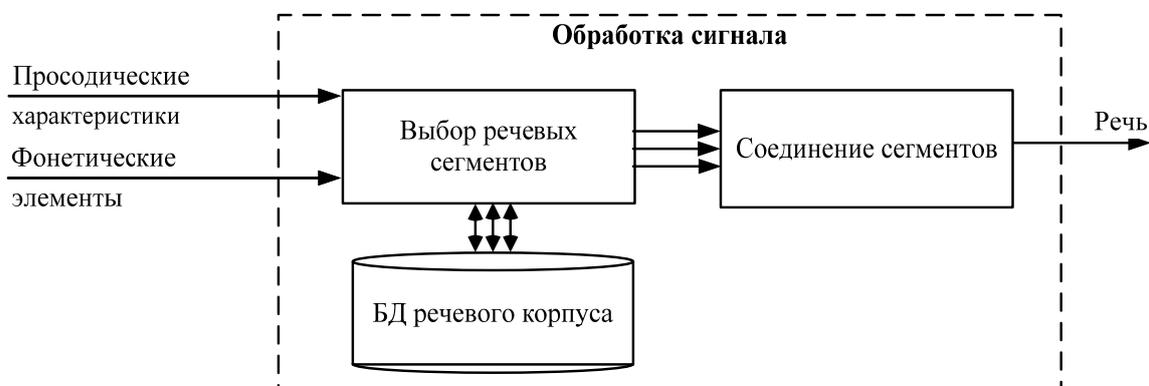


Рис 2.5. Схема корпусного синтезатора речи по тексту

Использование сегментов естественной речи и минимум модификации речевого сигнала потенциально обеспечивает при синтезе речи хорошую передачу индивидуальных акустических характеристик. Однако при этом БД должна содержать максимальное количество всех возможных комбинаций фонетических элементов, необходимых для синтеза речи в соответствии с используемыми фонетической и просодической моделями. Создание такой БД, её фонетическая и просодическая маркировка сопряжены с огромными трудностями. Кроме того, объём создаваемой БД может стать очень большим (до нескольких часов звучащей речи), что неприемлемо для ряда практических приложений.

2.1.5. Выбор метода синтеза речи по тексту для передачи индивидуальности голоса и манеры чтения

Особенности различных методов синтеза речи по тексту, показанные в разделах 2.1.1 – 2.1.4, позволяют утверждать, что наиболее подходящими для передачи индивидуальных характеристик речи являются компиляционный и корпусный методы синтеза. Предпочтение отдаётся компиляционному методу синтеза речи, обладающему, по сравнению с корпусным, следующими преимуществами:

- значительно меньший объём создаваемого речевого корпуса, что позволяет упростить процедуру звукозаписи, минимизировать неизбежную вариативность акустических характеристик голоса диктора в процессе записи, а также упростить процедуру коррекции неточностей автоматической сегментации речевого корпуса;

- значительно меньший объём создаваемой БД естественно-речевых сегментов, что способствует их более тщательному отбору и создаёт предпосылки к использованию синтезатора речи в мобильных устройствах с ограниченным объёмом памяти.

2.2. Методы моделирования просодических характеристик речи

Просодика играет важную роль как при восприятии смысла речи, так и при восприятии индивидуальности голоса и дикции личности. Поэтому просодическая модель, используемая при синтезе речи по тексту, должна адекватно отражать как языко-зависимые, так и дикторозависимые характеристики.

Существует достаточно большое число просодических моделей, предложенных для использования в системах синтеза речи по тексту. По методу представления интонации просодические модели можно разделить на следующие основные категории:

- автосегментная модель (АМ-модель) [77];
- ПРО-модель [78];
- суперпозиционная модель (СП-модель) [79];

- непрерывная параметрическая модель (НП-модель)[80];
- модель портретов акцентных единиц (ПАЕ) [81, 69].

2.2.1. Автосегментная просодическая модель

АМ-модель представляет интонацию как линейную последовательность уровней основного тона. При этом контур основного тона описывается конечно-автоматной грамматикой, состоящей из необязательного начального граничного тона, одного или более тоновых ударений, фразового ударения и конечного граничного тона, как показано на рис. 2.6.

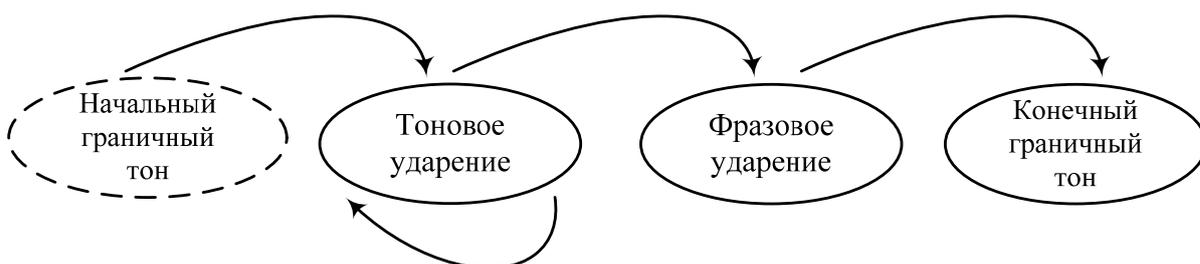


Рис 2.6. Конечно-автоматная грамматика контура основного тона

Три основных интонационных события: тоновое ударение, фразовое ударение и граничный тон – описываются с помощью инвентаря, составляющего систему транскрипции интонации ToBI (Tone and Break Indices) [77], которая включает два базовых уровня тона: высокий - H (high) и низкий - L (low), а также диакритические знаки для указания типа интонационного события: «*» – тоновое ударение, « $\bar{}$ » – фразовое ударение и несколько дополнительных диакритических знаков для описания направления движения и уровня тона. Полный фонологический инвентарь системы ToBI включает 12 элементов.

Автоматические системы анализа и синтеза интонации, использующие методологию ToBI [82, 83], основаны на статистических моделях. Местоположение и тип тона для каждого тонового ударения предложения, используя деревья принятия решений и Марковские цепи.

Для обучения модели используется корпус естественной речи, размеченный на слоги. Параметры модели, используемые для обучения, включают информацию о местоположении фразы и предложения внутри параграфа, о длине предложения, о знаках препинания, о местоположении просодической фразы внутри предложения, о длительности просодической фразы, о позиции слога внутри фразы, о позиции фразового и слогового ударения, грамматическую информацию о слове, содержащем данный слог, информацию о количестве безударных слогов перед данным слогом. Эта информация кластеризуется и используется для предсказания уровня подъема тона.

2.2.2. Просодическая модель IPO

IPO-модель [78], так же, как и AM-модель, представляет интонацию как последовательность дискретных событий. Но инвентарь IPO-модели состоит из движений, а не из уровней основного тона. Базовые предположения IPO-модели следующие:

- при моделировании интонации необходимо учитывать только перцептивно релевантные движения;
- аппроксимация F_0 прямыми линиями не ухудшает перцептуальное качество интонации [84].

Модель описывает движения основного тона как линейные изменения логарифма F_0 во времени. Процесс моделирования интонации конкретного языка или конкретного диктора с использованием IPO-метода включает следующие этапы:

- стилизация контуров основного тона;
- создание инвентаря движений основного тона;
- объединение элементов инвентаря в структуры и создание грамматики контуров;
- разработка генерирующего множества правил синтеза интонации.

Стилизация контуров основного тона осуществляется интерактивно: экспериментатор заменяет небольшие секции контуров основного тона прямой линией и ре-синтезирует фразу. Процесс продолжается до тех пор, пока не будет получено минимальное количество линий, не изменяющих воспринимаемое качество интонации. Для создания инвентаря на основе экспериментальных данных определяется минимальный возможный набор движений основного тона, который может быть использован для адекватного учёта всех значимых подъёмов и падений F_0 . При этом каждое движение характеризуется направлением (подъём или падение), позицией относительно ударного слога (раннее, среднее или позднее), длительностью (краткое – занимающее один слог – или протяжённое – занимающее два и более слогов), и степенью изменения F_0 (полное, половинное). Такой инвентарь был разработан, в частности, для голландского [78], британского английского [85] и русского [86] языков.

Наиболее сложная проблема IPO-моделирования – процесс стилизации, который требует взаимодействия между человеком и компьютером, многих часов интенсивного прослушивания и повторного синтеза речи для получения удовлетворительного набора движений основного тона для конкретного языка или конкретного диктора.

2.2.3. Суперпозиционная просодическая модель

СП-модель [79] основана на физиологической модели речеобразования, объединенной с иерархической теорией просодической фонологии [87], и представляет интонацию как

последовательность событий с перекрывающимися областями. При этом события являются результатом интерактивных интонационных «команд», которые представляют слоги, акцентные и ударные группы, фразы и более крупные просодические единицы. Команды комбинируются таким образом, что каждый уровень иерархии представлен в генерируемом контуре F_0 .

Каждая команда состоит из пика и экспоненциального угасания. Пик может представляться выходом фильтра, который возбуждается прямоугольной входной командой [79], или задаваться параметрами пика и угасания [88]. Фразовая команда, которая является максимальной единицей в модели, обычно устанавливает интонационный регистр для фразы. Акцентная команда – это острый пик с коротким временем угасания, представляющий движение F_0 при интонационном событии.

Комбинация команд при формировании интонационного контура показана на рис. 2.7 [79].

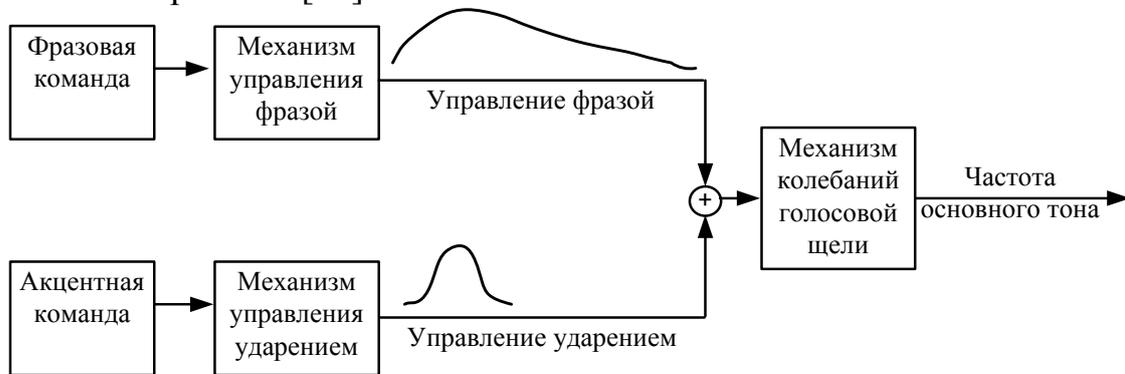


Рис. 2.7. Суперпозиционная модель генерации контура F_0

СП-модель является основой для интонационных моделей нескольких языков в многоязычном синтезаторе, описанном в [89]. При этом реализация синтеза интонации использует гибридную модель, сочетающую предсказание уровней тона (с использованием сокращённого инвентаря тонов ToBI) и предсказанием, на основе полученных уровней тона, параметров акцентных и фразовых команд.

СП-модели апробированы при синтезе интонации для нескольких языков [89, 90]. Однако при анализе интонации естественной речи в рамках СП-модели возникают некоторые неоднозначности. В частности, главной проблемой является представление различных движений основного тона суммой акцентных и фразовых тонов.

2.2.4. Непрерывная параметрическая просодическая модель

НП-модель [80] интерпретирует частоту основного тона в акустической области, описывая интонацию в терминах движения F_0 во времени. Одной из наиболее удачных реализаций НП-моделей является модель Tilt [80].

В модели Tilt существует 4 основных интонационных элемента: тоновое ударение, граничный тон, связка и пауза. Связка и пауза являются интонационными элементами, описываемыми лишь одним параметром: значением F_0 в начале события. Тоновое ударение и граничный тон описываются пятью параметрами: значением F_0 в начале события, длительностью, амплитудой подъёма, позицией пика (точки, в которой заканчивается подъём и начинается падение), и амплитудой падения, как показано на рис. 2.8.

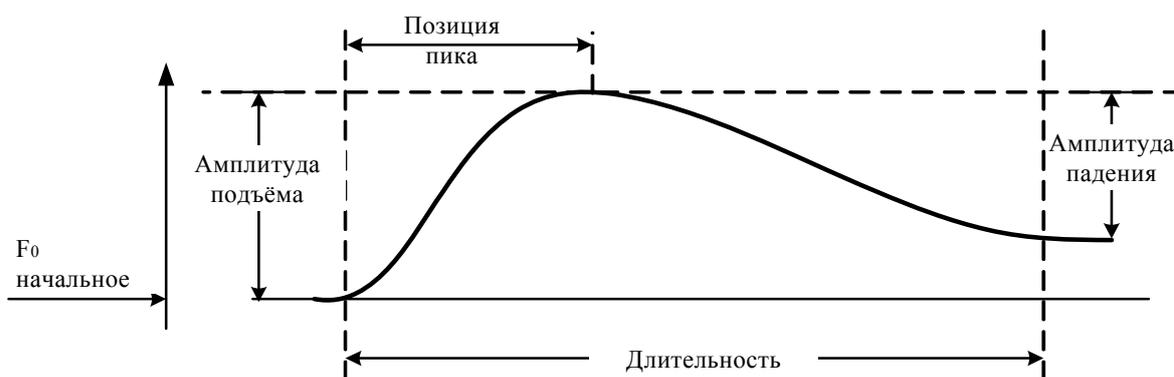


Рис. 2.8. Параметры модели Tilt

Прикладные системы, использующие Tilt модель, обучаются на данных с использованием Марковской модели [80]. Модель обучается для нахождения позиций ударений, границ, связок и пауз. Для обучения используются нормализованные значения F_0 и среднего квадрата энергии, а также их производные. Реализация Tilt – модели, так же, как и других методов моделирования, может быть автоматизирована и построена на основе аннотированной БД.

2.2.5. Просодическая модель портретов акцентных единиц

ПАЕ-модель основана на представлении интонации фразы последовательностью просодических портретов акцентных единиц (ПАЕ). В общих чертах ПАЕ-модель описана в разделе 1.4 (Глава 1). ПАЕ-модель была предложена одним из авторов этой книги более 20 лет назад [81] и с тех пор успешно использовалось во многих моделях синтеза речи по тексту [65, 69, 91, 92], в том числе многоязычных [31].

В соответствии с ПАЕ-моделью, минимальной просодической единицей является акцентная единица (АЕ), состоящая из одного или более слов, и имеющая в своём составе только один, полноударный слог. АЕ, в свою очередь, состоит из ядра (полноударный слог), предъядра (все фонемы, предшествующие полноударному слогу) и заядра (все фонемы за полноударным слогом). Главное предположение ПАЕ-модели состоит в том, что топологические свойства просодических параметров для определенного типа интонации фразы не изменяются

(или изменяются незначительно) с изменениями фонетического контекста и числа слогов в пред-и заядре АЕ. Этот факт иллюстрируется рис. 2.9, где показаны контуры F_0 для однословных вопросительных фраз с различным положением словесного ударения.

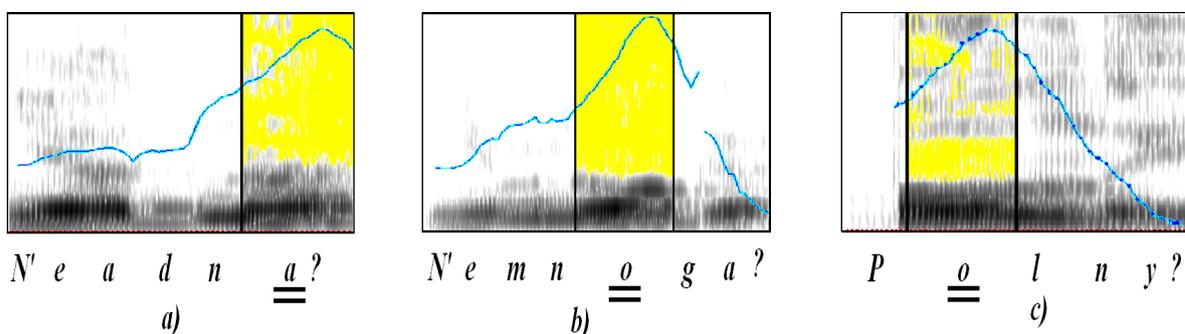


Рис. 2.9. Контурсы F_0 вопроса для однословных фраз: а) “Не одна?”, б) “Не много?”, в) “Полный?” (ударные гласные подчеркнуты двойной чертой)

АЕ может состоять также и из более чем одного слова в случае, когда фраза имеет только одно главноударное слово. Это иллюстрируется на рис. 2.10, где представлены контуры F_0 для трёхсловных фраз с тремя различными положениями главноударного слова во фразе. Фраза “Мама мыла малину?” была произнесена диктором три раза с вопросительным типом интонации, с тремя различными положениями главного ударения.

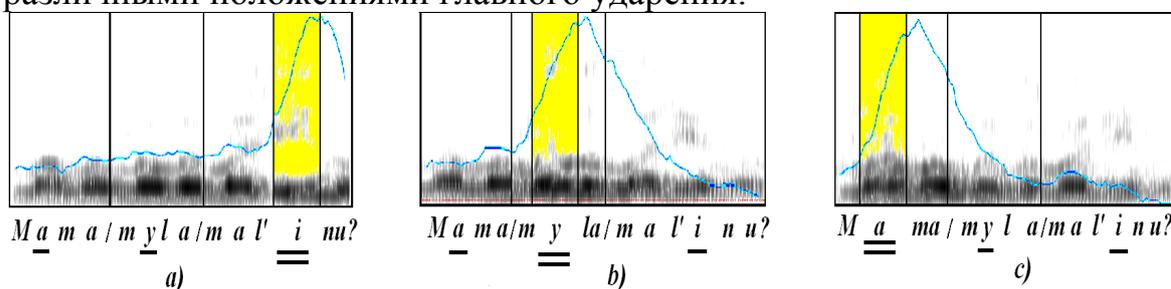


Рис. 2.10. Контурсы F_0 синтагмы вопросительного типа “Мама мыла малину?”, в которой главноударное слово: а) ”малину”, б) “мыла”, в) ”мама”

Как видно из рис. 2.10, каждая из этих фраз состоит только из одной АЕ, а поведение контура F_0 подобно поведению на ядре, пред-и за-ядре однословной фразы, показанной на рис. 2.9. Это позволяет нам представить F_0 -ПАЕ в нормированном пространстве «частота-время» с равной относительной длительностью трёх его частей - ядра, предъядра и заядра. На рис. 2.11а показан F_0 -ПАЕ однословной фразы словом, а на рис. 2.11б – трёхсловной фразы, полученные на основе, соответственно, рис. 2.9 и 2.10.

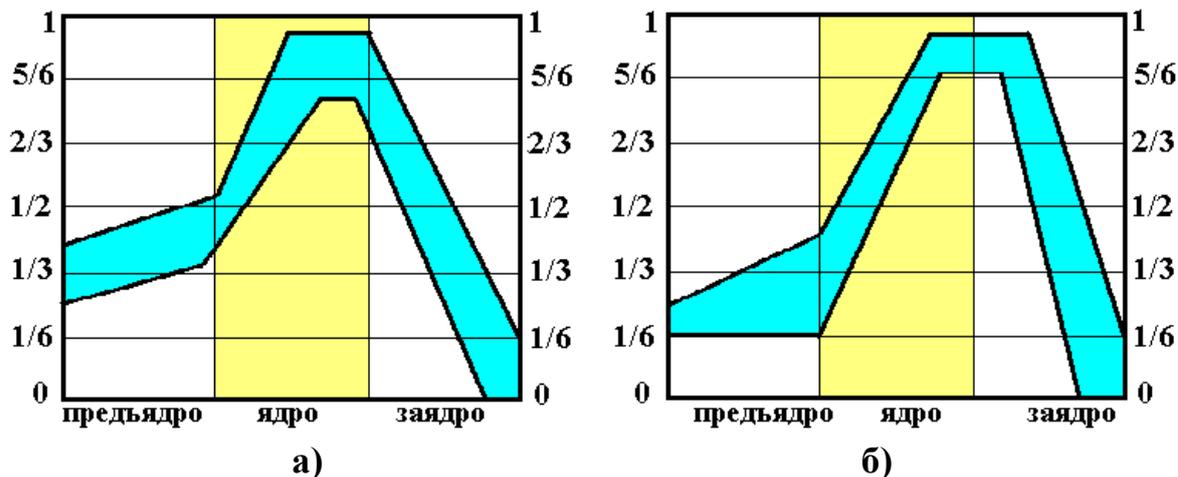


Рис.2.11. ПАЕ для вопросительного типа интонации для а) однословных фраз и б) трёхсловных фраз

Как видно из рис. 2.11а и 2.11б, различие в их ПАЕ малосущественно, что позволяет говорить о применимости ПАЕ модели как для однословных, так и для многословных АЕ.

Отмеченные выше на примере вопросительной интонации закономерности представления F_0 -ПАЕ справедливы также для интонации завершённости, незавершённости и других интонационных типов. Подобное заключение может быть также сделано относительно возможности использования ПАЕ для описания динамических (A_0 -ПАЕ) и ритмических (T_0 -ПАЕ) характеристик просодики речи.

Рассмотренные примеры касались только одноакцентных синтагм. Однако синтагма может состоять также из 2-х и более АЕ. В этих случаях обобщённый просодический портрет синтагмы составляется из последовательности ПАЕ, которые используются затем системой синтеза речи по тексту независимо от фонетического содержания конкретных АЕ, входящих в синтагму.

ПАЕ-модель, как и НП-модель, описывает интонацию в терминах движения F_0 во времени, но характер движения не ограничивается фиксированными параметрами, а может быть представлен какой угодно непрерывной кривой, в соответствии с персональными особенностями интонирования речи. Таким образом, созданные F_0 -ПАЕ совместно с A_0 -ПАЕ и T_0 -ПАЕ не только описывают просодический комплекс различных интонационных типов и их вариантов, но и характеризуют персональные просодические характеристики речи диктора.

Подробно процедуры использования ПАЕ-модели в системах компьютерного синтеза и клонирования просодических характеристик речи будут рассмотрены в последующих главах.

2.3. Методы выбора базовых речевых единиц для синтеза речи

В процессе синтеза речи по тексту тип базовых речевых единиц влияет на формирование текстовых и речевых корпусов, на методы

сегментации и маркировки сигнала и на степень сохранения индивидуальных речевых характеристик в персонализированной речевой БД. В основу классификации базовых речевых единиц положены такие понятия, как фонема, аллофон и слог, рассмотренные в первой главе.

При выборе речевых единиц для создания БД существует несколько подходов, в соответствии с которыми могут быть сформированы базовые сегменты различной длины. При выборе сегментов той или иной длины используются следующие критерии:

- объём работы, необходимый для создания речевого корпуса, последующей сегментации и маркировки;
- степень сохранения эффектов взаимодействия звуков, реализующиеся в естественном потоке речи;
- степень сохранения специфики межзвуковых переходов между выбранными элементами в естественном потоке речи.

При использовании звуковых единиц бóльшей длины в значительной степени сохраняется естественность эффектов взаимодействия звуков и характеристик межзвуковых переходов, но при этом резко возрастает количество звуковых единиц и, соответственно, объём работы для создания корпуса, его сегментации и маркировки. При использовании коротких речевых единиц меньше времени и усилий тратится на создание индивидуализированной речевой БД, но естественность проявления эффектов взаимодействия звуков и характеристик межзвуковых переходов могут быть представлены в недостаточной степени.

В [69] предложено в качестве базовых речевых единиц минимальной длительности использовать отрезки аллофонов, равные периоду основного тона для гласных, звонких и вокализованных согласных – микроволны (МВ-метод). При этом каждый аллофон описывается тремя сегментами: начальным, серединным и конечным, а переходы между ними аппроксимируются с помощью алгоритма «плавной сшивки» микроволн. МВ-метод обеспечивает минимально возможный объём БД для синтеза речи.

При таком подходе можно избежать изменения физических параметров звуков в процессе персонализированного синтеза речи, так как базовые речевые единицы обеспечивают широкий выбор периодов основного тона. С другой стороны, в процессе синтеза речи может потребоваться значительное вмешательство во внутреннюю структуру периодов между двумя граничащими сегментами, что приводит к уменьшению естественности синтезируемой речи.

В работах [93 – 95], и во многих других, в качестве базовых речевых единиц предлагается использовать дифоны – отрезки речевой волны, заключённые между серединами соседних фонем (дифонный метод). Преимущества такого подхода заключаются в сохранении в

речевых сегментах естественного переходного участка между фонемами, а также в сравнительной лёгкости вычленения дифонов при сегментации естественной речи, так как в этом случае, в отличие от элементов аллофонной размерности, нет необходимости точно определять границы сегментов. Недостатком такого подхода является увеличение числа базовых единиц, поскольку в этом случае невозможно объединить предшествующие и последующие фонемы в сходные по артикуляторным признакам группы, как при аллофоном представлении сегментов, и, следовательно, время и усилия на создание текстовых и речевых корпусов существенно увеличиваются.

Преимущества выбора аллофонов в качестве базовых единиц [65, 96 – 98] состоят в том, что, во-первых, речевые единицы сохраняют эффекты взаимодействия звуков, во-вторых, количество базовых единиц относительно небольшое и варьируется в различных системах от 450 до 1500. Определённым недостатком такого подхода является требование прецизионной разметки аллофонов при сегментации естественного речевого сигнала.

Существуют системы, объединяющие преимущества использования одновременно аллофонов и дифонов [99]. В таких системах в качестве базовой речевой единицы принимается половина аллофона (или, что то же самое, половина дифона): от левой границы до середины аллофона и от середины до правой границы. К недостаткам данного подхода можно отнести необходимость дополнительной, более детальной, чем при аллофонном подходе, классификации звуков речи.

Удачным и обоснованным выглядит выбор слогов в качестве базовых речевых сегментов, поскольку слог считается минимальной речеобразующей единицей с сильным эффектом взаимодействия звуков между составляющими его элементами [100]. Эффекты взаимодействия звуков между слогами проявляется гораздо меньше. Такие системы реализованы, в частности, для японского [101], чешского [102], индийского [103] и английского [104] языков. Количество различных по фонемному содержанию наиболее частотных слогов относительно небольшое, но с учётом количественной и качественной редукции гласных оно возрастает в несколько раз. Если же учитывать, так называемые межсловные слоги, образующиеся в слитной речи, то, очевидно, в этом случае базовыми речевыми единицами должны быть все теоретически возможные слоги языка, что приводит к очень большому объёму БД элементов компиляции. Поэтому при использовании слогов в качестве базовых единиц либо создаются дополнительные единицы, такие как пары аллофонов гласный-гласный [101] или сегменты гласный – согласный+гласный [105], либо количество используемых слогов уменьшается, а в процессе синтеза речи осуществляется сглаживание акустических параметров различными методами. При этом

естественность синтезируемой речи ощутимо снижается, что практически может уничтожить все преимущества, достигнутые использованием слогов в качестве элементов компиляции.

При любом из рассмотренных выше наборов базовых речевых единиц возможно их дальнейшее расширение путём добавления в БД идентичных сегментов с различными просодическими характеристиками: частоты основного тона, амплитуды и длительности.

Подходы, основанные на фонемно-аллофоном представлении элементов компиляции, можно назвать *экспертно-фонетическим*, в отличие от ещё одного, достаточно распространённого подхода, реализующего выбор базовых единиц на основании *формально-лингвистического* анализа текстов большого объёма и соответствующих им фонограмм речи.

При формально-лингвистическом подходе формирование базовых речевых единиц осуществляется на основе корпуса текстов большого объёма, представляющие различные жанры: художественные и научные тексты, журнальные статьи, сводки новостей и т.д. [106, 107]. Используется предположение, что при достаточно большом объёме корпуса в нём представлены все наиболее часто встречающиеся комбинации фонем языка, причём в множестве просодических окрасок. Поэтому при этом подходе зачастую ставится задача отразить не только фонетико-акустические, но и просодические индивидуальные характеристики речи.

При формально-лингвистическом подходе в качестве базовых речевых единиц могут использоваться все сегменты, представленные в сформированных корпусах. Длина сегмента для синтеза речи при этом не является строго определённой. Весь созданный речевой корпус размечается на речевые единицы достаточно малой длины (фонемы, полу-дифоны, дифоны), для каждой из которых вычисляются просодические характеристики: частота основного тона, амплитуда и длительность. При этом речевая БД содержит, как правило, несколько экземпляров лингвистически идентичных сегментов с различными просодическими характеристиками. При синтезе речи с использованием данного подхода входной текст транскрибируется и каждой фонеме назначаются целевые просодические характеристики. Затем при синтезе речи осуществляется поиск по речевому корпусу последовательностей сегментов, соответствующих сформированной последовательности фонем. При этом должны выполняться следующие условия:

– лингвистическое содержимое сегментов в речевой БД должно быть идентично лингвистическому содержимому последовательности фонем, используемому при синтезе речи,

– несоответствие просодических характеристик сегментов и целевых просодических характеристик последовательности фонем должно быть минимальным,

– при одинаковых условиях предпочтение отдаётся последовательности сегментов максимальной длины.

При формально-лингвистическом подходе за счёт хранения в БД нескольких экземпляров лингвистически идентичных сегментов с различными просодическими характеристиками модификация параметров звука при синтезе может не потребоваться или она будет минимальной [106]. За счёт этого предполагается достичь более высокой естественности синтезируемой речи.

К недостаткам формально-лингвистического подхода к формированию базовых сегментов можно отнести следующее:

– негарантированное покрытие аллофонного состава языка базовыми сегментами;

– наличие большого количества «избыточных» сегментов, т.е. нескольких экземпляров сегментов с совпадающими фонетическими, акустическими и просодическими характеристиками;

– недостаточный учёт эффектов редукции, коартикуляции и ассимиляции гласных [108];

– слишком большой размер БД элементов компиляции (БД элементов компиляции совпадает с БД исходного речевого корпуса и содержит, как правило, порядка нескольких часов записей речи). Это влечёт за собой определённые трудности при сегментации и маркировке сигнала;

– большой объём памяти и вычислительные затраты на хранение и оперативный поиск требуемых элементов компиляции в процессе синтеза речи.

Последнее обстоятельство делает метод синтеза речи на основе формально-лингвистического подхода практически непригодным для использования в компьютерах малой производительности (карманные ПК) и в мобильных устройствах (телефоны).

2.4. Методы формирования речевых и текстовых корпусов для создания индивидуализированных речевых баз данных

Обоснованное формирование текстового и речевого корпусов, наряду с лингвистически обоснованной классификацией и выбором базовых сегментов для синтеза речи, во многом определяет степень разборчивости и естественности синтезируемой речи. Формирование корпусов по заданным базовым сегментам, кроме основной цели – покрытия всех сегментов, должно удовлетворять следующим требованиям [109]:

– результирующий корпус должен быть фонетически полным, т.е. в фонетической транскрипции текста должны встречаться все основные фонемы речи и их варианты;

– объём корпуса должен быть, по возможности, минимизирован;

– созданный корпус должен быть фонетически сбалансированным, т.е. распределение частот встречаемости фонем и других фонетических единиц в сформированном корпусе должно быть близким к теоретическому, полученному на достаточно представительных и больших по объёму выборках.

Последнее требование к формированию корпусов предъявляется в случае, если при выборе базовых сегментов не учитывались фонетические контексты.

Один из подходов к формированию текстовых корпусов – экспертный подход – состоит в подборе слов или фраз, составляющих корпус, опытными экспертами-фонетистами «вручную» [110] или полуавтоматически, с использованием программ-транскрайберов. Преимущества такого подхода состоят в гарантированном покрытии всего необходимого множества и в подборе наиболее подходящих для произношения слов и фраз. Экспертный подход при определённых условиях гарантирует получение минимального или близкого к минимальному объёма корпуса. К недостаткам такого подхода можно отнести большой объём «ручной» работы и требования высокой квалификации к эксперту-фонетисту, выполняющему данную работу.

Нестандартный подход к созданию текстовой БД предложен в работах [111 – 113]. В отличие от традиционных подходов, в качестве текстового корпуса предлагается использовать бессмысленные слова, «псевдослова». Преимущества данного подхода: не требуется производить выборку из набора текстов слов или фраз, удовлетворяющих заданным условиям, а объём созданной речевой БД будет гарантированно минимальным (поскольку псевдослова могут содержать *только* базовые сегменты). Однако данный подход имеет ряд существенных недостатков, а именно: неоднозначным и трудным для диктора является произношение псевдослов, что ухудшает качество записанной речи; записи речи будут просодически непредсказуемыми (скорее всего, не нейтральными), что также влечёт потерю качества содержимого речевого корпуса.

При автоматическом, статистическом подходе формирование текстового корпуса происходит на основании наборов текстов различных жанров [94, 114], включающих газетные статьи, научные тексты, художественные произведения. В качестве исходного набора текстов используются стенограммы записей [115]. Предполагается, что стенограммы текстов большого объёма с высокой степенью вероятности

включают все возможные в языке сегменты речи. Далее осуществляется выбор из набора этих текстов его элементов (слов или фраз), из которых формируется новый текстовый корпус, так чтобы этот корпус был минимального размера и покрывал максимально возможное множество базовых сегментов для синтеза речи. Решение этой задачи формулируется как решение задачи минимизации покрытия множества.

Задача минимизации покрытия множества является *NP*-сложной [116], и для её решения используются различные оптимизационные алгоритмы, в частности, так называемый поглощающий («жадный») алгоритм, описанный в [117], и его различные модификации, а также генетический алгоритм, описанный в [118]. Суть «жадного» алгоритма заключается в определении наибольшего количества базовых сегментов в каждом элементе текста (такими элементами могут являться слова или предложения) и добавлении на каждом шаге алгоритма в формируемый новый текстовый корпус фрагментов исходного текста, содержащих наибольшее количество требуемых сегментов. Процесс добавления элементов продолжается до тех пор, пока формируемый текстовый корпус не будет включать все необходимые сегменты. Такой подход используется, в частности, в [119-123]. В некоторых системах в сформированный корпус вручную добавляются также элементы, содержащие редко встречающиеся сегменты [94] или же наоборот, только наиболее часто встречающиеся сегменты [114]. Такой подход фактически является комбинацией экспертного и статистического подхода.

При использовании генетического алгоритма выбирается начальный набор предложений, покрывающий все базовые элементы (в качестве такого набора могут использоваться все предложения, составляющие исходный корпус) – исходная популяция. При этом каждое предложение – индивидуум популяции. Оценка предложения происходит на основе количества базовых сегментов, которые в нём представлены. Затем итеративно осуществляются операции мутации индивидуумов, их скрещивания, селекции и отбора наиболее элитных, представительных. После каждой итерации формируется новое поколение с «лучшими» характеристиками, т.е. формируется подмножество предложений, мощность которого меньше мощности подмножества на предыдущей итерации, но покрывает набор базовых сегментов. Такой алгоритм используется, в частности, в [124].

Во всех подходах для решения задачи дальнейшей минимизации набора базовых сегментов постулируется следующее:

- исходный корпус уже покрывает все необходимые для синтеза базовые сегменты;
- в корпусе присутствуют элементы, излишние с точки зрения фонетики и просодики;

– в корпусе присутствуют элементы, которые, в силу своих нестандартных акустических и просодических характеристик никогда не будут выбраны для синтеза.

На основании двух последних постулатов и происходит уменьшение речевой БД [125-128]. Для максимального уменьшения объёма речевой БД выбирается лишь один экземпляр для каждого базового сегмента, оптимальный по просодическим характеристикам [128].

Другой подход, используемый для уменьшения речевой БД, преследующий цель сохранения уровня естественности синтезированной речи, основан на удалении из БД неиспользуемых экземпляров, а также экземпляров, просодические характеристики которых схожи с просодическими характеристиками других присутствующих в базе экземпляров того же сегмента. Определение оптимального размера речевой БД в этом случае осуществляется путём создания нескольких БД различного объёма и перцептивной оценки речи, синтезированной на основе созданных БД.

Для проверки выполнения условия 3 (обеспечение фонетической репрезентативности корпуса) вычисляется статистика встречаемости фонем в исходном наборе текстов и в результирующем текстовом корпусе. Если она приблизительно совпадает, сформированный корпус считается фонетически репрезентативным. Такую оценку, пожалуй, нельзя назвать корректной, поскольку эксперименты показывают [120], что распределение частот одних и тех же единиц в текстах довольно сильно варьирует в зависимости от жанра текста (газеты, технические описания, телефонные переговоры). Кроме того, даже если сравнивать распределения частот единиц в текстах одинакового жанра, коэффициент корреляции между такими распределениями в некоторых случаях оказывается менее 0,5.

С технической точки зрения формирование речевых корпусов осуществляется путём записи в цифровом виде естественной речи диктора. Для достижения высокого качества речи на этом этапе должно обеспечиваться выполнение следующих технических условий [129]: запись должна происходить в идентичных студийных условиях, длительность сессии записи не должна превышать «критического» для диктора времени, после которого ухудшаются акустические характеристики его голоса, расстояние до микрофона в течение всего периода записи должно быть постоянным, темп речи диктора должен быть средним или менее среднего. Следует заметить, что чем больше объём корпуса, тем сложнее выполнение этих условий.

2.5. Методы сегментации и маркировки естественного речевого сигнала

Сегментация речевого сигнала на фонетические элементы применяется в различных областях речевых технологий, в частности, в системах синтеза речи, в фонемно-ориентированных системах распознавания речи, в системах идентификации и верификации диктора. В контексте синтеза речи, когда известна последовательность фонем, составляющих речевой корпус, и, как правило, речевой сигнал размечен на периоды основного тона, задача сегментации сигнала упрощается, за счёт чего точность сегментации может быть существенно повышена.

Для систем синтеза речи, использующих в составе БД элементы компиляции, содержащие наряду с фонетико-акустическими также и просодические характеристики задача состоит не только в сегментации потока речи на базовые речевые единицы, но и в вычислении для каждой речевой единицы просодических параметров: длительности, энергии и частоты основного тона.

Во многих случаях процесс сегментации и маркировки речевой БД осуществляется вручную экспертом-фонетистом с использованием полуавтоматических средств просмотра осциллограмм, спектрограмм и сонограмм сигнала [130-132]. «Ручной» метод сегментации и маркировки требует много времени и усилий, но обеспечивает при достаточно высокой квалификации эксперта, в отличие от автоматической сегментации, достаточно точную разметку речевого корпуса.

Требования к точности разметки зависят от типа базовых речевых единиц. Так, при использовании в качестве базовых речевых единиц дифонов или слогов допускается некоторая погрешность разметки, которая затем компенсируется в процессе синтеза речи, в то время как при использовании в качестве элементов компиляции аллофонов требования к точности разметки существенно возрастают.

Автоматическая сегментация и маркировка включает следующую последовательность действий:

- параметрическое представление речевого сигнала;
- обучение, «настройка» модели;
- сегментация и маркировка;
- пост-коррекция результатов сегментации.

Вне зависимости от метода сегментации используются, в частности, следующие типы параметрического представления речевого сигнала: кепстральные коэффициенты [133, 134], энергия и дельта-энергия сигнала [135, 136], линейные спектральные пары [137].

Основные подходы, используемые для автоматической сегментации и маркировки сигнала – это нейро-сетевые модели (НС-модели),

скрытые марковские модели (СММ), и методы динамического программирования (ДП-методы).

Использование НС-моделей требует предварительной процедуры обучения для настройки нейронной модели, накопления достаточного количества статистических данных, которое осуществляется на базе уже размеченных речевых корпусов большого объёма. Тем не менее, при использовании НС-модели не достигается требуемая точность разметки, поэтому, как правило, он применяется в комбинации с другими методами [138, 139].

Методы сегментации и маркировки, использующие СММ [140-143], учитывают не только акустические, но и фонетические свойства речевого сигнала. Получение определённых спектральных характеристик фонем достигается путём анализа больших по объёму БД. Однако содержащейся в них информации о фонемных переходах часто недостаточно. Для решения этой проблемы используется контексто-зависимая фонемная модель. Существует несколько подходов для определения фонемных переходов: использование методов нечёткой логики, правила которой представляют фонетические знания об образцах изменений на фонетических переходах [144], нейронных сетей, обучающихся на статистике [145], и моделей гауссовых смесей [134]. В любом случае, однако, нет уверенности, что границы фонем, вычисленные в процессе обучения СММ модели, будут соответствовать действительным границам фонем в речевом сигнале. Для достижения большей точности данный подход требует дополнительного обучения, т. е. наличия очень большого (порядка нескольких часов) размеченного речевого корпуса, что не всегда возможно.

ДП-метод заключается в динамическом сопоставлении двух векторов и нахождении оптимального пути соответствия между ними [146, 147]. В качестве таких векторов используются параметрически представленный синтезированный речевой сигнал – вектор-эталон, и параметрически представленный естественный речевой сигнал – вектор-реализация [148, 149].

Блок-схема системы сегментации и маркировки речевого корпуса ДП-методом представлена на рис. 2.12.

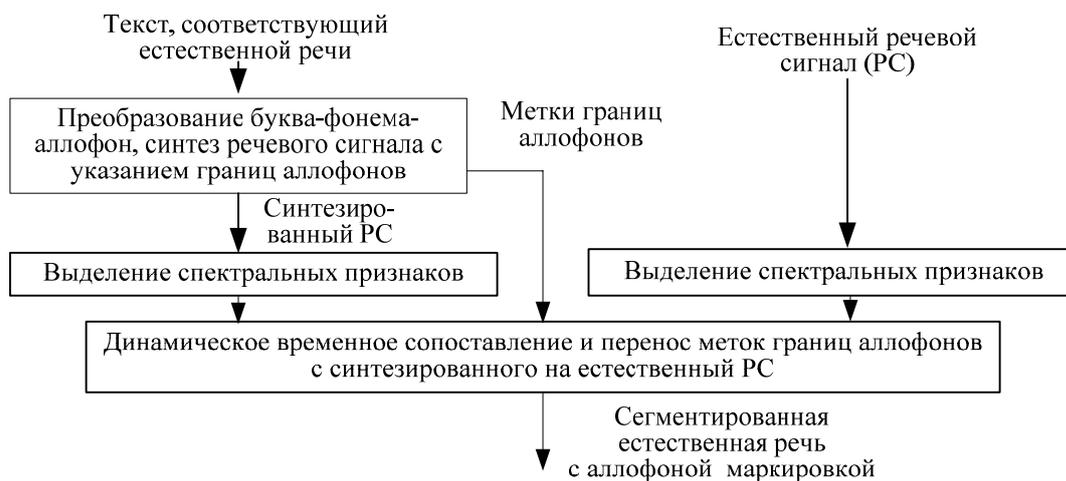


Рис. 2.12. Блок-схема системы сегментации и маркировки речевого корпуса ДП-методом

Для синтеза речевого сигнала содержимое текстового корпуса транскрибируется и преобразуется к последовательности символов, обозначающих базовые речевые единицы, акустическая реализация которых извлекается из уже существующей БД элементов компиляции. Затем осуществляется конкатенация акустических единиц. При этом в синтезированном речевом сигнале отмечаются границы речевых сегментов. Найденный в процессе ДП-сопоставления оптимальный путь соответствия указывает положение границ базовых сегментов в естественном речевом сигнале. Использование ДП-метода не требует процедуры обучения модели, кроме того, является в значительной степени дикторонезависимым.

Глава 3

Экспериментальные исследования фонетико-акустических и просодических характеристик речи

Данная глава посвящена описанию экспериментальных исследований, направленных на выявление особенностей реализации фонетико-акустических и просодических явлений в естественной речи различных дикторов. Результаты проведённых исследований используются при разработке алгоритмических основ синтеза и клонирования речи путём создания общеязыковых и индивидуализированных баз данных.

В *первом разделе* приводятся методика и результаты сравнительных исследований внутрисловных и межсловных явлений в естественной русской речи на фонетическом и акустическом уровнях [150]. Полученные результаты позволили сформулировать алгоритмы преобразования «буква-фонема», «фонема-аллофон» и обосновать выбор расширенного набора аллофонов и мультифонов, используемых в процессе синтеза и клонирования фонетико-акустических явлений [151] в речевом потоке.

Во *втором разделе* приводятся результаты исследования фонетико-акустических особенностей сегментации последовательности слов на слоги [152, 153]. Предлагаются три различных способа сегментации речи на слоговые комплексы и обосновывается целесообразность их совместного использования в процессах синтеза и клонирования речи.

В *третьем разделе* описана методика анализа межъязыковых и междикторских различий интонации речи [154]. Обсуждаются особенности реализации интонационных контуров завершенности и незавершенности на примере русского и польского языков, показаны их межъязыковые и междикторские различия.

В *четвёртом разделе* описана методика исследования особенностей синтагматического членения речи на примере фонограмм речи трёх профессиональных дикторов [155]. Получены количественные оценки [156] числа акцентных единиц в синтагмах, длительности межсинтагменных пауз, вероятности сочетаемости синтагм и др. Проведенное исследование позволило сформулировать алгоритмические основы технологии клонирования синтагматических характеристик речи [157] и применения результатов при синтезе персонализированной речи по тексту [158, 159].

3.1. Исследование внутрисловных и межсловных фонетико-акустических явлений в русской речи

Современные требования, предъявляемые к синтезаторам речи по тексту, делают актуальной задачу высококачественного воспроизведения слитной речи не только для полного стиля речи, но и для стиля, близкого к разговорному. Одной из особенностей слитной речи является наличие специфических явлений на стыках слов внутри синтагмы, иногда существенно отличающихся от подобных внутрисловных явлений. Эти явления проявляются на фонетическом уровне (особенности преобразования «буква-фонема») и на акустическом (особенности преобразования «фонема-аллофон», или «фонема-мультифон»).

Внутрисловные закономерности изменения звуков русской речи к настоящему времени изучены достаточно подробно как на фонетическом, так и на акустическом уровне [160–163]. Построенные на их основе правила преобразования «буква-фонема» и «фонема-аллофон» успешно используются в существующих системах синтеза речи по тексту [65, 90, 163].

В данном разделе исследуются сравнительные особенности фонетических и акустических явлений внутри и на стыках слов слитной речи [150]. Учёт этих особенностей позволяет конкретизировать состав элементов компиляции синтезатора, необходимый для качественного клонирования фонетико-акустических свойств речи диктора.

3.1.1. Методика эксперимента

Для проведения исследования был создан специальный набор отдельных слов и набор пар слов, которые должны наиболее полно отображать внутрисловные и межсловные фонетико-акустические явления. В эти наборы вошли слова и пары слов, в которых реализованы звукосочетания различных пар фонем и их аллофонов, а именно: «гласная-гласная», «гласная-согласная», «согласная-гласная» и «согласная-согласная». Затем слова и пары слов из указанных наборов помещались в одну из двух стандартных фраз:

«Прочитайте слово **ИНТУИЦИЯ** в обычном темпе»

«Прочитайте словосочетание **ПИСЬМО ИВАНА** в обычном темпе» (подчёркиванием отмечено изучаемое звукосочетание).

Список таких фраз предлагалось зачитать нескольким дикторам. По полученным фонограммам с помощью программного пакета PRAAT [165] исследовались спектральные характеристики изучаемых звукосочетаний.

Для гласных фонем в слове выделены следующие изучаемые типы позиционных аллофонов:

- V_0 – полноударный;
- V_1 – частично-ударный;
- V_2 – первый предударный;
- V_3 – не первый предударный;
- V_4 – первый заударный;
- V_5 – не первый заударный.

Здесь V обозначает аллофон любой гласной (A, O, U, Y, E, I), V_0 – полноударный аллофон гласной, который употребляется в большинстве знаменательных слов, а V_1 – частично-ударный, который употребляется, как правило, в многосложных служебных словах, местоимениях, а также в сложных двухакцентных знаменательных словах. Употребление V_1 вместо V_0 может быть обусловлено особенностями акцентной структуры синтагмы или фразы.

Для согласных фонем выделяются следующие изучаемые типы аллофонов:

- C_{01} – глухие твёрдые губные: P, F ;
- C_{02} – глухие твёрдые не губные: T, C, S, SH, K, H ;
- C_{03} – глухие мягкие: $P', F', T, S', SH', CH', K', H'$;
- C_{11} – звонкие твёрдые губные: B, V, M ;
- C_{12} – звонкие твёрдые не губные: D, Z, N, L, ZH, R, G ;
- C_{13} – звонкие мягкие: $B', V', M', D', Z', N', L', R', G', J$.

3.1.2. Гласные в сочетаниях «гласная-гласная»

В отличие от внутрисловной позиции, где сочетания «гласная-гласная» встречаются достаточно редко, в позиции на стыках слов это явление проявляется весьма часто (например, на стыках с часто употребляемыми служебными словами: *на, о, и, не* и др.).

В соответствие с вышеприведенной классификацией позиционных аллофонов гласных внутри знаменательных слов возможны следующие сочетания:

$$\begin{array}{ll}
 V_0, V_4 & V_3, V_2 \\
 V_1, V_4 & V_3, V_3 \\
 V_2, V_0 & V_4, V_5 \\
 V_2, V_1 & V_5, V_5
 \end{array} \quad (3.1)$$

На стыке служебного и знаменательного слов (внутри фонетического слова) возможны следующие комбинации гласных (где символом «—» обозначается стык служебного и знаменательного слов):

$$\begin{array}{ll}
 \text{для проклитиков} & \text{для энклитиков} \\
 V_2-V_0 & V_4-V_5 \\
 V_2-V_1 & V_5-V_5
 \end{array} \quad (3.2)$$

$$\begin{matrix} V_3-V_2 \\ V_3-V_3 \end{matrix}$$

Сравнивая списки (3.1) и (3.2), можно отметить, что каждая из комбинаций гласных на стыке слов внутри фонетического слова (3.2) присутствуют также и в списке (3.1). На рис 3.1 (а, б) для сравнения приведено два примера реализации акустических явлений внутри и на стыке слов для комбинаций V_2V_0 , V_2-V_0 и V_3V_2 , V_3-V_2 . Как видно из приведенных примеров, динамика спектральных переходов внутри слова и на стыке служебного и знаменательного слов практически совпадает, что даёт основание к использованию в таких случаях последовательности одних и тех же аллофонов. Это правило подтверждается также и для других наборов слов и пар слов, в которых реализованы исследуемые сочетания «гласная-гласная».

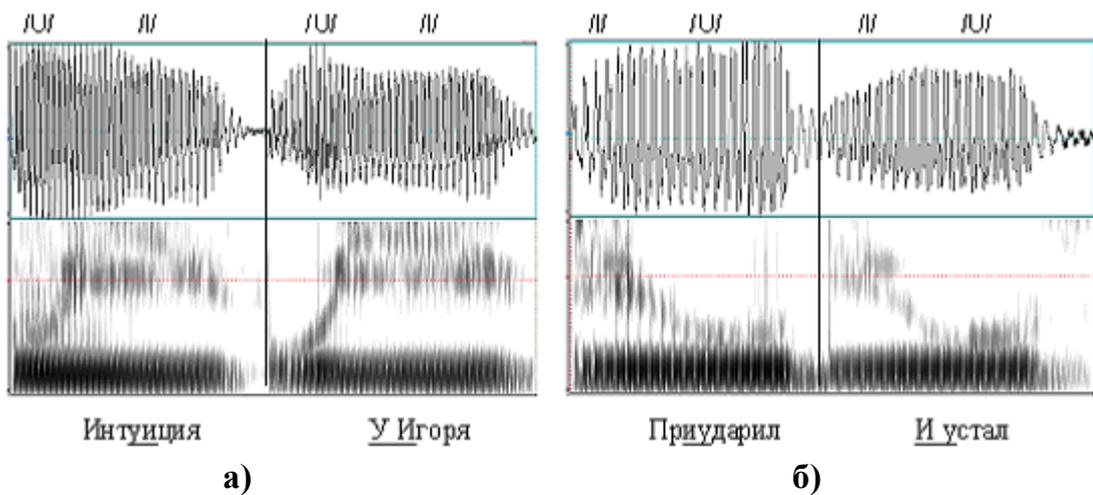


Рис. 3.1. Примеры реализации акустических явлений на стыке служебного и знаменательного слов а) для сочетаний гласных V_2, V_0 ; V_2-V_0 ; б) для сочетаний гласных V_3, V_2 ; V_3-V_2

На стыке двух знаменательных слов возможны следующие сочетания комбинаторных аллофонов гласных (где символом « » обозначается стык двух знаменательных слов):

$$\begin{matrix} V_0_V_0 & V_4_V_0 \\ V_0_V_1 & V_4_V_1 \\ V_0_V_2 & V_4_V_2 \\ V_0_V_3 & V_4_V_3 \\ V_1_V_0 & V_5_V_0 \\ V_1_V_1 & V_5_V_1 \\ V_1_V_2 & V_5_V_2 \\ V_1_V_3 & V_5_V_3 \end{matrix} \quad (3.3)$$

Сравнивая списки (3.3) и (3.1), необходимо отметить, что ни одна из комбинаций гласных на стыке слов из списка (3.3) не присутствуют в списке (3.1), что выдвигает требование использования при синтезе речи

специфических межсловных аллофонов.

На рис. 3.2 приведены примеры реализации акустических явлений на стыке слов для комбинаций $V_0_V_0$, $V_4_V_0$, $V_0_V_2$, $V_4_V_2$. При рассмотрении этих и других примеров становится очевидным, что на стыке двух знаменательных слов отсутствует даже какой-нибудь намёк на межсловную паузу, а динамика спектральных переходов вполне непрерывна. Это ещё раз подтверждает, что на стыках слов необходимо использовать последовательность специфических межсловных аллофонов гласных.

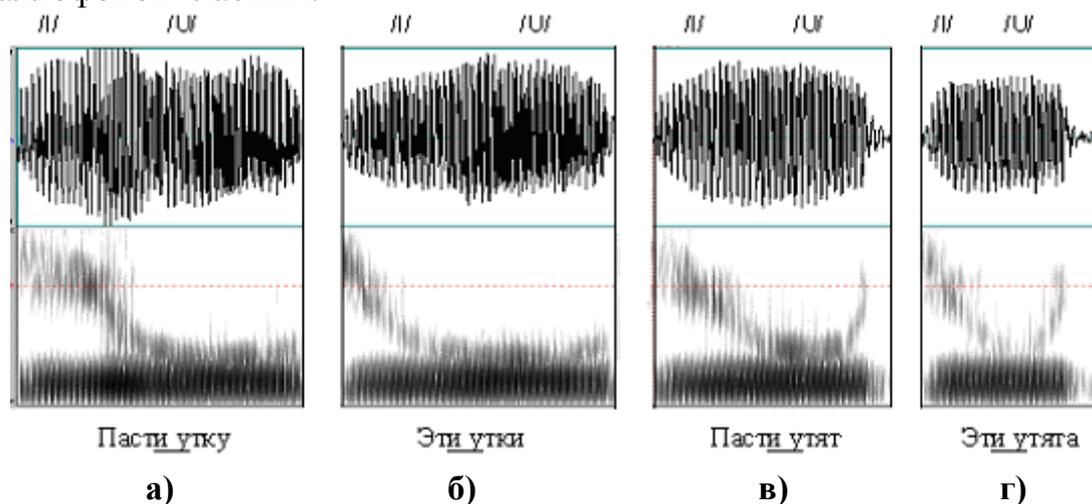


Рис. 3.2. Примеры реализации акустических явлений на стыке двух знаменательных слов а) для сочетаний гласных $V_0_V_0$; б) для сочетаний гласных $V_4_V_0$; в) для сочетаний гласных $V_0_V_2$; г) для сочетаний гласных $V_4_V_2$

3.1.3 Гласные в сочетаниях «согласная – гласная», «гласная - согласная»

На рис. 3.3 приведены примеры реализации акустических характеристик гласной, следующей после согласной, внутри слова (а, в) и на стыке двух знаменательных слов (б, г). Как видно из приведенных примеров, влияние предшествующей согласной на гласную внутри слова существенно сильнее, чем на стыке слов. Однако это влияние всё-таки присутствует, что проявляется не только при сравнении спектральных характеристик, но и на уровне слухового восприятия синтезированной речи.

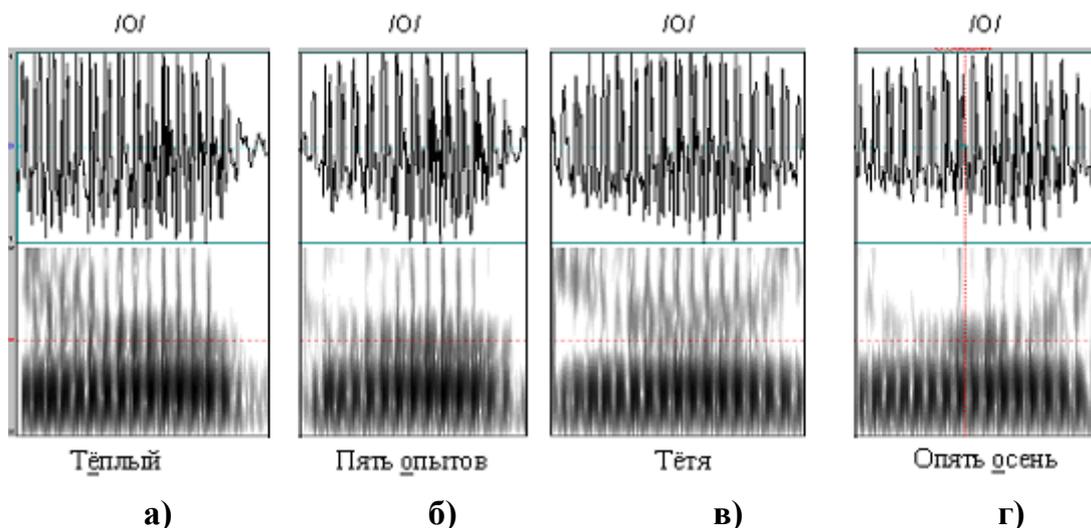


Рис. 3.3. Примеры акустических характеристик гласной внутри слова (а, в) и на стыке двух знаменательных слов (б, г) при реализации а) гласной V_{041} ; б) гласной V_{04_1} ; в) гласной V_{043} ; г) гласной V_{04_3}

Подобно рассмотренному выше случаю, акустические характеристики гласной изменяются также и под воздействием последующей согласной, как внутри, так и на стыке слов. На рис. 3.4 приведены примеры реализации акустических явлений на стыке слов (а, в) и внутри слова (б, г). Как видно из приведенных и других исследованных примеров, влияние последующей согласной на гласную внутри слова существенно сильнее, чем на стыке слов.

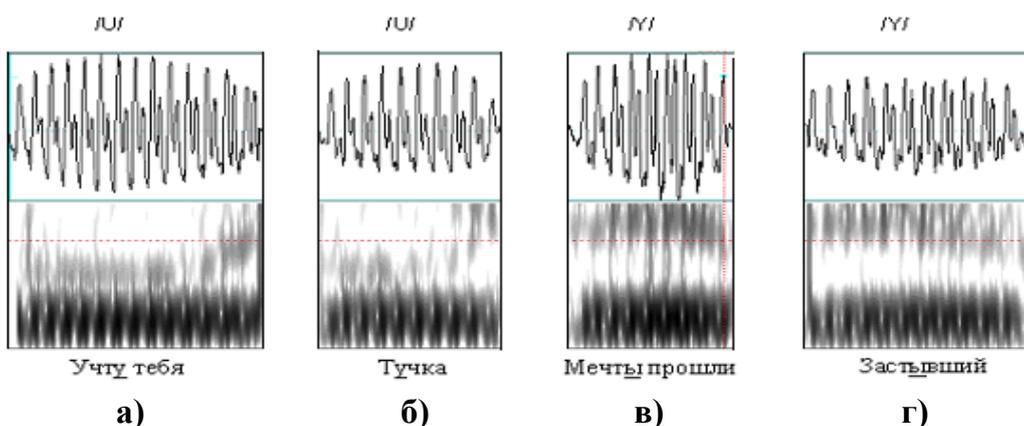


Рис. 3.4. Примеры акустических характеристик гласной внутри слова (а, в) и на стыке двух знаменательных слов (б, г) при реализации а) гласной V_{01_3} ; б) гласной V_{013} ; в) гласной V_{01_2} ; г) гласной V_{012}

Как показали эксперименты, при добавлении в акустическую базу синтезатора речи специфичных межсловных аллофонов для сочетаний «согласная-гласная», «гласная-согласная» синтезированная речь становится более естественной, «гладкой», менее «отрывистой».

3.1.4 Акустические характеристики согласных

Как показали исследования, акустические характеристики согласных внутри и на стыках слов различаются, в отличие от гласных, в значительно меньшей степени, так что межсловный переход не приводит к появлению сколько-нибудь значимых для восприятия специфических аллофонов. Это утверждение иллюстрируется рис. 3.5 – 3.7 на примерах различных межсловных и внутрисловных сочетаний согласных.

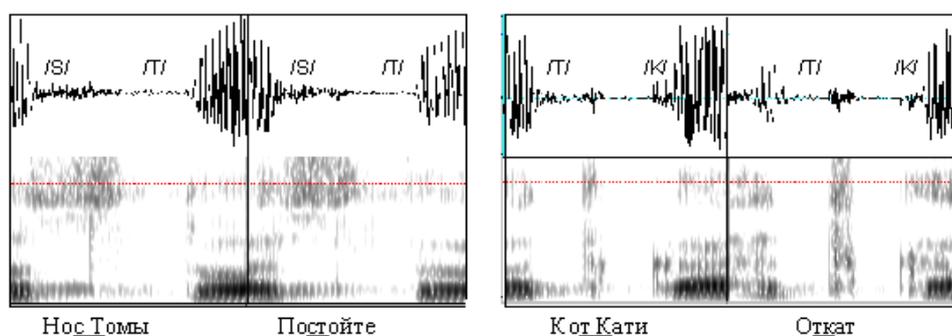


Рис. 3.5. Примеры реализации акустических явлений на стыке двух знаменательных слов и внутри слова для сочетаний согласных фонем «глухая – глухая»

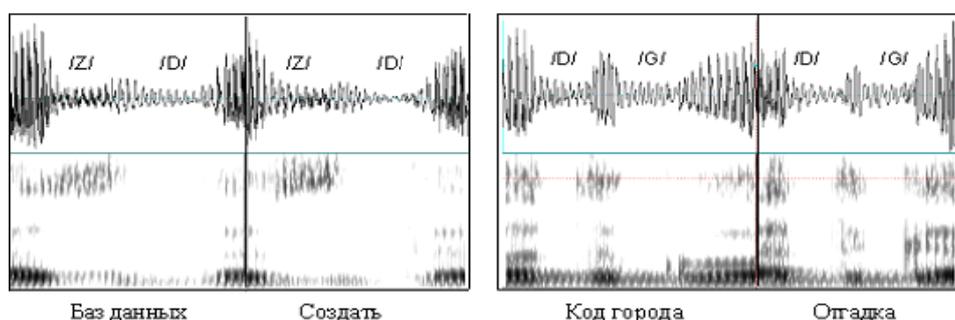


Рис. 3.6. Примеры реализации акустических явлений на стыке двух знаменательных слов и внутри слова для сочетаний согласных фонем «звонкая – звонкая»

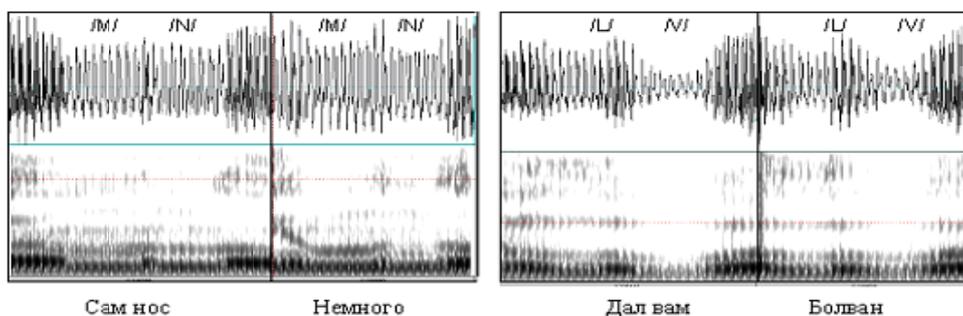


Рис. 3.7. Примеры реализации акустических явлений на стыке двух знаменательных слов и внутри слова для сочетаний согласных фонем «сонорная – сонорная»

3.2. Исследование фонетико-акустических особенностей сегментации последовательности слов на слоги

С точки зрения образования слог представляет собой звук или несколько звуков, произносимых одним выдохательным толчком. Со стороны акустической слог - это звуковой отрезок речи, в котором один звук выделяется наибольшей звучностью в сравнении с соседними - предшествующим и последующим. Гласные звуки, как наиболее звучные, являются слогообразующими, притягивая к себе соседствующие, менее звучные, согласные.

Согласно одному из определений слога [18], звуки в нём располагаются от наименее звучного к наиболее звучному. Закон восходящей звучности проиллюстрирован на приводимых ниже примерах, где степень звучности обозначена цифрами: 3 – гласные, 2 – сонорные согласные, 1 – остальные согласные и паузу.

Примеры: Во-да: 1-3/1-3; лод-ка: 2-3-1/1-3; ма-сло: 2-3/1-2-3; вол-на: 1-3-2/2-3; ин-тер-на-ци-о-наль-ный: 3-2/1-3-2/2-3/1-3/3/2-3-2/2-3-2.

Описанный закон слога деления, впервые предложенный в работах Л.В. Щербы [166], вызывал много споров относительно истинного положения границы слогораздела. В работах Л.В. Бондарко [167] предложено и фонетически обосновано более простое определение слога, когда граница слогораздела всегда устанавливается непосредственно после гласного звука, а сам слог получил название открытого слога.

Ввиду относительной автономности слогов, составляющих слово, они могут быть использованы наряду с аллофонами и диаллофонами как эффективные элементы компиляции при синтезе речи [152]. При этом, на наш взгляд, целесообразно использовать понятие открытого слога, предложенное Л.В. Бондарко, ввиду простоты и однозначности его определения.

Определение открытого слога, однако, может быть положено в основу разметки речевого корпуса на слоговые комплексы только с некоторыми уточнениями. Действительно, основным аргументом при выборе слога в качестве элемента компиляции является предположение об относительной лёгкости его вычленения в речевом потоке, т.е. об относительной большой величине контраста акустических параметров звуков (их спектрально-временных характеристик) на слоговых границах. Однако это не всегда так. Ослабление межслоговых контрастов может вызываться следующими факторами:

- взаимной ассимиляцией звуков;
- редукцией, вплоть до полного исчезновения, некоторых сонорных согласных, находящихся между гласными;

– редукцией, вплоть до полного исчезновения, безударных гласных, находящихся между согласными.

Для уточнения условий проявления указанных явлений были проведены экспериментальные исследования акустических характеристик слоговых составляющих на материале фонограмм речи нескольких дикторов, которыми был начитан текстовый корпус [168], подготовленный для создания БД элементов компиляции и клонирования речи (см. приложение 1). По полученным фонограммам определялись спектральные характеристики изучаемых слогов, по которым затем рассчитывались значения межслоговых контрастов (конечных разностей спектров) на границах различных типов слогов в соответствии со следующей формулой:

$$\Delta S(i) = \frac{1}{K} \sum_{k=0}^{K-1} [S(i+k) - S(i-1-k)] \quad (3.4)$$

где $\Delta S(i)$ – усредненная конечная разность спектра по времени;

K – интервал усреднения конечных разностей спектра;

$S(i)$ – i -й спектральный отсчет по времени.

В результате проведенного анализа оказалось, что для большинства типов слогов значения межслоговых контрастов имеют значимую величину, по которой межслоговая граница может быть определена с достаточной для практики точностью. Однако имеется определённая часть типов слогов, для которых, вследствие действия указанных выше 3-х факторов, межслоговая граница определяется не достаточно точно. К таким типам относятся слоги, в которых проявляется значительное ослабление межслогового контраста вследствие следующих причин:

– взаимной ассимиляции в сочетаниях «гласный+гласный» (см. пример на рис 3.8);

– взаимной ассимиляции в сочетаниях «гласный + /J'/'» (см. пример на рис 3.9);

– редукции межслогового сонанта /J'/' , вплоть до полного исчезновения, находящегося между безударными гласными (см. пример на рис 3.10).

– редукции безударных гласных, вплоть до полного исчезновения, находящихся между согласными, по крайней мере один из которых является глухим согласным (см.рис. 3.11 а,б,в).

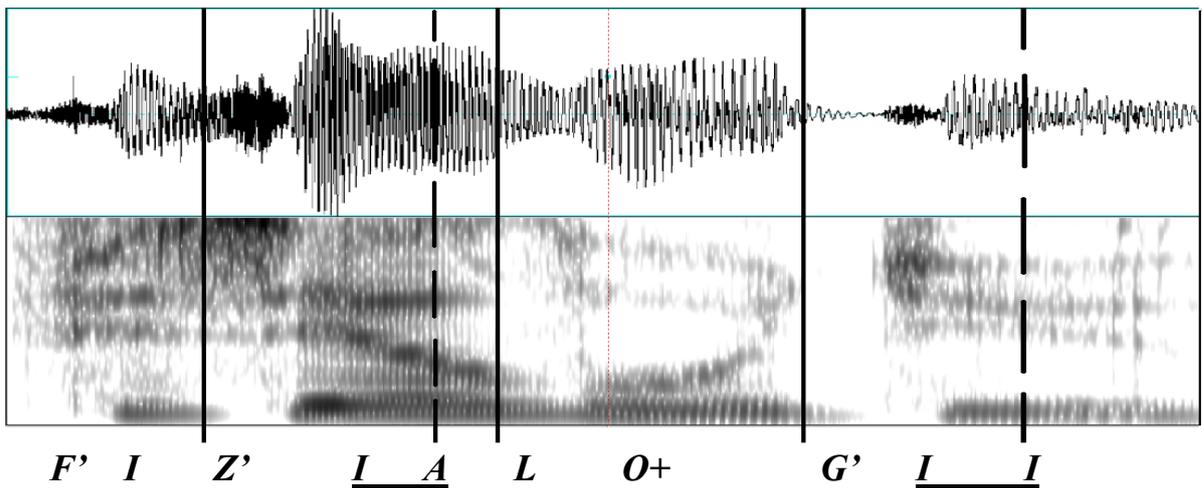


Рис.3.8. Пример взаимной ассимиляции звуков в сочетании «гласный+гласный»

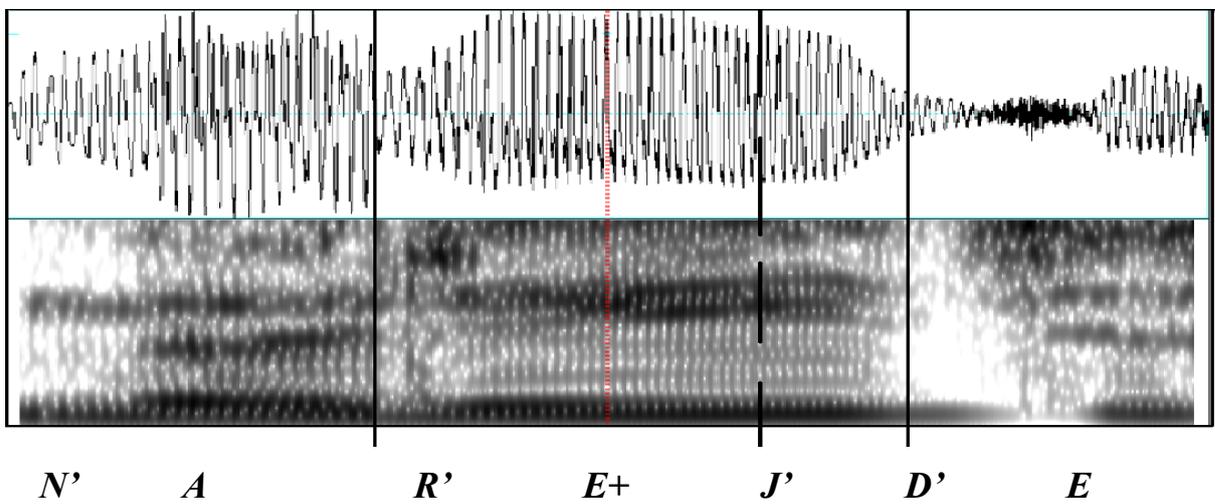


Рис.3.9. Пример взаимной ассимиляции звуков в сочетании «гласный+J'»

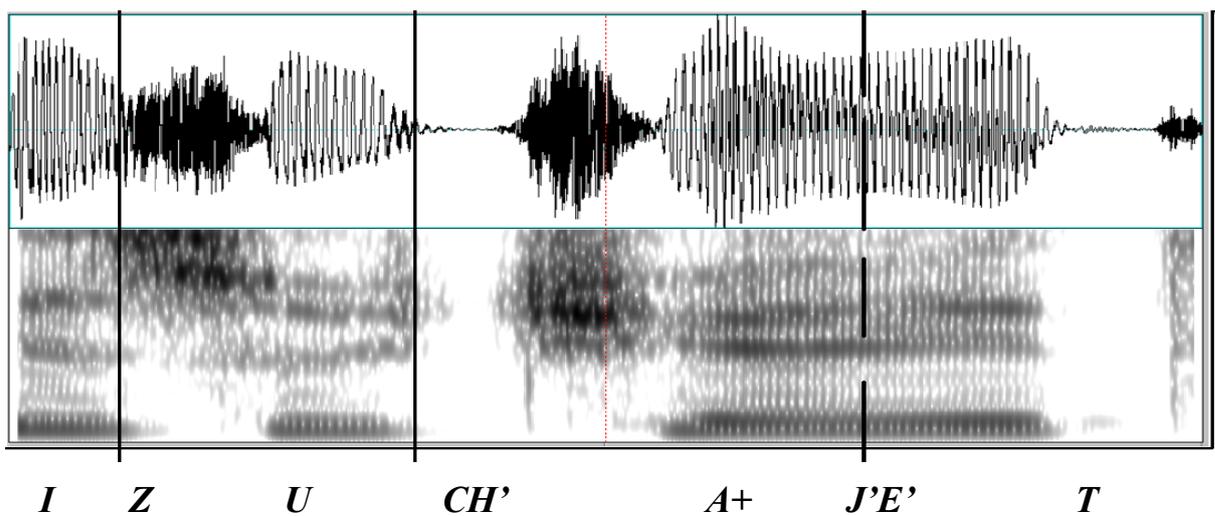
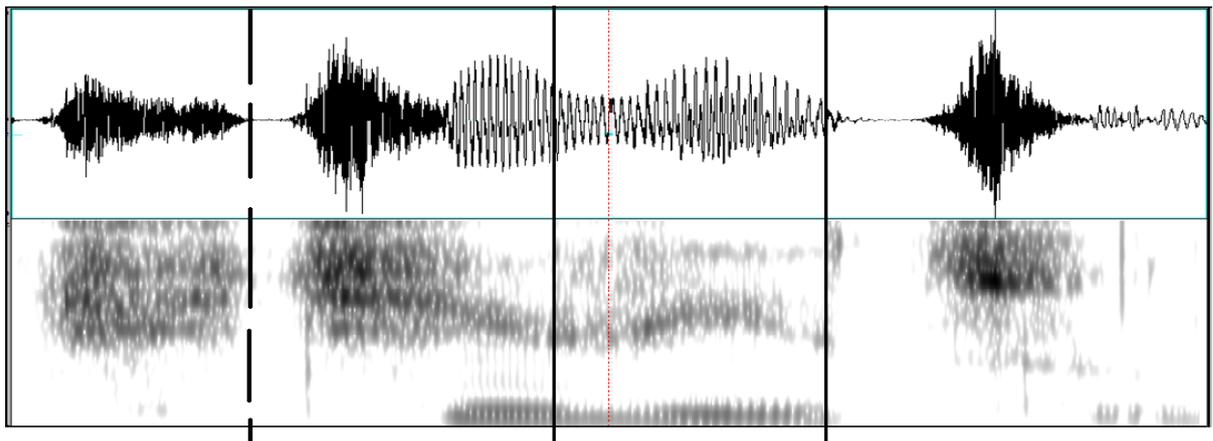
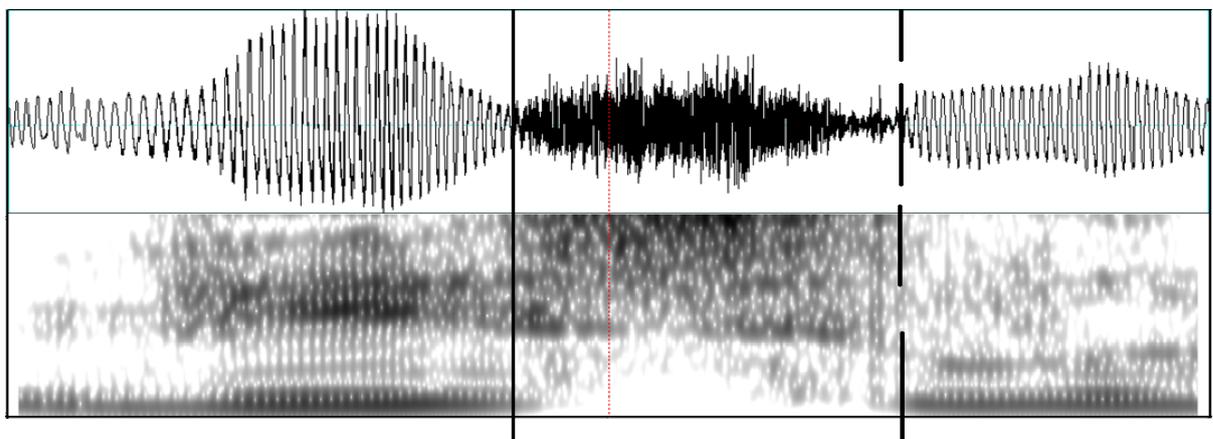


Рис.3.10. Пример редукции сонанта J' между безударными гласными



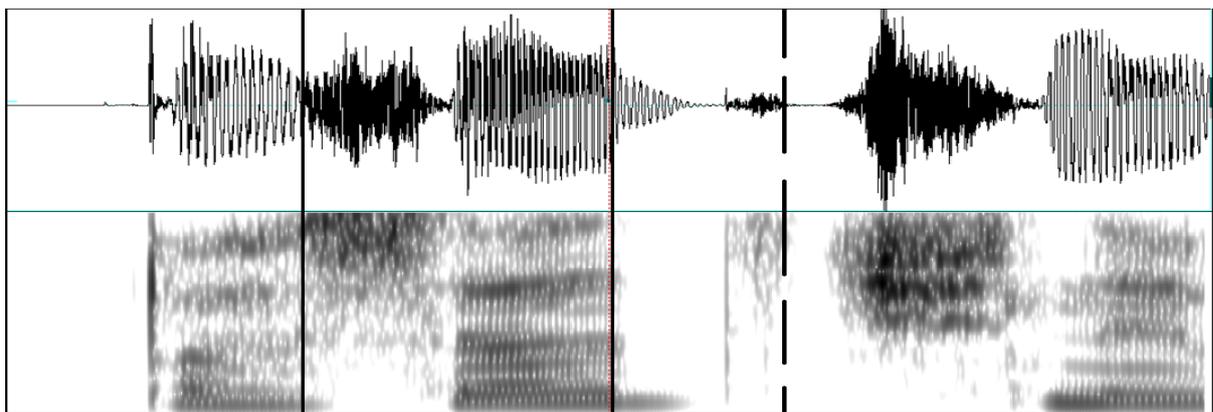
CH'E CH' E V' I+

а)



V' E+ S' E L A

б)



P A S A+ D A CH'' N Y J'

в)

Рис.3.11. Пример редукции безударного гласного, находящегося:
а) между глухими согласными, б) между глухим и звонким
согласными, в) между звонким и глухим согласными

При синтезе речи, основанном на компиляции аллослогов, целесообразно ввести определения слоговых комплексов 3-х типов.

1. Слоговым комплекс 1-го типа определяется как открытый слог в соответствии с правилом:

– граница слогораздела всегда устанавливается непосредственно после гласного звука.

Примеры:

В слове «физиологии» – «F', I, Z', I, A, L, O+, G', I, I» – будут выделены слоги: <F'I>, <Z'I>, <A>, <LO+>, <G'I>, <I> (здесь и в последующих примерах границы слога помечаются символами «<»», «>», ударная гласная – знаком +).

В слове «на рейде» – «N, A, R', E+, J', D', E» – будут выделены слоги: <NA>, <R'E+>, <J'D'E>.

В слове «изучает» – «I, Z, U, CH', A+, J', E, T» – будут выделены слоги: <I>, <ZU>, <CH'A+>, <J'ET>.

В слове «чечевицу» – «CH', E, CH', E, V', I+, C, U» – будут выделены слоги <CH'E>, <CH'E>, <V'I+>, <CU>.

В слове «весело» – «V', E+, S', E, L, A» – будут выделены слоги <V'E+>, <S'E>, <LA>.

В слове «посадочный» – «P, A, S, A+, D, A, CH', N, Y, J» – будут выделены слоги <PA>, <SA+>, <DA>, <CH'NYJ>.

2. Слоговым комплекс 2-го типа определяется в соответствии с правилом пункта 1 за исключением тех случаев, когда в слоге проявляются описанные выше эффекты ассимиляции гласных и согласных звуков. При этом действуют следующие 2 правила:

– если за гласным, определяющим конец слога, находится гласный, он присоединяется к текущему слогу.

– если за гласным, определяющим конец слога, следует не менее двух согласных, первый из которых – сонант J', а последующий – любой согласный, то граница определяется после первого из них.

Примеры:

В слове «физиологии» – «F', I, Z', I, A, L, O+, G', I, I» – будут выделены слоги: <F'I>, <Z'IA>, <LO+>, <G'II>.

В слове «на рейде» – «N, A, R', E+, J', D, E» – будут выделены слоги: <NA>, <R'E+J'>, <D'E>.

3. Слоговым комплекс 3-го типа определяется в соответствии с правилами пунктов 1, 2 за исключением тех случаев, когда в слог проявляются описанные выше эффекты редукции согласных и гласных звуков. При этом действуют следующие 2 правила:

– если за гласным, определяющим конец слога, находится последовательность «J' – безударный гласный», вся последовательность присоединяется к текущему слогу».

– безударный слог, содержащий гласный второй степени редукции, находящихся между согласными, по крайней мере один из которых является глухим согласным, присоединяется к предыдущему или последующему слогу, содержащему гласный меньшей степени»

Примеры:

В слове «изучает» – «I, Z, U, CH', A+, J', E, T» – будут выделены слоги: <I>, <ZU>, <CH'A+J'ET>.

В слове «чечевицу» – «CH', E, CH', E, V', I+, C, U» – будут выделены слоги <CH'ECH'E>, <V'I+>, <CU>.

В слове «весело» – «V', E+, S', E, L, A» – будут выделены слоги <V'E+S'E>, <LA>.

В слове «посадочный» – «P, A, S, A+, D, A, CH', N, Y, J» – будут выделены слоги <PA>, <SA+DA>, <CH'NYJ>.

На рис. 3.8 – 3.11 межслоговые границы, соответствующие строгому определению открытого слога показаны пунктирными линиями, а сплошными линиями – новые границы в соответствии с описанными выше правилами.

Разметка на слоговые комплексы может проводиться двумя способами: пословно и посинтагменно. В первом случае разметка осуществляется в отдельности для каждого фонетического слова, входящего в синтагму. Во втором случае осуществляется разметка всей последовательности слов в синтагме, рассматриваемой как единый речевой поток. С использованием такой разметки при синтезе речи автоматически учитываются межсловные фонетико-акустические явления, описанные в разделе 3.1.

3.3. Исследование межъязыковых и междикторских различий просодических характеристик речи

Исследование проведено в рамках создания многоголосой системы синтеза речи по тексту на славянских языках [31, 169, 170]. Целью исследования являлось изучение особенностей реализации интонационных контуров завершенности и незавершенности в речи нескольких дикторов на примере русского и польского языков для одного из интонационных стилей – чтения научного текста. Представление мелодических контуров осуществлялось в рамках интонационной ПАЕ-модели [81], принципы функционирования которой описаны в Главах 1, 2.

3.3.1. Методика создания мелодических портретов акцентных единиц

Для создания речевого материала несколькими русскоязычным и польскоязычным дикторам предлагалось начитать тексты

приблизительно одинакового научного содержания. Объем текстов на каждом из языков составлял около 300 слов. Ниже приведены фрагменты используемых текстов.

Фрагмент русскоязычного текста:

«Спектр результирующего речевого сигнала может быть представлен в виде произведения спектра источника звука на передаточные функции речевого тракта и каналов передачи звуковых сигналов, плюс – спектры различного рода сигналов акустических помех».

Фрагмент польскоязычного текста:

«Widmo sygnału mowy obliczamy przez pomnożenie aparatu wokalnego, źródła mowy, stylu mowy, mnożymy to też razy mikrofon, czyli jakie występują tam wypaczenia i dodajemy do tego wszystkiej lokalną akustykę».

Полученные в результате записи звуковые файлы в формате WAVE PCM подвергались дальнейшему анализу.

Этапы исследования интонационных характеристик для создания набора мелодических портретов показаны на рис. 3.12.

Прежде всего, речевой материал прослушивается, из него удаляются неинформативные участки: шумы, вдохи, «лишние» звуки (например, «э-э-э»). Затем каждая фраза фонограммы и соответствующий ей текст предложения стенограммы обрабатывается опытными экспертами-фонетистами: разбивается на синтагмы, акцентные единицы и определяется, к какому интонационному типу принадлежит синтагма: завершённому или незавершённому. В фонограмме границы синтагм отмечаются регионами, в стенограмме – знаком «//», границы АЕ отмечаются, соответственно, маркерами и знаком «/», интонационный тип завершенности обозначается точкой, незавершенности - запятой. В стенограмме, кроме того, для каждого слова АЕ указываются сильное (обозначается знаком /+/) или слабое (обозначается знаком /=/) ударения.

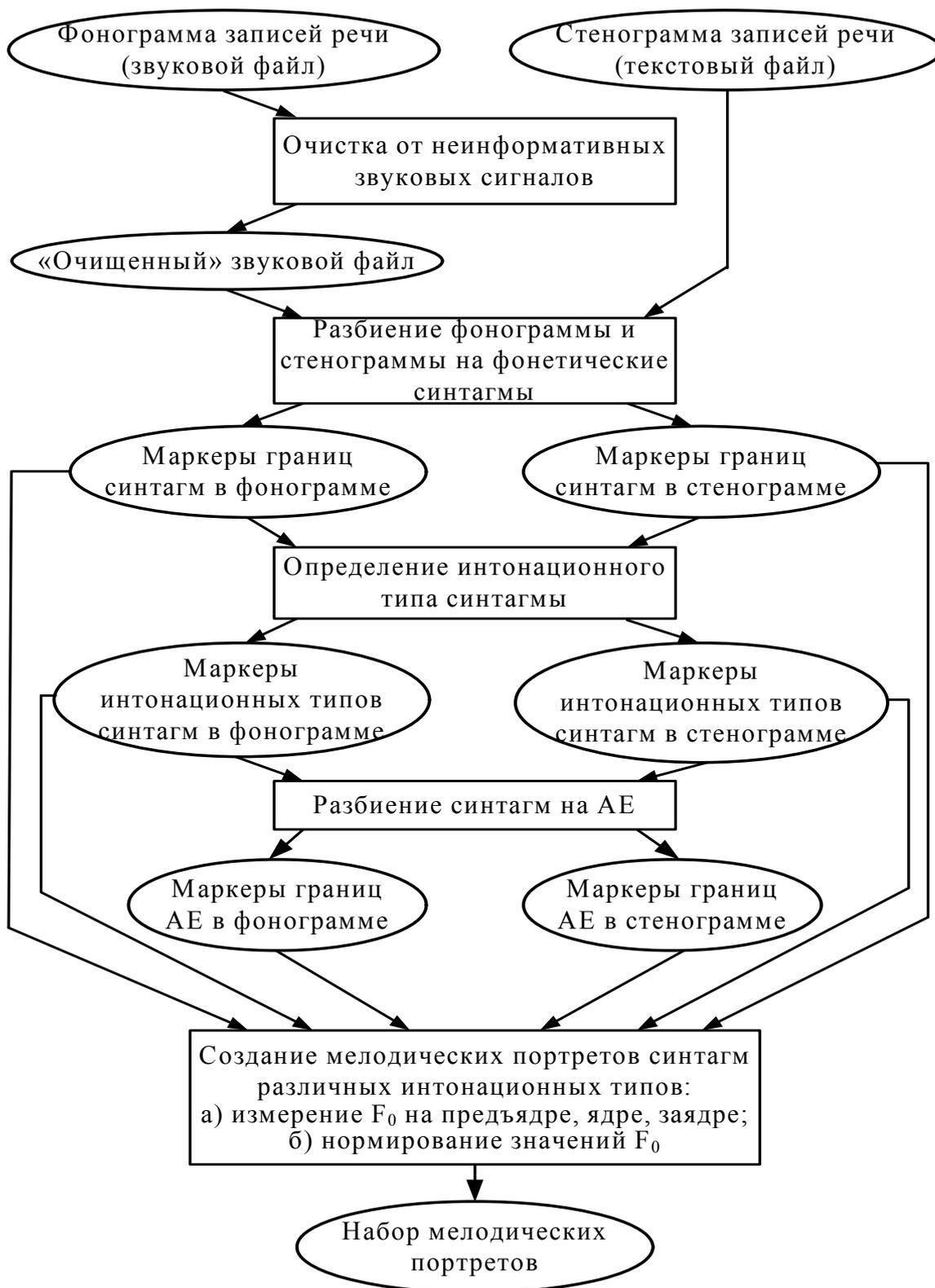


Рис.3.12. Процедура создания мелодических портретов

Фрагменты обработанных таким образом стенограмм для русского и польского языков показаны в табл. 3.1 и 3.2 соответственно.

Таблица 3.1

Фрагмент русского текста с делением на синтагмы, АЕ, с указанием интонационного типа

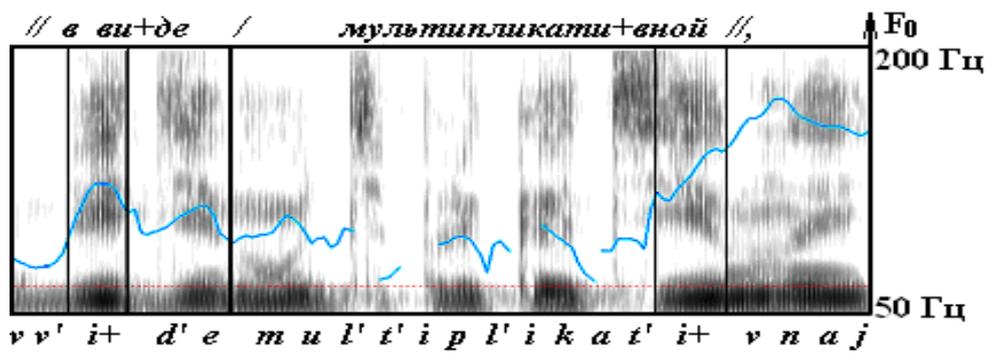
Маркированный текст синтагмы	Количество АЕ	Интонационный тип
/таки=м о+бразом //	1	(,)
/спе+ктр / речево=го сигнала //	2	(,)
/зави+сит / не то+лько //	2	(,)
/от его= смыслово+го / содержа+ния //	2	(,)
/но= та+кже / от ви+да / ре+чи //	3	(,)
/индивидуа=льных осо+бенностей / ди+ктора //	2	(,)
/и= характери+стик / акусти=ческой среды+ //	2	(,)
/в кото=рой распростра+няется / речево+й / сигнала //	3	(.)

Таблица 3.2

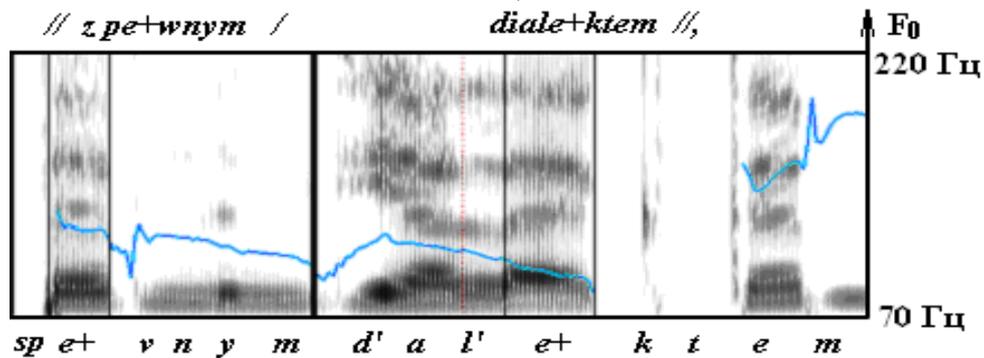
Фрагмент польского текста с делением на синтагмы, АЕ, с указанием интонационного типа

Маркированный текст синтагмы	Количество АЕ	Интонационный тип
/oso+ba //	1	(,)
/któ=ra mó+wi //	1	(,)
/mo+że / mó+wić //	2	(,)
/z re+wnym / diale+ktem //	2	(,)
/lu=b pod wpły+wem / jaki+ś / emo+cji //	3	(.)

Следующий этап обработки – получение интонационных контуров синтагм – значений частоты основного тона F_0 на вокализованных участках речи – может осуществляться с использованием программного пакета PRAAT [165]. Каждая АЕ синтагмы предварительно делится на ядро – ударный гласный, предъядро – участок АЕ перед ударным гласным, и заядро – участок АЕ после ударного гласного. На рис. 3.13, 3.14 приведены сравнительные примеры мелодических контуров (значений F_0) 2-хакцентных синтагм с интонацией незавершённости (рис. 3.13) и завершённости (рис. 3.14) для русского (рис. 3.13а, рис. 3.14а) и польского (рис. 3.13б, рис. 3.14б) языков. На рисунках указаны границы АЕ, а также области предъядра, ядра и заядра каждой АЕ.

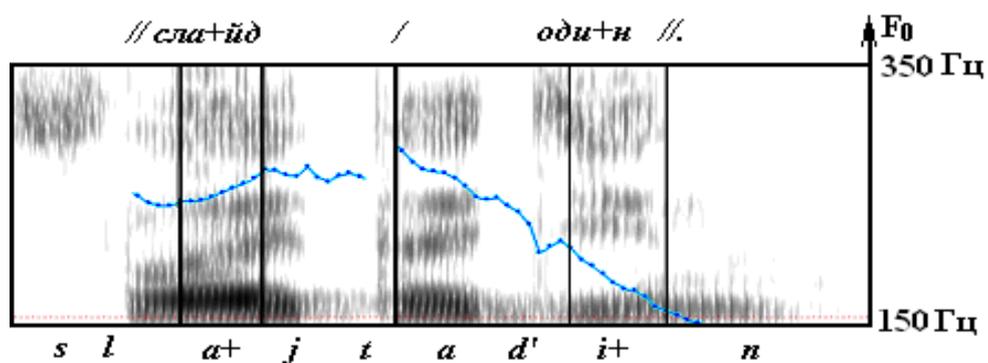


а)

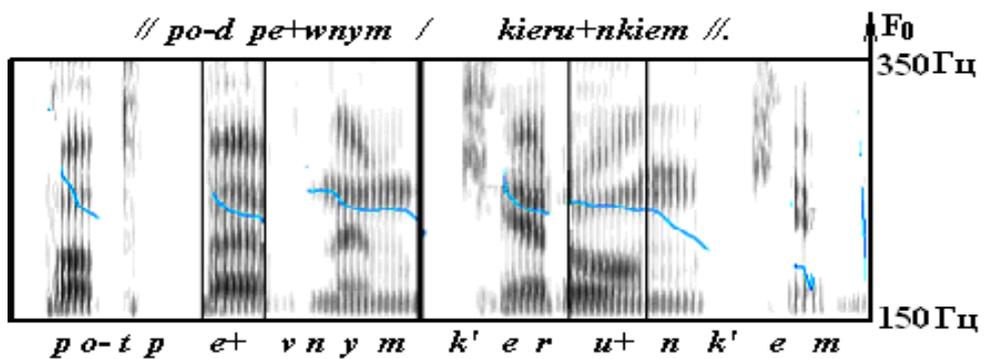


б)

Рис. 3.13. Примеры интонационных контуров незавершённости для синтагм: а) «в виде мультипликативной», б) «зрелым диалектом»



а)



б)

Рис. 3.14. Примеры интонационных контуров завершённости для синтагм: а) «слайд один», б) «по-прежнему kierunkiem»

Согласно интонационной ПАЕ-модели, мелодический портрет представляется нормированными значениями F_0 на участках предъядра, ядра и заядра каждой АЕ синтагмы.

Процесс создания мелодического портрета синтагмы показан на рис. 3.15 на примере фразы “которые могут быть представлены”, произнесённой мужским голосом. Данная фраза является синтагмой с интонацией незавершённости, состоящей из 3 АЕ.

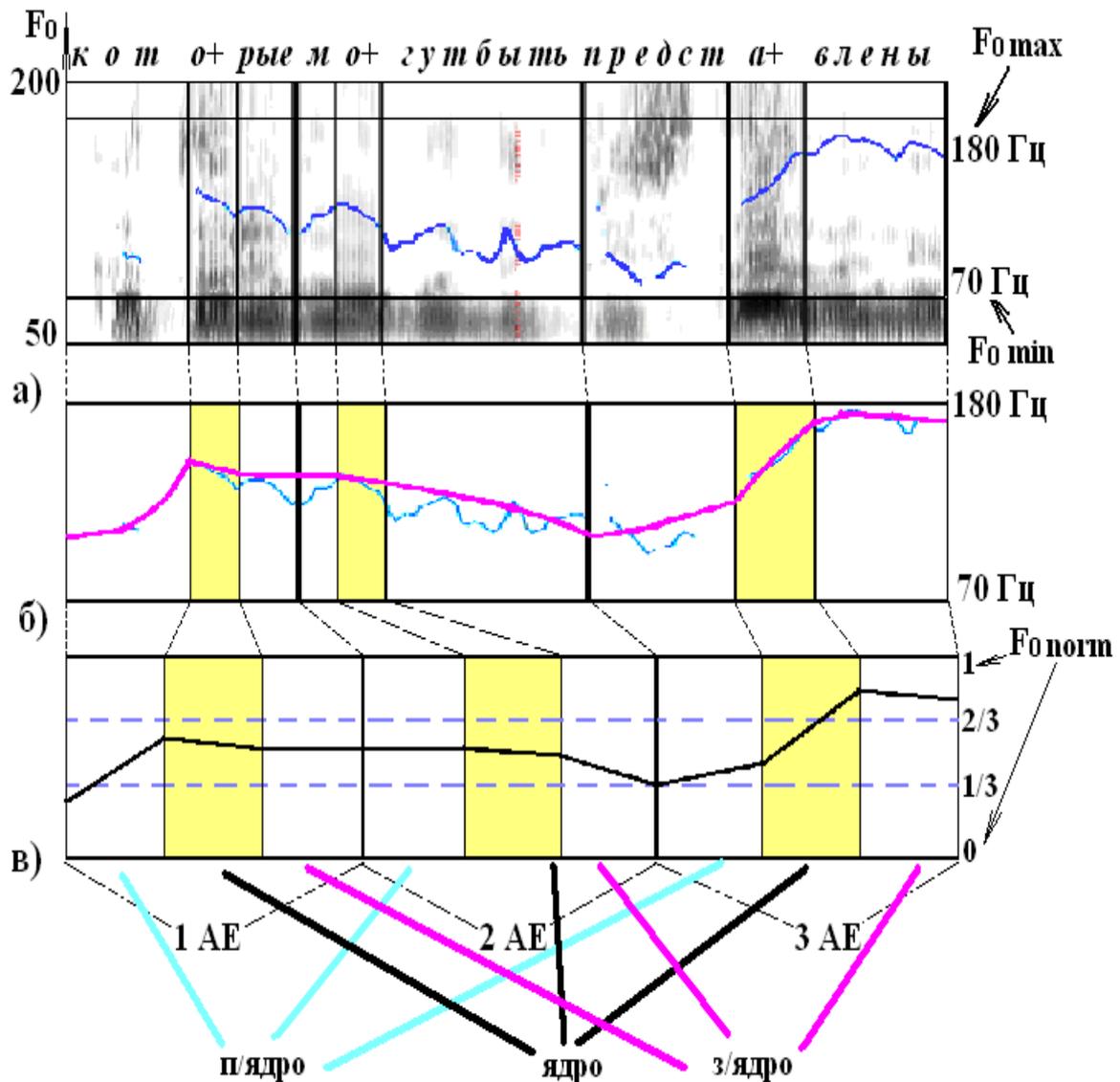


Рис. 3.15. Процесс создания мелодического портрета синтагмы

- а) спектрограмма и интонограмма речевого сигнала, границы АЕ и участков предъядра, ядра, заядра; б) экстраполяция контура F_0 ;**
- в) нормированный «портрет» контура F_0**

Для речевого сигнала вычисляются значения F_0 на всех вокализованных участках (рис. 2.15а), осуществляется разметка границ АЕ и границ областей предъядра, ядра, заядра для каждой АЕ, а также

интерполяция значений на невокализованные участки (рис. 2.15б), и, наконец, осуществляется нормализация каждой из областей по длительности и частоте.

Нормализация по длительности осуществляется уравниванием длительности областей предъядра, ядра и зядра. Для нормализации по частоте определяются минимальное – $F_{0\ min}$ – и максимальное – $F_{0\ max}$ – значения на всей исследуемой фонограмме. Нормированное значение F_0 вычисляется согласно формуле:

$$F_{0\ norm} = \frac{F_0 - F_{0\ min}}{F_{0\ max} - F_{0\ min}} \quad (3.5)$$

В данном примере $F_{0\ min}$ и $F_{0\ max}$ были определены равными, соответственно, 70 Гц и 180 Гц.

Результатом описанных операций является создание последовательности нормированных мелодических портретов АЕ, составляющих синтагму (см. рис. 2.15в).

3.3.2. Результаты исследования межъязыковых различий

Специфика исследуемого интонационного стиля – чтение научного текста – такова, что главное внимание было уделено наиболее «массовым» явлениям в речи – интонационным типам незавершённости и завершённости. Другие интонационные типы, такие как вопрос или восклицание, остались вне рассмотрения. В соответствии с описанной выше методикой были построены мелодические портреты различных подтипов интонаций незавершённости и завершённости для русской и польской речи. На рис. 3.16 отражены наиболее частотные области локализации мелодических контуров конечной АЕ для интонаций незавершённости и завершённости в русской и польской речи, полученные в процессе реализации описанной выше методики построения ПАЕ.

Известно [171, 172], что наибольшую информационную нагрузку несёт интонационный контур конечной АЕ синтагмы, в котором наиболее ярко проявляются особенности того или иного интонационного типа. Как видно из рисунков, мелодические ПАЕ интонационных типов незавершённости и завершённости весьма существенно отличаются в русской и польской речи.

Интонация незавершённости, характеризующаяся в обоих случаях восходящим тоном, реализуется в русском языке, в основном, на ядерном участке АЕ, в то время как в польском – на зядерном участке.

Интонация завершённости, характеризующаяся в общем случае нисходящим тоном, также реализуется в русском языке, в основном, на ядерном участке АЕ, в то время как в польском – на зядерном участке.

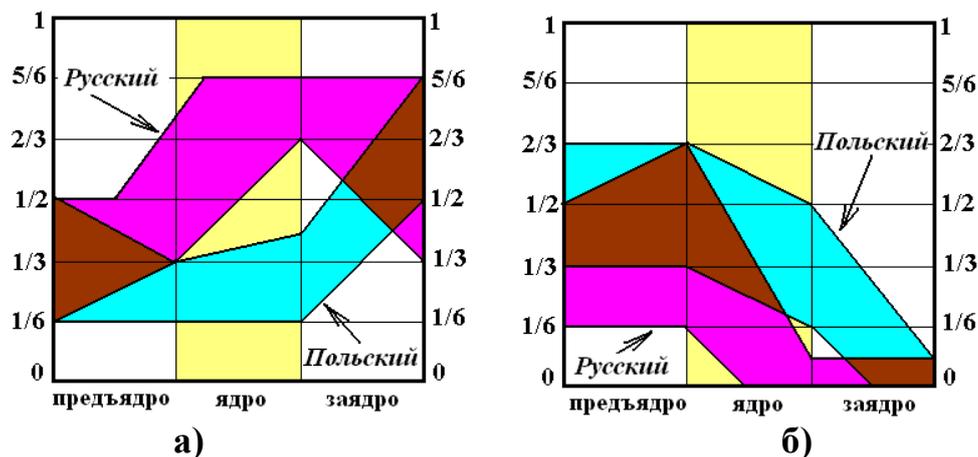
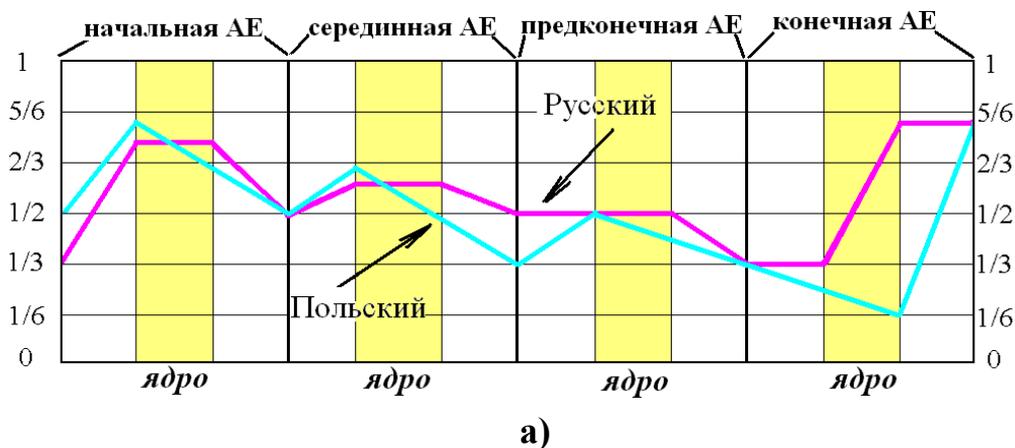


Рис. 3.16. Закономерности реализации мелодических контуров ПАЕ для конечной АЕ: а) для интонации незавершённости, б) для интонации завершённости

Такое явление может быть легко интерпретировано, исходя из того факта, что в польском языке практически всегда присутствует заядерный участок слова (как правило, ударение падает на предпоследний слог), в то время как в русском языке заядерный участок слова очень часто может вообще отсутствовать (свободная позиция ударения).

Закономерности реализации мелодических контуров для других АЕ (начальной, срединной и предконечной) в четырёхакцентных синтагмах для интонаций незавершённости и завершённости в русской и польской речи представлены на рис. 3.17.



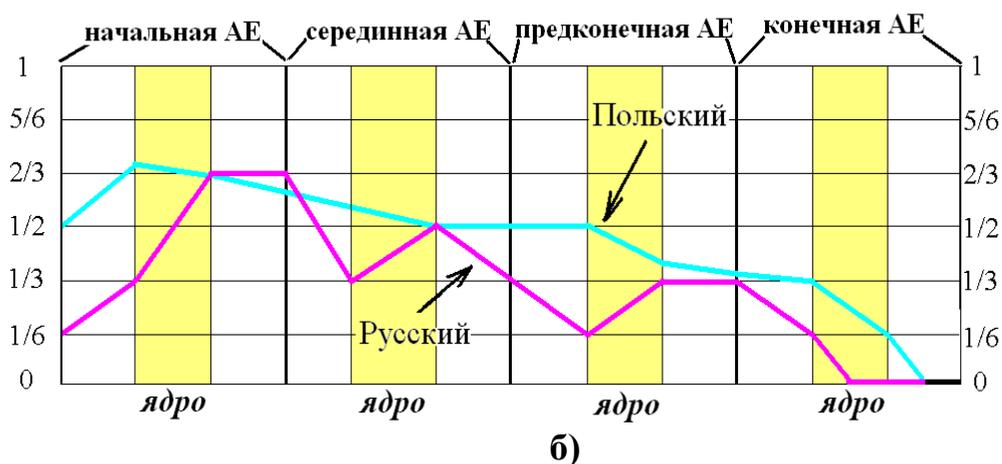


Рис. 3.17. Закономерности реализации мелодических контуров для четырёхакцентных синтагм: а) для интонации незавершённости, б) для интонации завершённости

Из представленных графиков видно, что мелодические контуры русской и польской речи различаются не только на конечной АЕ синтагмы, но, хотя и не в столь сильной степени, на начальной, срединной и предконечной АЕ.

3.3.3. Результаты исследования междикторских различий

Изображенные на рис. 3.16 интонационные портреты получены на базе анализа фонограмм чтения описанных выше текстов 3-мя русскоязычными и 3-мя польскоязычными дикторами. Индивидуальные дикторские особенности реализации интонаций незавершённости и завершённости в русской и польской речи, при сохранении отмеченных выше межъязыковых различий, укладываются достаточно точно в пределы областей, отмеченных на рисунке 3.16. Выход за пределы этих областей и их пересечение может в определённой степени интерпретироваться как проявление межъязыковой интерференции или акцента.

Для русского языка были проведены более детальные дополнительные исследования междикторских различий. Ниже дана иллюстрация различий на примере интонационного типа «вопрос-уточнение». Исследовались динамические характеристики интонационного контура F_0 в одноакцентной синтагме «Художник?» в контексте: «Кто её нарисовал? Художник?». Речевой сигнал исследуемой синтагмы «вырезался» из фонограмм одного и того же текста, начитанного 2-мя профессиональными дикторами (Д1-М - мужчина и Д1-Ж – женщина) и 2-мя непрофессиональными (Д2-М и Д2-Ж). На рис. 3.18 приведены их интонограммы, полученные с помощью программного пакета PRAAT [165].

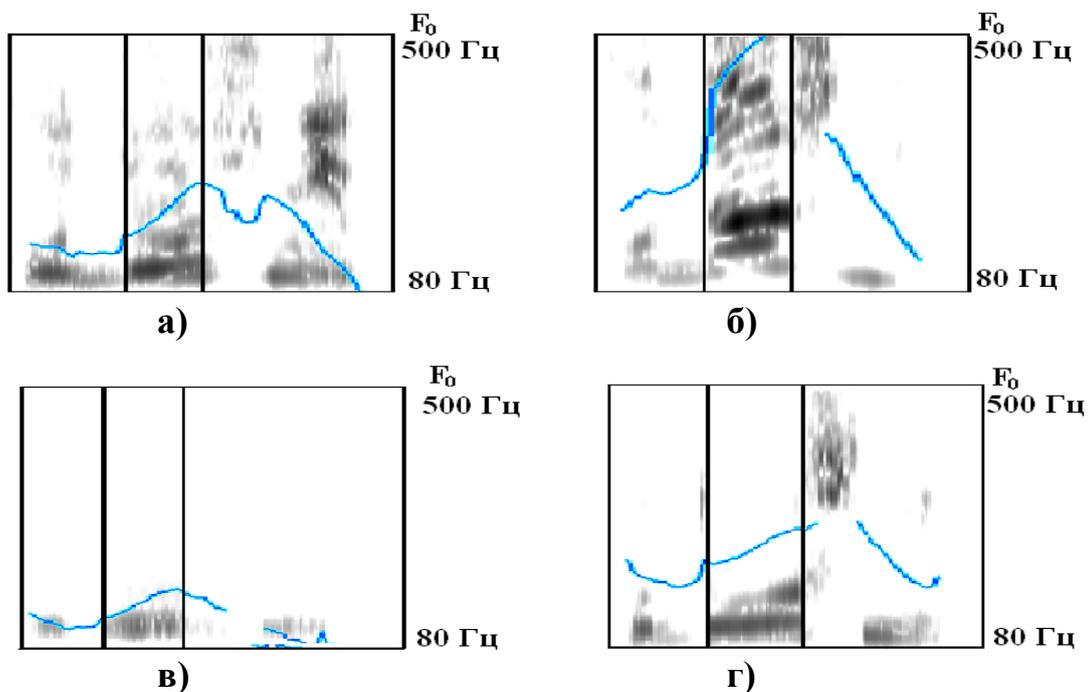


Рис. 3.18. Интонограммы одноакцентной синтагмы вопросительного типа, произнесённой дикторами: а) Д1-М, б) Д1-Ж, в) Д2-М, г) Д2-Ж

Как видно из рис. 3.18, наиболее яркие междикторские различия проявляются в абсолютных значениях $F_{0\max}$ и $F_{0\min}$. Более определённо эти различия представлены на плоскости $\{F_{0\max}; F_{0\min}\}$ (см. рис. 3.19).

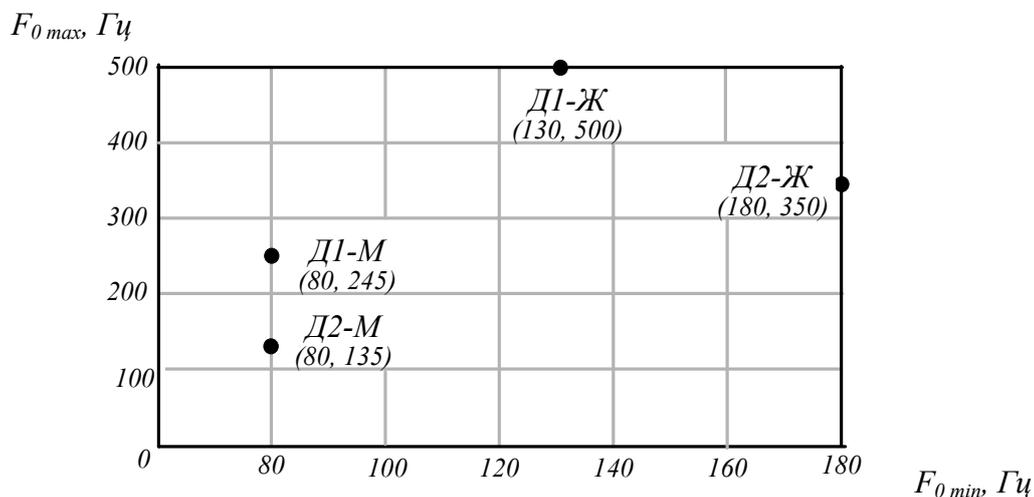
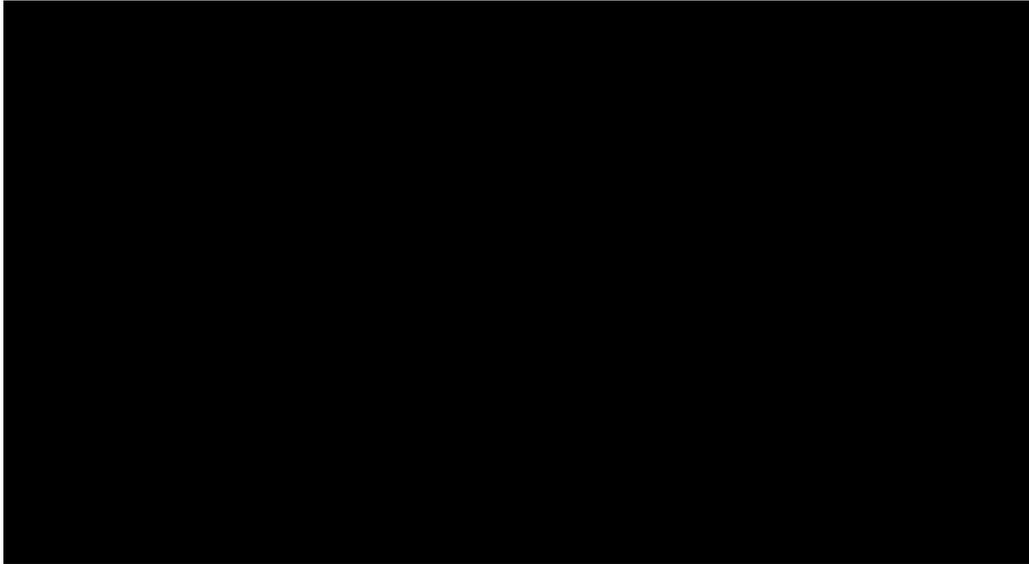


Рис. 3.19. Абсолютные значения $F_{0\min}$ и $F_{0\max}$ четырёх дикторов

На рис. 3.20 показаны также междикторские различия, проявляющиеся при анализе относительного диапазона изменений F_0 у дикторов, рассчитанного в соответствие с выражением:



$$D_n = 1 - \frac{F_{0 \min}}{F_{0 \max}} \quad (3.6)$$

Рис.3.20. Относительный диапазон изменения F_0 для четырёх дикторов

Анализируя рис. 3.20, необходимо отметить существенно более высокий диапазон изменений F_0 у профессиональных дикторов в сравнении с непрофессиональными.

Междикторские различия отчётливо проявляются также в нормированных интонационных портретах (рис. 3.21), построенных в соответствии с методикой, описанной выше.

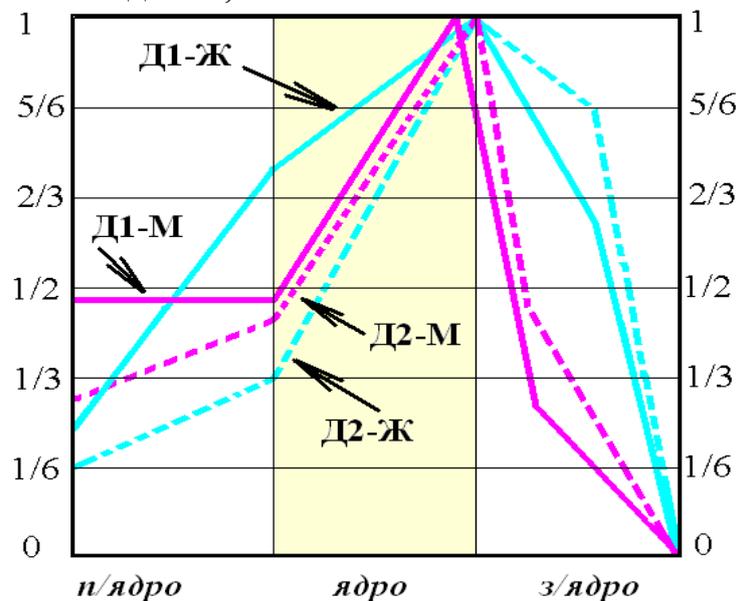


Рис. 3.21. Мелодические портреты одноакцентной вопросительной синтагмы для четырёх дикторов

Для исследования персональных особенностей реализации ПАЕ для четырёх различных интонационных типов: незавершённость, завершённость, восклицание и вопрос, были проведены дополнительные исследования в соответствии со следующей методикой. Два профессиональных диктора радио (Олег и Светлана) и три непрофессиональных (Борис, Елена, Лилия) зачитали один и тот же отрывок художественного текста. Затем в фонограммах записей каждого диктора были выделены одни и те же участки речи, на которых ими были реализованы указанные четыре интонационных типа, и на основании анализа контуров F_0 построены ПАЕ в соответствии с разработанной методикой. На рис. 3.22 представлены полученные мелодические портреты конечных АЕ для четырёх интонационных типов 5-ти дикторов.

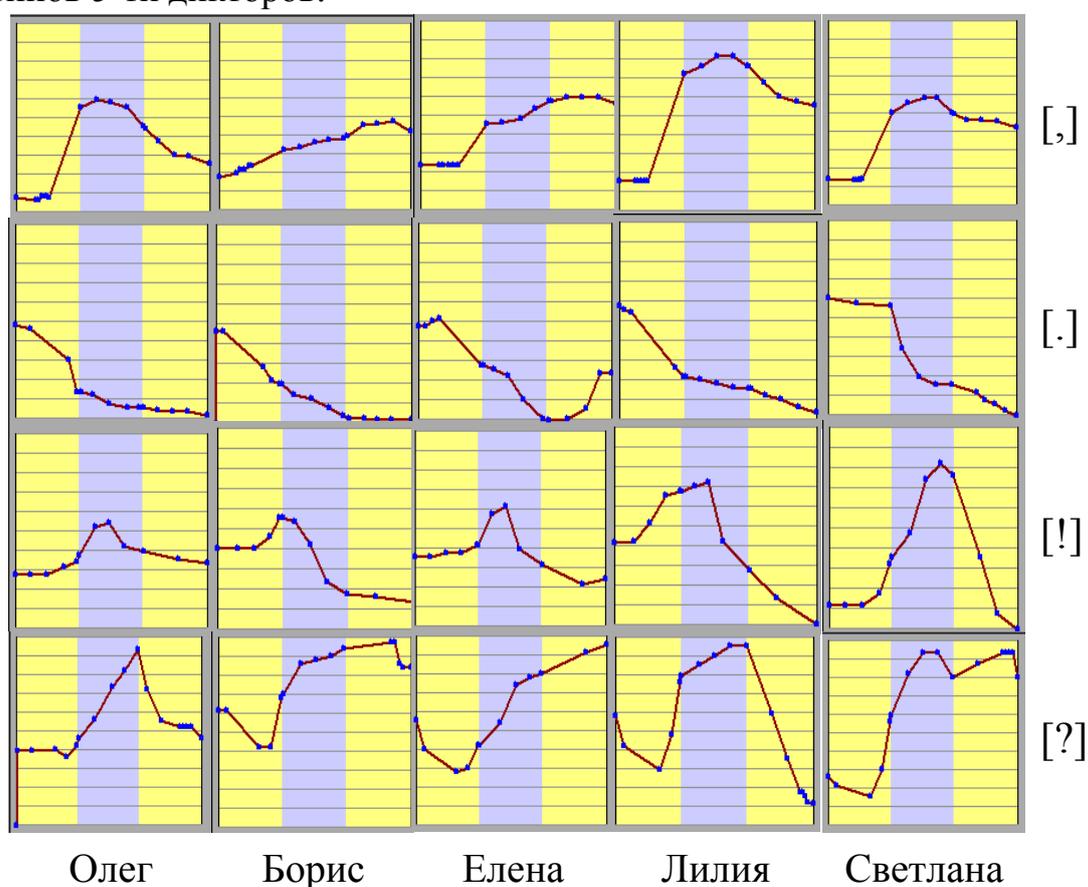


Рис. 3.22. Мелодические портреты конечной АЕ для четырёх интонационных типов 5-ти дикторов

Как видно из рис. 3.22, полученные ПАЕ имеют достаточно ярко выраженные персональные особенности. Причём в наибольшей степени индивидуальные дикторские различия в ПАЕ проявляются на предъядре и заядре, в то время как на ядерных участках они менее значительны. В целом же, однако, сохраняется рисунок портретов, характерный для каждого исследуемого интонационного типа.

Проведенное исследование позволило создать основной набор мелодических портретов интонации и разработать основы клонирования персональных просодических характеристик речи.

3.4. Исследование персональных особенностей синтагматического членения речи

Под синтагмой, как уже было сказано ранее, понимаются элементы фразы, которые обладают определенной самостоятельностью, а также определенной ритмической и мелодической структурой, и которые допускают некоторую паузу после того, как они были произнесены. Членение одного и того же предложения на синтагмы может быть различным в зависимости от контекста, ситуации, индивидуальной экспрессивной окраски, придаваемой высказыванию говорящим. В данном разделе описываются результаты экспериментального исследования персональных особенностей синтагматического членения речи 3-х дикторов: телеведущего Ю. Сенкевича (Ю.С.), проведенного на базе фонограмм ТВ-передач «Клуба кинопутешественников», и двух других профессиональных дикторов (Д1-М и Д2-Ж) – на базе студийных записей чтения ими специальных, фонетически сбалансированных текстов, описанных в работе [173]. Длительность записи для каждого из дикторов составляла порядка 15 минут, что соответствует примерно 1000 словам орфографического текста стенограммы. Основная цель исследования заключалась в создании алгоритмических основ персонализированного членения на синтагмы при клонировании персональных просодических характеристик речи диктора. Другая цель заключалась в сравнительной оценке степени различий в характеристиках синтагматического членения речи для разных дикторов.

3.4.1. Методика эксперимента

Проведение эксперимента осуществляется в несколько этапов.

1-й этап – предварительная обработка фонограмм и текста.

Прежде всего, расшифровывается и записывается дословный текст фонограммы. По результатам расшифровки производится коррекция текста и фонограммы. Из них убираются ошибочно произнесённые слова и звуки, участки с разного рода помехами (вдохи, шум, музыка, слова с малым уровнем звука и др.). Проводится при необходимости корректировка акустических характеристик фонограммы (выравнивания звуковых уровней, корректировка амплитудно-частотных характеристик звуков).

2-й этап – Разбиение фонограмм и текста на синтагмы Путём последовательного прослушивания фонограммы экспертом-фонетистом

на фонограмме и в тексте отмечаются границы синтагм. Решение о наличии конца синтагмы принимается на основе ряда признаков, таких как: присутствие дыхательной паузы, комплексная реализация одного из возможных интонационных типов синтагмы, наличие определённой динамической структуры (контура силы звука) и определённой ритмической структуры (контура длительности звуков). При членении фонограммы на синтагмы во внимание принимается также присутствие знаков препинания в соответствующем ей тексте, а также некоторых других формальных признаков текста.

3-й этап – Просодическая маркировка синтагм

После членения фонограммы на синтагмы каждая из них последовательно прослушивается и маркируется следующим образом. Каждое слово синтагмы маркируется ударением, т.е. указывается место ударения и его тип: сильное (+), слабое (=) или ударение отсутствует. Слова, в которых ударение отсутствует, объединяются в тексте с одним из соседних слов в единое фонетическое слово. Затем слова со слабым ударением объединяются в единую акцентную единицу (АЕ) с одним из слов с сильным ударением и в тексте указываются границы АЕ. После этого каждой синтагме аудитором присваивается определённый интонационный тип с указанием количества АЕ в синтагме (например, 2-С, что означает: 2-х акцентная синтагма с интонацией незавершённости), а также длительность синтагматической паузы. В данном эксперименте исследовались особенности синтагматического членения 2-х интонационных типов: С (comma) - незавершённость, Р (point) - завершённость.

В табл. 3.3 приведен пример разметки текста на синтагмы и маркировки словесных ударений для части фонограммы Ю. Сенкевича после её аудитивного анализа. Знаком « / » отмечены границы АЕ.

Таблица 3.3

Пример синтагматической разметки текста

Текст, маркированный ударениями, размеченный на АЕ и синтагмы	Интонационный тип	Длительность паузы, мс
/Здра+вствуйте/,	1-С	200
/дороги+е/ /люби+тели/ /путеше+ствий/.	3-Р	1100
/Сего+дня/	1-С	75
/мы+ /отпра+вимся/ /с ва+ми/	3-С	50
/в Финля+ндию/,	1-С	750
/и+/,	1-С	150
/мне+ /ка+жется/,	2-С	0
/что= э+то/ /путеше+ствие/ /бу+дет/ /для вас+ /	4-С	400

/интере+сным/,	1-С	0
/поско+льку/	1-С	400
/путеше+ствие/ /э+то/ /не то+лько/ /в простра+нстве/,	4-С	50
/но+/,	1-С	120
/и= во вре+мени/.	1-Р	650
/Мы+/, /раска+жем/ /в+ам/	1-С	120
/о стари+нных /	1-С	900
/фи+нских/ /крепостя+х/.	2-Р	400
/Крепостя+х/ /за+мках/.	2-Р	2400
/В простра+нстве/,	1-С	0
/потому+ что=/ /они+/, /располо+жены/ /в ра+зных/ /частя+х/	5-С	460
/Финля+ндии/,	1-С	170
/а= во вре+мени/,	1-С	0
/потому+ что=/	1-С	330
/мы+/, /коснее+мся/ /исто+рии/,	3-С	150
/есте+ственно/.	1-Р	850

3.4.2. Статистические характеристики синтагматического членения

Статистическая обработка результатов экспериментальных исследований фонограмм речи проводилась с целью получения некоторых количественных характеристик, полезных с точки зрения персонализации синтагматического членения синтезируемой речи. К таким характеристикам относятся: сравнительные частоты встречаемости синтагм с различным количеством АЕ, частоты встречаемости пауз различной длительности, частоты встречаемости пар синтагм с различным количеством АЕ. Основные количественные результаты статистической обработки приведены на рисунках 3.23 – 3.25.

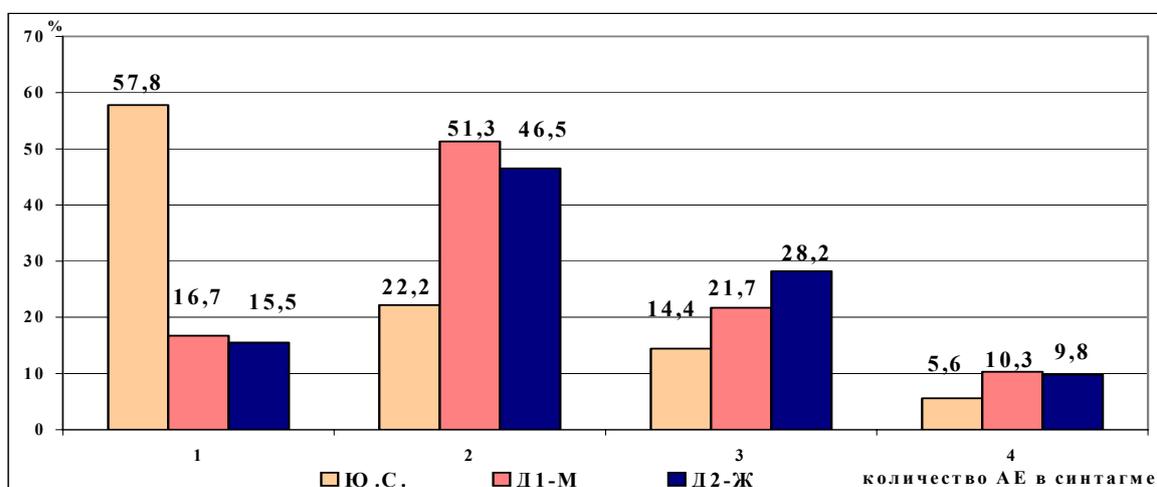
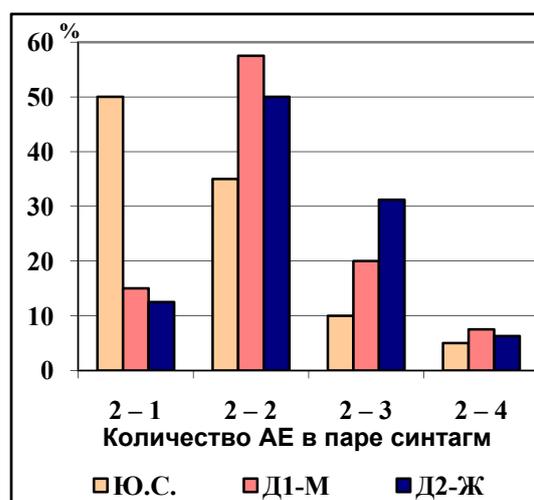
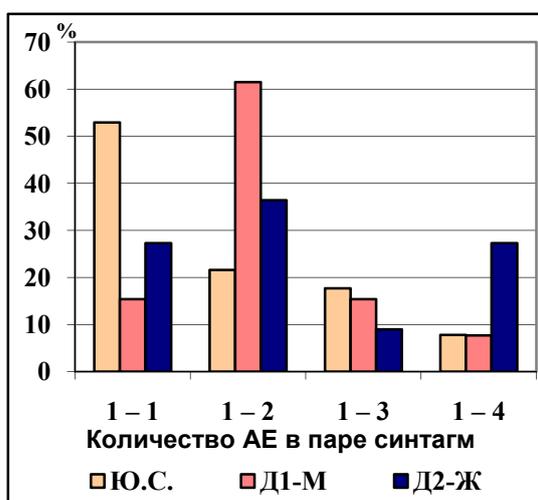
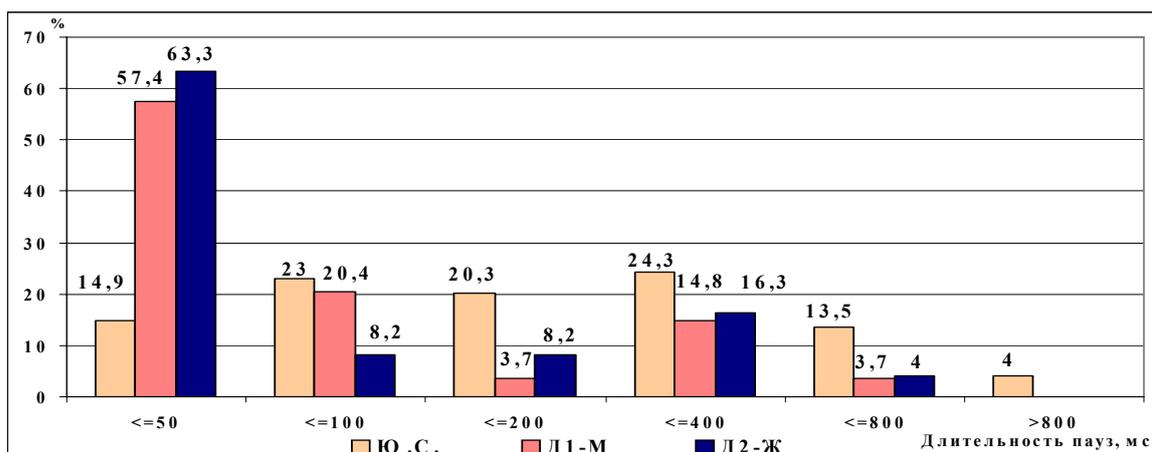
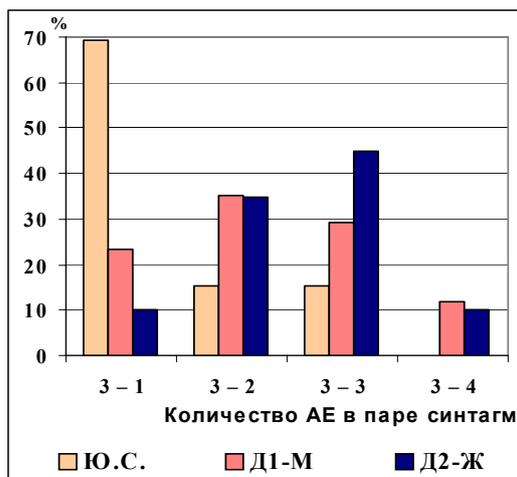


Рис.3.23. Сравнительная частота встречаемости синтагм с

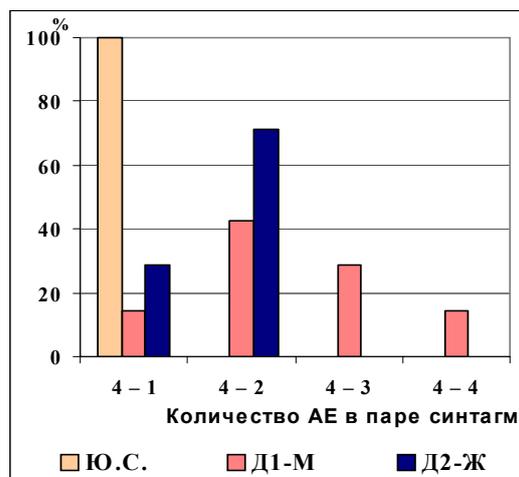


различным количеством АЕ

Рис.3.24. Сравнительная частота встречаемости пауз различной длительности



а)



б)

б)

г)

Рис 3.25. Сравнительная частота встречаемости пар синтагм с различным количеством АЕ

Кроме перечисленных выше чисто статистических, внеязыковых характеристик: частот встречаемости синтагм с различным количеством АЕ, пауз различной длительности, пар синтагм с различным количеством АЕ, проведены исследования сравнительной частоты встречаемости пар слов с различными грамматическими категориями на стыке синтагм.

Подсчёт частоты встречаемости пар слов с различными грамматическими категориями на стыке синтагм производился следующим образом. Для каждой пары слов текста, не разделенных знаком препинания, определялась их принадлежность к конкретным частям речи и вычислялось общее количество пар различных частей речи в тексте. Затем вычислялось количество различных пар частей речи, разделенных границей синтагмы. По результатам обработки составлялся двумерный массив частот появления конца синтагмы между различными частями речи. Частота появления конца синтагмы между i -той и j -той частями речи R_{ij} вычислялась по формуле:

$$R_{ij} = \frac{K_{ij}}{N_{ij}} * 100\% \quad (3.7)$$

где N_{ij} – общее количество пар частей речи ($i - j$) в исследуемой фонограмме;

K_{ij} – количество пар частей речи ($i - j$) в исследуемой фонограмме, разделённых границей синтагмы.

В процессе анализа результатов было принято решение выделять 8 основных частей речи: глагол, существительное, местоимение, наречие, прилагательное, союз, предлог, а все остальные объединить как «другая часть речи».

В табл. 3.4 приведены результаты анализа частоты встречаемости пар различных частей речи на стыке синтагм для 2-х дикторов – Ю.С. и Д-1М. В строках и в столбцах таблицы указаны части речи, на пересечении i -ой строки и j -ого слобца указаны значения R_{ij} (в %) для дикторов Ю.С. и ДМ-1, отделённые символом «—». Значение R_{ij} , равное 0, показывает, что пара частей речи $i - j$ либо вообще не встретилась в анализируемых текстах (в пределах 10-ти процентной статистической

достоверности результатов анализа), либо ни разу не была разделена границей синтагмы в речи данного диктора.

Таблица 3.4

Частота встречаемости (в %) границ синтагм между различными частями речи у двух дикторов: Ю.С. – ДМ-1

	Глагол	Наречие	Прилагательное	Существительное	Местоимение	Союз	Предлог
Глагол	25 – 15	55 – 0	75 – 0	55 – 5	35 – 40	0 – 0	40 – 0
Наречие	35 – 30	0 – 0	20 – 35	0 – 10	15 – 0	0 – 0	45 – 25
Прилагательное	0 – 0	0 – 0	75 – 50	40 – 0	0 – 0	0 – 0	0 – 0
Существительное	70 – 30	80 – 70	80 – 50	55 – 0	20 – 0	90 – 0	80 – 30
Местоимение	35 – 5	0 – 0	55 – 0	15 – 0	0 – 0	0 – 0	45 – 25
Союз	45 – 0	10 – 20	75 – 0	55 – 0	45 – 0	0 – 0	50 – 0
Предлог	0 – 0	0 – 0	30 – 0	30 – 0	20 – 0	0 – 0	0 – 0

3.4.4. Обсуждение результатов эксперимента

Приведенные на рисунках 3.23-3.25 статистические характеристики особенностей синтагматического членения устной речи для 3-х дикторов показывают, что речь диктора Ю. Сенкевича обладает ярко выраженными отличиями от речи 2-х других дикторов. Из рисунка 3.23 видно, что только у него наблюдается значительное преобладание количества одноакцентных синтагм и сравнительно равномерное распределение 2-х и 3-х акцентных. С другой стороны, из рисунка 3.24 видно, что только у Ю. Синкевича распределение длительностей межсинтагменных пауз относительно равномерно, в то время как у двух других дикторов наблюдается ярко выраженный максимум при нулевой длительности. Подобные выводы можно сделать также из анализа распределений на рисунке 3.25.

Приведенные в таблице 3.4 статистические характеристики особенностей встречаемости пар слов с различными грамматическими категориями на стыке синтагм для 2-х дикторов также показывают, что речь диктора Ю.С. обладает определённо выраженными отличиями от речи диктора Д1-М. Из таблицы 3.4 видно, что у этого диктора есть характерная тенденция наличия границы синтагмы после предлогов и союзов, а также наблюдается значительное преобладание количества границ синтагм перед существительными, после глаголов и после местоимений.

Полученные экспериментальные характеристики распределений используются далее при клонировании речи диктора в рамках разработанной стохастической модели синтагматического членения в синтезаторах речи по тексту. Опыт, приобретённый в процессе аудитивного анализа фонограмм, активно используется также при создании автоматизированной системы персонализированного синтагматического членения аудио-файлов речи дикторов.

Проведенные статистические исследования явлений синтагматического членения являются далеко не исчерпывающими. Вне рассмотрения остались пока закономерности членения, связанные с наличием некоторых морфологических или синтаксических указателей в тексте, таких как определённые части речи или знаки препинания, перед которыми в какой-то степени допускается или запрещается наличие синтагматической границы.

Глава 4

Компьютерная модель мультиволнового синтеза речи по тексту

Описываемая в данной главе компьютерная модель синтеза речи базируется на результатах многолетних исследований авторов по созданию лингво-акустических основ синтеза речи по тексту [51, 65, 31]. В модели аккумулированы теоретико-экспериментальные сведения о специфике лингвистической обработки текстов, фонетической и просодической структуре русской речи, артикуляторно-акустических явлений процесса речеобразования. Основные положения этих сведений изложены в предыдущих главах книги.

Отличительной особенностью описываемой модели, нашедшей отражение в её названии – «мультиволновый синтез», является использование в качестве элементов компиляции речи отрезков естественной речевой волны, соотносимой с элементами различной фонетической длины: аллофонами, диаллофонами и аллослогами.

4.1. Структура синтезатора речи по тексту

Синтез устной речи по тексту осуществляется на основе лексико-грамматического анализа входного текста путём моделирования процессов речеобразования с учётом правил произношения звуков и интонирования, свойственных данному языку. Орфографический текст документа (книги, статьи, веб-страницы и т.п.) поступает на вход синтезатора и далее подвергается последовательной обработке рядом специализированных процессоров в соответствии с общей структурой синтезатора речи по тексту, представленной на рис. 4.1. Синтезатор включает четыре основных модуля: текстовый процессор, просодический процессор, фонетический процессор и акустический процессор. Каждый из этих модулей поддерживается наборами соответствующих БД и правил. Рассмотрим основные функции этих модулей.

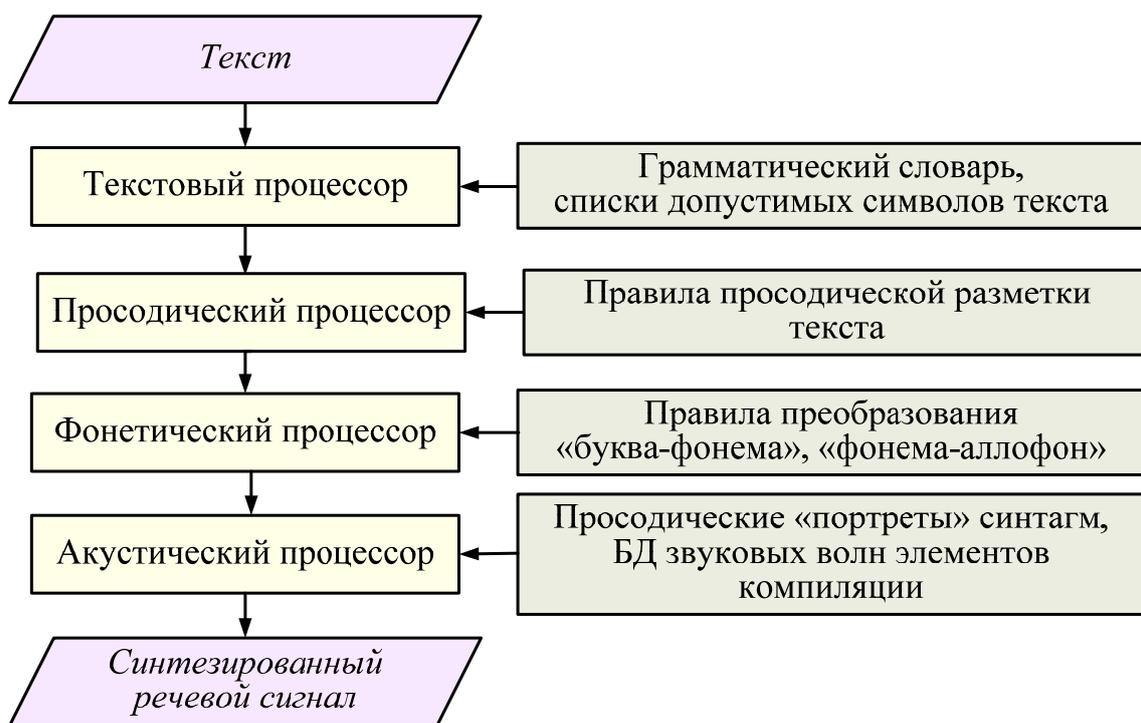


Рис. 4.1. Структура системы синтеза речи по тексту

4.2. Текстовый процессор

Текстовый процессор (рис. 4.2) включает два основных блока, которые поддерживаются соответствующими базами данных, словарями и правилами. Он выполняет предварительную обработку входного текста, а также морфологическую и акцентную маркировку слов текста.

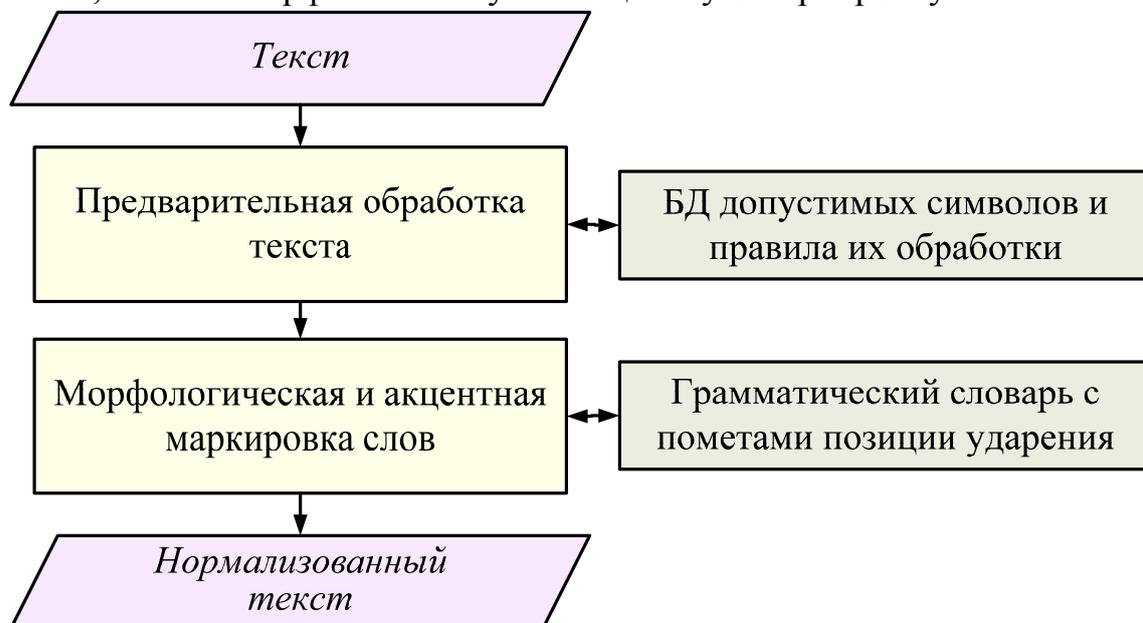


Рис. 4.2. Структура текстового процессора

На вход системы синтеза речи могут поступать тексты, взятые из разных источников и зачастую содержащие графические объекты, ссылки, числовые значения, формулы, а также другие объекты и символы, непригодные для синтеза речи. Основной задачей первого блока - блока предварительной обработки - является нормализация текста, т.е. приведение его к такому виду, когда текст состоит из последовательности слов русского языка. Следующий блок – блок морфо-фонетической маркировки – осуществляют маркировку каждого слова входного текста, необходимую для адекватного синтеза звуков и интонации речи. Для такой маркировки используется грамматический словарь, в котором каждое слово снабжено пометой позиции ударения.

4.2.1. Блок предварительной обработки

Структура блока предварительной обработки показана на рис. 4.3.

Очистка текста. Очистка текста осуществляется с целью удаления из входного текста графических объектов, ссылок, различных маркеров и других неинформативных для синтеза речи символов. Для реализации этой задачи необходимо иметь БД допустимых символов и объектов, содержащую русские и латинские буквы, знаки пунктуации, цифры, математические символы, а также специальные символы, такие как «@», «^» и т.д. Вообще, в данной БД должны содержаться только те символы, которые могут быть «озвучены» синтезатором речи. Например, если в БД содержатся римские цифры или сложные математические символы, такие как « Σ », « \int », то на последующих этапах обработки текста должны быть блоки, преобразующие последовательности этих символов в слова.

Примечательно, что данный блок с точки зрения разработчика не представляет ни особой трудности, ни особого интереса, и в большинстве случаев при разработке систем синтеза речи реализуется в последнюю очередь. Для пользователей же системы синтеза речи этот блок, напротив, очень важен, поскольку от алгоритмов его работы зависит полнота «озвучивания» входного текста.

Дешифровка чисел. Задача этого блока – преобразовать числа, встретившиеся в тексте, в числительные. При этом необходимо учитывать, что числа, встретившиеся в тексте, могут обозначать целые, десятичные и дробные количественные числительные, порядковые числительные (которые могут быть записаны как арабскими, так и римскими цифрами), дату, время, номера телефонов и т.д. Для корректного преобразования чисел необходимо использовать правила преобразования число – числительное, учитывающие не только число, но и окружающие его слова, сокращения, которые позволяют определить характеристики числа.

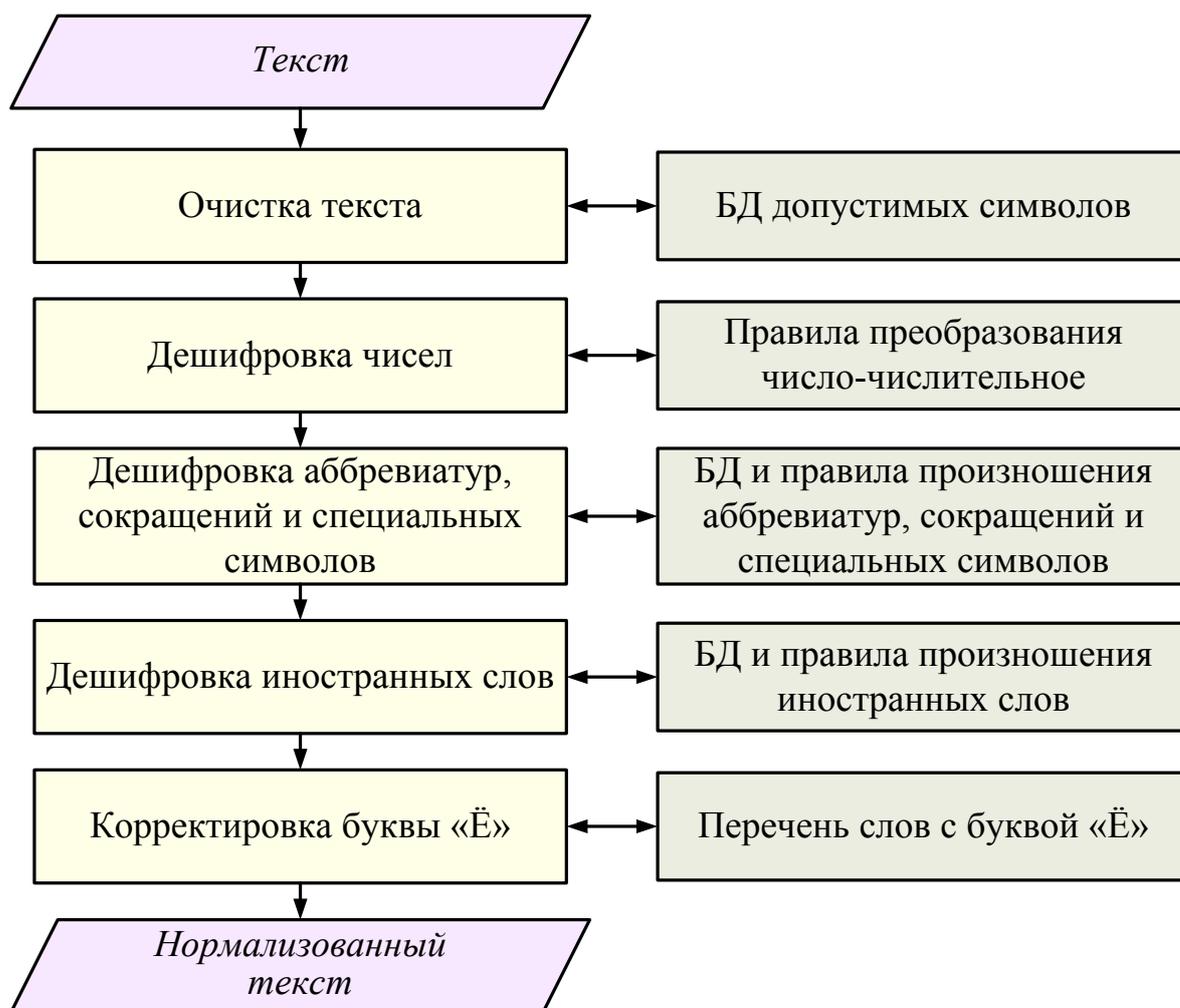


Рис. 4.3. Структура блока предварительной обработки текста

Кроме того, необходимо учитывать, что знаки «.» и «,» могут использоваться как для разделения разрядов в целых числах, так и для отделения целой части от дробной. Например, в записи числа 53,45 запятая отделяет целую часть от дробной, а в записи 378,812,547 служит для разделения разрядов.

Дешифровка аббревиатур, сокращений и специальных символов. При синтезе речи необходимо учитывать, что правила чтения аббревиатур, сокращений и специальных символов отличаются от соответствующих правил для слов русского языка. Для решения этой задачи необходимо преобразовать аббревиатуры, сокращения и специальные символы в слова, для которых применимы стандартные правила, используемые на этапах фонетической и просодической обработки текста. При дешифровке необходимо учитывать следующие факторы:

1. Аббревиатуры в текстах не всегда пишутся заглавными буквами. Это характерно в первую очередь для текстов электронных писем, блогов и других текстов, полученных из различных интернет-ресурсов.

2. Некоторые аббревиатуры и сокращения могут расшифровываться по-разному в зависимости от предметной области, от контекста, например «г.» может означать «город» или «год», «т.» - «товарищ» или «тонн».

3. Некоторые аббревиатуры читаются не в соответствии со стандартными правилами дешифровки, например «США» по правилам расшифровывается как «эс-ше-а», тем не менее общепринятое произношение – «сэ-ше-а».

4. Специальные символы могут преобразовываться по-разному, например «%» – «процент», «процента» или «процентов», «\$» - «доллар», «доллара», «долларов».

Для решения этих задач необходимо использовать БД и правила произношения аббревиатур, сокращений и специальных символов. Содержащийся в БД перечень аббревиатур русского языка позволит обнаружить в тексте аббревиатуру даже в случае, если она записана прописными символами. Перечень сокращений и варианты их расшифровки, а также анализ контекста сокращения позволят корректно преобразовать сокращение в слово.

Аббревиатуры произносятся, как правило, по буквам, например «КГБ» – «ка-гэ-бэ», «ФРГ» – «эф-эр-гэ», при этом каждый слог является ударным. Однако наиболее употребительные аббревиатуры, а также аббревиатуры, содержащие большое количество гласных, произносятся, как правило, в одно слово, например, «ЮНЕСКО» – «юнэско». Это должно учитываться правилами произношения аббревиатур и сокращений. Правила произношения специальных символов для корректного преобразования должны учитывать контекст символа.

Дешифровка иностранных слов. В текстах на русском языке могут встречаться интернет-адреса, адреса электронной почты, названия организаций, записанные латинскими символами. Для преобразования таких слов в последовательность русских букв, читаемых по общим правилам, используется блок дешифровки иностранных слов. Этот блок использует БД и правила дешифровки латинских символов. В БД должны содержаться наиболее употребительные иностранные слова, а также их эквиваленты на русском языке, например «Microsoft» – «ма́йкросо́фт», «www» – «три да́блью». Кроме того, правила дешифровки латинских символов должны содержать русские эквиваленты каждой латинской букве. Тогда в случае, если встретившееся в тексте слово, записанное латинскими буквами, не будет найдено в БД, каждая буква будет преобразована по соответствующим правилам.

Корректировка буквы «ё». Проблема расстановки точек над «ё» - это, пожалуй, проблема только русского языка. Интересно, что человек при чтении текста не задумывается, как правильно прочитать слово, с

буквой «ё» или «е», используя для коррекции свои знания о языке. Если же при синтезе вместо, например, слова «ёлка» прозвучит «елка» или вместо «весёлый» – «веселый», такая неточность будет сразу же замечена пользователем. В подавляющем большинстве случаев для корректировки буквы «ё» достаточно лексической информации, а именно БД, содержащей наиболее полный перечень слов с буквой «ё» в русском языке. Тогда в каждом слове текста, содержащем одну или несколько букв «е», каждая из них последовательно заменяется на «ё» и осуществляется поиск соответствующего слова в БД. Однако в некоторых случаях такой информации недостаточно, например, как корректно прочитать слово «все»: «*Все в машине?*» или же «*Всё в машине?*». Очевидно, что в этом случае необходимо использовать не только лексический и синтаксический, но и семантический и прагматический анализ. Однако такие ситуации встречаются в текстах довольно редко.

Функционирование блока предварительной обработки текста.

Орфографический текст подаётся на вход лингвистического процессора в виде множества символов. На первом шаге поступившее множество необходимо очистить от элементов, не принадлежащих списку допустимых символов. Список допустимых символов состоит из следующих конечных множеств (где все символы указаны в кавычках):

– множество русских заглавных и строчных букв – {«А», «Б», «В», «Г», «Д», «Е», «Ё», «Ж», «З», «И», «Й», «К», «Л», «М», «Н», «О», «П», «Р», «С», «Т», «У», «Ф», «Х», «Ц», «Ч», «Ш», «Щ», «Ъ», «Ы», «Ь», «Э», «Ю», «Я», «а», «б», «в», «г», «д», «е», «ё», «ж», «з», «и», «й», «к», «л», «м», «н», «о», «п», «р», «с», «т», «у», «ф», «х», «ц», «ч», «ш», «щ», «ъ», «ы», «ь», «э», «ю», «я»};

– множество латинских заглавных и строчных букв – {«A», «B», «C», «D», «E», «F», «G», «H», «I», «J», «K», «L», «M», «N», «O», «P», «Q», «R», «S», «T», «U», «V», «W», «X», «Y», «Z», «a», «b», «c», «d», «e», «f», «g», «h», «i», «j», «k», «l», «m», «n», «o», «p», «q», «r», «s», «t», «u», «v», «w», «x», «y», «z»};

– множество цифр – {«1», «2», «3», «4», «5», «6», «7», «8», «9», «0»};

– множество математических знаков – {«-», «+», «/», «*», «=», «<», «>», «%»};

– множество знаков пунктуации – {«.», «,», «:», «;», «(», «)», «!», «?», «-»};

– множество специальных знаков – {«~», «`», «№», «#», «@», «\$», «^», «&», «|», «\», «{», «}», «[», «]», «"», «'»}.

Блок предварительной обработки осуществляет преобразование в текст следующих последовательностей символов:

– многозначные и дробные числа;

- даты из римских цифр;
- обозначения времени и даты;
- телефонные номера;
- знаки пунктуации;
- математические выражения;
- слова с буквой «е» вместо «ё»;
- аббревиатуры;
- сокращения;
- интернет-адреса;
- иностранные слова.

Правила преобразования цифровых символов можно разделить на несколько типов: преобразование количественных и порядковых числительных, десятичных и дробных целых чисел, а также преобразование времени, даты, телефонных номеров.

Опознать порядковое числительное в тексте можно по знаку дефиса, стоящему после цифры, и характерному родовому окончанию. Например, 1-й класс, 2-го класса, 3-му классу и т.д. Для расшифровки порядковых числительных создаётся специальный набор базовых числительных и список родовых окончаний.

Информационный маркер времени в тексте можно определить по следующим признакам:

- [число] – массив целых десятичных чисел.
- [ключевое слово] – группа: час, часы, часов; час.; ч.;
- [ключевое слово] – группа: минут, минута, минуты; м.; мин.

Алгоритм начинает работать, если текущее слово принадлежит массиву целых чисел, а следующее слово принадлежит группе ключевых слов.

Информационный маркер даты можно определить по следующим признакам:

- массив целых десятичных чисел;
- массив римских чисел в диапазоне I – XXX;
- группа слова [ключевое слово]:
- названия месяцев (январь, февраль, март и т.д.);
- год (г.);
- век (в.);
- наша эра, до нашей эры (н.э., д.н.э.);
- столетие (столетия) (ст.).

Если число опознано, но оно не принадлежит ни одному из перечисленных выше классов, система автоматически преобразует его как количественное числительное. Расшифровка количественных числительных происходит с помощью набора базовых числительных: [0, 1, ... , 9], [10, 11, ... , 19], [20, 30, ..., 90], [100, 200, ..., 900], [1000, 1000000, 1000000000, ...].

В массовых пользовательских системах синтеза речи функция озвучивания знаков пунктуации является излишней. Однако для людей со слабым зрением такая возможность оказывается достаточно полезной. При включении такой опции система будет автоматически конвертировать все знаки препинания в произносительный вариант, который также задан в лингвистической базе.

Интернет-ресурсы опознаются по списку маркеров, к которым относятся ключевые слова: *www*, *ftp*, *url*, *@*, *http*. Каждому такому маркеру ставится в соответствие вариант русского произношения, например, *www* – *даблью – даблью - даблью*, *ftp* – *эф - ти - пи* и т.д. После информационного маркера следует, как правило, английские буквы, отражающие положение Интернет-ресурса (исключение составляет *@*, когда маркер стоит в середине). Английские буквы в данном случае читаются отдельно, т.е. каждой английской букве сопоставлено русское произношение, например www.vox.com = *даблью-даблью-даблью, точка, ви-оу-экс, точка, ком*.

Одной из ключевых проблем при автоматическом синтезе речи является опознание и озвучивание аббревиатур. Существующие в современном русском языке аббревиатуры можно классифицировать по следующим основным группам: 1) буквенные; 2) звуковые; 3) буквенно-звуковые; 4) аббревиатуры смешанного типа. Структурная схема расшифровки аббревиатур и сокращений приведена на рис. 4.4 а. База, составленная из интернет-источников, содержит порядка 1000 аббревиатур и 500 сокращений.

Все слова, состоящие из латинских букв, система идентифицирует как иностранные, расшифровка которых происходит по следующей схеме. Если иностранное слово присутствует в лингвистической базе, то вместо него ставится русский произносительный вариант. Как правило, в такой список входит часто используемые иностранные слова (например, наименование компаний, информационно-компьютерных терминов и др.). В случае, если слово отсутствует в базе, оно расшифровывается побуквенно. Каждой латинской букве ставится соответствующее ей русское произношение.

В электронном варианте русского текста достаточно часто встречается ситуация, когда вместо буквы Ё стоит буква Е. По правилам русской орфографии и пунктуации буква Ё обязательно пишется только в следующих случаях:

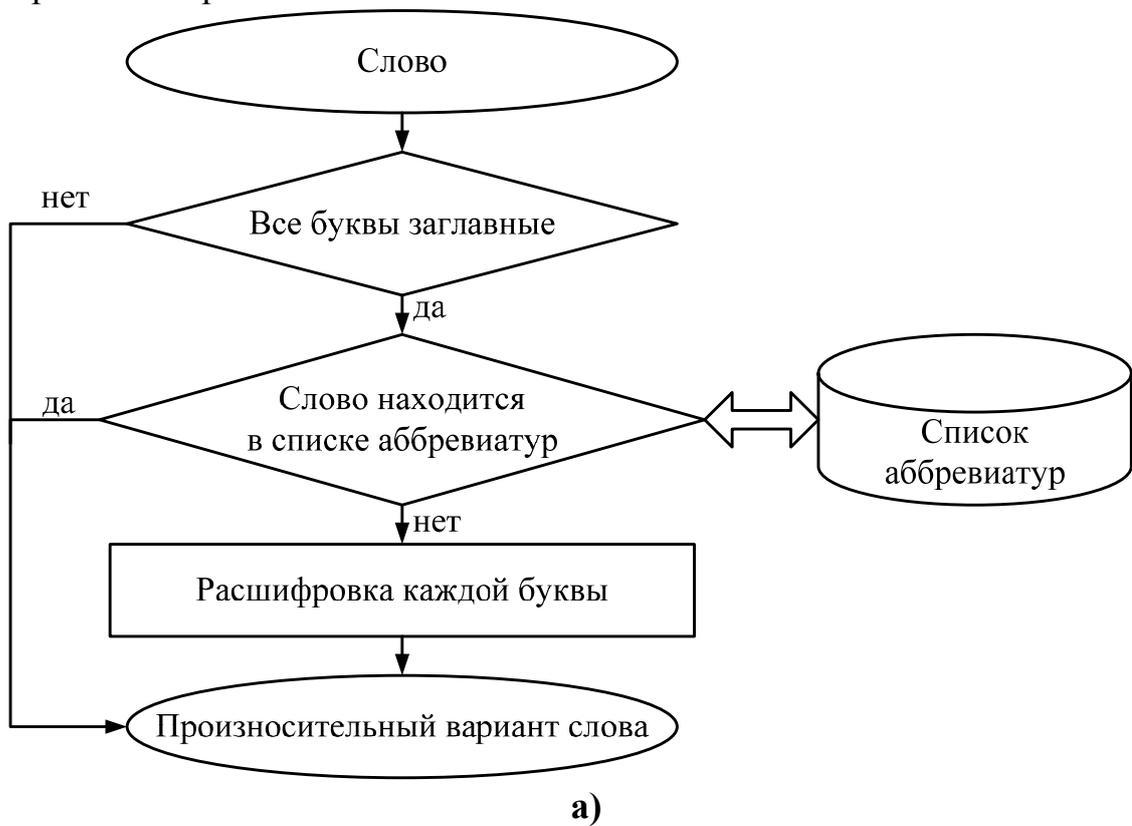
– когда необходимо предупредить неверное чтение и понимание слова, например: *узнаём* в отличие от *узнаем*, *совершённый* (причастие) в отличие от *совершенный* (прилагательное) и т. д.

– когда надо указать произношение малоизвестного слова, например: *река Олёкма*.

– в специальных текстах: букварях, учебниках русского языка для иностранцев, учебниках орфоэпии и т. п., а также в словарях для указания места ударения и правильного произношения.

В остальных случаях написание буквы Ё обязательно.

Проблема буквы Ё решается следующим образом. Создаётся специальный словарь – «словарь Ё», содержащий по возможности все слова с буквой Ё (количество таких слов в словаре порядка 7000). Обращение к этому словарю происходит в случаях, когда в основном словаре под ударением находится буква Е, а затем проверяется, содержится ли это слово в «словаре Ё». Если слово содержится в словаре, замена происходит автоматически, в противном случае буква Е остаётся без изменения. Схема алгоритма замены буквы "Е" на "Ё" изображена на рис. 4.4 б.



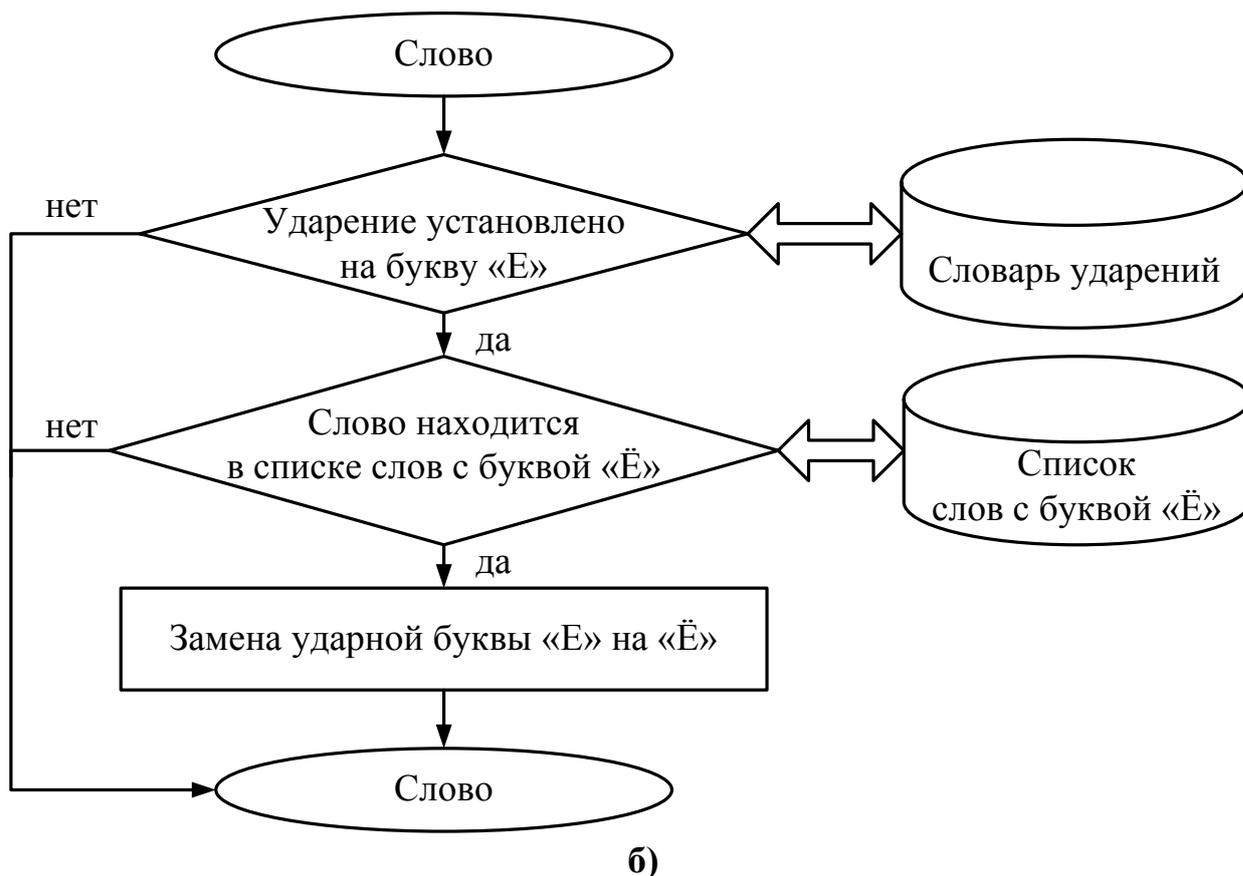


Рис. 4.4. Структурные схемы расшифровки аббревиатур (а), определения буквы «ё» (б)

4.2.2. Блок морфологической и акцентной маркировки слов

Структура блока морфо-акцентной маркировки слов представлена на рис.4.5.

Морфологическая маркировка заключается в указании для каждого слова входного текста его принадлежности к той или иной части речи, а также дополнительных морфологических характеристик, определённых для данной части речи.

Акцентная маркировка заключается в маркировке ударений в словах с сильным или слабым ударением с последующим присоединением к ним безударных слов, образующих вместе с ударным словом одно фонетическое слово.

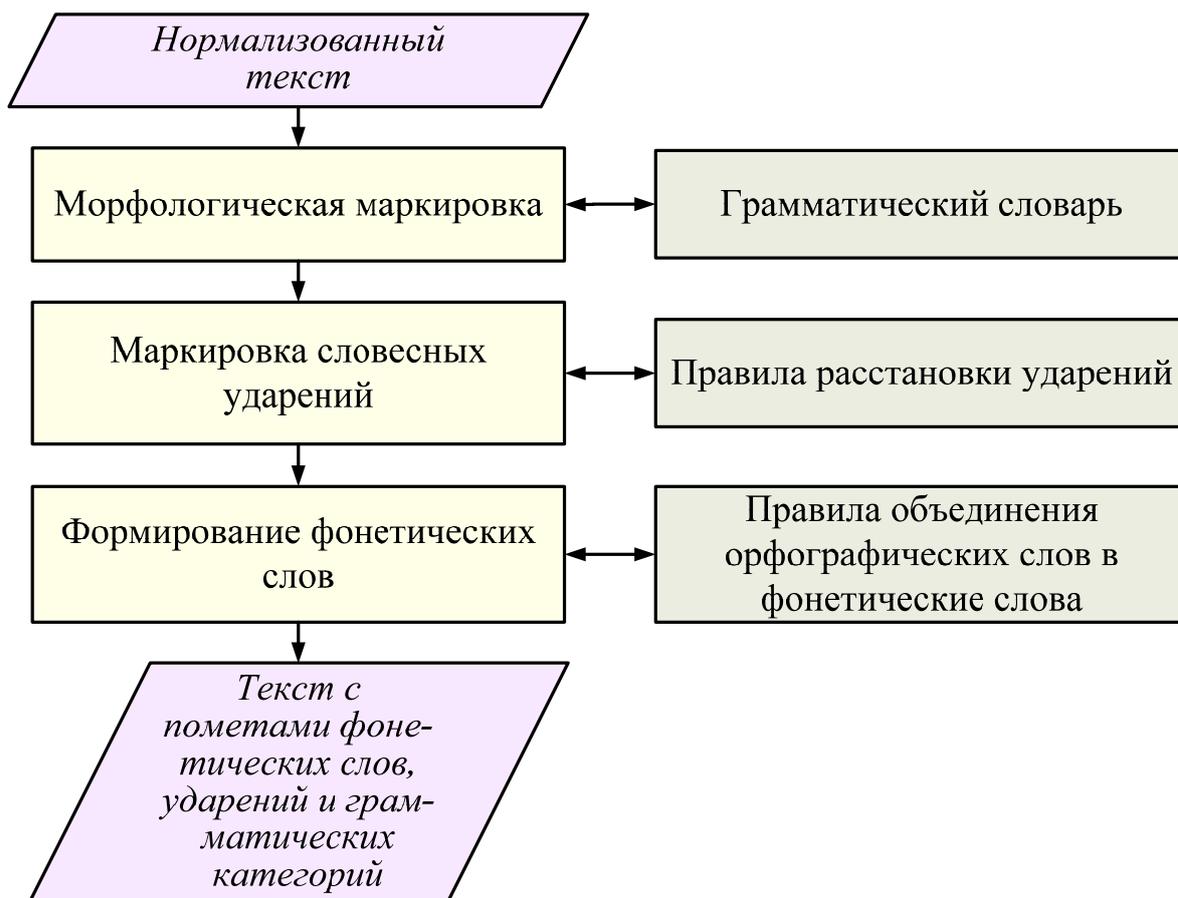


Рис. 4.5. Структура блока морфологической маркировки, расстановки ударений и формирования фонетических слов

Морфологическая маркировка. При морфологической маркировке каждое слово текста должно снабжаться пометами названия части речи, а также дополнительными грамматическими категориями, характерными для данной части речи. Для определения этой информации используется грамматический словарь.

В русском языке десять основных частей речи: имя существительное, местоимение-существительное, имя прилагательное, имя числительное, глагол, наречие, предлог, союз, частица, междометие. Некоторые глагольные формы, такие как причастие и деепричастие, также имеет смысл отнести к самостоятельным частям речи.

Дополнительные грамматические категории для частей речи могут быть словоизменяемыми, когда члены этих категорий могут быть представлены формами одного и того же слова. Например: категория падежа и числа существительного, категории лица, числа, времени и наклонения глагола, категория степени сравнения наречия и т.д. Категории для частей речи могут несловоизменяемыми, когда члены этих категорий не могут быть представлены формами одного и того же слова. Например: категория рода существительных и категория вида

глагола. Знаменательные части речи: имя существительное, местоимение-существительное, имя прилагательное, имя числительное, глагол и наречие – обладают как словоизменительными, так и несловоизменительными категориями. Служебные части речи: предлог, союз, частица, междометие – обладают только несловоизменительными категориями. Определение грамматических категорий частей речи необходимо в дальнейшем для адекватной просодической маркировки текста.

Наиболее полно все категориальные характеристики и свойства лексических единиц (около 100 тыс. знаков) для русского языка отражены в «Грамматическом словаре русского языка» А.А. Зализняка [174]. Здесь в единстве представлены все совокупности форм того или иного слова, что позволяет найти необходимые сведения об изменяемости отдельно взятой лексемы. В данном словаре информация о грамматической парадигме слова (где под парадигмой понимается совокупность всех грамматических форм некоторого слова) даётся с помощью системы условных обозначений и индексов.

В качестве иллюстрации приведём полную словоизменительную парадигму имени существительного **<конкурс м 1а>** и глагола **<выбрасывать нсв 1а>**. Символ **<м>** указывает на целый ряд несловоизменительных признаков, характеризующих слово «конкурс», а именно: существительное, неодушевлённое, мужского рода, субстантивного склонения; **<1>** обозначает тип склонения, зависящий от окончания основы слова; **<а>** – схема ударения (постоянное ударение на основе). Для глагола **выбрасывать <нсв>** выступает характеристикой вида (глагол несовершенного вида); цифра индекса указывает на тип спряжения, в зависимости от которого выбираются способы построения форм глагольной парадигмы. В данном случае **<1>** показывает, что инфинитив заканчивается на **-ать, -ять** или **-еть**, а формы глагола 1-го и 3-го лица единственного числа настоящего времени соответственно: **-аю, -ает; -яю, -яет; -ею, -еет**. Наконец, по индексу **<а>** узнаём схему ударения (постоянное ударение на основе). Помимо указанной информации, глагол также обладает специфическими формами причастия, деепричастия и целым набором категориальных значений, что оказывает существенное влияние на словоизменительные характеристики слов.

Словоизменительные категории существительного и прилагательного показаны на рис. 4.6, 4.7 соответственно. Количество слов в грамматическом словаре А.А. Зализняка, относящихся на каждую часть речи, а также количество словоформ в парадигмах представлено в табл. 4.1.

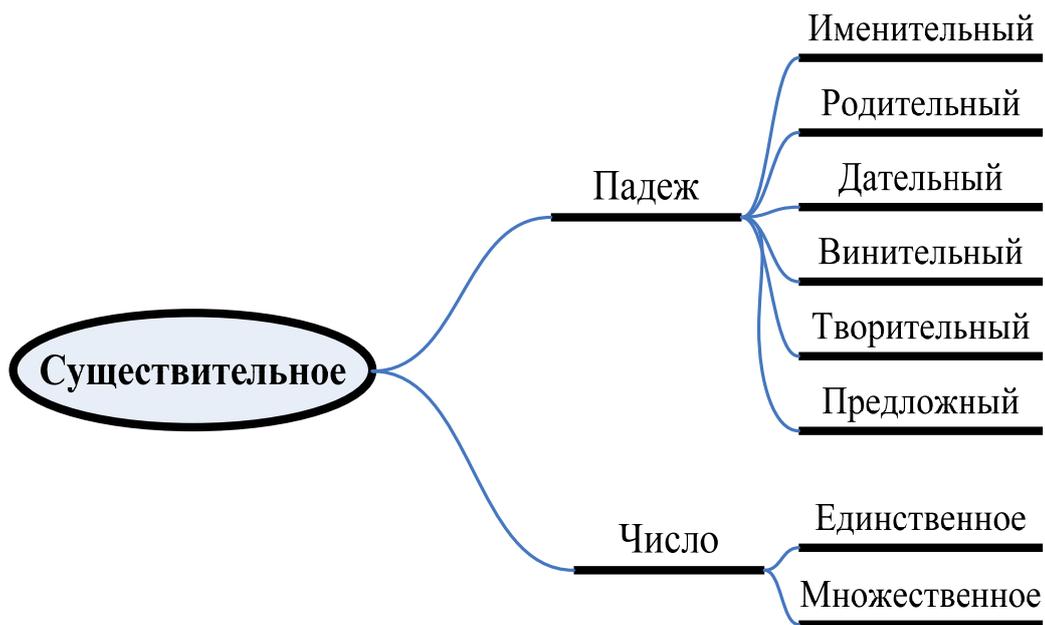


Рис. 4.6. Словоизменительные категории существительных

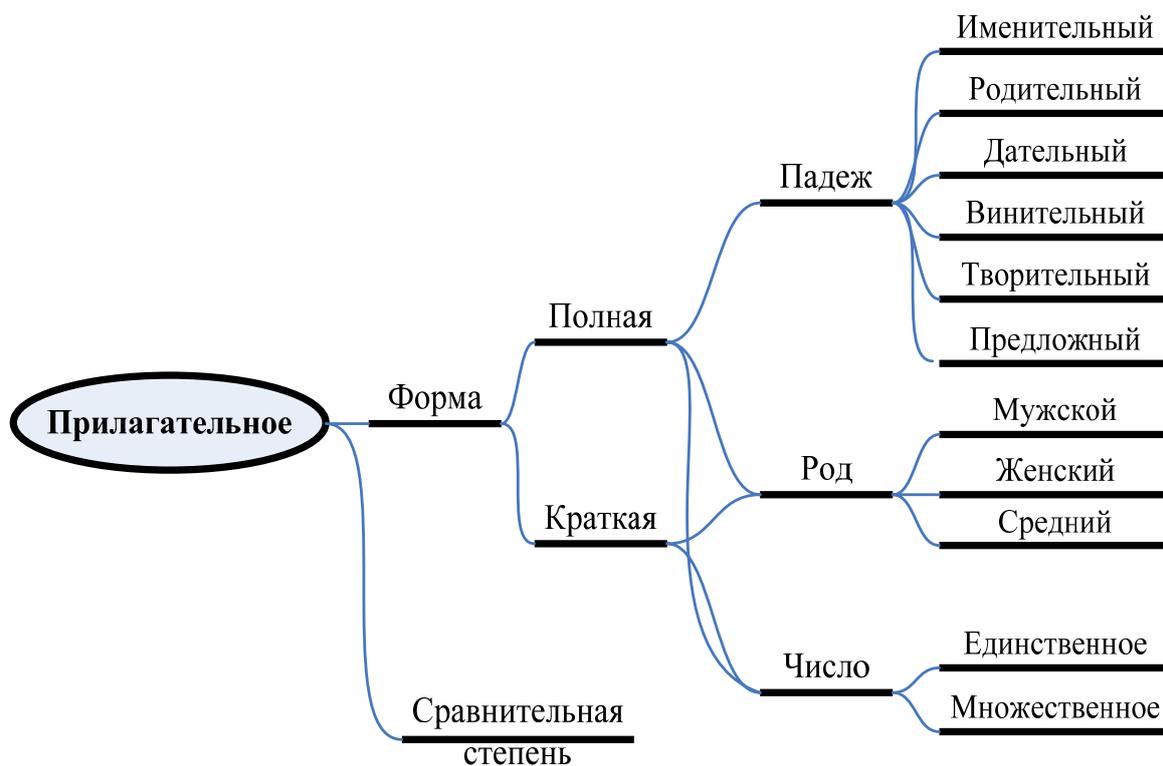


Рис. 4.7. Словоизменительные категории прилагательных

Таблица 4.1

Количество слов и словоформ в грамматическом словаре А.А. Зализняка

Часть речи	Количество лексем в словаре	Количество словоформ в парадигме	ВСЕГО словоформ
Существительное	46523	12	558276
Глагол	27474	35	961590
Союз	68	1	68
Наречие	1353	1	1353
Междометие	180	1	180
Прилагательное	20622	29	599198
Предлог	87	1	87
Частица	86	1	86
Числительное	100	10	1000
ВСЕГО:	98 222	-	2 121 838

Общее количество слов в исходном словаре равно 98 222, из которых в русском языке порождается более двух миллионов словоформ.

Маркировка словесных ударений. Для расстановки ударений используется грамматический словарь с пометами позиции ударения в словоформе. Знаменательные части речи, как правило, относятся к полнударным словам с одним ударением. Однако некоторые полнударные слова могут иметь наряду с одним сильным (полным) ударением, помечаемым знаком (+), одно или несколько слабых (частичных), помечаемым знаком (=). К таким словам относятся, в частности, составные прилагательные и существительные, например «ра=диолокацио+нный», «мо=тове=лозаво+д».

Следует отметить, что определённое количество знаменательных слов - *омографов* - могут иметь различное ударение в разных грамматических категориях при одинаковом буквенном составе. Например, некоторые существительные имеют одинаковое написание в родительном падеже единственного числа и в именительном падеже множественного числа: «руки+», «ру+ки». Для разрешения таких ситуаций, т.е. для определения точной грамматической категории слова, необходим более глубокий анализ текста.

На этапе расстановки ударений необходимо учитывать также, что какой бы ни был большой словарь, в тексте вполне может встретиться слово, которое отсутствует в нём. Для маркировки ударения в таком слове (поскольку без ударения его озвучить нельзя) можно применять один из двух способов: устанавливать полное ударение на основе статистической информации о словесных ударениях либо устанавливать

частичное ударение на каждом слоге. Во втором случае слово будет «прочитано» по слогам. При использовании первого способа существует вероятность того, что ударение будет установлено неверно, и такое слово не будет воспринято слушающим правильно. Поэтому второй способ представляется более разумным.

Формирование фонетических слов. Многие служебные слова могут произноситься без явно выраженного ударения. К безударным словоформам относятся неслоговые предлоги *в, к, с* и частицы *б, ж, ль*, а также односложные предлоги *без, во, для, за* и др. и частицы *де, ка, ан, бы* и др. Например, в словосочетаниях «*доехать до Киева*», «*прибыл бы вовремя*» предлог *до* и частица *бы* произносятся, как правило, без ударения и присоединяются соответственно к последующему и предшествующему словам.

После маркировки словесных ударений необходимо присоединить каждое безударное слово к близлежащему (предыдущему или последующему) ударному слову. Операцию присоединения безударных слов к ударным выполняет блок формирования фонетических слов. Под фонетическим словом понимается одно или несколько орфографических слов, имеющих одно общее ударение. Для присоединения безударных слов к частично- или полноударным словам используется набор правил, учитывающих грамматические характеристики безударного слова, а также окружающих его слов. Причём к предыдущему слову присоединяются лишь частицы «*бы*», «*-де*», «*дескать*», «*-ли*», «*-же*», «*мол*», «*-то*», «*-ка*», «*-либо*», «*-нибудь*».

Например, в предложении: «*Мальчик успел бы вовремя, если бы не остановился поболтать с другом*», - предлог «*с*» присоединяется к последующему слову (существительному): «*сЪдругом*», а частицы «*бы*» – к предыдущим словам (глаголу и союзу): «*успелЪбы*», «*еслиЪбы*». Здесь буква *Ъ* используется как символ присоединения.

4.3. Просодический процессор

Синтез речи по тексту предполагает наличие автоматической процедуры формирования текущих контуров мелодии, силы звука, фонемной длительности и длительности пауз на основе анализа определенных свойств входного текста и его просодической разметки. Просодическая разметка текста заключается в его членении на синтагмы, разметке синтагм на акцентные единицы и маркировке интонационного типа синтагм в соответствии с определёнными правилами.

Как было указано (см. раздел 1.5), под синтагмой понимается самостоятельная в интонационном смысле часть предложения или всё предложение. Установка границ синтагм влияет на передачу

интонационных характеристик при синтезе речи, а также на передачу смыслового содержания. При разбиении текста на синтагмы важно не поставить границу синтагмы там, где она может нарушить смысловое восприятие речи (или передачу смыслового содержания текста), например, между предметом и его признаком. Для установки границ синтагм используются определённые правила синтагматического членения, базирующиеся на пунктуационном, морфологическом и синтаксическом анализе текста, а также на статистическом анализе синтагматического членения в естественной речи.

Синтагмы в речи отделяются, как правило, паузами. Паузы принимают участие в передаче определённых синтаксических и смысловых отношений. Кроме того, временные интервалы, создаваемые паузами, позволяют слушателю производить лингвистическую обработку текста, запоминать её результаты и строить смысловую структуру, необходимую для восприятия текста. В естественной речи различают грамматические паузы, отделяющие друг от друга интонационно-оформленные части фразы, выделительные паузы и паузы хезитации (неуверенности). Граница синтагмы может быть промаркирована не только физическим перерывом в речевом сигнале, но и резкой сменой высоты тона и (или) других просодических характеристик, которые воспринимаются как нарушение плавного течения речи.

Важно заметить, что процесс синтагматического членения должен удовлетворять решению двух основных задач: установить границы синтагм в тех местах, где они обязательно должны присутствовать, и не устанавливать границу синтагмы там, где она может нарушить смысловое восприятие речи.

Структура просодического процессора представлена на рис. 4.8.

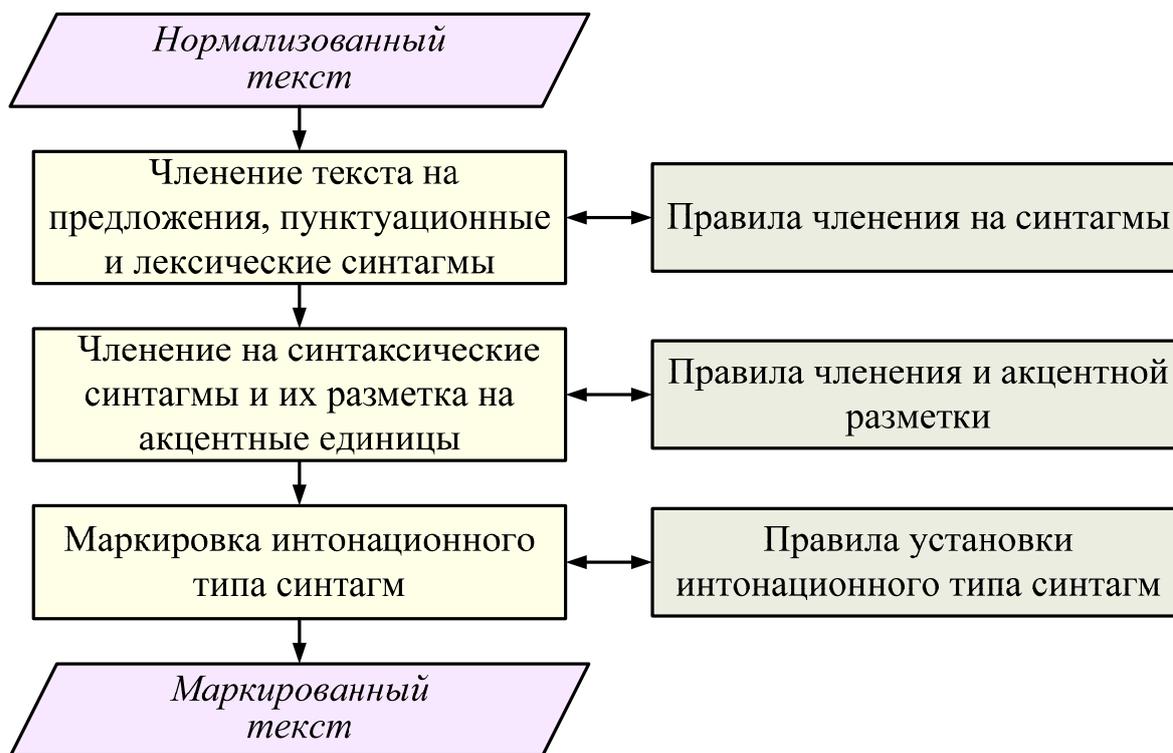


Рис. 4.8. Структура просодического процессора

4.3.1. Блок членения текста на предложения, пунктуационные и лексические синтагмы

Структура блока членения входного текста на предложения, пунктуационные и лексические синтагмы представлена на рис. 4.9.

Членение текста на предложения. Синтез речи осуществляется по предложениям, которые характеризуются достаточной степенью интонационной автономности в тексте и допускают наличие достаточно длительной паузы между ними (0,5 – 1,5 сек.). Предложение - это минимальная единица речи, представляющая собой грамматически организованное соединение слов (или только одно слово), обладающее определённой смысловой и интонационной законченностью.

Предложением считается отрезок текста, ограниченный знаками [., [?, [?!], [!], [!!!]. *Конец предложения* может быть обозначен также знаком [...], при условии, что следующее за ним слово начинается с большой буквы.

Предложением будем считать также заголовок всего текста или его части, в конце которого знак [.] может отсутствовать. Конец такого предложения обозначим знаком [*]. Кроме того, в отдельный тип выделяется *предложение*, ограниченное точкой в конце абзаца. Конец абзаца обозначим знаком [#].

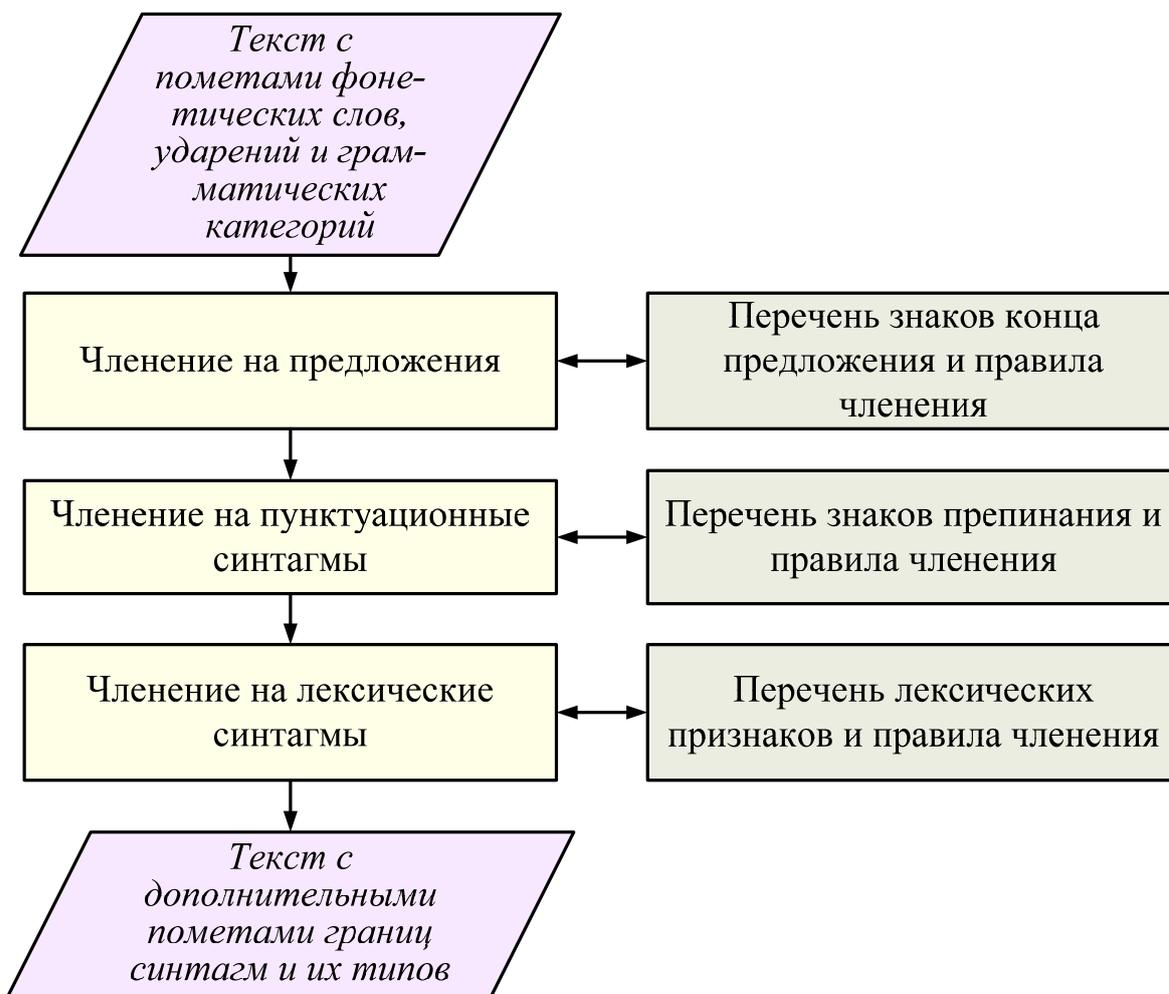


Рис. 4.9. Структура блока членения текста на предложения, пунктуационные и лексические синтагмы

Членение предложения на пунктуационные синтагмы. Индикаторами пунктуационных синтагм (ПС) являются знаки препинания. *Пунктуационными синтагмами* будем считать предложение (при отсутствии в нём знаков препинания) или части предложения, ограниченные следующими знаками:

- точка с запятой [;],
- двоеточие [:],
- запятая [,],
- тире [–],
- открывающая скобка [(],
- закрывающая скобка [)],
- комбинация знаков [, –],
- знаки открытия кавычек [«], [“],
- знаки закрытия кавычек [»], [”]

Таким образом, если предложение включает n знаков препинания (включая знак конца предложения), то оно разбивается на n пунктуационных синтагм ($n=1,2,3,\dots$). Определённым исключением из

этого правила может служить ситуация, когда знак препинания стоит после сочинительного союза: *и, да, но и, так и, а, но, однако, зато, или, либо, то* и др. В этом случае предпочтительнее будет отказаться от установки синтагматической границы на месте этого знака препинания, хотя она и допустима для некоторого индивидуального стиля речи.

Пример: «Он быстро вошел и, увидя нас, внезапно остановился».

Членение пунктуационных синтагм на лексические синтагмы. Очевидно, что пунктуационная синтагма может быть самой разнообразной длины, не включая внутри себя ни одного знака препинания. Если длина синтагмы слишком большая (например, более 4-х слов), то следует убедиться, не содержит ли она некоторые простые лексические признаки (определённые слова или словосочетания), которые позволили бы разбить её на более мелкие **лексические синтагмы (ЛС)**. Экспериментальные исследования показали, что во многих случаях к таковым может быть отнесено присутствие следующих лексических признаков:

– соединительного союза «И».

Пример: «Они посидели / *и* пошли гулять дальше». Раздел синтагмы – **перед «И»**.

– разделительного союза «ИЛИ».

Пример: «Стоит ли нам сейчас пообедать / *или* подождать до 3-х часов»? Раздел синтагмы – **перед «ИЛИ»**.

– имён собственных (ИС).

Пример: «Сегодня певица Алла Пугачёва / решила выступить в нашем городе». Раздел синтагмы – **после последнего из следующих подряд ИС**.

– аббревиатур (АБ).

Пример: «Возможность победы БНФ / вызывает большие сомнения». Раздел синтагмы – **после АБ**.

– названий разрядов чисел (РЧ).

Пример: «Два миллиона / десять тысяч / сто пять целых / двадцать пять сотых». Раздел синтагмы – **после каждого РЧ**.

– названий месяцев, слов «час, минута» при расшифровке даты и времени (ДВ).

Пример: «Десять часов / пять минут / десятого июня / седьмого года». Раздел синтагмы – **после ДВ**.

Указанный перечень не является полным и может быть расширен в процессе анализа всё более обширного текстового корпуса.

4.3.2. Блок членения ПС и ЛС на синтаксические синтагмы и их разметка на акцентные единицы

Даже после разбиения предложения на ПС и ЛС их длина может оказаться слишком большой.

Пример: «Но молодая жена упорно продолжала отстирывать белую в кровавых пятнах рубаху мужа посиневшими от холода руками в железном тазике с ледяной водой».

В приведенном предложении отсутствуют какие-либо признаки наличия в нём ПС или ЛС. Очевидно, что при отсутствии механизма дальнейшего членения таких предложений на более мелкие синтаксические синтагмы (СС) неизбежно возникнут затруднения в понимании смысла синтезированной речи. Идеальным решением проблемы дальнейшего членения такого рода ПС или ЛС на СС было бы использование комплекса правил их глубинного синтаксического разбора. Однако, в виду сложности и недостаточной разработанности таких правил приходится пока ограничиваться использованием процедуры поверхностного синтаксического анализа, опирающейся на доступную морфосинтаксическую информацию о словосочетаниях, составляющих ПС или ЛС.

Словосочетание рассматривается как пара по смыслу и грамматически связанных слов, выделяемая из предложения [7]. Являясь наряду со словом элементом построения предложения, словосочетание выступает в качестве одной из основных синтаксических единиц. Непосредственной целью рассматриваемой процедуры поверхностного синтаксического анализа является предварительное разбиение ПС или ЛС на последовательность словосочетаний 2-х типов: устойчивые фразеологические словосочетания (ФРС) и грамматико-смысловые словосочетания (ГРС).

Структура блока членения ПС или ЛС на синтаксические синтагмы и их разметки на акцентные единицы, основанная на анализе словосочетаний, представлена на рис. 4.10.

Выделение фразеологических словосочетаний. В анализируемой синтагме отмечаются фразеологические словосочетания (ФРС), найденные в словаре устойчивых словосочетаний. К фразеологическим словосочетаниям относятся:

– фразеологические сращения – «попасть впросак», «бить баклуши», «ничтоже сумняшеся», «собаку съест» и др.

– фразеологические единства – «зайти в тупик», «бить ключом», «плыть по течению», «брать в свои руки», «прикусить язык» и др.

– фразеологические сочетания – «потупить взор», «щекотливый вопрос», «бархатный сезон», «поголовные аресты» и др.



Рис. 4.10. Структура блока разметки ПС и ЛС на синтаксические синтагмы

Выделяются следующие типы компонентного состава фразеологизмов:

– сочетание прилагательного с существительным: *краеугольный камень, заколдованный круг, лебединая песня*;

– сочетание существительного в именительном падеже с существительным в родительном падеже: *точка зрения, камень преткновения, бразды правления, яблоко раздора*;

– сочетание имени существительного в именительном падеже с существительными в косвенных падежах с предлогом: *кровь с молоком, душа в душу, дело в шляпе*;

– сочетание предложно-падежной формы существительного с прилагательным: *на живую нитку, по старой памяти, на короткой ноге*;

– сочетание глагола с существительным (с предлогом и без предлога): *окинуть взором, посеять сомнения, взять в руки, взяться за ум, водить за нос*;

– сочетание глагола с наречием: *попасть впросак, ходить босяком, видеть насквозь*;

– сочетание деепричастия с именем существительным: *спустя рукава, скрепя сердце, сломя голову*.

Позиции слабых и сильных словесных ударений в устойчивых сочетаниях могут быть определены в словаре сочетаний, при этом одно из слов обязательно несёт сильное ударение. При отсутствии помет слабых и сильных ударений вполне допустима установка сильного ударения на каждом из слов устойчивого словосочетания.

Объединение слов в грамматико-смысловые словосочетания. Грамматико-смысловым словосочетанием (ГрС) является пара по смыслу и грамматически связанных слов, выделяемая из предложения [7]. Например: «*нужная книга*», «*лекция по литературе*», «*бежать опрометью*», «*два студента*», «*несколько книг*».

Основной смысл выделения в ПС или ЛС словосочетаний типа ФрС и ГрС заключается в том, что теперь свобода разделения ПС или ЛС на СС ограничивается. Граница синтагмы может находиться только за пределами ФрС или ГрС, но не внутри их.

В зависимости от того, какое слово является первым в словосочетании, различаются основные лексико-грамматические группы словосочетаний. Классификация словосочетаний по признаку первого слова имеет следующую схему.

(1) Группа прилагательных словосочетаний. Эта группа включает прилагательные, местоимения-прилагательные, порядковые числительные и причастия, которые сочетаются:

- ***С существительным*** (*полезная книга, зелёную листву, свою находку, унесённые ветром*). Признаки словосочетания: прилагательное + существительное (в одном падеже).

- ***С инфинитивом*** (*способный работать, готовый учиться*). Признаки: прилагательное + инфинитив.

(2) Группа наречных словосочетаний. Эта группа сочетается:

- ***С инфинитивом*** (*безнаказанно игнорировать, хорошо петь*). Признаки: наречие + инфинитив.

- ***С наречием*** (*очень удачно, по-прежнему хорошо*). Признаки: наречие + наречие

- **С существительным** (*далеко от дома, наедине с сыном, незадолго до экзаменов*). Признаки: наречие + существительное (в косвенном падеже с предлогом).

- **С местоимением-существительным** (*недалеко от них, наедине с ней, незадолго до неё*). Признаки: наречие + местоимение-существительное (в косвенном падеже с предлогом).

(3) **Группа глагольных словосочетаний**. Глагольная группа сочетается:

- **С инфинитивом** (*предложил выучить, просит взять*). Признаки: глагол (в любой форме) + глагол (инфинитив).

- **С деепричастием** (*идёт оглядываясь, говорить улыбаясь*). Признаки: глагол (в любой форме) + деепричастие.

- **С наречием** (*поступал справедливо, заниматься вдвоем*). Признаки: глагол (в любой форме) + наречие.

- **С существительным** (*искать покоя, писал брату, стоять у дороги, подъехал к дому, встретиться с друзьями*). Признаки: глагол (в любой форме) + существительное в косвенном падеже.

- **С местоимением-существительным** (*искать их, писал ему, стоять около неё, подъехал к нему, встретился с ними*). Признаки: глагол (в любой форме) + местоимение-существительное в косвенном падеже.

(4) **Группа числительных словосочетаний**. Эта группа сочетается:

- **С существительным** (*две книги, оба друга, трое в шинелях, сто рублей*). Признаки: количественное или собирательное числительное + существительное, в одном падеже.

(5) **Группа существительных словосочетаний**. Эта группа включает существительные и местоимения-существительные и сочетается:

- **С существительным** (*письмо родителям, его доклада, оценку выступления, входом в театр*). Признаки: существительное + существительное (в отличающихся падежах без предлога или с предлогом).

- **С наречием** (*прогулка верхом, судак по-польски*). Признаки: существительное + наречие.

Перечисленные правила объединения слов в ГрС представлены в Табл. 4.2, где по горизонтали расположены типы групп словосочетаний в порядке степени (силы) его связности со вторыми в паре словами. В таблице указано также место предпочтительной установки полного (+) и частичного (=) ударений.

Таблица 4.2

Правила объединения слов в словосочетания

		Первое слово в паре				
		Прилагательное	Наречие	Глагол	Числительное	Существительное
Второе слово в паре		1	2	3	4	5
Инфинитив	1	(=) (+)	(=) (+)	(=) (+)	–	–
Деепричастие	2	–	–	(=) (+)	–	–
Наречие	3	–	(+) (=)	(=) (+)	–	(=) (+)
Существительн.	4	(=) (+)	(+) (=)	(+) (=)	(+) (=)	(+) (=)
Местоимение	5	–	(+) (=)	(+) (=)	–	–

Из таблицы видно, что в соответствии с правилами русской грамматики [7], допустимыми и наиболее частотными являются 14 различных типов ГрС. С учётом этого предлагается следующая последовательность действий по разметке синтагм на словосочетания.

1. В синтагме отыскиваются пары слов - прилагательные словосочетания, состоящие из слова группы прилагательных и стоящего справа от него существительного либо инфинитива глагола. Эти пары слов объединяются в словосочетания. Если такой пары не находится, то слово из группы прилагательных остаётся «одиноким».

2. Затем в синтагме рассматриваются оставшиеся слова, т.е. не объединённые в словосочетания по п. 1, и отыскиваются пары слов - наречные словосочетания, состоящие из двух наречий или наречия и стоящего справа от него инфинитива глагола, либо существительного или местоимения-существительного с предлогом. Если таковые находятся, то они объединяются в словосочетания, если нет, то наречие остаётся «одиноким».

3. Далее в синтагме рассматриваются оставшиеся слова, и отыскиваются пары слов - глагольные словосочетания, т.е. глагол в любой форме и стоящие справа от него наречие, инфинитив или деепричастие, которые объединяются в одно словосочетание. Если таковых не обнаружено, то глагол может быть объединён с существительным или с местоимением-существительным в косвенном падеже, стоящим справа. Если их нет, то глагол остаётся «одиноким».

4. В оставшихся необъединённых словах ищутся пары слов - числительные словосочетания, состоящие из количественного или собирательного числительного и стоящего справа от него существительного, согласованного с числительным по падежу, которые объединяются в словосочетания. Если нет, то числительное остаётся «одиноким».

5. Наконец, в оставшихся необъединённых словах ищутся соседние пары слов - существительные словосочетания, состоящие из слова группы существительных и стоящего справа от него наречия либо существительного или местоимения-существительного, которые объединяются в словосочетания. Если таких слов не находится, то существительное остаётся «одиноким».

Ниже приведены примеры разметки синтагм на словосочетания в соответствии с предложенной последовательностью действий. Словосочетания отмечены квадратными скобками, в круглых скобках указан тип словосочетания (см. таблицу 4.2).

Пример 1: Если Вам [необходимо активировать(2)] [услугу передачи(5)] данных для Вашего [мобильного номера(1)].

Пример 2: Но благодаря [разумному сочетанию(1)] лекарств он [смог остановить(3)] [развитие болезни(5)] [в большинстве случаев(5)].

Пример 3: Тогда тарификация [Ваших звонков(1)] [начинается с момента(3)] соединения [с телефоном абонента(5)].

Пример 4: [Идеальным решением(1)] [проблемы членения(5)] [такого рода(1)] предложений на синтагмы [было бы использование(3)] [комплекса правил(5)] разбора [на синтаксические компоненты(1)].

Расширение двухсловных ГрС до трёх и более словных сочетаний Если синтагма, обработанная в соответствии с указанной выше в п. 1 – 5 последовательностью действий по разметке синтагм на словосочетания, содержит слова, не вошедшие в созданные двухсловные сочетания, то рассматривается возможность их расширения до трёх- и более словных сочетаний по следующей схеме.

1. Рассматриваются полученные «прилагательные словосочетания».

а) если перед прилагательным словосочетанием стоит слово из группы прилагательных, то оно дополнительно включается в это словосочетание.

Пример 1: Если Вам [необходимо активировать(2)] {[услугу передачи(5)] данных} {для Вашего [мобильного номера(1)]}.

Здесь и далее в круглых скобках указан тип словосочетания, в квадратных – двухсловные сочетания, в фигурных – расширенное по п.1 трёхсловное сочетание.

б) если после прилагательного словосочетания стоит слово из группы существительных в родительном падеже, то оно дополнительно включается в это словосочетание.

Пример 2: Но благодаря {[разумному сочетанию(1)] лекарств} он [смог остановить(3)] [развитие болезни(5)] [в большинстве случаев(5)].

2. Если после наречного или глагольного словосочетания стоит слово из группы существительных в родительном падеже, то оно дополнительно включается в это словосочетание.

Пример 3: *Тогда тарификация [Ваших звонков(1)] {[начинается сЪмомента(3)] соединения} [сЪтелефоном абонента(5)].*

3. Если после числительного или существительного словосочетания стоит слово из группы существительных в родительном падеже, то оно дополнительно включается в это словосочетание.

Пример 4: *[Идеальным решением(1)] [проблемы членения(5)] {[такого рода(1)] предложений} наЪсинтагмы [былоЪбы использование(3)] {[комплекса правил(5)] разбора} [наЪсинтаксические компоненты(1)].*

В приведенных выше примерах 1 – 4, несмотря на реализацию полной последовательности действий по разметке синтагм на словосочетания и по их расширению, остаются ещё отдельные слова, не вошедшие ни в одно из словосочетаний. Следующим, последним шагом, является дополнительное расширение словосочетаний путём включения в их состав слабоударных слов, таких, как многосложные предлоги, союзы и местоимения. Если же и после такого расширения словосочетаний остаются отдельные слова, то они определяются как частный случай однословных сочетаний.

В результате описанных процедур примеры 1 – 4 переписутся в следующем виде.

Пример 1: *{Если Вам [необходимо активировать(2)]} {[услугу передачи(5)] данных} {дляЪвашего [мобильного номера(1)]}.*

Пример 2: *{Но благодаря [разумному сочетанию(1)] лекарств} {он [смог остановить(3)]} [развитие болезни(5)] [вЪбольшинстве случаев(5)].*

Пример 3: *{Тогда [тарификация]} [Ваших звонков(1)] {[начинается сЪмомента(3)] соединения} [сЪтелефоном абонента(5)].*

Пример 4: *[Идеальным решением(1)] [проблемы членения(5)] {[такого рода(1)] предложений} [наЪсинтагмы] [былоЪбы использование(3)] {[комплекса правил(5)] разбора} [наЪсинтаксические компоненты(1)].*

Расстановка сильных и слабых ударений и разметка на акцентные единицы. Для двухсловных сочетаний в таблице 4.2 указано место предпочтительной установки сильного и слабого ударений. Такое распределение позиций ударений не претендует, конечно, на универсальность. Оно характеризует среднестатистическую тенденцию для достаточно широкого набора различных текстов. При определённых условиях (индивидуальная манера чтения, стремление к определённой ритмической структуре и др.) знаки (+) (=) для данного словосочетания могут меняться местами, либо оба знака индицировать сильное ударение

– (+) (+). Важную роль может играть также наличие некоторых индикаторов потенциальной «слабости» или «силы» какого-либо из слов в словосочетании. В частности, к индикатору «слабости» может быть отнесена принадлежность слова к группе потенциально слабоударных слов, таких, как многосложные предлоги и частицы, союзы и местоимения (список потенциально слабоударных слов приведен в Приложении 6). К индикатору «силы» может быть отнесено наличие перед словом усилительной или отрицательной частицы.

После применения указанных правил расстановки сильных и слабых ударений примеры 1 – 4 переписываются в следующем виде.

Пример 1: {Е=сли Ва=м [необходи=мо активи+ровать(2)]} { [услу+гу переда=чи(5)] да=нных} {дляЪва=шего [моби=льного но+мера(1)]}.

Пример 2: {Но= благодаря= [разу=мному сочета+нию(1)] лека=рств} {о=н [смо=г остано+вить(3)]} [разви+тие боле=зни(5)] [вЪбольшинстве+ слу=чаев(5)].

Пример 3: {Тогда= [тарифика+ция]} [Ва=ших звонко+в(1)] { [начина+ется сЪмоме+нта(3)] соедине=ния} [сЪтелефо+ном абоне=нта(5)].

Пример 4: [Идеа=льным реше+нием(1)] [пробле+мы члене=ния(5)] { [тако=го ро+да(1)] предложе=ний} [наЪсинта+гмы] [бы+лоЪбы испо=льзование(3)] { [ко+мплекса пра=вил(5)] разбо=ра} [наЪсинтакси=ческие компоне+нты(1)].

На заключительном этапе целесообразно провести окончательную корректировку позиций сильных и слабых ударений с точки зрения приближения к оптимальной организации ритмической структуры синтагм. При этом уточняются ситуации, когда в ФрС или в ГрС имеется более одного слова со слабым ударением. Окончательная корректировка осуществляется, исходя из необходимости соблюдения следующих условий:

– в ФрС или в ГрС не должно быть двух следующих подряд слов со слабым ударением. В этом случае в одном из этих слов, например во втором, слабое ударение заменяется на сильное.

– в ФрС или в ГрС количество слов со слабым ударением не должно быть больше количества слов с сильным ударением. Например, последовательность (=) (+) (=) заменяется на последовательность (=) (+) (+).

Следует заметить, что приведенные здесь правила отражают лишь среднестатистические закономерности. Окончательные условия особой выделенности того или иного слова могли бы быть адекватно определены только в результате глубокого синтаксического и семантического анализа предложений, что в настоящий момент пока недостижимо.

После применения указанных правил корректировки примеры 1– 4 переписываются в следующем виде.

Пример 1: {Е=сли Ва+м [необходи=мо активи+ровать(2)]} {услу+гу переда=чи(5)} {да+нных} {дляЪва=шего [моби+льного но+мера(1)]}.

Пример 2: {Но= благодаря+ [разу=мному сочета+нию(1)] лека+рств} {о=н смо+г остано+вить(3)} [разви+тие боле=зни(5)] [вЪбольшинстве+ слу=чаев(5)].

Пример 3: {Тогда= [тарифика+ция]} [Ва=ших звонко+в(1)] {начина+ется сЪмоме+нта(3)} соедине=ния} [сЪтелефо+ном абоне=нта(5)].

Пример 4: [Идеа=льным реше+нием(1)] [пробле+мы члене=ния(5)] {тако=го ро+да(1)} предложе+ний} [наЪсинта+гмы] [бы+лоЪбы испо=льзование(3)] {[ко+мплекса пра=вил(5)] разбо+ра} [наЪсинтакси=ческие компоне+нты(1)].

Разметка полученной последовательности слов на акцентные единицы осуществляется по следующим правилам:

1. Разметка на АЕ осуществляется отдельно для каждой ФрС или ГрС.

2. Если в ФрС или ГрС имеются слова со слабым ударением, то каждое из них объединяется в одну АЕ с сильноударным словом, стоящим слева или справа от него.

3. Оставшиеся слова с сильным ударением отмечаются как отдельные АЕ.

После применения указанных правил примеры 1 – 4 переписываются в следующем виде (где круглыми скобками отмечены полученные АЕ в каждом из ГрС, которые ограничены фигурными скобками и помещены на отдельных строках., причём справа от каждой строки указано количество АЕ в данной ГрС):

Пример 1:

{(е=сли Ва+м) (необходи=мо активи+ровать)}	[2]
{(услу+гу переда=чи) (да+нных)}	[2]
{(дляЪва=шего моби+льного) (но+мера)}	[2]

Пример 2:

{(но= благодаря+) (разу=мному сочета+нию) (лека+рств)}	[3]
{(о=н смо+г) (останови+ть)}	[2]
{(разви+тие боле=зни)}	[1]
{(вЪбольшинстве+ слу=чаев)}	[1]

Пример 3:

{(Тогда= тарифика+ция)}	[1]
{(Ва=ших звонко+в)}	[1]
{(начина+ется) (сЪмоме+нта соедине=ния)}	[2]

{(сЪтелефо+ном абоне=нта)} [1]

Пример 4:

{(Идея=льным реше+нием)} [1]

{(пробле+мы члене=ния)} [1]

{(тако=го ро+да) (предложе+ний)} [2]

{(наЪсинта+гмы)} [1]

{(бы+лоЪбы испо=льзование)} [1]

{(ко+мплекса пра=вил) (разбо+ра)} [2]

{(наЪсинтакси=ческие компоне+нты)} [1]

Разметка ПС и ЛС на синтаксические синтагмы. Как уже указывалось, основной смысл предварительного разбиения ПС или ЛС на ФрС и ГрС заключается в том, что теперь свобода разделения ПС или ЛС на СС ограничивается, т.к. граница между СС не может находиться внутри ФрС или ГрС. В простейшем случае границей каждой СС могут служить границы ГрС. В этом случае, как видно из примеров 1 – 4, каждая СС будет включать различное количество АЕ: от 1-й до 3-х. Если же требуемый стиль чтения предполагает, что СС должна включать по возможности не менее 2-х АЕ, то в этом случае получим схему членения, показанную на примерах 1 – 4 ниже, где справа от каждой строки указано количество АЕ в данной СС.

Пример 1:

{(е=сли Ва+м) (необходи=мо активи+ровать)} [2]

{(услу+гу переда=чи) (да+нных)} [2]

{(дляЪва=шего моби+льного) (но+мера)} [2]

Пример 2:

{(но=благодаря+) (разу=мному сочета+нию) (лека+рств)} [3]

{(о=н смо+г) (останови+ть)} [2]

{(разви+тие боле=зни)} {(вЪбольшинстве+слу=чаев)} [2]

Пример 3:

{(Тогда=тарифика+ция)} {(Ва=ших звонко+в)} [2]

{(начина+ется) (сЪмоме+нта соедине=ния)} [3]

{(сЪтелефо+ном абоне=нта)}

Пример 4:

{(Идея=льным реше+нием)} {(пробле+мы члене=ния)} [2]

{(тако=го ро+да) (предложе+ний)} {(наЪсинта+гмы)} [3]

{(бы+лоЪбы испо=льзование)} {(ко+мплекса пра=вил)} [3]

{(разбо+ра)}

{(наЪсинтакси=ческие компоне+нты)} [1]

4.3.3. Блок маркировки интонационного типа синтагм

Структура блока интонационной маркировки синтагм в предложении представлена на рис. 4.10.

Маркировка интонационного типа предложения. Как было сказано выше, при синтезе речи по тексту предложение может быть выделено по одному из следующих восьми знаков: [·], [?], [?!], [!], [!!!], [...], [*],[#]. Знаки конца предложения определяют три его основные интонационные категории:

- повествовательное – *Nr* (Narration),
- вопросительное – *In* (Interrogation),
- восклицательное (побудительное) – *Ex* (Exclamation).

Категория **повествовательных предложений** – *Nr* – характеризуется повествовательной или, иначе, завершённой интонацией – *F* (Finality). Категория распознаётся по знакам [·], [...], [*], [#], которые определяют её интонационный тип, обозначаемый при обработке текста, соответственно, символами:

- *F0* – интонация «точки» - [·],
- *F1* – интонация «многоточия» - [...],
- *F2* – интонация «заголовка» - [*],
- *F3* – интонация «абзаца» - [#].

Кроме указанных основных четырёх интонационных типов повествовательных предложений, выделяемых пунктуационно, возможно расширение их количества, например, путём дополнения их подтипами, характеризующими прямую или косвенную речь и др.

Категория **вопросительных предложений** – *In* – распознаётся по знакам: [?], [?!] и обозначается при обработке текста, соответственно, символами интонационных типов:

- *I0* – интонация «вопроса» - [?],
- *I1* – интонация «вопроса-восклицания» - [?!] .

Кроме указанных 2-х типов интонации вопросительных предложений, выделяемых пунктуационно, каждый из них может иметь несколько подтипов, таких как интонация:

- общего вопроса,
- частного вопроса,
- альтернативного вопроса,
- вопроса-переспроса,
- вопроса-отрицания и др.

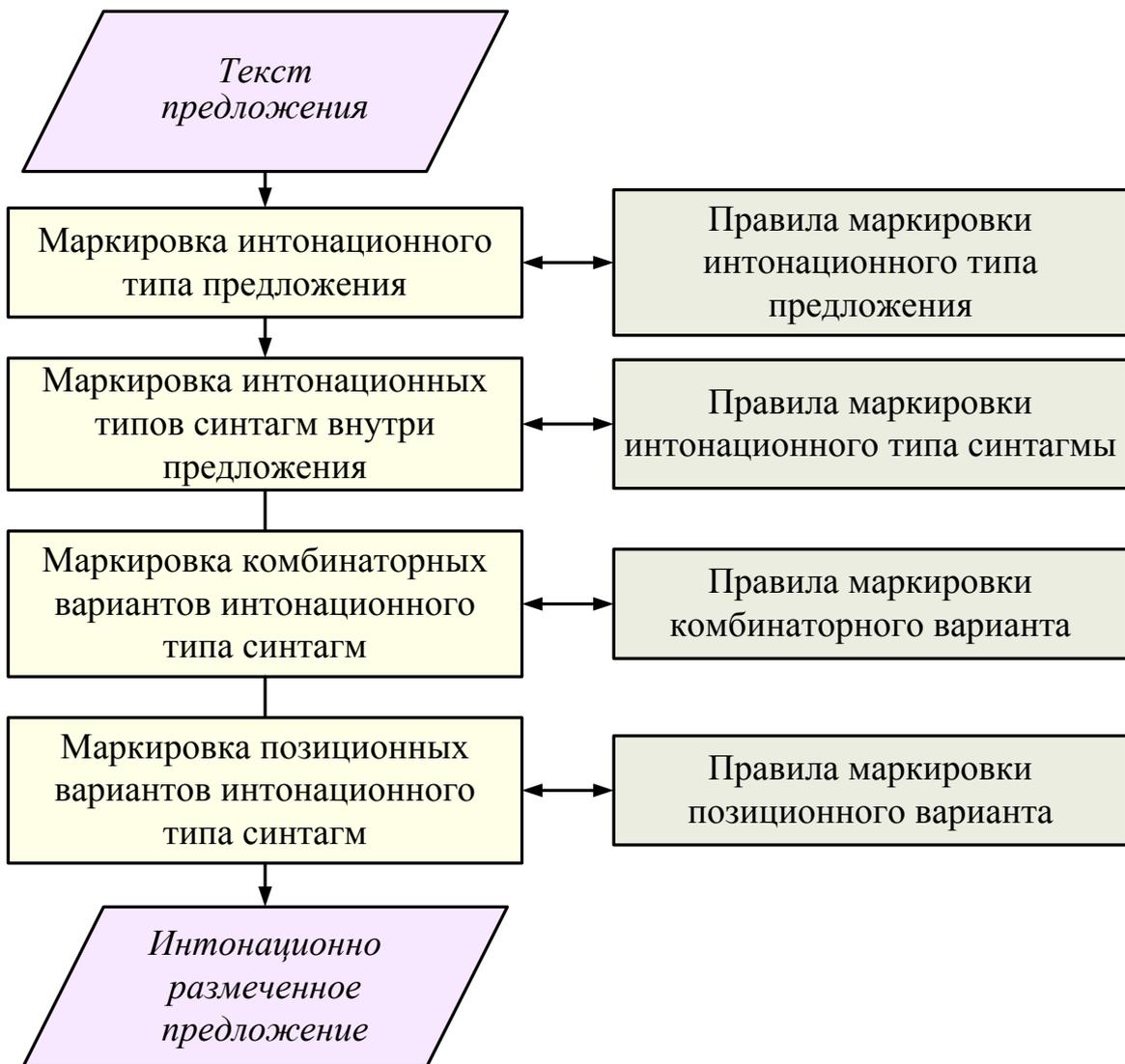


Рис. 4.10. Структура блока интонационной разметки синтагм в предложении

Категория *восклицательных и побудительных предложений* – *Ex* – распознаётся по знакам: [!], [!!!] и обозначается, соответственно, символами интонационных типов:

- *E0* – интонация восклицания-побуждения - [!],
- *E1* – интонация эмоционального восклицания - [!!!].

Кроме указанных 2-х типов интонации восклицательных предложений, выделяемых пунктуационно, каждый из них может иметь несколько подтипов, выражающих определённые чувства и побуждения с помощью различного рода междометий (см. Приложение 6)

Маркировка интонационных типов синтагм внутри предложения. Особенности маркировки интонационного типа синтагм внутри предложения рассмотрим на примере повествовательных предложений.

Кроме перечисленных выше основных пунктуационных типов интонации завершенности, реализующихся в последней ПС или ЛС предложения, внутри него могут присутствовать также два дополнительных пунктуационных типа интонации, характеризующихся различной степенью завершенности:

- **F4** – интонация «точки с запятой» – [;],
- **F5** – интонация «вводности» – [)], [,-], [-].

Интонация «водности» реализуется при условии, что указанным знакам непосредственно предшествует, соответственно, знаки [(], [,-], [-].

Внутри предложения могут присутствовать также 5 пунктуационных подтипов интонации, характеризующихся различной степенью незавершенности:

- **N0** – интонация «запятой»- [,],
- **N1** – интонация «тире» - [-],
- **N2** – интонация «двоеточия» - [:],
- **N3** – интонация «предвводности»- [(], [,-], [-].

Интонация «предводности» реализуется при условии, что за указанными знаками в тексте предложения следуют, соответственно, знаки [)], [,-], [-].

Пунктуационные синтагмы, в свою очередь, могут содержать лексические синтагмы с интонацией незавершенности следующих 3-х типов (см. раздел 4.3.1):

- **N4** – интонация «союза И»,
- **N5** – интонация «союза ИЛИ»,
- **N6** – интонация лексических синтагм – [ИС], [АБ], [РЧ], [НВ].

Далее, как само предложение, так и входящие в него пунктуационные и лексические синтагмы могут содержать неопределённое количество синтаксических синтагм с характерной для них интонацией незавершенности:

- **N7** – интонация синтаксических синтагм.

Примеры:

*Возможность победы **БНФ** [N6] вызывает большие сомнения[F0].*

*В пробирке оказалось **2 миллиона** [N6] **350 тысяч** [N6] молекул белка[F0].*

*Сегодня в **10 часов** [N6] **15 минут** [N6] **34 секунды**[F0].*

*Он приехал в **четверг** [N6] **20-го июня** [N6] **7-го года**[N6]навсегда[F0].*

Время от времени [N3] – для разрядки [F5] – он вставлял шутки[F0].

Любой народ [N3], - говорил он [F5], - достоин уважения[F0].

Маркировка комбинаторных вариантов интонационного типа синтагм. Многие из рассмотренных выше интонационных типов ПС и

ЛС могут иметь определённые комбинаторные варианты. Это замечание в наибольшей степени касается интонационных типов: *N0* - [.] и *F0* - [./]. Причиной возникновения комбинаторных вариантов являются определённые различия в левом и правом контекстах анализируемой синтагмы, определяемые типом союзного слова, используемого наряду со знаком запятой для разделения синтагм интонационных типов *N0*, *F0*. При этом комбинаторные варианты интонационного типа *N0* образуются за счёт различий в правом контексте ПС, а *F0* - за счёт различий в левом контексте. Запятой и союзом могут отделяться однородные члены внутри предложения, а также сложносочинённые и сложноподчинённые предложения. Рассмотрим подробнее особенности возникновения комбинаторных вариантов интонационных типов *N0* и *F0*.

Варианты интонационного типа *N0* и *F0* с сочинительными союзами. (Список сочинительных союзов приведён в Приложении 6). Можно выделить следующие два основных варианта интонирования синтагм в зависимости от способа отделения однородных членов и сложносочинённых предложений друг от друга:

1. Отделяемые запятой и следующими за ней соединительными или разделительными союзами: *и, ни...ни, или, либо, ли...ли, то...то*, и др.
Комбинаторный вариант (0) - *N0.0, F0.0*

Примеры:

И пращ [N0.0], и стрела [N0.0], и лукавый кинжал [F0.0].

За дождем не видно было ни моря [N0.0], ни неба [F0.0].

Гаврила либо сбежал [N0.0], либо утонул [F0.0].

Стало совсем темно [N0.0], и улица мало-помалу опустела [F0.0].

Слышался звон бутылки [N0.0], или начинал шуметь механизм [F0.0].

2. Отделяемые запятой и следующими за ней противительными союзами: *а, но, да* (в значении «но»), *однако* и др.

Комбинаторный вариант (1) - *N0.1, F0.1*

Примеры:

На смелого собака лает [N0.1], а трусливого кусает [F0.1].

Он был силен [N0.1], да не умен [F0.1].

Я звал тебя [N0.1], но ты не оглянулась [F0.1].

Варианты интонационного типа *N0* и *F0* с обособленными членами предложения. Обособлением называется смысловое и интонационное выделение второстепенных членов с целью придать им некоторую самостоятельность в предложении. Выделяются следующие виды обособленных членов предложения:

3. Обособление, выраженное причастием.

Комбинаторный вариант (2) - *N0.2, F0.2*

Пример:

Внезапно он улетел [N0.2], встревоженный вихрем [F0.2].

4. Обособление, выраженное деепричастием.

Комбинаторный вариант (3) - N0.3, F0.3.

Пример:

Длинная стружка лезла из рубанка [N0.3], завиваясь штопором [F0.3].

Вариант интонационного типа N0, F0 в сложноподчиненном предложении. Сложноподчиненным называется сложное предложение, части которого связаны подчинительными союзами или относительными (союзными) словами. Придаточная часть от главной части отделяется подчинительными союзами места, причины, образа действия, сравнения, цели и др. Список подчинительных союзов приведен в Приложении 6.

5. Сложноподчинённое предложение.

Комбинаторный вариант (4) - N0.4, F0.4.

Примеры:

Все заглядывали вперед [N0.4], где качалось красное знамя [F0.4].

Базар кишел народом [N0.4], потому что было воскресенье [F0.4].

Потом он потянулся так [N0.4], что хрустнули кости [F0.4].

Не жалко сил [N0.4], если они дают результаты [F0.4].

Вдоль дороги вился извилистый след [N0.4], какой оставляет зверь [F0.4].

Замечание. При отсутствии признаков, определяющих указанные выше интонационные варианты 1 – 5, второму индексу присваивается значение «0».

Пример:

Впереди виднелись горы [N0.0], их вершины блестели [F0.0].

Маркировка позиционных вариантов интонационного типа синтагм. Многие из рассмотренных выше интонационных типов могут иметь определённые варианты, зависящие от позиции предложения в абзаце, от позиции пунктуационной или лексической синтагмы в предложении и от позиции синтаксической синтагмы в ПС или в ЛС.

В процессе синтеза речи особенно важно избежать так называемой «монотонности второго рода». Этот вид монотонности проявляется при использовании одних и тех же интонационных контуров для двух или более идущих подряд синтагм одного интонационного типа. В естественной речи говорящий, как правило, стремится избежать такого рода монотонности путём варьирования интонационных параметров. Это замечание в наибольшей степени касается интонационных подтипов **F0.0**, **N0.0**, и **N7.0**, для которых частота последовательного появления в текстах весьма значительна. Определим минимально необходимый набор позиционных вариантов указанных интонационных типов.

Позиционные варианты интонации завершённости – F0.0:

Позиционный вариант (0)

– *F0.0.0*, при условии, что интонационный подтип *F0.0* встретился **в абзаце** впервые или в третий, пятый и т.д. нечётный раз подряд, и так - вплоть до конца абзаца;

Позиционный вариант (1)

– *F0.0.1*, при условии, что интонационный подтип *F0.0* встретился **в абзаце** во второй, четвёртый и т.д. чётный раз подряд, и так вплоть до конца абзаца.

Позиционные варианты интонации незавершённости – N0.0:

Позиционный вариант (0)

– *N0.0.0* при условии, что этот интонационный тип встретился **в предложении** впервые или в третий, пятый и т.д. нечётный раз подряд раз подряд, и так вплоть до конца предложения;

Позиционный вариант (1)

– *N0.0.1*, при условии, что интонационный тип *N1* встретился **в предложении** во второй, четвёртый и т.д. чётный раз подряд, и так вплоть до конца предложения.

Позиционные варианты интонации незавершённости – N7.0:

Позиционный вариант (0)

– *N7.0.0*, при условии, что этот интонационный тип встретился **в пунктуационной или лексической синтагме** впервые или в четвёртый, седьмой и т.д. раз подряд, и так - вплоть до конца ПС или ЛС предложения;

Позиционный вариант (1)

– *N7.0.1* при условии, что интонационный тип *N7.0* встретился **в пунктуационной или лексической синтагме** во второй, пятый и т.д. раз подряд, и так - вплоть до конца ПС или ЛС предложения;

Позиционный вариант (2)

– *N7.0.2* при условии, что интонационный тип *N7.0* встретился **в пунктуационной или лексической синтагме** в третий, шестой и т.д. раз подряд, и так - вплоть до конца ПС или ЛС предложения.

При необходимости подобным же образом можно создавать позиционные варианты других интонационных типов, рассмотренных выше.

Пример интонационной маркировки синтагм повествовательного предложения:

Уже давно известно [N0.4.0], что математическая логика [N7.0.0] как научная дисциплина [N7.0.1] возникла на базе [N7.0.2] так называемой формальной логики [N4.0.0] и первое время [N7.0.0] развивалась в связи со стремлением [N7.0.1] формализовать некоторые стороны [N7.0.2] интеллектуальной деятельности человека [F0.0.0].

В общем случае интонационные тип, подтип, комбинаторный и позиционный варианты каждой синтагмы в повествовательном

предложении обозначаются следующим набором букв и цифровых индексов:

1. Интонационный тип синтагмы: завершённый – F , или незавершённый – N .

2. Интонационный подтип синтагмы: первый индекс - $i = 0, 1, 2, \dots$

3. Комбинаторный вариант подтипа синтагмы: второй индекс - $j = 0, 1, 2, \dots$

4. Позиционный вариант подтипа синтагмы: третий индекс - $k = 0, 1, 2, \dots$

Рассмотренный подход к маркировке интонационных подтипов синтагм в повествовательных предложениях, их комбинаторных и позиционных вариантов может быть использован также и при маркировке вопросительных и восклицательно-побудительных предложений.

4.4. Фонетический процессор

Задачей фонетического процессора является преобразование орфографического текста в последовательность аллофонов, которая используется на этапе акустической обработки при синтезе речевого сигнала.

В фонетическом процессоре заложены правила преобразования орфографического текста в последовательность фонем (преобразование буква-фонема) и правила преобразования последовательности фонем в аллофонную последовательность (преобразование фонема-аллофон). Общая структура фонетического процессора представлена на рис. 4.11.

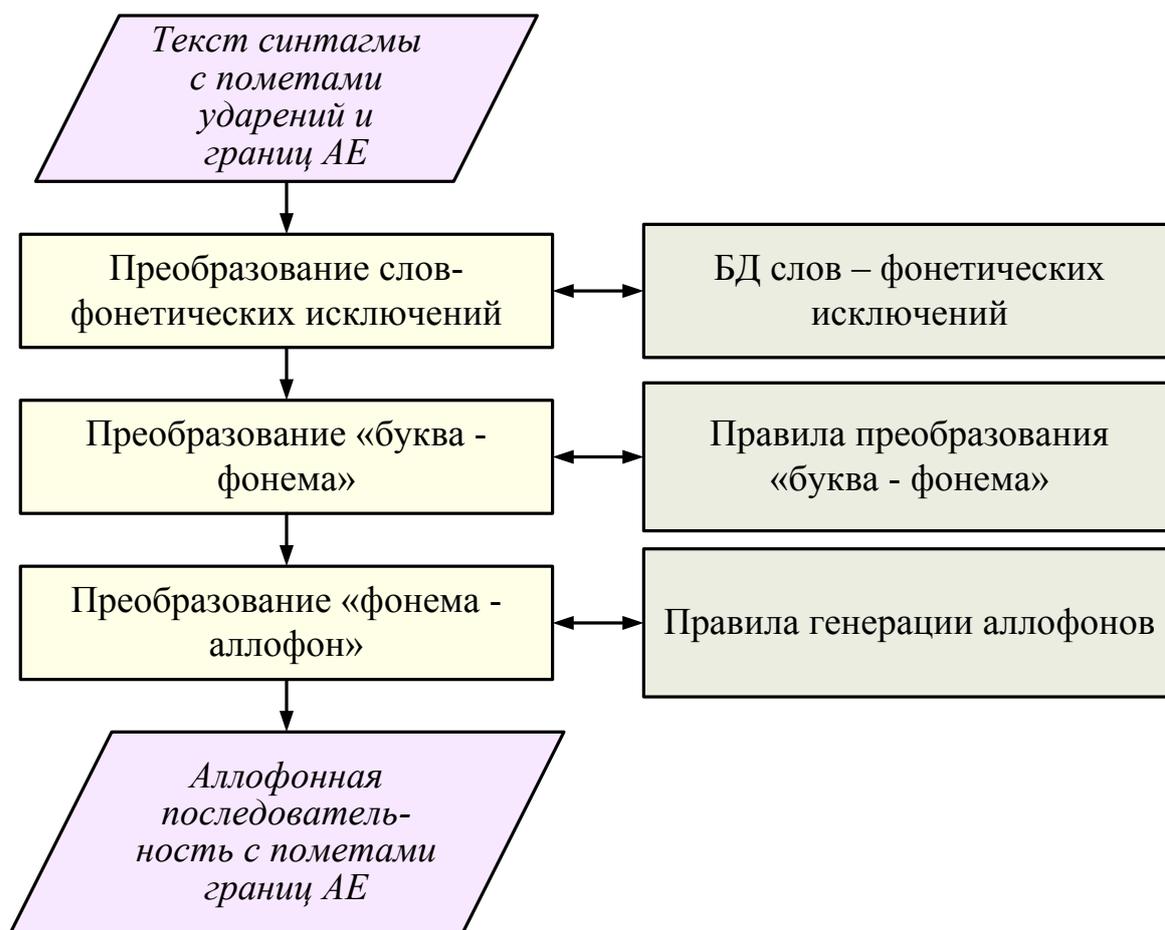


Рис. 4.11. Структура фонетического процессора

4.4.1. Блок преобразования слов – фонетических исключений

На вход процессора подаётся орфографический текст синтагмы с пометами словесных ударений и границ акцентных единиц. На начальном этапе осуществляется поиск каждого слова синтагмы в БД слов – фонетических исключений. В случае, если слово найдено, оно заменяется на соответствующий эквивалент.

В слова – фонетические исключения – входит большое количество иностранных слов или слов с иностранным корнем, например, «*ателье*», «*варьете*», «*декольте*», «*интервью*», «*кабаре*», «*кафе*», «*кашне*», «*моделировать*», «*филателист*», в которых согласный перед «*е*» не смягчается. Слова-исключения преобразуются в их эквиваленты, подчиняющиеся стандартным правилам преобразования буква-фонема. Для приведенных примеров эквивалентами будут, соответственно: «*атэльэ*», «*варьэтэ*», «*декольтэ*», «*интэर्वью*», «*кабарэ*», «*кафэ*», «*кашнэ*», «*модэлировать*», «*филатэлист*». В состав слов – фонетических исключений - входят также частоупотребительные слова, такие как «*пожалуйста*» (эквивалент – «*пожалуста*»), «*здравствуйте*»

(эквивалент «здравствуйте»), «**что**» (эквивалент – «**што**»), «**чувства**» (эквивалент «**чуста**»).

Список слов – фонетических исключений – приведен в Приложении 6.

4.4.2. Блок преобразования буква-фонема

На следующем этапе по стандартным правилам осуществляется преобразование буква – фонема, учитывающее произносительные особенности для русского языка. Основные регулярные правила преобразования буква-фонема были описаны в разделе 1.4. Здесь мы отметим некоторые особенности преобразования для согласных букв, не отражённые приведенными выше регулярными правилами, а также опишем правила преобразования буква-фонема на стыке слов.

Дополнительные внутрисловные правила преобразования буква-фонема. В русском языке существуют трёхчленные сочетания согласных, в которых один из согласных не произносится. К таким сочетаниям относятся «стн», «стл», «нтг», в которых не произносится «т»; «здн», «здц», «ндц», «рдц», «ндш», «гдт», в которых не произносится «д»; сочетание «лнц», в котором не произносится «л».

Примеры: **частный** – *ch', a, s, n, y, j'*; **счастливый** – *sh', a, s, l', i, v, y, j'*; **рентген** – *r', e, n, g', e, n*; **поздно** – *p, o, z, n, a*; **под уздцы** – *p, a, d, u, s, c, y*; **голландцы** – *g, a, l, l, a, n, c, y*; **сердце** – *s', e, r, c, e*; **ландшафт** – *l, a, n, sh, a, f, t*; **ягдташ** – *j', a, g, t, a, sh*; **солнце** – *s, o, n, c, e*.

Сочетания букв «сч», «жч» преобразуются в фонему «sh'».

Примеры: **счастье** – *sh', a, s', t', j', e*; **перебежчик** – *p', e, r', e, b', e, sh', i, k*.

Сочетания букв «тс», «тьс» преобразуются в фонему «с», а сочетания «тц», «дс» и «дц» - две фонемы «сс».

Примеры: **переволноваться** – *p', e, r', e, v, a, l, n, a, v, a, c, c, a*; **ручается** – *r, u, ch', a, j', e, c, c, a*; **блюдец** – *b, l', u', c, c, e*; **отца** – *a, c, c, a*.

Сочетание «гк» преобразуются в последовательность фонем «hk» или «h'k'».

Примеры: **лёгкий** – *l', o, h', k', i, j'*; **легко** – *l, e, h, k, o*.

Сочетания букв «ого», «его» в конце слова преобразуются в сочетания фонем «ava», «j'eva» за исключением словоформ производных от слов «**много, дорого**».

Примеры: **красного** – *k, r, a, s, n, a, y, a*; **синего** – *s', i, n', e, y, a*.

Сочетание букв «сш» преобразуется в удвоенную фонему «sh», а букв «ЗЖ» – в удвоенную /zh/.

Примеры: **расшиб** – *r, a, sh, sh, y, p*; **въезжать** – *v, j', e, zh, zh, a, t'*.

Правила межсловного преобразования буква-фонема на стыке служебного и знаменательного слов. Объединение 2-х или 3-х

орфографических слов в одно фонетическое слово осуществляется путём присоединения служебных (предлоги и частицы) к знаменательным словам. При этом предлоги {без, безо, близ, в, во, вне, для, до, за, из, изо, к, ко, меж, на, над, надо, о, об, обо, от, ото, по, под, подо, пред, предо, при, про, с, со, у, чрез} и частица {не} присоединяются к последующему слову, а частицы {бы, де, ли, же, -то, -ка, -либо, -нибудь, -таки} - к предшествующему слову.

Приведём правила межсловного преобразования для букв, разделённых пробелом, на стыке служебного и знаменательного слов внутри фонетического слова.

Правило 1. Если последняя буква первого слова

а) гласная:

– после объединения полностью сохраняются внутрисловные правила.

Примеры: *по дороге* [p, a, d, a, r, o, g', e], *на столе* [n, a, s, t, a, l', e].

б) согласная:

– внутрисловные правила ассимиляции по глухости-звонкости сохраняются полностью.

Примеры: *в саду* [f, s, a, d, u], *в лесу* [v, l', e, s, u], *с дороги* [z, d, a, r, o, g', i], *с Васей* [s, v, a, s', e, j'].

– в отличие от внутрисловных правил, твёрдая согласная не переходит в мягкую фонему перед гласными Я, Ё, Ю, Е, И.

Примеры: *под ёлкой* [p, a, d, j', o, l, k, a, j'], *Но: потёмки* [p, a, t', o, t, k', i].

– в отличие от внутрисловных правил, отсутствует ассимиляция с последующей мягкой согласной, одинаковой по месту образования.

Примеры: *под сеном* [p, a, t, s', e, n, a, t], *без деда* [b', e, z, d', e, d, a]. *Но: везде* [v', e, z', d', e].

Правило 2. Если первая буква второго слова

а) гласная:

– внутрисловные правила сохраняются, но не полностью.

– в отличие от внутрисловных правил, О и И после Ъ не переходят в сочетания фонем [j', o] и [j', i] соответственно. Гласная О переходит в [o] в ударной позиции и в [a] в безударной, гласная И переходит в [i].

Примеры: *иль осень наступила* [i, l', o, s', e, n', ...], *иль иволга поёт* [i, l', i, v, o, l, g, a, ...], *но чьи* [ch', j', i], *бульон* [b, u, l', j', o, n].

– гласная И после всех согласных, кроме Ъ, переходит в фонему [y].

Примеры: *под ивой* [p, a, d, y, v, a, j'], *с Иваном* [s, y, v, a, n, o, t]. *Но: ведь Иван* [v, e, d', i, v, a, n].

б) согласная:

– внутрисловные правила сохраняются полностью.

Правила межсловного преобразования буква-фонема на стыке двух знаменательных слов.

Правило 1. Если последняя буква первого слова

а) гласная:

– внутрисловные правила сохраняются полностью.

б) согласная:

– внутрисловные правила ассимиляции по глухости-звонкости сохраняются, но не полностью.

– озвончение глухих согласных П(ь), Т(ь), К(ь), Ф(ь), С(ь), Ш(ь), Щ, Ц, Ч(ь) происходит, когда первой буквой второго слова являются Б, Д, Г, З или Ж (как и для внутрисловного преобразования). Особенностью реализации межсловной ассимиляции в этом случае является появление «неканонических» фонем русской речи, возможных только на стыках слов при озвончении согласных Щ, Ц, Ч, а именно: Щ > [zh'], Ц > [dz], Ч > [dzh'].

Примеры: *борщ горячий* [b, o, r, zh', g, a, r', a, ch', i, j'], *заяц беляк* [z, a, j', a, dz, b', e, l', a, k], *дочь гуляет* [d, o, dzh', g, u, l', a, j', e, t].

– в отличие от внутрисловных правил ассимиляции, звонкие согласные Б(ь), В(ь), Г, Д(ь), Ж(ь), З(ь) не оглушаются только в тех случаях, когда первой буквой второго слова являются Б, Д, Г, З или Ж..

Примеры: *дуб зелёный* [d, u, b, z', e, l', o, n, y, j'], *бровь Зои* [b, r, o, v', z, o, i]. Но: *бровь Нины* [b, r, o, f', n', i, n, y].

– в отличие от внутрисловных правил твёрдая согласная не переходит в мягкую фонему перед гласными Я, Ё, Ю, Е, И.

Примеры: *сад ёлок* [s, a, t, j', o, l, a, k], *город Ереван* [g, o, r, a, t, j', e, r', e, v, a, n]. Но: *хотели* [x, a, t', e, l', i].

– в отличие от внутрисловных правил отсутствует ассимиляция с последующей мягкой согласной, одинаковой по месту образования.

Примеры: *город Симбирск* [g, o, r, a, t, s', i, m', b', i, r, s, k]. Но: *вместе* [v, t', e, s', t', e].

Правило 2. Если первая буква второго слова

а) гласная:

– внутрисловные правила сохраняются, но не полностью.

– в отличие от внутрисловных правил, О и И после Ъ не переходят в сочетания фонем [j', o] и [j', i] соответственно. Гласная О переходит в [o] в ударной позиции и в [a] в безударной, гласная И переходит в [i].

Примеры: *щадить оленя* [sh', a, d', i, t', a, l', e, n', a], *любит игру* [l', u, b', i, t', i, g, r, u].

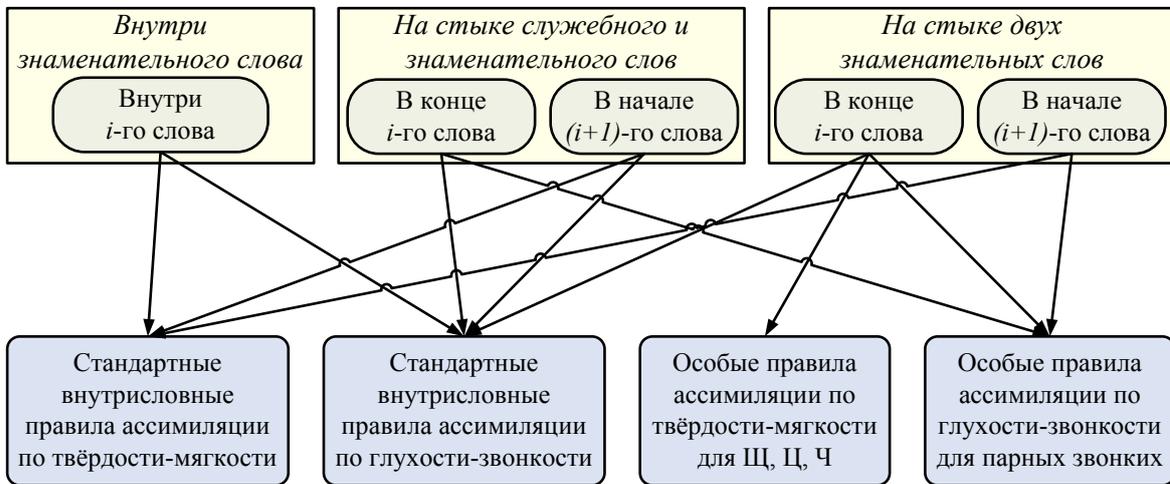
– гласная И после всех согласных, кроме Ё, переходит в фонему [y].

Примеры: *лист ивы* [l', i, s, t, y, v, y], *нос Ивана* [n, o, s, y, v, a, n, a]. Но: *мать Ивана* [m, a, t', i, v, a, n, a].

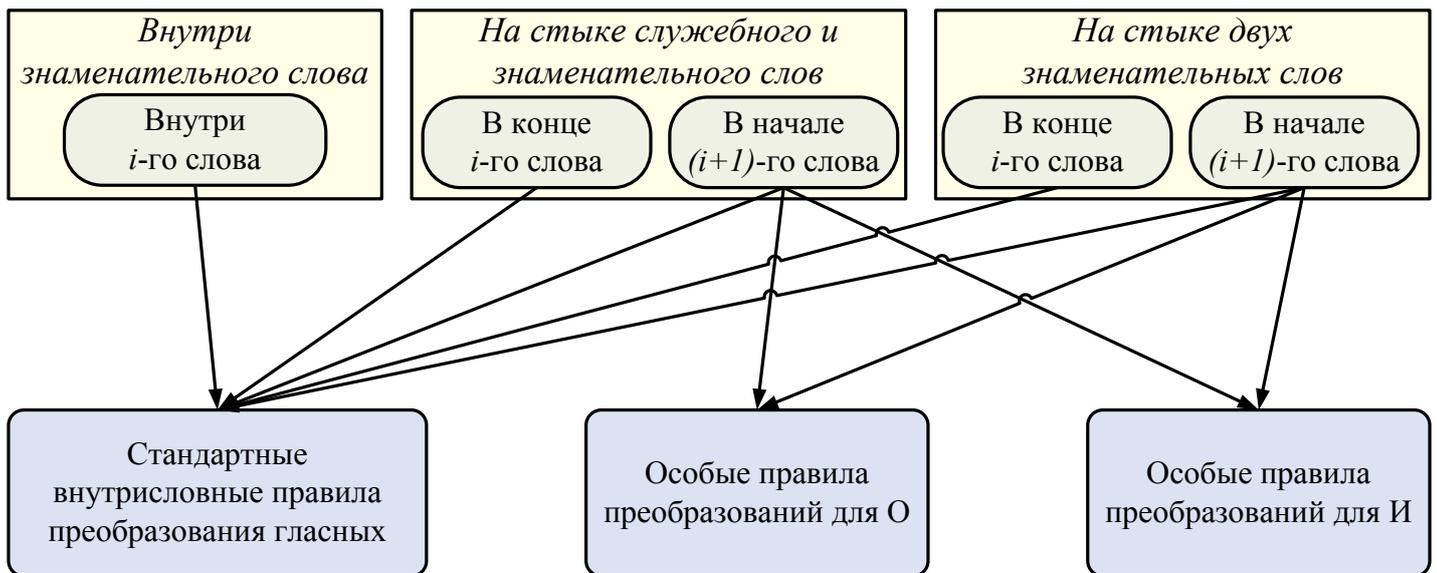
б) согласная:

– внутрисловные правила сохраняются полностью.

Описанные особенности преобразования буква-фонема внутри знаменательного слова, на стыке служебного и знаменательного слова и на стыке двух фонетических слов показаны на рис. 4.12 а для согласных и на рис. 4.12 б для гласных.



а)



б)

Рис.4.12. Особенности преобразования буква-фонема внутри и на стыках слов: а) для согласных, б) для гласных.

4.4.3. Блок преобразования фонема-аллофон

Преобразование фонема-аллофон осуществляется в два этапа. На первом этапе происходит преобразование фонем в позиционные аллофоны, на втором – преобразование позиционных аллофонов в позиционно-комбинаторные.

Преобразование фонем в позиционные аллофоны (которые показывают позицию фонемы по отношению к словесному ударению) особенно важно для гласных, поскольку они в большой степени подвержены количественной и качественной редукции. Наибольшую длительность и силу звука имеют сильноударные гласные, несколько меньшую – слабоударные. Следующие по длительности и силе звука – гласные первой степени редукции, и, наконец, наименьшую выразительность имеют гласные второй степени редукции.

Как показали проведённые исследования [151], первую степень редукции имеют гласные, находящиеся в фонетическом слове непосредственно перед ударным гласным (т.е. первые предударные) либо являющиеся первым или последним звуком фонетического слова. К гласным, имеющим вторую степень редукции, относятся не первые предударные и заударные гласные (в случае, если они не являются первым или последним звуком фонетического слова).

Эти факторы учитываются правилами генерации позиционных аллофонов гласных, представленными на рис. 4.13. Обозначив позиционный аллофон индексом i , следующим за именем фонемы, будем иметь следующие значения i для гласных: 0 – сильноударный, 1 – слабоударный, 2 – гласный первой степени редукции, 3 – гласный второй степени редукции.

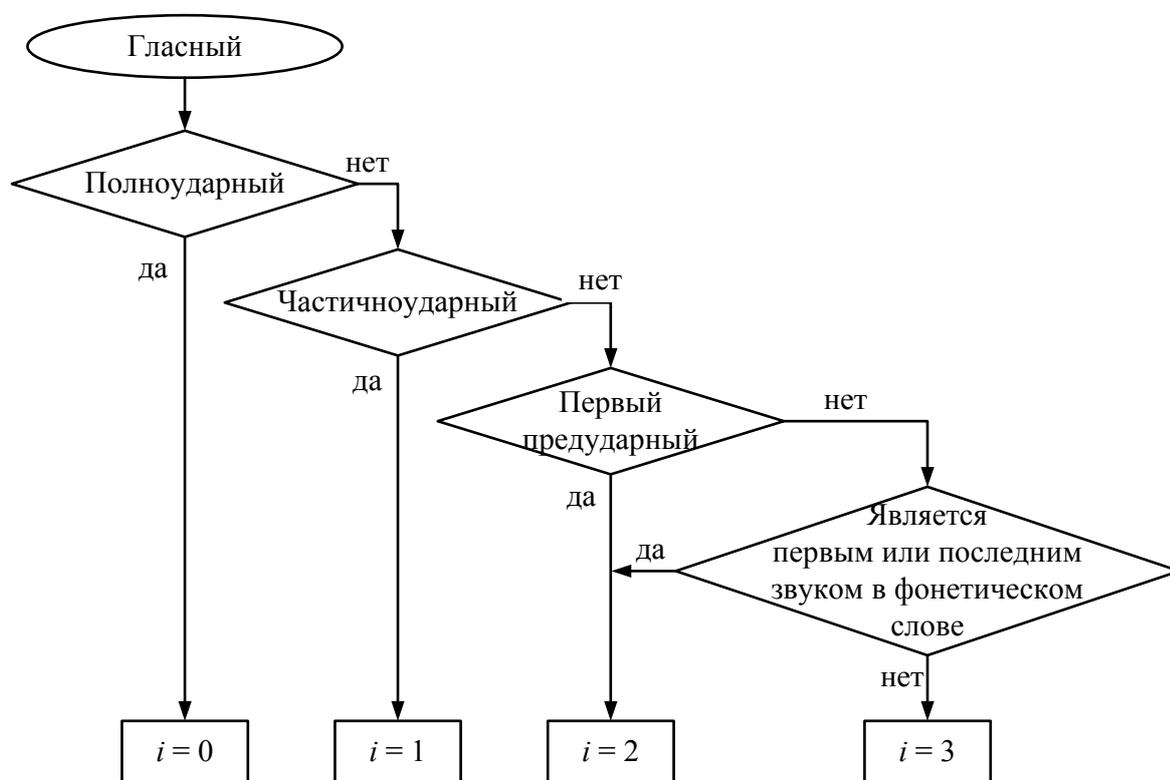


Рис.4.13. Правила формирования позиционного индекса гласных

Редукция согласных в естественной речи по сравнению с гласными мала, и здесь с учётом слухового восприятия нужно различать две ситуации: согласный находится в ударном слоге и в безударном слоге. Такое разделение, однако, увеличивает количество аллофонов согласных в два раза, что влечёт увеличение размера речевого корпуса, времени на его запись, а также времени на подготовку фонетико-акустической БД. Принимая во внимание эти факторы, можно отказаться от учёта позиционной аллофонии согласных. Такое упрощение в какой-то мере компенсируется на последующих этапах синтеза, когда из фонетико-акустической БД выбираются аллослоговые сегменты, содержащие требуемые для синтеза позиционные аллофоны согласных.

Для идентичности обозначения аллофонов для согласных также вводится позиционный индекс, но он в отличие от гласных показывает удвоение фонемы и означает следующее: 0 – обычная фонема, 1 – удвоенная фонема. Правила формирования позиционного индекса согласных приведены на рис. 4.14.

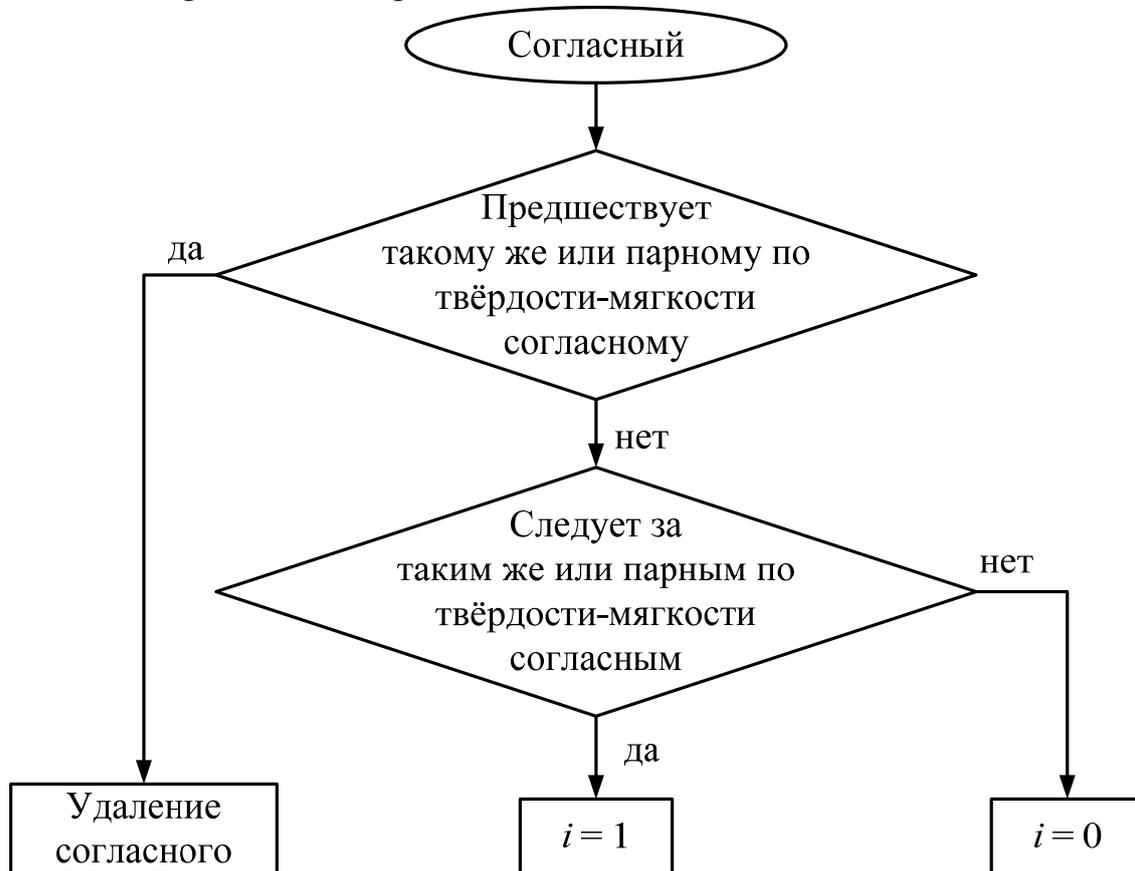


Рис.4.14. Правила формирования позиционного индекса согласных

Примечательно, что при использовании таких правил позиционный аллофон, формируемый по двум одинаковым согласным фонемам, будет совпадать с позиционным аллофоном, формируемым по двум согласным фонемам, парным по твёрдости-мягкости. Например, в фонемной

последовательности «*p, a, d, d', e, r', e, v, a, m*» (орфографический текст «под деревом») пара фонем *d, d'* (твёрдая и мягкая) преобразуется в позиционный аллофон *d'₁*; в фонемой последовательности «*h, o, d', d', e, r', e, v, a*» (орфографический текст «хоть дерево») пара фонем *d', d'* (обе мягкие) также преобразуется в позиционный аллофон *d'₁*. В таких ситуациях твёрдость-мягкость фонемы учитывается на последующем этапе (описанном ниже) – при преобразовании позиционных аллофонов в комбинаторные, причём будут различными комбинаторные индексы аллофона, находящегося перед парой одинаковых либо парных по твёрдости-мягкости согласных.

Следующий этап блока преобразования фонема-аллофон – преобразование к позиционно-комбинаторным аллофонам. Комбинаторный фактор учитывает ближайшее окружение фонемы, т. е. левый контекст – фонему, расположенную непосредственно перед данной, и правый контекст – фонему, расположенную непосредственно за данной фонемой. Комбинаторные характеристики обозначаются индексами *j* и *k*, при этом индекс *j* указывает группу левого контекста, индекс *k* – группу правого контекста.

Как показали проведённые исследования (см. раздел 3.1), правила генерации комбинаторных аллофонов различны для фонем, разных *по способу образования*. Для генерации комбинаторных аллофонов всё множество фонем разбивается на следующие классы:

- негубные гласные {*a, i, e, y*},
- губные гласные {*u, o*},
- большинство глухих согласных {*p, p', t, t', k', c, ch', f, f', s, s', sh, sh', h'*},
- твёрдые аспиративные согласные {*k, g, h*},
- звонкие взрывные, щелевые и сонорные согласные {*b, b', d, d', g', z, z', zh, l, l', m, m', n, n', r, r'*},
- плавные сонорные согласные {*v, v', j'*}.

Левый и правый фонемные контексты группируются *по месту образования*, но они различны для разных классов фонем. Наибольшее количество групп левых контекстов (шесть) имеют гласные, а наименьшее (один левый контекст) – все согласные, кроме плавных сонорных. Наибольшее количество правых контекстов (четыре) приходится на гласные, а также на звонкие и плавные сонорные, наименьшее (два контекста) – на глухие согласные.

Группы левых и правых контекстов для разных по способу образования фонем, а также соответствующие позиционные индексы аллофонов приведены в Приложении 2.

Группировка фонем в левые и правые контексты имеет следующие особенности:

– согласная фонема *l* включается в группу губных согласных при определении левого и правого контекстов негубных гласных; при определении левого и правого контекстов губных гласных фонема *l* включается в группу переднеязычных согласных;

– все группы левых контекстов (и соответствующие позиционные индексы) для гласных совпадают внутри слова и на границах слов, за исключением группы мягких согласных. В случае, если гласная находится внутри слова после мягкой согласной, индекс *j* принимает значение 4, если же гласная находится в начале слова, а предыдущее слово заканчивается на мягкий согласный, индекс *j* принимает значение 5.

Реальное число используемых аллофонов меньше, чем полученное теоретически, поскольку некоторые комбинации фонем не встречаются в русской речи. Полный перечень мини-набора аллофонов, используемых в системе синтеза речи, приведен в Приложении 2. Общее количество аллофонов полноударных гласных – 122, частично-ударных гласных – 117, гласных первой степени редукции – 95, гласных второй степени редукции – 36.

Общее количество аллофонов согласных – 181, из которых 100 – одиночных, и 81 – удвоенных.

Общее количество аллофонов мини-набора равно 551, из которых 370 – гласных и 181 – согласных.

4.5. Акустический процессор

Общая структура акустического процессора представлена на рис. 4.15.

Задачей первого блока акустического процессора является преобразование просодически размеченной последовательности аллофонов синтагмы в последовательность их звуковых волн со значениями ЧОТ – F_0 , амплитуды – A и длительности – T , задаваемыми БД просодических портретов. Во втором блоке осуществляется синтез речевого сигнала путём выбора из БД звуковых волн мультифонов (единичных аллофонов, диаллофонов, аллослогов), соответствующих входному аллофонному тексту, и их конкатенации (соединения).

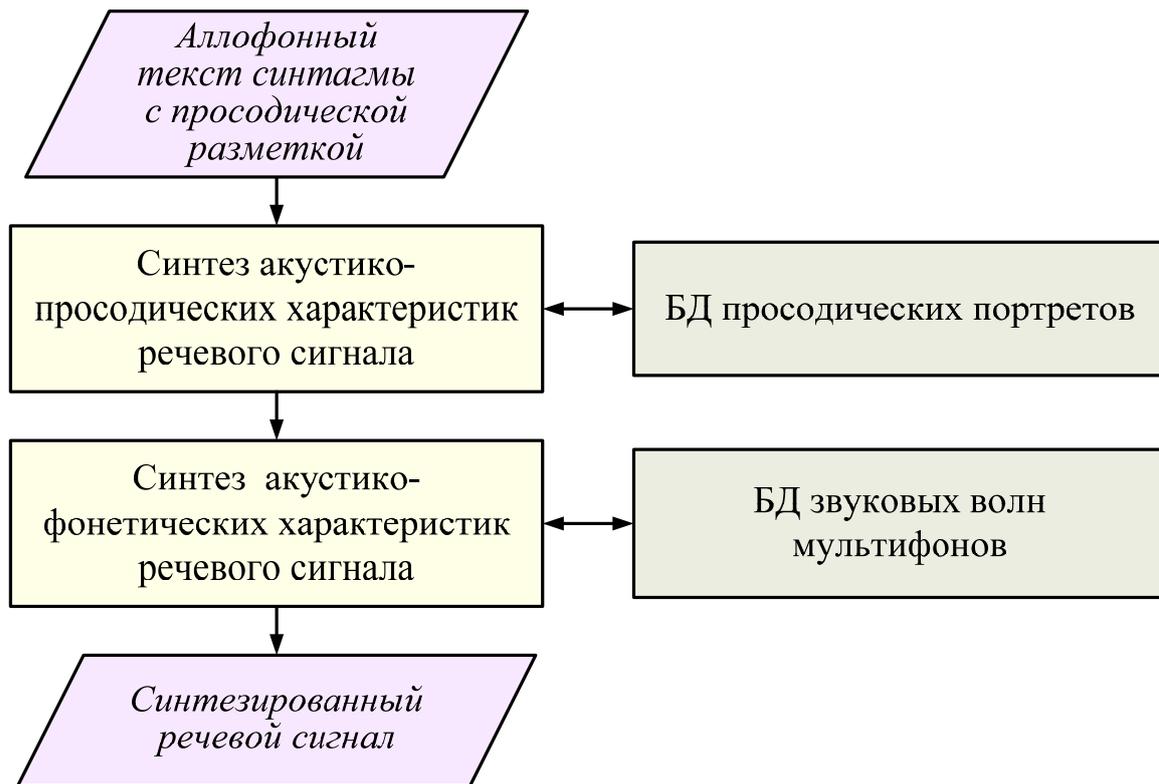


Рис. 4.15. Структура акустического процессора

4.5.1. Блок синтеза акустико-просодических характеристик речевого сигнала

Функциональная схема блока синтеза просодических характеристик речи представлена на рис. 4.16. Синтез просодических характеристик осуществляется последовательно для каждой синтагмы. На первом этапе осуществляется разметка каждой синтагмы на АЕ, каждую АЕ на элементы акцентной единицы (ЭАЕ): предъядро, ядро, заядро. Ядром АЕ, согласно используемым правилам, является полноударный гласный; все аллофоны, предшествующие полноударному гласному, являются предъядерным участком, все следующие за полноударным гласным аллофоны – заядерным участком.

Затем для каждой синтагмы необходимо выбрать соответствующие её интонационному типу просодические контуры: интонационный, ритмический, динамический. Для этого используется БД просодических портретов акцентных единиц (ПАЕ), содержащая просодические «портреты» для каждого используемого интонационного типа. БД ПАЕ может содержать несколько наборов просодических портретов, каждый из которых характеризует определённый стиль «озвучиваемого» текста (научный, официально-деловой, публицистический стиль, стиль художественной литературы, разговорный), индивидуальные просодические характеристики того или иного диктора, выражение

различных эмоций в речи и т.д. В просодических портретах сохраняется также длительность межсинтагменной паузы.

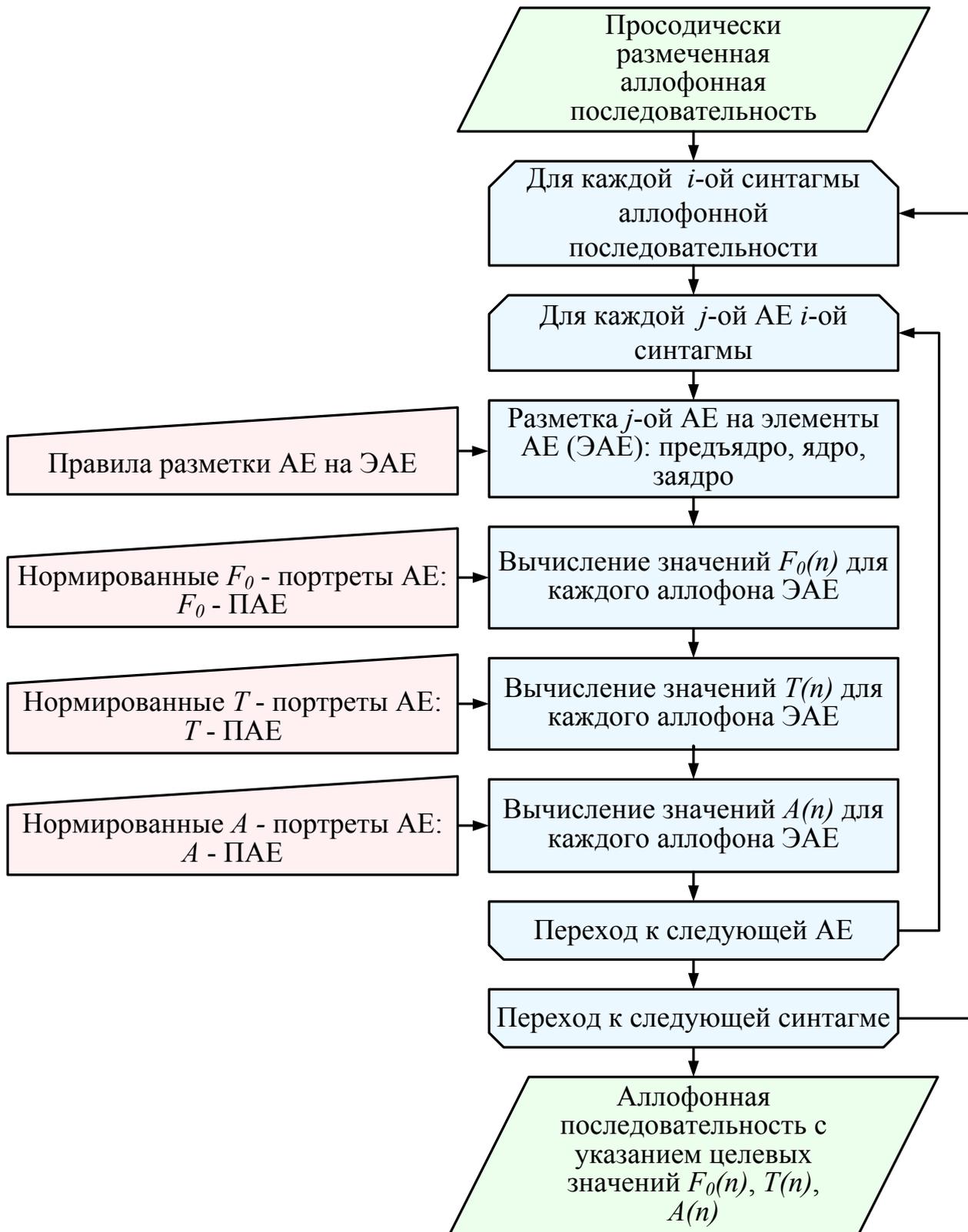


Рис. 4.16. Функциональная схема блока синтеза акустико-просодических характеристик

Далее с использованием нормированных портретов F_0 -ПАЕ, А-ПАЕ, Т-ПАЕ для синтагмы соответствующего интонационного типа осуществляется вычисление значений $F_0(n)$, $A(n)$, $T(n)$ для каждого n -го аллофона элементов предъядра, ядра, заядра j -й АЕ.

Алгоритмы выбора и создания БД просодических портретов рассматриваются далее в 5-й главе в связи с решением общей задачи компьютерного клонирования персональных характеристик речи.

Процесс вычисления абсолютных значений просодических параметров для каждого аллофона синтагмы показан на рис. 4.17 на примере вычисления F_0 для фразы «*Мариана приехала?*», фонемная запись которой «*m a r' i a n a p r' i j' e h a l a*». Эта фраза является синтагмой вопросительного типа, состоящей из двух АЕ. Соответствующий мелодический портрет, выбранный из БД просодических портретов, показан на рис. 4.17 (а), где ось абсцисс T_N соответствует нормированному времени, ось ординат F_N – нормированному значению F_0 .

Следующий этап – разметка каждой АЕ синтагмы на предъядро, ядро, заядро и разбиение мелодического портрета в соответствии с количеством фонем на предъядре и заядре – показан на рис. 4.17 (б). Рассматриваемая синтагма состоит из 2 АЕ: «*m a r' i a n a*» и «*p r' i j' e h a l a*». Предъядро первой АЕ содержит четыре фонемы: «*m, a, r', i*», ядром АЕ, как указывалось выше, является ударный гласный, в данном случае это «*a*», заядро первой АЕ содержит фонемы «*n, a*». Предъядро, ядро и заядро второй АЕ имеют, соответственно, следующий состав: «*p, r', i, j'*», «*e*», «*h, a, l, a*». На рис. 4.17 (б) ось абсцисс соответствует так называемому «фонемному» времени T_{ph} , когда все фонемы синтагмы полагаются одинаковой длительности.

На этом этапе разметки необходимо учитывать такие ситуации, как отсутствие предъядра или заядра в АЕ, а также отсутствие вокализованных фонем на предъядре и заядре. Действительно, если в таких случаях использовать «усечённый» портрет, т.е. без предъядра или без заядра, мелодический контур не будет полностью реализован, и произойдёт искажение интонации. Чтобы этого избежать, необходимо размечать начальную или конечную часть ударной гласной, соответственно, как предъядро или заядро.

Следующим этапом (см. рис. 4.17 (в)) является разметка интонационного портрета в соответствии с собственной длительностью фонем. Как видно из рисунка, сравнительно большую собственную длительность имеют ядра АЕ (ударные гласные), а также конечная гласная синтагмы; наименьшую собственную длительность в данном случае имеет согласный «*j'*». Ось абсцисс на рис. 4.17 (в) соответствует реальному времени T .

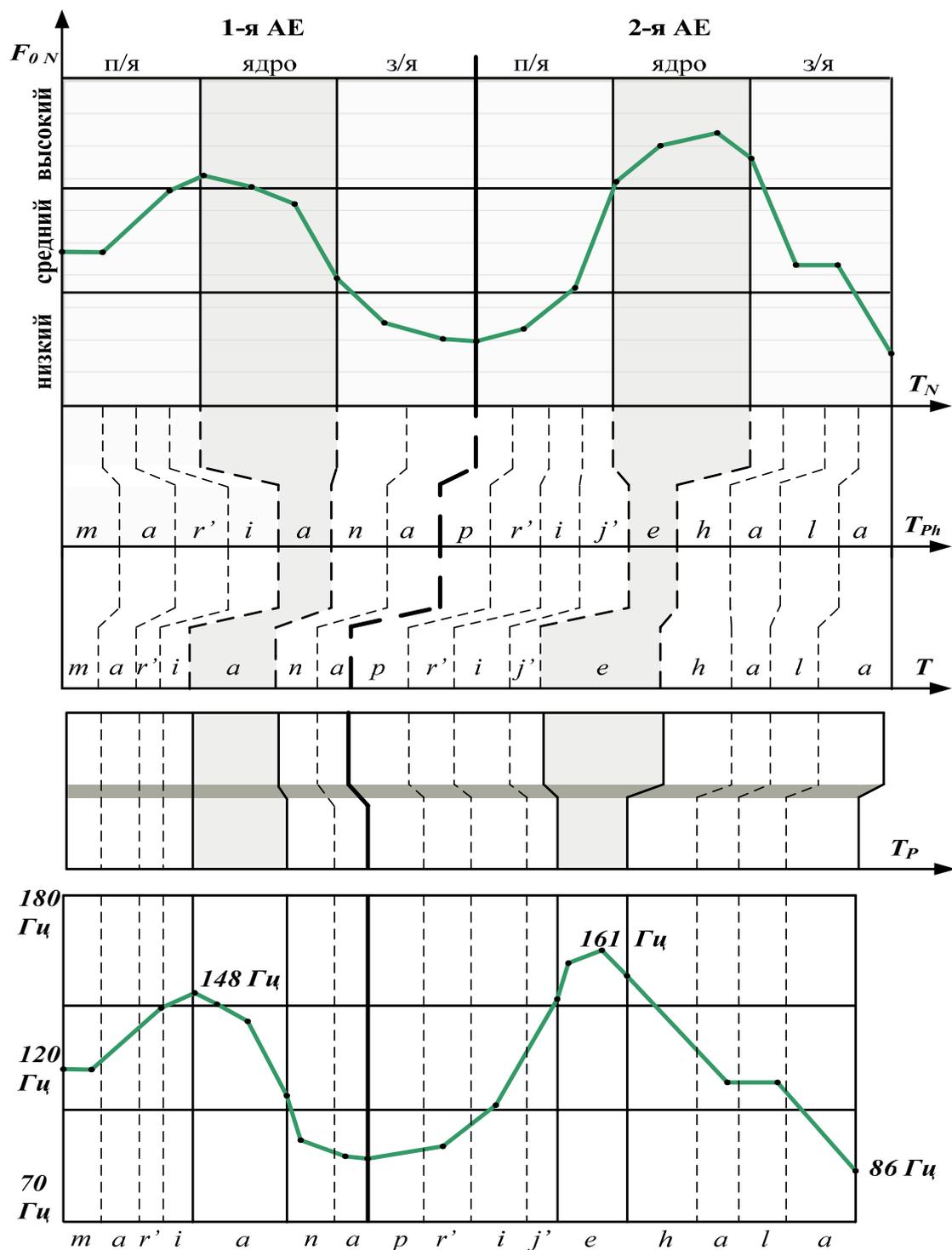


Рис. 4.17. Процесс вычисления абсолютных значений длительности звуков T и частоты основного тона F_0

На следующем этапе, показанном на рис. 4.17 (г), осуществляется корректировка собственных длительностей фонем в соответствии с ритмическим портретом синтагмы данного типа, выбранным из БД просодических портретов. В верхней части рис. 4.17 (г) показаны собственные длительности фонем синтагмы, в нижней части – доли, в

соответствии с которыми происходит корректировка. Как видно из рисунка, длительность ядра первой АЕ двухакцентной синтагмы вопросительного типа увеличивается, а ядра второй АЕ – уменьшается. По оси абсцисс на рис. 4.17 (г) отложено время T_p , нормированное в соответствии с ритмическим портретом синтагмы.

Последним этапом (см. рис. 4.17 (д)) является вычисление абсолютных значений частоты основного тона F_0 для каждой фонемы синтагмы. Поскольку мелодический портрет задаёт нормированные значения, для вычисления абсолютных значений необходимо задать диапазон изменения F_0 . При этом на диапазон изменения просодических параметров будет влиять тип синтезируемого голоса: женский, мужской или детский; стиль текста, выражение эмоций и т.д. Абсолютное значение F_0 вычисляется по формуле:

$$F_0 = F_{0 \min} + F_{0 \text{norm}}(F_{0 \max} - F_{0 \min}) \quad (4.1)$$

В рассматриваемом случае $F_{0 \min} = 70$ Гц, $F_{0 \max} = 180$ Гц. Полученные абсолютные значения F_0 показаны на рис. 4.17 (д); максимальное значение, достигаемое в ядре второй АЕ, получилось равным 161 Гц, минимальное, достигаемое в конце конечной гласной синтагмы, – равным 86 Гц.

Аналогичные преобразования производятся в просодическом процессоре для вычисления абсолютных значений амплитуды и длительности каждого звука синтагмы.

4.5.2. Блок синтеза акустико-фонетических характеристик речевого сигнала

Структура блока синтеза акустико-фонетических характеристик речевого сигнала представлена на рис. 4.18.

Задачей этого блока является синтез речевого сигнала в соответствии с выходными данными блока формирования акустико-просодических характеристик речи (см рис. 4.16) путём выбора элементов из БД звуковых волн мультифонов, их конкатенации, синтеза модифицированных звуковых волн аллофонов и формирования длительности звуков и темпа речи.

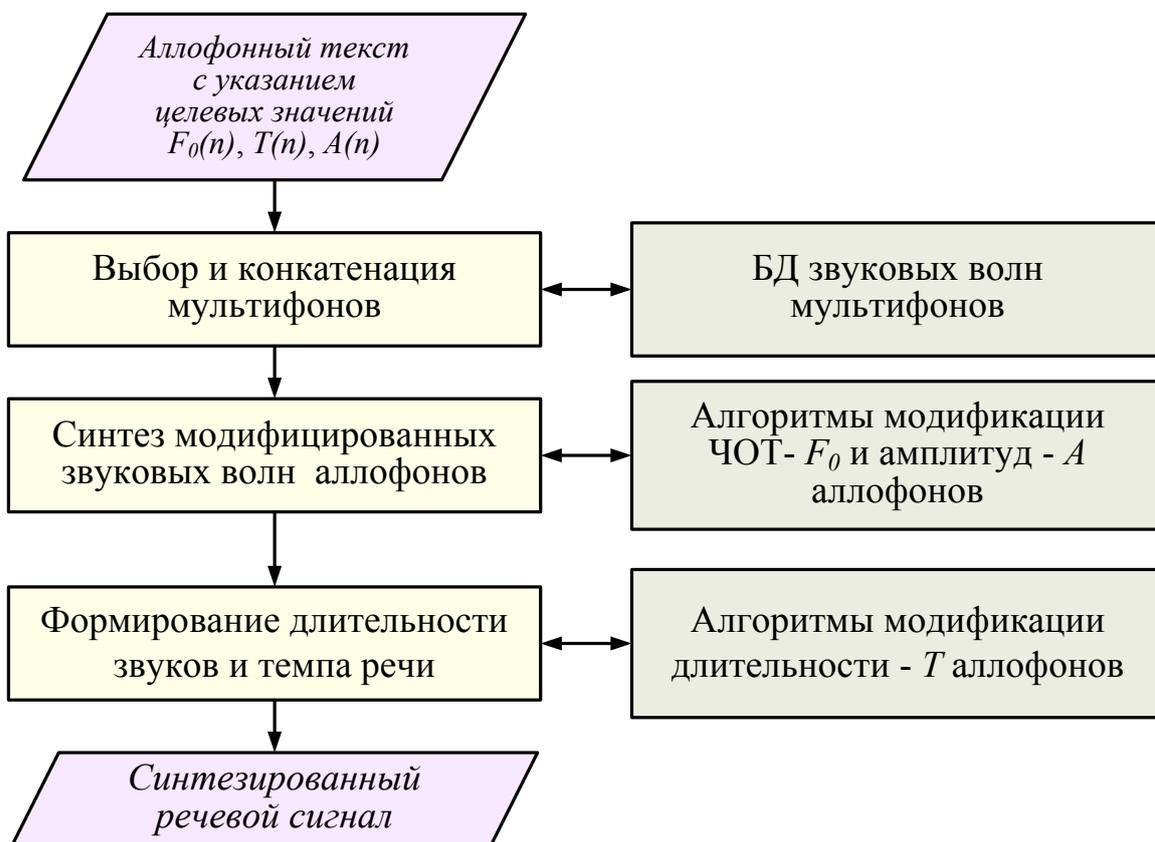


Рис. 4.18. Структура блока синтеза акустико-фонетических характеристик

Выбор и конкатенация аллофонов и мультифонов.

Использование базового набора аллофонов обеспечивает синтез вполне разборчивой речи по произвольному тексту, однако качество речи может оказаться недостаточно высоким. Это объясняется тем, что реальное разнообразие оттенков фонем при их взаимодействии в потоке речи несоизмеримо большее, чем это обеспечивается используемым минимальным набором аллофонов. Кроме того, взаимовлияние соседних аллофонов в некоторых случаях может быть настолько сильным, что провести чёткую границу между ними зачастую просто невозможно. К таким случаям относятся, в частности, сочетания аллофонов гласный-гласный, гласный-сонорный, сонорный-сонорный. Существенное повышение качества и естественности речи может быть достигнуто, если в качестве элементов компиляции использовать не только аллофоны, но также и более крупные фонетические сегменты – диаллофоны (последовательность двух подряд идущих аллофонов) и аллослоги (слоговые сегменты с учётом позиционной и комбинаторной аллофонии). Следует, однако, иметь в виду, что платой за достижение более высокого качества может стать резкое возрастание объёма фонетико-акустической БД. Действительно, грубый подсчёт потенциального количества диаллофонов оценивается числом: $N_{da} = N_a^2$

$= 561^2 = 314\,721$. Далеко не все комбинации аллофонов возможны, но как показывает опыт, их количество в слитной речи может достигать десятков тысяч.

Для преодоления указанных противоречий при синтезе речи целесообразно использовать фонетико-акустическую БД, содержащую не только минимальный набор аллофонов, но и дополнительный набор из наиболее часто встречающихся диаллофонов и аллослогов, извлекаемых из представительного корпуса естественной речи.

Алгоритм выбора речевых сегментов должен быть сформирован таким образом, чтобы для произвольной аллофонной последовательности, подаваемой на вход блока, выбиралось наибольшее возможное число мультифонных сегментов.

Для решения этой задачи вначале осуществляется поиск в БД слоговых комплексов (см. раздел 3.2) в соответствии со следующим приоритетом: внутрисинтагменные и затем внутрисловные слоговые комплексы 3-го типа, которые имеют максимальную длительность, далее, соответственно, слоговые комплексы 2-го типа, имеющие среднюю длительность, и, наконец, слоговые комплексы 1-го типа, имеющие минимальную среди всех типов длительность.

На каждом шаге поиска в случае, когда в БД не найден внутрисинтагменный слоговый комплекс 3-го типа, осуществляется последовательный поиск составляющих его слоговых комплексов низших уровней в соответствии с указанным выше приоритетом.

В случае, когда в БД не найден ни один из сформированных типов аллослогов, осуществляется поиск составляющих его диаллофонов. При этом всё множество диаллофонов разбивается в порядке уменьшения взаимовлияния соседних аллофонов и, как следствие, важности их вклада в качество синтезируемой речи на 4 группы: ГГ, СГ, СС, ГС (где Г обозначает гласный, С – согласный). Указанный порядок задаёт приоритет их выбора. В случае, когда необходимые диаллофоны отсутствуют в БД элементов компиляции, происходит выбор соответствующих аллофонов.

Последовательность выбора диаллофонов при синтезе речи для синтагмы *«Издёрганная скрипка плакала в дуэте с виолончелью»* показана в таблице 4.3, где границы диаллофонов помечены значками «<», «>». Диаллофоны, добавляемые на каждом шаге выбора, выделены жирным шрифтом.

Выбор диаллофонов при синтезе речи

Группа диаллофонов	Размеченная на диаллофоны синтагма
ГГ	$I_{203}, Z'_{001}, D'_{002}, O_{042}, R_{001}, G_{002}, A_{332}, N_{102}, A_{323}, J'_{012}, A_{342}, S_{001}, K_{002}, R'_{002}, I_{041}, P_{001}, K_{002}, A_{331}, P_{001}, L_{002}, A_{012}, K_{002}, A_{331}, L_{002}, A_{311}, V_{013}, D_{002}, \langle U_{222}E_{013} \rangle, T'_{001}, E_{342}, S_{001}, V'_{002}, \langle I_{342}A_{341} \rangle, L_{002}, A_{212}, N_{003}, CH'_{001}, E_{043}, L'_{001}, J'_{002}, U_{340}$
ГГ+СГ	$I_{203}, Z'_{001}, \langle D'_{002}O_{042} \rangle, R_{001}, \langle G_{002}A_{332} \rangle, \langle N_{102}A_{323} \rangle, \langle J'_{012}A_{342} \rangle, S_{001}, K_{002}, \langle R'_{002}I_{041} \rangle, P_{001}, \langle K_{002}A_{331} \rangle, P_{001}, \langle L_{002}A_{012} \rangle, \langle K_{002}A_{331} \rangle, \langle L_{002}A_{311} \rangle, V_{013}, D_{002}, \langle U_{222}E_{013} \rangle, \langle T'_{001}E_{342} \rangle, S_{001}, V'_{002}, \langle I_{342}A_{341} \rangle, \langle L_{002}A_{212} \rangle, N_{003}, \langle CH'_{001}E_{043} \rangle, L'_{001}, \langle J'_{002}U_{340} \rangle$
ГГ+СГ+СС	$I_{203}, Z'_{001}, \langle D'_{002}O_{042} \rangle, R_{001}, \langle G_{002}A_{332} \rangle, \langle N_{102}A_{323} \rangle, \langle J'_{012}A_{342} \rangle, \langle S_{001}K_{002} \rangle, \langle R'_{002}I_{041} \rangle, P_{001}, \langle K_{002}A_{331} \rangle, P_{001}, \langle L_{002}A_{012} \rangle, \langle K_{002}A_{331} \rangle, \langle L_{002}A_{311} \rangle, \langle V_{013}D_{002} \rangle, \langle U_{222}E_{013} \rangle, \langle T'_{001}E_{342} \rangle, \langle S_{001}V'_{002} \rangle, \langle I_{342}A_{341} \rangle, \langle L_{002}A_{212} \rangle, N_{003}, \langle CH'_{001}E_{043} \rangle, L'_{001}, \langle J'_{002}U_{340} \rangle$
ГГ+СГ+СС+Г С	$\langle I_{203}Z'_{001} \rangle, \langle D'_{002}O_{042} \rangle, R_{001}, \langle G_{002}A_{332} \rangle, \langle N_{102}A_{323} \rangle, \langle J'_{012}A_{342} \rangle, \langle S_{001}K_{002} \rangle, \langle R'_{002}I_{041} \rangle, P_{001}, \langle K_{002}A_{331} \rangle, P_{001}, \langle L_{002}A_{012} \rangle, \langle K_{002}A_{331} \rangle, \langle L_{002}A_{311} \rangle, \langle V_{013}D_{002} \rangle, \langle U_{222}E_{013} \rangle, \langle T'_{001}E_{342} \rangle, \langle S_{001}V'_{002} \rangle, \langle I_{342}A_{341} \rangle, \langle L_{002}A_{212} \rangle, N_{003}, \langle CH'_{001}E_{043} \rangle, L'_{001}, \langle J'_{002}U_{340} \rangle$

В приведённом примере в результате разметки на диаллофоны с учётом указанного приоритета формируется 19 диаллофонов, из которых 2 – типа ГГ, 13 – типа СГ, 3 – типа СС и 1 – типа ГС. Четыре аллофона остаются не сгруппированными (не присоединёнными к соседним): $R_{001}, P_{001}, N_{003}, L'_{001}$.

В результате указанной стратегии приоритетов элементы БД аллофонов, составляющих мини-набор, будут использоваться только в тех крайних случаях, когда необходимые для синтеза элементы верхних уровней – мультифоны – отсутствуют в имеющейся БД аллофонов и мультифонов.

Синтез модифицированных звуковых волн аллофонов.

Формирование целевых значений ЧОТ, которое влечёт модификацию периодов естественного речевого сигнала, должно осуществляться с максимально возможным сохранением индивидуальности и качества звучания речи. Для формирования мелодического контура $F_0(t)$ используется SL-алгоритм [69], который позволяет осуществлять “щадящую” модификацию ЧОТ путём “плавной сшивки” (“Soft Lacing”)

соседних периодов естественного сигнала на интервалах открытой голосовой щели, сохраняя речевой сигнал неизменённым на остальных участках.

Согласно теории речеобразования, наибольшую информацию о звуке несёт участок закрытой голосовой щели, на котором реализуются наиболее интенсивные формантные колебания. Поэтому модифицируемый речевой сигнал (аллофон) должен быть размечен на периоды основного тона таким образом, чтобы граница периода указывала на момент времени, непосредственно предшествующий началу смыкания голосовых связок. При такой разметке для модификации периодов используется вторая половина периода, которая соответствует участку открытой голосовой щели. Начальная же половина остаётся неизменной.

Пример такого аллофона, извлечённого из БД на этапе выбора и компиляции отрезков естественной речевой волны, представлен на рис. 4.19. Границы периодов основного тона, отмеченные на рисунке вертикальными штриховыми линиями, установлены в точках перехода сигнала через ноль, которые соответствуют моменту смыкания голосовых связок. Длительность одного периода основного тона T_0 представленного аллофона равна 10 мс, а частота основного тона F_0 – 100 Гц.

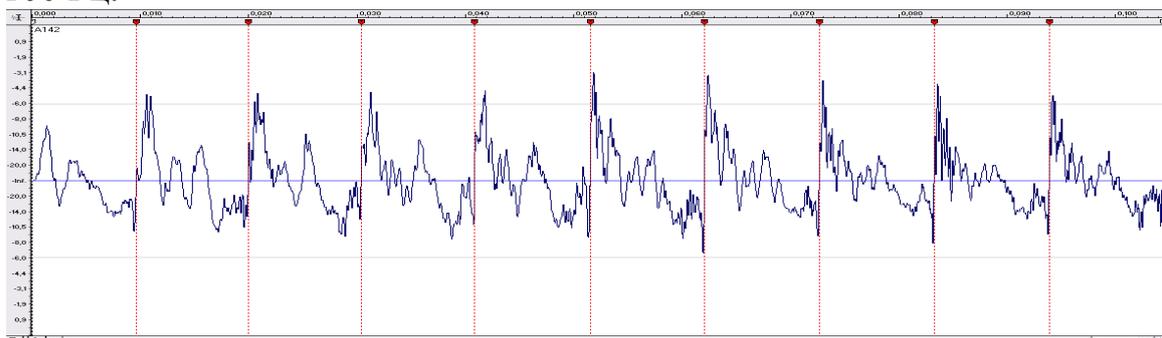


Рис. 4.19. Фрагмент вокализованного аллофона A_{142} с маркерами периодов основного тона

Для изменения значений частоты основного тона необходимо увеличить или уменьшить длительности каждого периода аллофона.

Если осуществлять процедуру уменьшения длительности периода простым отсечением «лишнего» участка, то возникнут искажения сигнала и качество звучания речевого сигнала существенно ухудшится. Пример такого изменения длительности периода продемонстрирован на рис. 4.20, 4.21. На рис. 4.20 показаны два подряд идущих периода основного тона сигнала, собственная длительность периода основного тона сигнала T_0 , целевая длительность периода T_0' , а также участок сигнала, который необходимо удалить. Результат удаления показан на рис. 4.21.

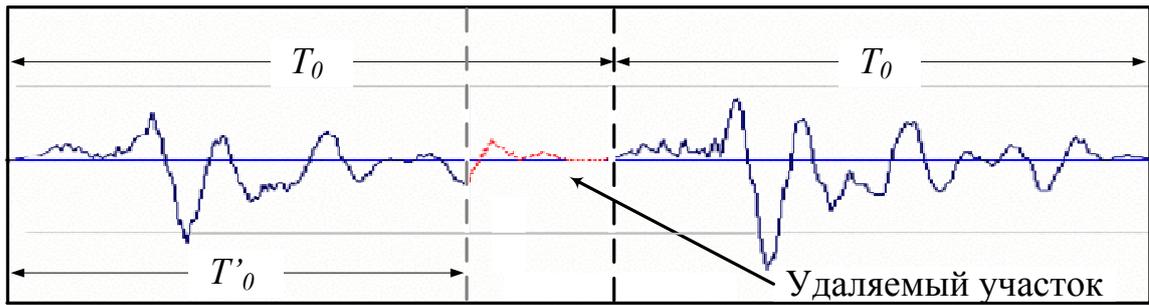


Рис. 4.20. Процесс уменьшения периода основного тона

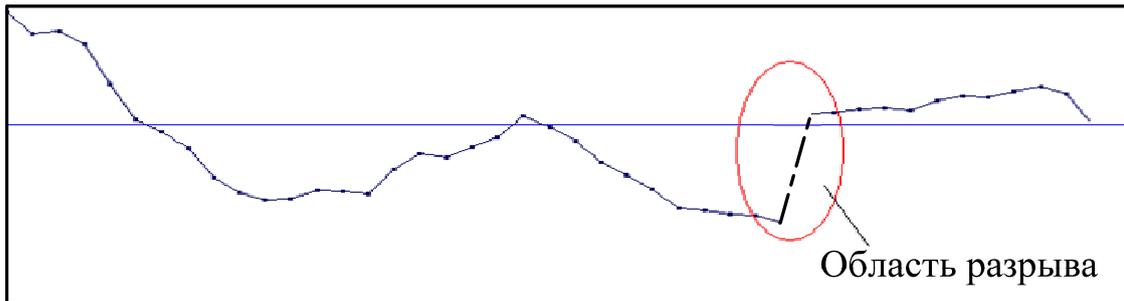


Рис. 4.21. Разрыв сигнала

Такие разрывы сигнала воспринимаются в синтезированной речи как характерные щелчки, наличие которых заметно ухудшает качество речи.

Для исключения разрывов применяется плавная сшивка участков двух граничащих периодов. При этом удаляемый участок «перемещается» влево и «накладывается» на предшествующий участок этого же периода, как показано на рис. 4.22, 4.23.

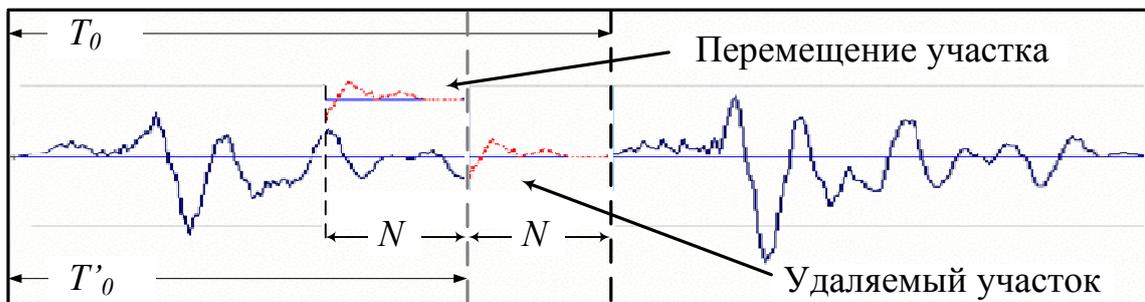


Рис. 4.22. Перемещение удаляемого участка

«Наложение» двух участков (рис. 4.23) происходит путём умножения каждого из них на характеризующие линии L_1 и L_2 , причём значение L_1 в начальной точке равно 1, в конечной точке – 0, а значение L_2 в начальной точке равно 0, в конечной точке – 1.

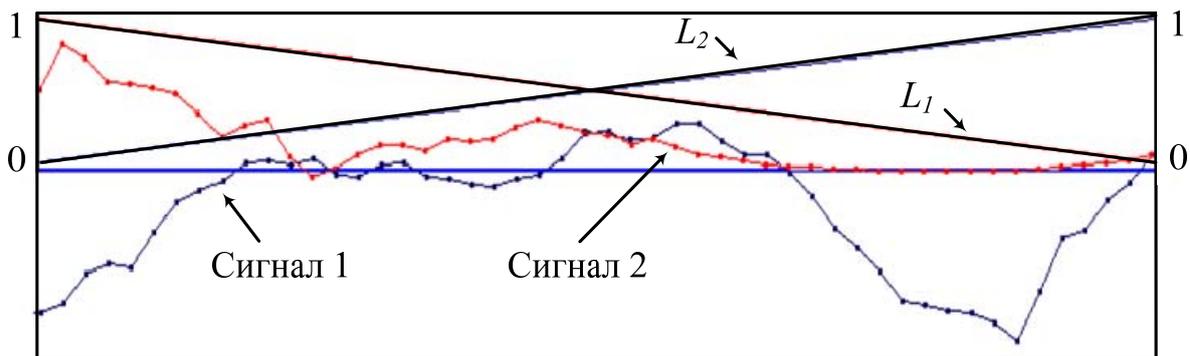


Рис. 4.23. «Наложение» участков двух сигналов

Математически такая модификация сигнала выражается формулой:

$$\tilde{s}(n) = s(n) \times L_1(n) + s(n + T_0 - T'_0) \times L_2(n) \quad (4.2)$$

где $\tilde{s}(n)$ – результирующий речевой сигнал, $s(n)$ – исходный сигнал, T_0 – период основного тона исходного сигнала, T'_0 – результирующий период основного тона, $L_1(n)$, $L_2(n)$ – линейные функции, задаваемые, соответственно, формулами:

$$L_1(n) = \begin{cases} 1, & n \leq (T'_0 - N) \\ 1 - \frac{n}{T'_0 - N}, & (T'_0 - N) < n \leq N \end{cases} \quad (4.3)$$

$$L_2(n) = \begin{cases} 0, & n \leq (T'_0 - N) \\ \frac{n}{T'_0 - N}, & (T'_0 - N) < n < N \end{cases} \quad (4.4)$$

где N – коэффициент сшивки, зависящий от результирующего значения T .

В практических приложениях N принимается равным $0,5T$.

Процесс уменьшения значения F_0 (и соответственно увеличение длительности периода основного тона) показан на рис. 4.24, 4.25.

Для добавления участка длиной N , где $N = T'_0 - T_0$, в исходный сигнал добавляется “пустой” участок длины N , а также формируется дополнительный сигнал путём вставки в начало периода “пустого” участка длины N (рис. 4.24).

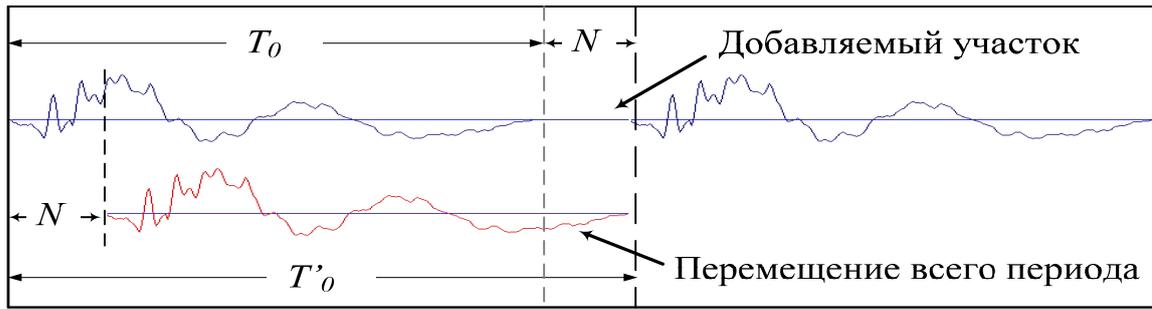


Рис. 4.24. Перемещение периода для добавления участка сигнала

Затем, так же, как и в случае удаления участка, происходит «наложение» двух сигналов путём умножения каждого из них на характеризующие линии L_1 и L_2 , но в данном случае длина каждого из сигналов равна T_0 . (рис. 4.25).

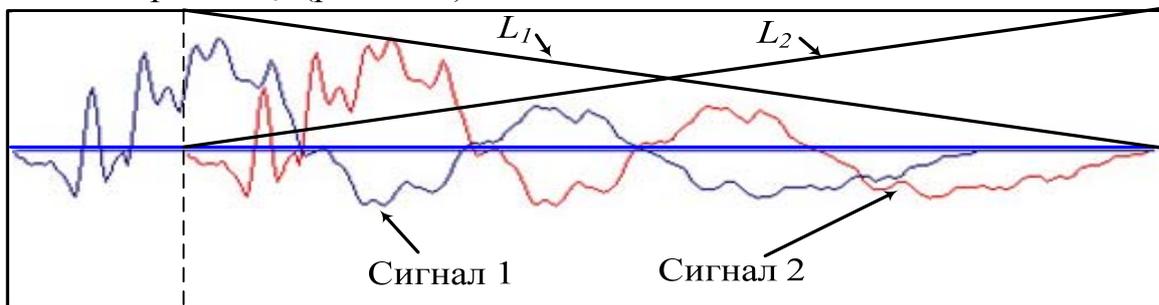


Рис. 4.25. Умножение сигналов на характеризующие линии

Математически такая модификация сигнала выражается формулой:

$$\tilde{s}(n) = s(n) \times L_1(n) + s'(n) \times L_2(n) \quad (4.5)$$

где $\tilde{s}(n)$ – результирующий речевой сигнал, $s(n)$ – исходный сигнал, $s'(n)$ – дополнительный сигнал, $L_1(n)$, $L_2(n)$ – линейные функции, задаваемые, соответственно, формулами:

$$L_1(n) = \begin{cases} 1, & n \leq (T'_0 - T_0) \\ 1 - \frac{n}{T'_0}, & (T'_0 - T_0) < n \leq T'_0 \end{cases} \quad (4.6)$$

$$L_2(n) = \begin{cases} 0, & n \leq (T'_0 - T_0) \\ \frac{n}{T'_0}, & (T'_0 - T_0) < n \leq T'_0 \end{cases} \quad (4.7)$$

Описанный SL-способ сшивки периодов при модификации F_0 показывает наилучшие результаты, если требуемая степень изменения периода основного тона находится в диапазоне 0,5 – 2. С увеличением диапазона изменения ЧОТ естественность полученного синтезированного сигнала постепенно снижается.

Формирование длительности звуков и темпа речи. Установка значений длительностей звуковых волн аллофонов - $T(t)$ осуществляется в соответствии с заданными целевыми значениями длительности звуковых элементов АЕ и корректируется затем с учетом её качественного и количественного состава. Регулировка темпа речи осуществляется путём корректировки длительности звуковых элементов АЕ и межсинтагменных пауз с учётом коэффициента "податливости" каждого конкретного звука темповым изменениям.

Для ориентировочной оценки пределов изменения средней длительности звуков и пауз в зависимости от темпа речи была проведена серия следующих экспериментов. Текст, содержащий примерно 1000 знаков, был прочитан несколькими дикторами в нормальном (среднем) темпе, замедленном (в стиле «диктант») и в максимально быстром темпе. Результаты измерения относительных средних длительностей различных звуков приведены в таблице 4.4.

Таблица 4.4

Относительная длительность звуков при изменении темпа речи

№ п/п	Тип звуковых единиц	Медленн. темп	Нормальн. темп	Быстрый темп
1.	Паузы	250%	100%	20%
2.	Ударные гласные	200%	100%	50%
3.	Предударные гласные	200%	100%	70%
4.	Заударные гласные	200%	100%	80%
5.	Сонанты	130%	100%	80%
6.	Звонкие взрывные и щелевые	130%	100%	80%
7.	Глухие взрывные	120%	100%	80%
8.	Глухие щелевые	110%	100%	85%

Процедура установки длительности звуковых волн в последовательности аллофонов основана на предварительном вычислении функции приращения времени – dT_a , которая определяет, какое количество отсчётов сигнала необходимо добавить или удалить в каждом конкретном аллофоне исходя из множества просодических факторов синтезируемой речи. Вычисление осуществляется в соответствии со следующей формулой

$$\partial T_a = (K_p(TMP * K_{a\ min} + (1 - TMP) * K_{a\ max}) - 1) * T_a \quad (4.8)$$

где K_p – просодический коэффициент, задаваемый ритмическим портретом; T_a – количество отсчётов сигнала в аллофоне; TMP – желаемый темп речи, задаваемый на интервале 0–1; $K_{a\ min}$ – коэффициент минимально-возможного укорочения аллофона; $K_{a\ max}$ – коэффициент максимально-возможного удлинения аллофона.

Формула (4.8) пригодна для вычисления приращения длительности аллофонов глухих согласных и пауз. Для гласных и звонких согласных функция приращения длительности должна определять, какое количество питчей необходимо добавить или удалить в каждом конкретном аллофоне исходя из множества просодических факторов синтезируемой речи. Вычисление осуществляется в соответствие со следующей формулой:

$$\partial P_a = (K_p(TMP * K_{a\ min} + (1 - TMP) * K_{a\ max}) - 1) * Np_a \quad (4.9)$$

где Np_a – количество питчей в аллофоне.

Кроме рассмотренных выше просодических факторов, определяющих длительность аллофонов гласных и звонких согласных, их длительность существенно зависит также от текущего значения частоты основного тона (ЧОТ) - F_0 , задаваемого мелодическим портретом синтезируемой синтагмы. Это связано с тем, что при модификации F_0 изменяется длительность каждого питча аллофона при неизменном их общем числе. Длительность питча $Tr_i = 1 / F_{0i}$ может изменяться под действием интонационного фактора весьма существенно (более чем в 2 раза).

Приращение длительности аллофона под действием интонационного фактора определяется отношением (Tr_a / Tr_i) , где Tr_a – исходная средняя длительность питча аллофона. С учётом действия этого фактора формула (4.9) преобразуется к виду:

$$\partial P_a = (K_p(Tr_a/Tr_i) * (TMP * K_{a\ min} + (1 - TMP) * K_{a\ max}) - 1) * Np_a \quad (4.10)$$

При этом Tr_i определяются мелодическим портретом синтезируемой АЕ, а также задаваемым диапазоном изменения ЧОТ - $F_{0\ min}, F_{0\ max}$.

4.5.3. Алгоритм просодической обработки речевого сигнала в акустическом процессоре

В соответствии с текстом текущей синтагмы из БД звуковых волн выбирается требуемая последовательность аллофонов и (или) мультифонов, осуществляется их последовательное соединение (конкатенация). По заданным просодическим характеристикам синтагмы выбираются требуемые просодические портреты АЕ и далее формируются текущие значения $F_0(t)$, $A(t)$, $T(t)$ звуковых волн аллофонов.

Общая структура алгоритма представлена на рис. 4.26.

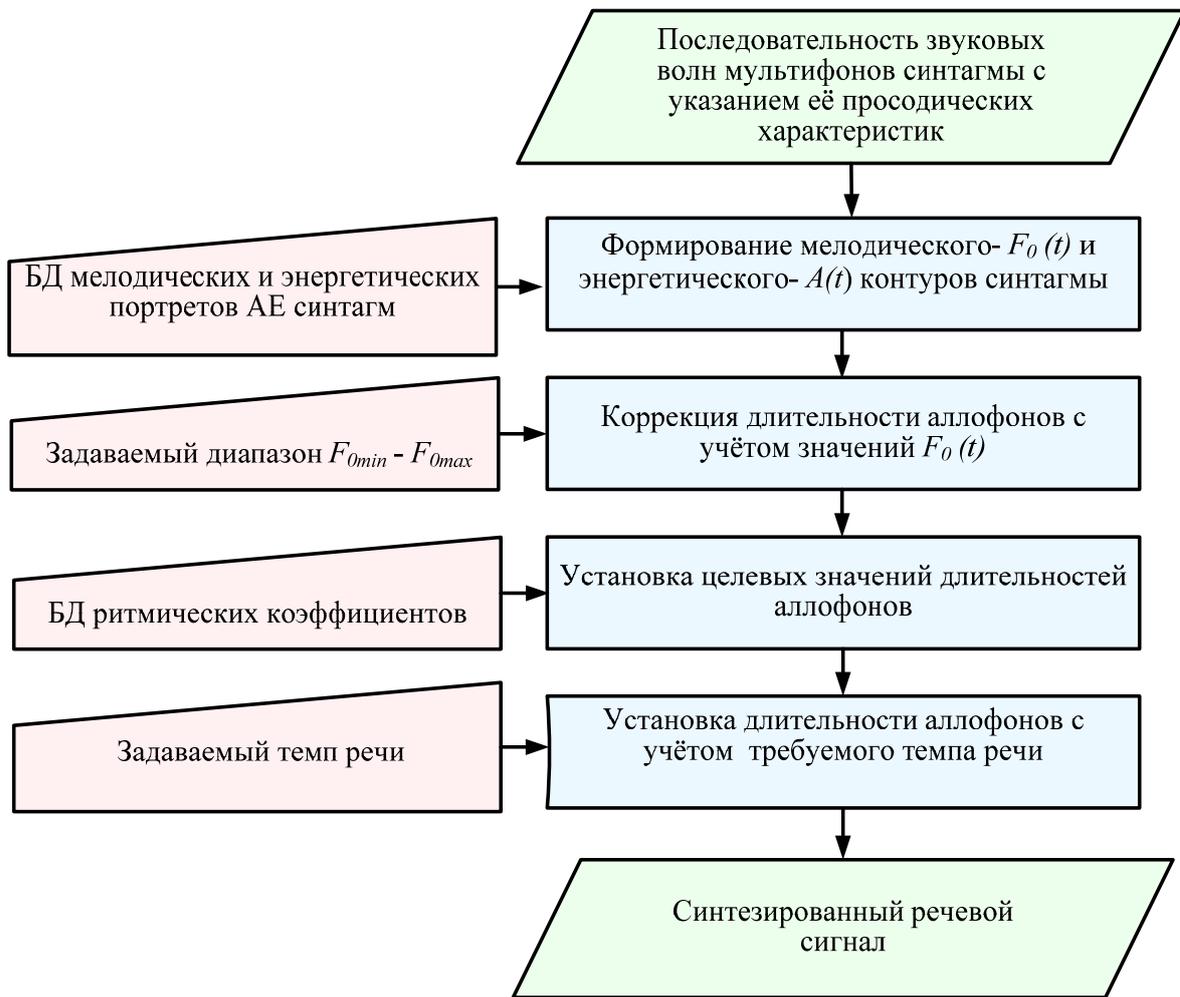


Рис. 4.26. Структурная схема алгоритма просодической обработки речевого сигнала

Формирование мелодического - $F_0(t)$ и энергетического - $A(t)$ контуров. Работу алгоритма проиллюстрируем на примере формирования мелодического- $F_0(t)$ и энергетического- $A(t)$ контуров синтагмы текста «*Машенька открыла глаза*», состоящей из трёх АЕ, интонационный тип которой – СЗ.

Аллофонная запись этой синтагмы имеет вид:

1-я АЕ: $M_{002}, \mathbf{A}_{012}, SH_{001}, E_{323}, N'_{003}, K_{002}, A_{232}$

2-я АЕ: $A_{222}, T_{001}, K_{002}, R_{002}, \mathbf{Y}_{021}, L_{002}, A_{212}$

3-я АЕ: $G_{002}, L_{002}, A_{212}, Z_{002}, \mathbf{A}_{020}$

Ядро каждой АЕ помечено жирным шрифтом. Ядрами являются полноударные гласные, распознаваемые по принадлежности к списку {А, Е, I, О, U} и первому цифровому индексу {0}.

Предъядром каждой АЕ являются аллофоны, стоящие от ядра слева, а заядром – стоящие справа от ядра. В 3-й АЕ заядро отсутствует.

В каждой АЕ подчёркиванием помечены **шумные** аллофоны (в отличие от **звонких**), которые не обрабатываются в процессе синтеза мелодического контура. Такие аллофоны распознаются по принадлежности к списку: $\{b, d, g, p, t, k, z, zh, f, s, sh, h, b', d', g', p', t', k', z', f', s', sh', h', c, ch'\}$ независимо от их цифровых индексов.

На рис.4.27 представлен мелодический портрет трёхакцентной синтагмы интонационного типа – СЗ, взятый из БД интонационных контуров. Каждый портрет АЕ содержит три одинаковых по длине участка: предядро, ядро и заядро, каждый из которых состоит из 100 равноотстоящих отсчётов мелодической кривой.

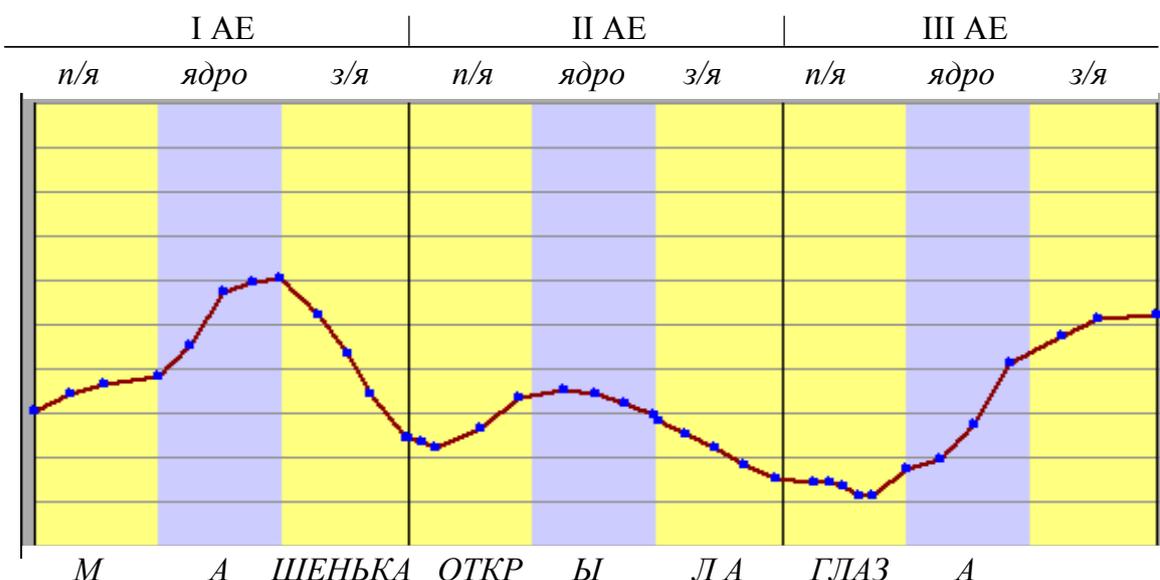


Рис. 4.27. Мелодический портрет трёхакцентной синтагмы интонационного типа – СЗ

Мелодический портрет описывает движение частоты основного тона (ЧОТ), нормированное - (F_0^N) от 0 до 1. Для того, чтобы найти абсолютные значения ЧОТ, необходимо задать минимальное - $F_{0\ min}$ и максимальное - $F_{0\ max}$ значения ЧОТ, характерные для синтезируемого голоса и произвести расчёт по формуле:

$$F_0 = F_0^N (F_{0\ max} - F_{0\ min}) + F_{0\ min} \quad (4.11)$$

При использовании БД аллофонов мужского голоса выбираем $F_{0\ max} = 200\text{Гц}$, а $F_{0\ min} = 70\text{Гц}$. При желании, изменяя эти значения, можно для того же голоса изменять диапазон ЧОТ (вплоть до монотонной речи), либо смещать высоту голоса.

Формула (4.11) используется далее для расчёта длительности каждого периода звонких (не глухих) аллофонов. Длительность

текущего периода (в количестве отсчётов сигнала - N_0) определяется по формуле:

$$N_0 = \frac{F_d}{F_0} \quad (4.12)$$

где F_d - частота дискретизации речевого сигнала.

Работу алгоритма проиллюстрируем на примере синтеза первой АЕ синтагмы - «Машенька».

Шаг 1. Из БД интонационных контуров считываем мелодический портрет 1-й АЕ (рис. 4.28а).

Шаг 2. Из БД звуковых волн мужского голоса считываем звуковые волны аллофонов, соответствующие предъядру - M_{002} , ядру - A_{012} и заядру - SH_{001} , E_{323} , N'_{003} , K_{002} , A_{232} . В БД звуковых волн каждый из аллофонов и мультифонов предварительно размечен на питчи, указывающие границы периодов основного тона.

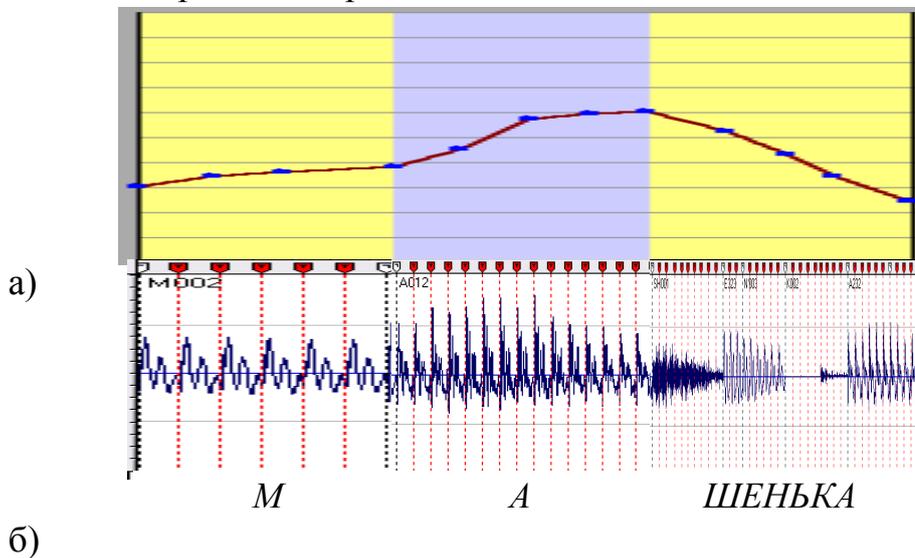


Рис. 4.28. Иллюстрация к алгоритму считывания нормированных ЧОТ

Шаг 3. Выравниваем длительности аллофонов, входящих в предъядро, ядро и заядро путём их растяжения или сжатия (рис. 4.28 б).

Шаг 4. Переносим метки питчей с сигналов звонких аллофонов предъядра, ядра и заядра на мелодический портрет АЕ и считываем с него соответствующие значения нормированных частот - F_0^N .

Шаг 5. По формулам (4.11, 4.12) вычисляем требуемые в соответствии с мелодическим портретом длительности периодов звуковых волн аллофонов.

Шаг 6. Изменяем длительность каждого периода звонких аллофонов в соответствии с SL-алгоритмом плавной сшивки периодов, описанном в ранее в разделе 4.5.2.

На рис. 4.29 приведена осциллограмма первого слога «МА» 1-й АЕ синтагмы, полученная путём непосредственной её компиляции из БД аллофонов (вверху) и осциллограмма АЕ после модификации длительности периодов каждого из звонких аллофонов в соответствии с мелодическим портретом АЕ.

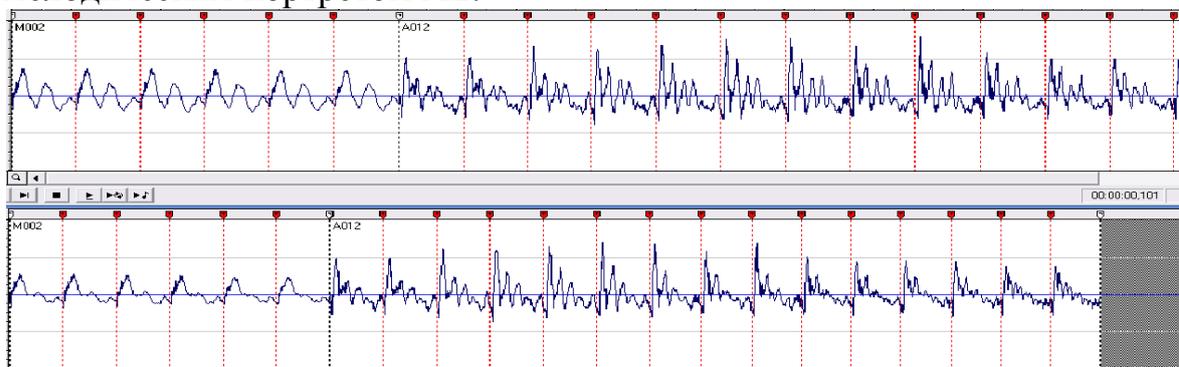


Рис. 4.29. Иллюстрация изменения длительности периодов

Если заядерный участок АЕ отсутствует полностью, как в третьей АЕ рассматриваемого примера: G_{002} , L_{002} , A_{212} , Z_{002} , A_{020} , либо в заядерной части отсутствуют звонкие аллофоны (как, например в слове *мост*), то для полной реализации интонационного контура АЕ необходимо искусственно создать заядерный участок на ударном гласном, выступающем как ядро АЕ. Для этой цели выполняются следующие шаги алгоритма.

Шаг 4'. Увеличиваем длительность ядерной гласной на 30%.

Шаг 4''. Отступаем от конца полученного аллофона 1/3 его длительности и помечаем этот участок как заядро АЕ.

Дальнейшие процедуры осуществляются в полном соответствии с описанным выше алгоритмом.

Синтез энергетического контура синтагмы осуществляется путём задания соответствующих значений амплитуд аллофонов, из последовательности которых она состоит. Установка значений амплитуды - A_i i -го аллофона синтагмы осуществляется в соответствии с энергетическим портретом АЕ. Пример ритмического портрета 3-х акцентной синтагмы представлен на рис. 4.29.

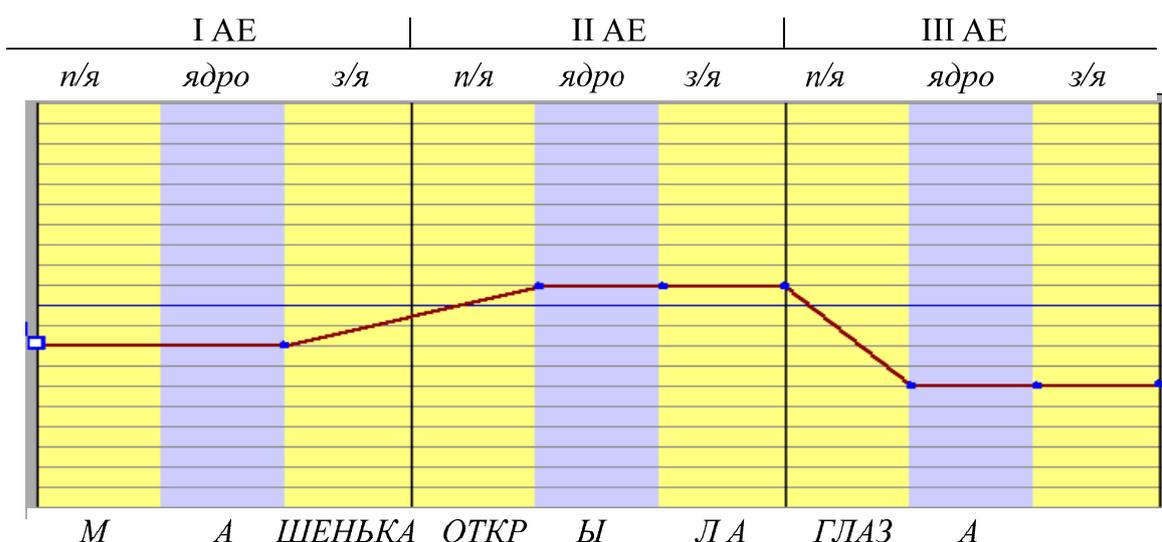


Рис. 4.29. Энергетический портрет 3-х акцентной синтагмы

Установка амплитуд аллофонов осуществляется усилением (ослаблением) сигналов аллофонов, составляющих данную АЕ, путём умножения значений сигнала на коэффициент, задаваемый энергетическим портретом АЕ.

В приведенном на рис. 4.29 примере аллофонные сигналы первой АЕ умножаются на коэффициент 0,8, второй АЕ – на 1,1 и третьей АЕ – на 0,6.

Коррекция длительности аллофонов с учётом значений ЧОТ.

На рис. 4.30 а приведена осциллограмма первой АЕ синтагмы, полученная путём непосредственной её компиляции из БД аллофонов, а на рис. 4.30 б – её осциллограмма после модификации длительности периодов каждого из звонких аллофонов в соответствии с мелодическим портретом АЕ.

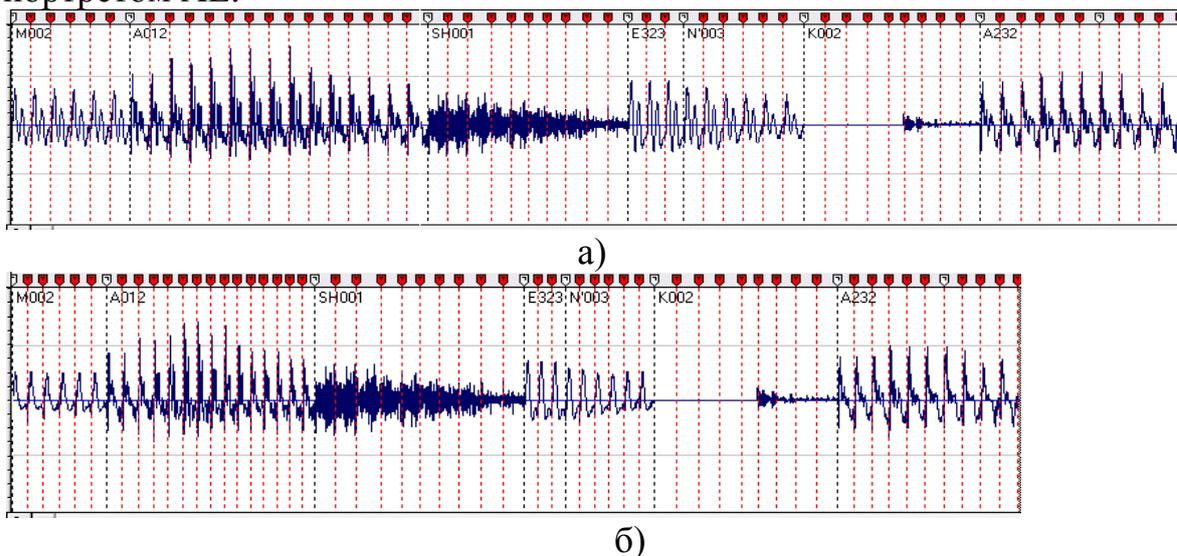


Рис. 4.30. Иллюстрация изменения длительности аллофонов

Как видно из сравнения рис. 4.30 а и 4.30 б, длительность глухих аллофонов при модификации ЧОТ в соответствии с описанным алгоритмом осталась неизменной, однако длительность звонких аллофонов изменилась в ряде случаев весьма существенно.

Для того, чтобы длительность звонких аллофонов осталась неизменной, необходимо осуществить следующие операции.

Шаг 1. Подсчитываем суммарную длительность питчей $-Ta_i^m$ i -го модифицированного аллофона и сравниваем её с длительностью исходного аллофона – Ta_i . Определяем разность

$$dT = Ta_i^m - Ta_i \quad (4.13)$$

Если dT оказалась <0 , то модифицированный аллофон следует удлинить, а если >0 , то укоротить на определённое целое число питчей. Удлинение или укорочение осуществляется до тех пор, пока dT не станет менее длительности одного питча.

Шаг 2. Осуществляется процедура удлинения (укорочения) модифицированного аллофона.

На рис. 4.31 представлена ситуация **удлинения** гласного аллофона A_{112} на 2 периода колебаний.

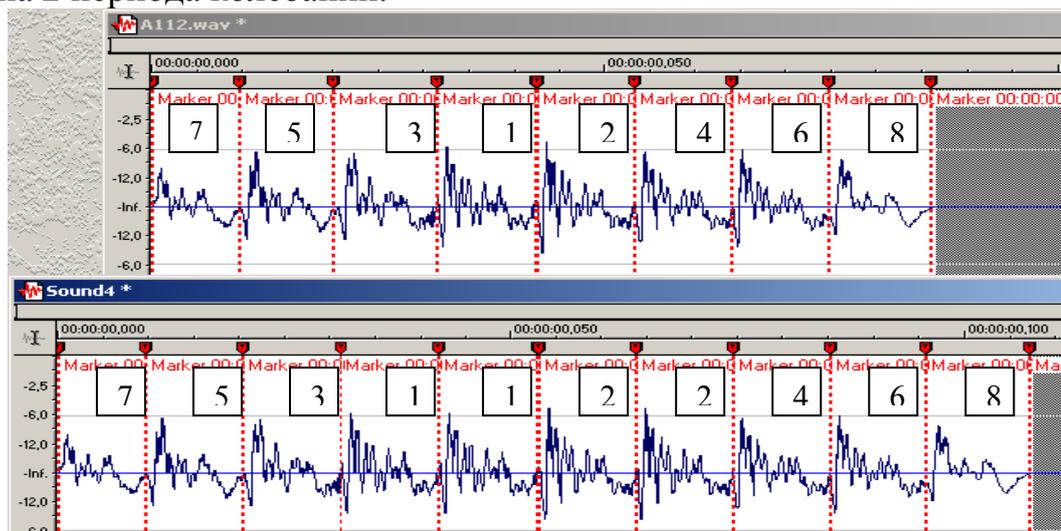


Рис. 4.31. Иллюстрация процедуры удлинения гласного аллофона

Вначале определяется положение срединного питча (1) (см. верхнюю осциллограмму рис 4.31). **Удлинение** на 1 питч осуществляется путём его однократного повторения. Для удлинения на 2 питча осуществляется повторение его правого соседа – питча (2). Результат удлинения гласного аллофона A_{112} на 2 периода колебаний показан на нижней осциллограмме рис. 4.31. Для дальнейшего удлинения на 3 периода осуществляется повторение левого «соседа», т.е. питча (3), на 4 периода – следующего правого соседа, т.е. питча (4) и т.д. После добавления каждого нового питча его длительность суммируется с текущим значением Ta_i^m . Описанная процедура

повторяется до тех пор, текущая длительность Ta_i^m и требуемая длительность аллофона $T^N a_i$ не станут равны с точностью до одного питча.

Для **укорочения** гласного аллофона осуществляется операция «прореживания». Для укорочения на один период убирается питч (2), на два периода – питчи (2, 3), на три периода – питчи (2, 3, 6) и т.д., не затрагивая, однако, **срединный** питч (1), а также **первый** (7) и **последний** (8) питчи, которые должны быть обязательно сохранены. Процедура сокращения прекращается, когда текущая длительность Ta_i^m и требуемая длительность аллофона $T^N a_i$ не станут равны с точностью до одного питча, либо когда в аллофоне останется всего три питча – два крайних и один срединный.

На рис.4.32 вверху показана осциллограмма первой АЕ синтагмы, полученная путём непосредственной её компиляции из БД аллофонов, а внизу – осциллограмма АЕ после изменения длительности периодов в соответствии с мелодическим портретом АЕ и после проведения описанной выше процедуры выравнивания длительностей исходных и модифицированных аллофонов.

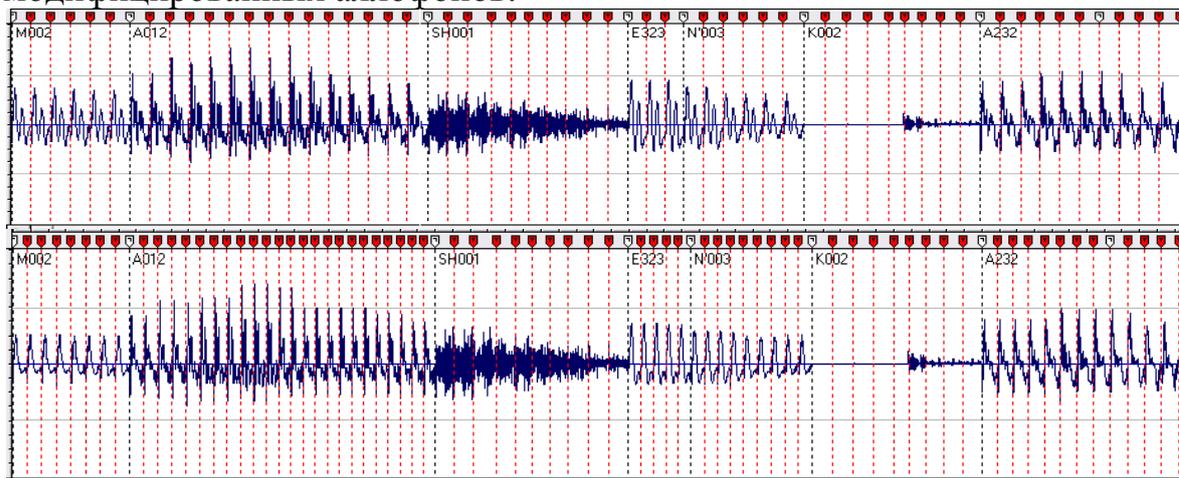


Рис. 4.32. Исходная (вверху) и модифицированная (внизу) осциллограммы первой АЕ синтагмы

Установка целевых значений длительностей аллофонов.

Формирование ритмического контура синтагмы осуществляется путём задания соответствующих значений длительности каждого аллофона, из последовательности которых она состоит. Установка значений длительности - T_i i -го аллофона синтагмы осуществляется в соответствии с коэффициентами, задаваемым ритмическими портретами акцентных единиц (АЕ) синтагмы, и корректируется затем с учетом требуемого темпа речи.

На рис. 4.33 показан пример ритмического портрета синтагмы текста «*Машенька открыла глаза*», состоящей из трёх АЕ, интонационный тип которой – СЗ. Нижний участок рисунка показывает

изменённые под действием ритмических факторов относительные длительности ядер первой, второй и третьей АЕ синтагмы.

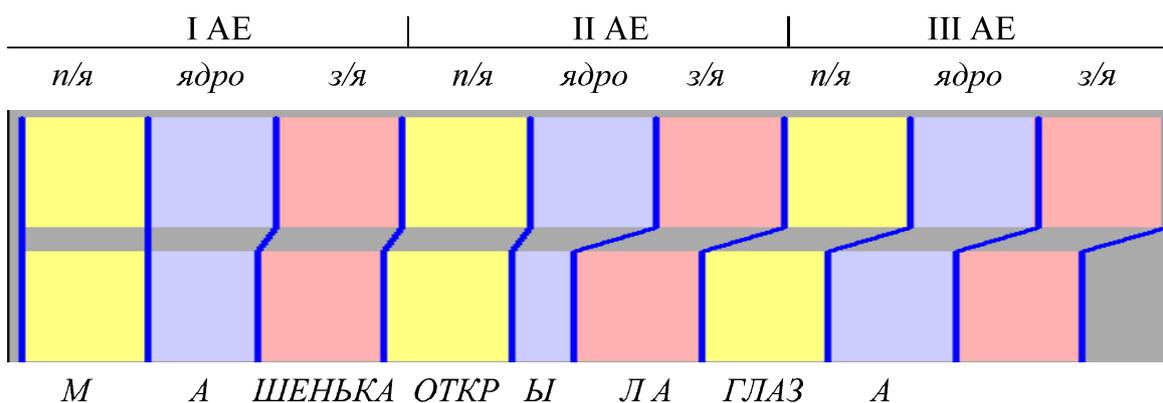


Рис. 4.33. Пример ритмического портрета 3-х акцентной синтагмы

Из рис. 4.33 видно, что в данном примере наибольшую длительность имеет ядро третьей (конечной) АЕ, наименьшую – ядро второй АЕ (серединной) и среднюю длительность – ядро первой (начальной) АЕ. Если не принимать во внимание ярко выраженных индивидуальных особенностей диктора или стиля речи, то в соответствии с известными для русской речи данными можно определить следующие факторы, влияющие на ритмическое оформление синтагмы:

1. Интонационный тип – **I** (незавершённый, завершённый, вопросительный, восклицательный);
2. Позиция АЕ в синтагме – **P** (серединная, начальная, конечная);
3. Количество АЕ в синтагме – **M** (три и более, две, одна);
4. Количество слогов в АЕ – **N** (три и более, два, один).

Здесь в скобках приведены градации каждого фактора в порядке их влияния на степень удлинения ядра АЕ. При создании БД аллофонов для синтеза речи использовались фонограммы фраз (см. главу 5), в которых для указанных четырёх ритмических факторов преобладали подчёркнутые обозначения их градаций. Эти подчёркнутые обозначения градаций принимаются равными 1. В первом приближении каждый из четырёх факторов может быть представлен тремя градациями, экспериментальные значения которых приведены в табл. 4.5. Набор этих значений описывает ритмический портрет синтагмы.

Таблица 4.5. Экспериментальные значения ритмических коэффициентов

№	Ритмический фактор	Обозначение	Градация	Значение K_r
1	Интонационный тип	I_1	Незавершённый	1
		I_2	Завершённый	1,1
		I_3	Вопросительный	1,2
		I_4	Восклицательный	1,3
2	Позиция АЕ в синтагме	P_1	Срединная	1
		P_2	Начальная	1,1
		P_3	Конечная	1,2
3	Количество АЕ в синтагме	M_1	Более двух	1
		M_2	Две	1,1
		M_3	Одна	1,2
4	Количество слогов в АЕ	N_1	Более двух	1
		N_2	Два	1,2
		N_3	Один	1,3

Общий ритмический коэффициент определяется как:

$$K_r = I * P * M * N \quad (4.14)$$

Для выбранных в табл.4.5 значений максимально-возможное удлинение ядерной гласной составит: $K_r = 1,3 * 1,2 * 1,2 * 1,3 = 2,4$.

Установка требуемого темпа речи. Регулировка темпа речи осуществляется путём корректировки длительности аллофонов и межсинтагменных пауз с учётом коэффициента "податливости" каждого конкретного звука темповым изменениям. Экспериментальная оценка пределов изменения относительной длительности аллофонов и пауз при изменении темпа приведены в таблице 4.6. Считается, что среднему темпу соответствует длительность аллофонов, помещённых в БД.

Таблица 4.6. Относительная длительность аллофонов при изменении темпа

№ пп	Звуковые единицы	Медлен н темп (K_{max})	Средний темп	Быстры й темп (K_{min})
1.	Паузы	2,5	1,0	0,2
2.	Ударные гласн. (индекс 0 или 1)	2,0	1,0	0,5
3.	Предударные гласн. (индекс 2)	2,0	1,0	0,7
4.	Заударные гласн. (индекс 3)	2,0	1,0	0,8
5.	Согласные	1,3	1,0	0,8

Синтез ритмического контура осуществляется посинтагменно, т.е. путём считывания в буфер последовательности Wav-файлов аллофонов, наполняющих синтагму.

Процедура синтеза ритмического контура основана на вычислении новых значений длительностей аллофонов - $T^N a_i$, исходя из множества перечисленных выше факторов. Вычисление осуществляется в соответствии со следующей формулой:

$$T^N a_i = K_p (K_{min} * TMP + (1 - TMP) * K_{max}) * T a_i \quad (4.15)$$

где: K_p – просодический коэффициент; TMP – желаемый темп речи, задаваемый на интервале (0 – 1); K_{min} – коэффициент минимально-возможного укорочения аллофона; K_{max} – коэффициент максимально-возможного удлинения аллофона; $T a_i$ – длительность i -го аллофона.

В формуле (4.15) значение коэффициента K_p берётся из таблицы 4.5 только для ударных гласных. Для остальных аллофонов его значение равно 1. Темп речи задаётся значениями на интервале (0 – 1). Нулю соответствует самый быстрый темп, единице – медленный. Коэффициенты K_{min} и K_{max} берутся из таблицы 4.6. Они должны быть различны для различных классов аллофонов. Длительность i -го аллофона $T a_i$ определяется БД аллофонов. Длительности: исходная - $T a_i$ и просодически модифицированная - $T^N a_i$, задаются количеством отсчётов в аллофоне сигнале.

Процесс модификации длительности глухих согласных аллофонов: $\{p, t, k, f, s, sh, h, p', t', k', z', f', s', sh', h', c, ch\}$ осуществляется следующим образом.

Шаг 1. Определяется значение разности

$$dT = T^N a_i - T a_i \quad (4.16)$$

Шаг 2. Если полученное значение – положительное число, то осуществляется удлинение аллофона. Для этого вначале определяется положение середины звука. Затем от неё откладывается отрезок требуемой длительности - dT , который повторяется (копируется).

Шаг 3. Если полученное значение – отрицательное число, то осуществляется укорочение аллофона путем вырезания участка звука требуемой длительности из середины аллофона.

На рис. 4.34 представлен пример процедуры удлинения глухого аллофона.

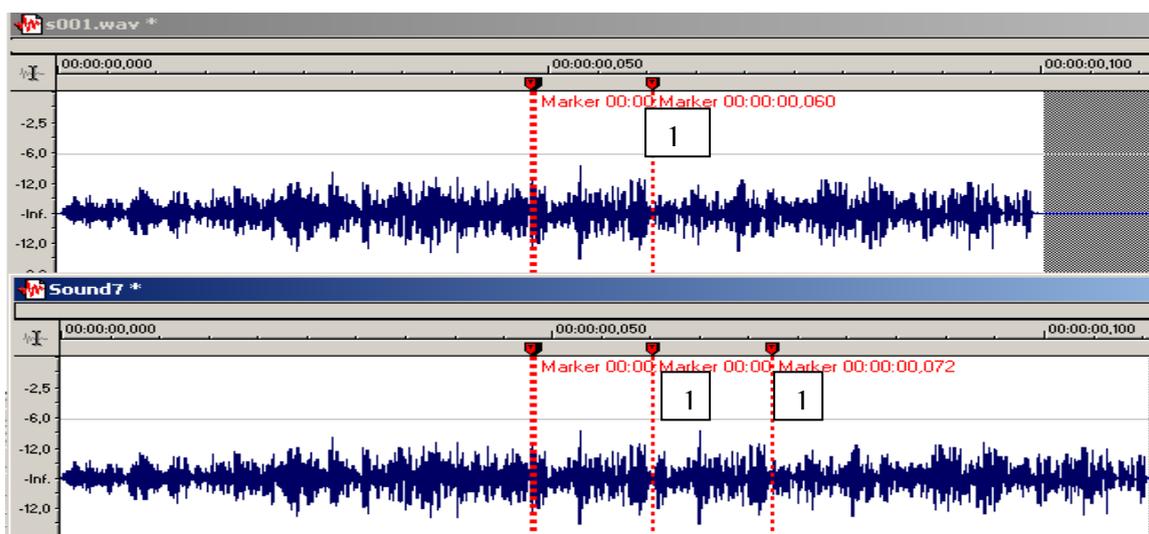


Рис. 4.34. Иллюстрация процедуры удлинения глухого щелевого аллофона S_{001} .

Процесс модификации длительности аллофонов гласных $\{a, e, i, o, u\}$ и звонких согласных $\{m, n, l, r, m', n', l', r', j', b, d, g, z, zh, b', d', g', z'\}$ осуществляется таким же образом, как и при рассмотренной ранее коррекции длительности аллофонов с учётом значений ЧОТ

4.6. Программная реализация системы синтеза речи “Мультифон”

4.6.1. Функциональная схема системы

Функциональная схема, входные и выходные данные, взаимодействие блоков системы синтеза речи представлены на рис 4.35. В системе реализованы описанные выше алгоритмы обработки текста и речевого сигнала.

Входные данные системы:

– орфографический текст, содержащийся в текстовом файле либо вводимый с клавиатуры.

Выходные данные системы:

– синтезированный речевой сигнал, подаваемый на устройство вывода звука либо сохраняемый в файл в формате WAVE PCM.

На первом этапе синтеза осуществляется очистка и морфосинтаксическая обработка текста. При этом из текста удаляются символы, не входящие в множество допустимых для синтеза русской речи, осуществляется расшифровка аббревиатур и сокращений, расстановка словесных ударений и указание морфологической категории (МК) для каждого слова текста. Для выполнения первого этапа обработки текста используются орфографический словарь, а также списки аббревиатур и сокращений.

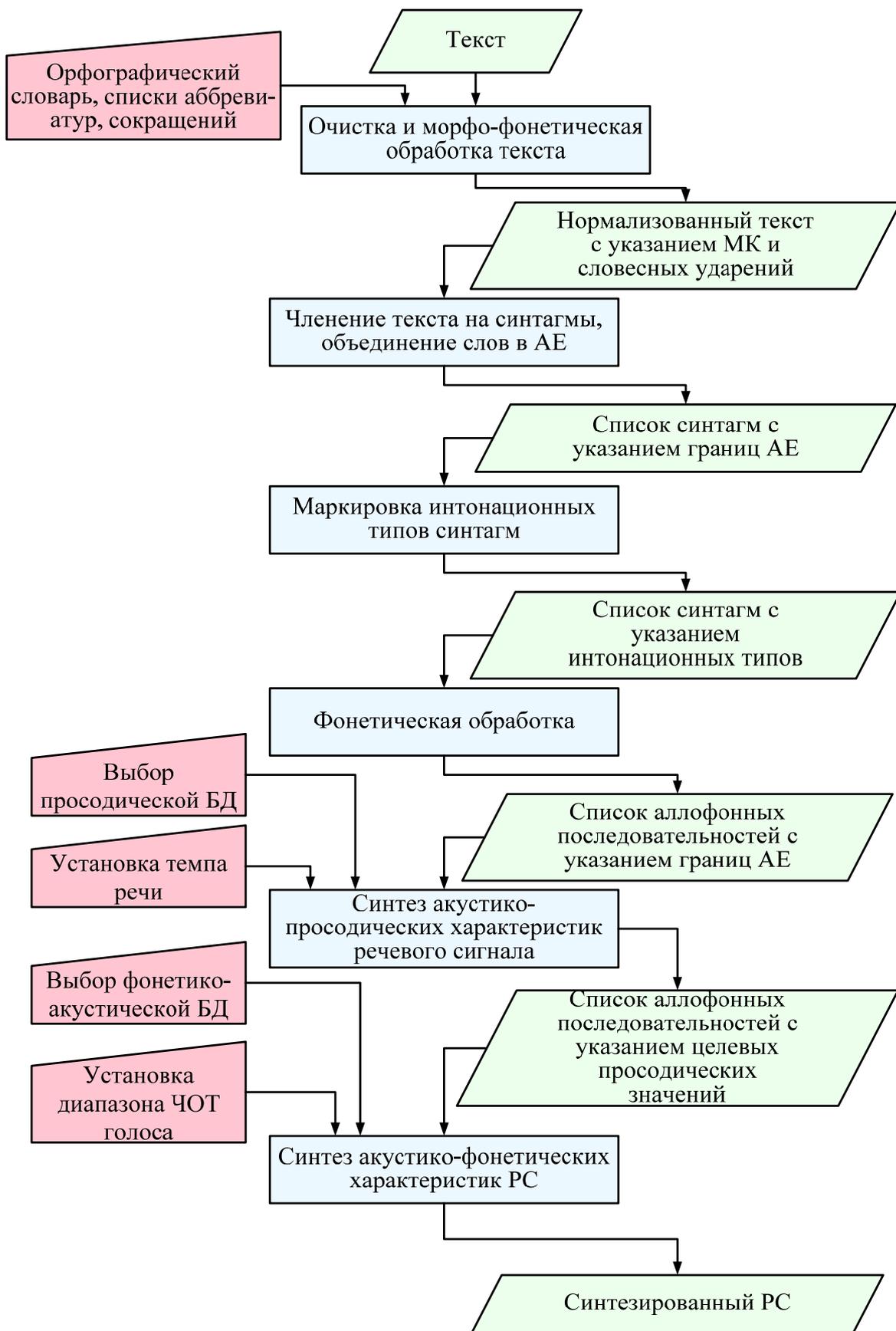


Рис. 4.35. Общая функциональная схема системы синтеза речи по тексту

Нормализованный текст с указанием морфологических категорий и словесных ударений поступает в блок членения на синтагмы и объединения слов в АЕ. В этом блоке на основании указанных МК осуществляется присоединение энклитиков и проклитиков к знаменательным частям речи, определение границ фонетических слов и АЕ, разбиение текста на синтагмы.

Следующий блок синтезатора речи осуществляет маркировку интонационных типов синтагм. Результат работы данного блока – список синтагм с указанием границ фонетических слов и АЕ, а также интонационног типа каждой синтагмы – подвергается фонетической обработке.

В процессе фонетической обработки, реализующей замену слов – фонетических исключений на эквиваленты, преобразования «буква-фонема» и фонема-аллофон, формируется список аллофонных последовательностей, в каждой из которых сохраняются пометы границ АЕ, полученные на предыдущих этапах обработки.

В следующем блоке – блоке синтеза акустико-просодических характеристик – на основании данных просодической БД осуществляется вычисление целевых значений просодических параметров (ЧОТ, амплитуды, длительности) каждого аллофона каждой синтагмы входного списка. На этом этапе обработки пользователь может выбрать просодическую БД, а также указать требуемый темп синтезируемой речи.

Синтез акустико-фонетических характеристик речевого сигнала осуществляется путём выбора из фонетико-акустической БД требуемых аллофонов или мультифонов, их конкатенации и модификации сигнала в соответствии с целевыми значениями просодических параметров. На этом этапе пользователь может выбрать фонетико-акустическую БД, а также указать диапазон изменения ЧОТ для выбранной БД.

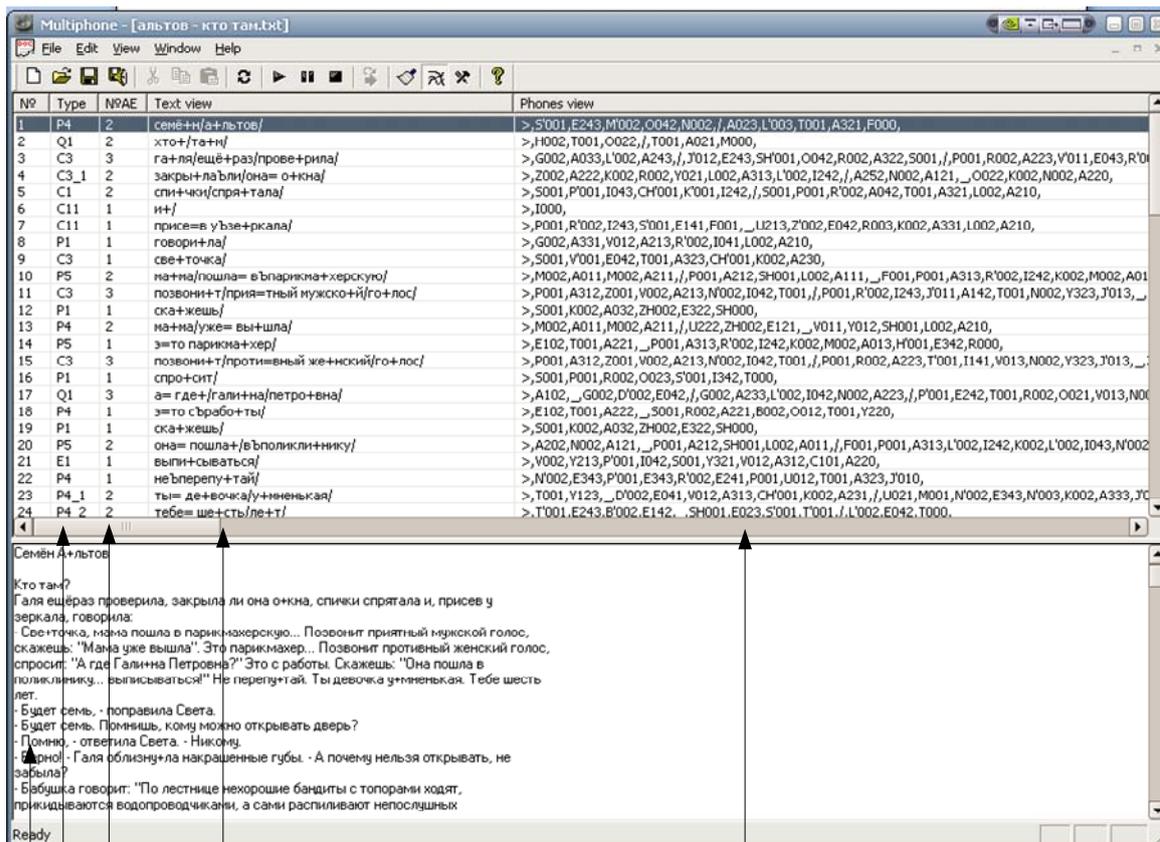
Алгоритмы работы системы «Мультифон» соответствуют основным положениям, рассмотренным в разделах 4.1 – 4.5.

4.6.2. Пользовательский интерфейс системы «Мультифон»

Пользовательский интерфейс системы «Мультифон» (рис. 4.36) включает следующие блоки:

- окно отображения входного текста;
- блок окон отображения результатов просодической и фонетической обработки текста, содержащий список синтагм, где для каждой синтагмы указаны её тип, количество АЕ, просодически обработанный текст – орфографический текст с указанием позиций словесных ударений и границ АЕ, и фонетически обработанный текст – последовательность аллофонов синтагмы;
- функциональные кнопки работы со звуком;

– диалоговые окна настроек параметров системы.



**Просодически
 обработанный текст**

**Фонетически
 обработанный текст**

**Количество
 АЕ в синтагме**

**Типы
 синтагм**

**Входной
 орфографический
 текст**

Рис. 4.36. Общий вид пользовательского интерфейса системы синтеза речи

Окно отображения входного текста (рис. 4.37) предназначено для отображения содержимого текстового файла и (или) ввода текстовой информации пользователем.

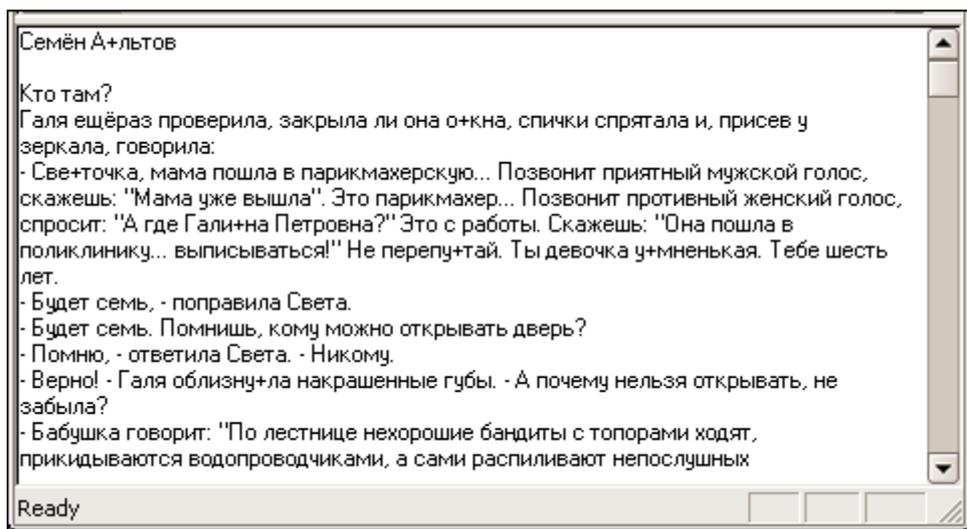


Рис. 4.37. Окно отображения входного текста

Блок окон отображения результатов просодической и фонетической обработки текста показан на рис. 4.38. Просодическая разметка текста каждой синтагмы включает метки границ синтагм (обозначенные символом «/»), а также словесных ударений: сильного (обозначенного символом «+» после ударного гласного) и слабого (обозначенного символом «=» после ударного гласного). Фонетически обработанный текст является последовательностью аллофонов, в которой сохранены метки границ АЕ, кроме того, установлены метки границ слоговых сегментов (символ «>»).

№	Type	№АЕ	Text view	Phones view
1	P4	2	се+м+н/а+льтов/	>, S'001E243, M'0020042, N002/A023, L'003, T001A321, F000,
2	Q1	2	кто+/та+м/	>, H002, T0010022, /, T001A021, M000,
3	C3	3	га+ля/ещё+раз/прове+рила/	>, G002A033, L'002A243, /, J'012, E243, SH'0010042, R002A322, S001, /P001R002A223
4	C3_1	2	закры+ла/бли/она= о+кна/	>, Z002, A222, K002, R002Y021, L002A313, L'002I242, /, A252, N002, A121, _O022, K002
5	C1	2	спички/спря+тала/	>, S001, P'001I043, CH'001, K'001I242, /, S001, P001R'002A042, T001A321, L002A210,
6	C11	1	и+т/	>, I000,
7	C11	1	присе=в у+бже+ркала/	>, P001R'002I243, S'001, E141, F001, _U213, Z'002, E042, R003, K002A331, L002A210,
8	P1	1	говори+ла/	>, G002A331, V012A213, R'002I041, L002A210,
9	C3	1	све+точка/	>, S001, V'001E042, T001A323, CH'001K002A230,
10	P5	2	ма+ма/пошла= в+парикма+херскую/	>, M002A011, M002A211, /, P001A212, SH001, L002A111, _F001, P001A313, R'002I242
11	C3	3	позвони+т/прия=тный мужско+й/го+лос/	>, P001A312, Z001, V002A213, N'002I042, T001, /P001R'002I243, J'011A142, T001, N002
12	P1	1	ска+жешь/	>, S001, K002A032, ZH002, E322, SH000,
13	P4	2	ма+ма/уже= вы+шла/	>, M002A011, M002A211, /, U222, ZH002, E121, _V011, V012, SH001, L002A210,
14	P5	1	э=то парикма+хер/	>, E102, T001A221, _P001A313, R'002I242, K002, M002A013, H'001E342, R000,

Рис. 4.38. Блок окон отображения результатов просодической и фонетической обработки текста

Функциональные кнопки работы со звуком (рис. 4.39) предназначены для воспроизведения синтезированного речевого сигнала и включают функции проигрывания звука, паузы и остановки.



Рис. 4.39. Функциональные кнопки работы со звуком

Диалоговое окно настроек системы (рис. 4.40) включает выбор фонетико-акустической БД (“Voice base”), просодической БД (“Prosody style”), указание диапазона изменения частоты основного тона (“Frequency range”), опции использования интонационных, ритмических и динамических портретов при синтезе просодических параметров речевого сигнала (“Use frequency”, “Use rhythm”, “Use energy”), а также указание значений темпа речи (“Tempo”) и уровня громкости (“Volume”).

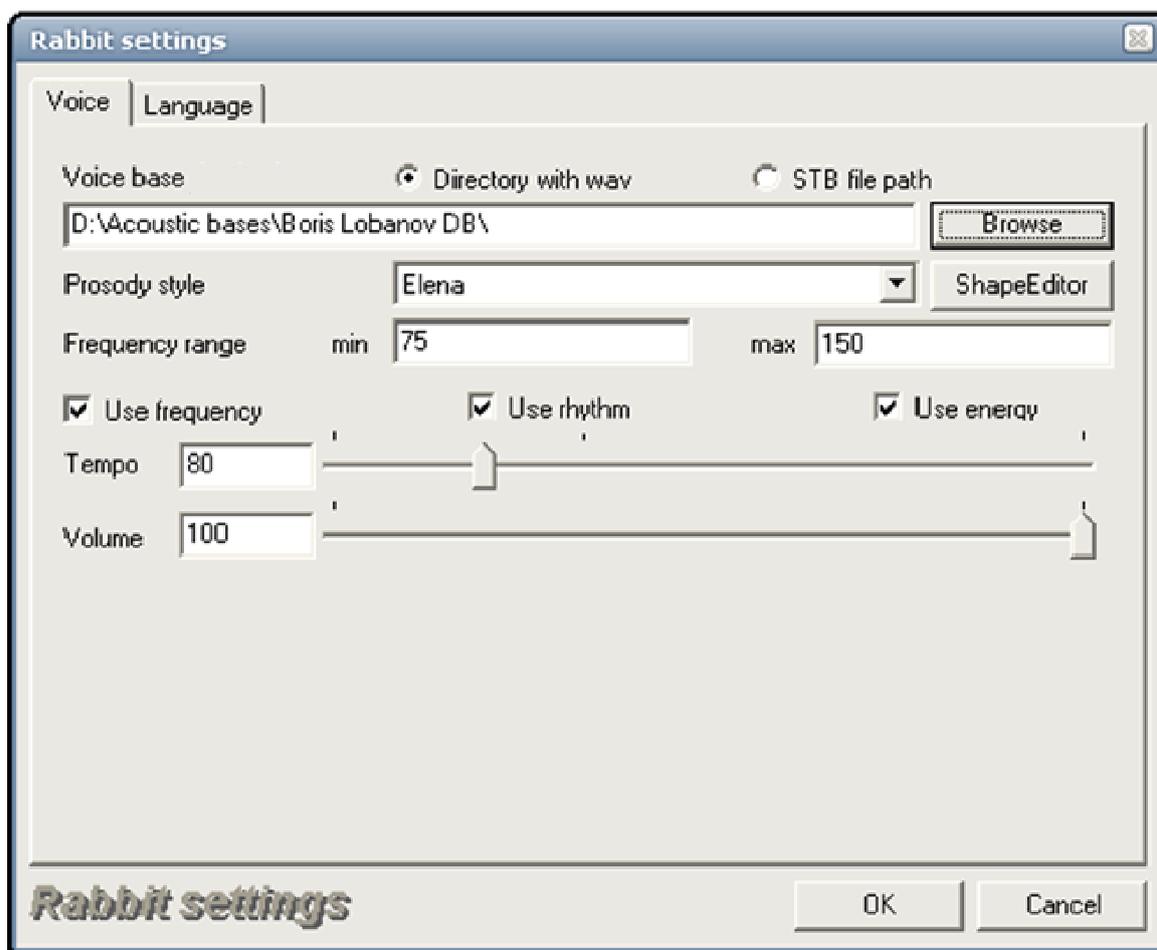


Рис. 4.40. Диалоговое окно настроек системы

4.6.3. Оценка разборчивости синтезированной речи

Для оценки разборчивости синтезируемой речи решаются следующие задачи:

- получить численную оценку слоговой разборчивости синтезированных речевых клонов;
- оценить вклад элементов компиляции различного фонетического количества (аллофоны, диаллофоны, аллослоги) в степень разборчивости синтезируемой речи;

– сравнить результаты слоговой разборчивости синтезированных речевых клонов с результатами разборчивости речи, синтезируемой с использованием наилучших известных синтезаторов русской речи [175].

В экспериментах по определению слоговой разборчивости речи вычислялась субъективная оценка. Методика проведения эксперимента основывалась на СТБ ГОСТ Р 50840-2000 «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости» [176].

Для проведения экспериментов использовались речевые клоны, синтезированные на основе созданных БД элементов компиляции: БД мужского голоса – БД-М, и БД женского голоса – БД-Ж. Содержимое БД, использованных для синтеза речи, приведено в таблице 4.7.

Речевым материалом для тестирования являлись слоговые артикуляционные таблицы [176], которые были синтезированы на основании баз БД-М и БД-Ж с соблюдением постоянного ритма и постоянного уровня речи и сохранены на цифровые носители в формате WAVE PCM с частотой дискретизации 22050 Гц и разрядностью 16 бит. Длительность пауз между слогами была приведена к значению 3 секунды.

Кроме того, слоговые артикуляционные таблицы были синтезированы с использованием русскоязычного синтезатора компании Nuance [175]. Значения ритма речи, уровня речи, длительности межслоговых пауз, а также формат сохранения были аналогичны указанным выше.

Таблица 4.7

Содержимое фонетико-акустических БД для тестирования слоговой разборчивости

Название БД	Содержимое БД
БД-М	БД диктора Д1 (мужской голос), содержащая аллофоны, диаллофоны, аллослоги в количестве 6818 единиц
БД-Ж	БД диктора Д2 (женский голос), содержащая аллофоны, диаллофоны, аллослоги в количестве 7073 единицы

Аудиторы – 2 мужчин и 3 женщины в возрасте от 22 до 30 лет, носители русского языка, без выявленных дефектов слуха – записывали принятые слоги в бланки, составленные в соответствии с СТБ ГОСТ Р 50840-2000. Тесты проходили в тихой комнате, длительность сессии не превышала 35 минут.

Результаты эксперимента по оценке слоговой разборчивости синтезированной речи. Степень слоговой разборчивости выражается через среднее значение разборчивости S для каждого измерения, которое вычисляется в несколько этапов. Начальное значение S вычисляется по формуле:

$$S = \frac{1}{N} * \sum_{i=1}^N S_i \quad (4.17)$$

где S_i – результат единичного измерения, % (диктор-таблица-аудитор), N – число единичных измерений.

Затем вычисляется среднее квадратическое отклонение (СКО) σ по формуле

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (S_i - S)^2}{N - 1}} \quad (4.18)$$

и единичные измерения S_i , для которых $|S_i - S| > 2\sigma$, исключаются. Окончательное значение S вычисляется по формуле:

$$S = \frac{1}{N - k} * \sum_{i=1}^{N-k} S_i \quad (4.19)$$

где N – число единичных измерений, k – число исключённых измерений.

Результаты вычислений значений S для каждого из типов синтезируемых голосов, а также соответствующие классы качества [176] представлены в таблице 4.8.

Таблица 4.8

Классы качества синтезируемых голосов

Название БД или системы синтеза	Слоговая разборчивость, %	Класс качества
Мультифон (БД-М)	91	Высший
Мультифон (БД-Ж)	78	1-й
Nuance [175]	55	2-й

Глава 5

Компьютерное клонирование индивидуальных характеристик речи

Данная глава посвящена вопросам компьютерной реализации технологии клонирования индивидуальных характеристик в системе синтеза речи по тексту, базирующейся на теоретических и экспериментальных результатах, описанных в предыдущих главах. Целью разрабатываемой технологии является создание методами компьютерного клонирования необходимого и достаточного набора минимальных речевых единиц конкретного диктора, позволяющего получить методами синтеза речи по тексту высококачественную речь, воспринимаемую на слух как максимально схожую с его манерой чтения, с его голосом, дикцией и интонацией.

Постановка проблемы клонирования речи впервые дана в работах одного из авторов [1, 5, 6]. Материал данной главы во многом основан на оригинальных результатах авторов, полученных за последние годы [91, 169].

5.1. Факторы вариативности речевого сигнала

Устная речь имеет двойственную природу: семантическую и акустическую. С одной стороны, она является результатом интеллектуальной деятельности говорящего, использующего средства языка для общения с другими членами данного языкового коллектива. С другой стороны, она представляет собой звуковую волну, передающую от говорящего к слушающему наряду со смысловой информацией комплекс экстралингвистической информации. Общеизвестным свойством речи является её вариативность, проявляющейся в особенностях формирования речевого сигнала, его информационной, кодовой и модуляционной структуры.

Учитывая, что устная речь, во-первых, является формой реализации языка, и, во-вторых, имеет свои специфические особенности, определяемые психо-акустическим процессом говорения, можно определить следующие уровни изменчивости речи (см. рис 5.1):

– лексико-синтаксический, заключающийся в параллельных возможностях формального выражения одних и тех же лингвистических значений (значимостей), связанных с одной и той же лексемой, словоформой или одной и той же синтаксической конструкцией;

– суперсегментный, или интонационно-просодический, который выражается в персональных особенностях интонирования речи, ритмического оформления и паузации;

- сегментно-фонетический, выражающийся в особенностях произнесения звуков (иностранный или региональный акцент, гнусавость, шепелявость, картавость);
- артикуляторно-акустический, который проявляется в различиях тембра голоса, значениях формантных частот, различиях шумовой/тональной составляющих сигнала, спектральных характеристиках сигнала.

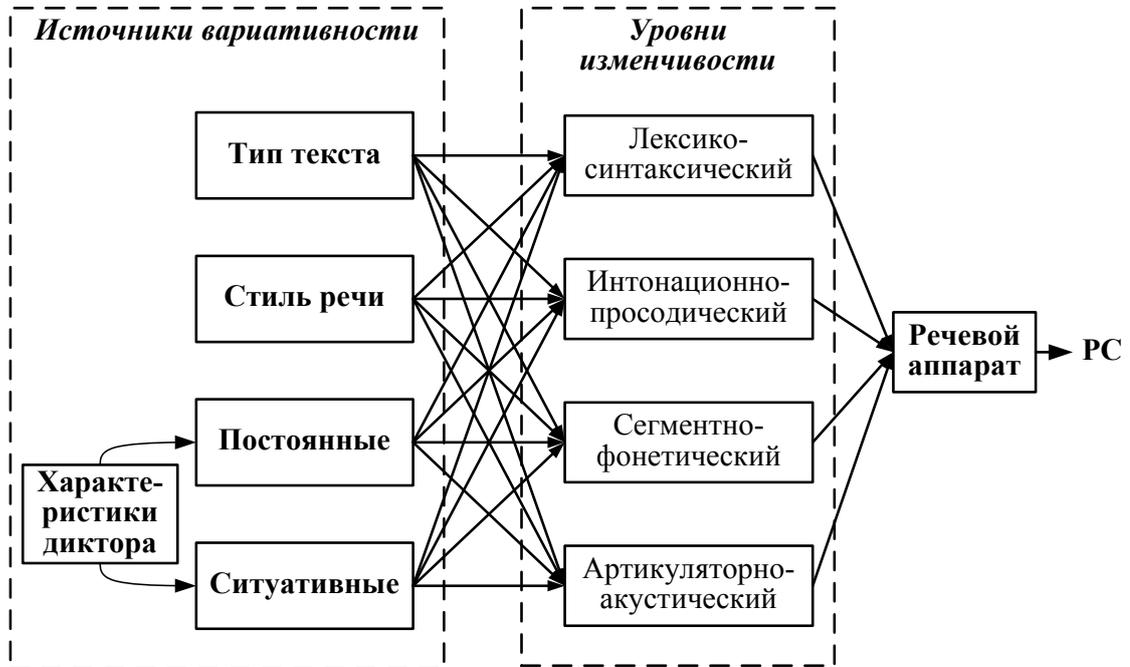


Рис. 5.1. Источники и уровни изменчивости речевого сигнала

К основным типам источников вариативности относятся вид текста, стиль речи, характеристики диктора – постоянные и ситуативные. Каждый из указанных типов является, в свою очередь, источниками изменчивости речевого сигнала на каждом из указанных выше уровней речи.

Вид произносимого текста и стиль речи в наибольшей степени влияют на лексико-синтаксическое и интонационно-просодическое оформление высказывания. К основным типам текста относятся деловой (доклад, техническое описание, официальное письмо и др.), характеризующийся наличием терминов и профессионального лексикона, и художественный (художественная проза, поэзия, пьесы и др.), включающий абстрактную и разговорную лексику.

К основным стилям речи относятся ораторский стиль, по отбору лексики и нормативности синтаксиса сближающийся со стилями письменной речи, и разговорный стиль, характеризующийся преобладанием обиходно-бытовой лексики и значительной свободой синтаксических построений.

Персональные характеристики диктора, как постоянные, так и ситуативные, являются источником вариативности на каждом из указанных выше уровней изменчивости речи. К постоянным характеристикам относятся, в частности, пол, возраст, речевая патология. К ситуативным характеристикам относится эмоционально-психологическое и физическое состояние диктора.

Необходимо подчеркнуть, что к источникам индивидуального варьирования, по мнению современных лингвистов [177], относятся также логические и психологические факторы, связанные с общими закономерностями общения и мышления, гибкостью и сложностью языкового поведения личности. Индивидуальное варьирование является результатом различий в языковой компетенции, причём эти различия обусловлены, в частности, интеллектуальным развитием человека.

Основным источником вариативности речевого сигнала на артикуляторно-акустическом уровне являются персональные особенности строения речевого аппарата человека. С акустической точки зрения голосовой тракт представляет собой неоднородную по своим характеристикам трубу от голосовой щели до губ. Персональные акустические характеристики голоса зависят от геометрических размеров и формы речевого тракта, а также от акустических параметров (упругость, эластичность) и динамики движения артикуляторных органов, в частности, голосовых связок.

Как следует из вышесказанного, решение задачи компьютерного клонирования речи личности в полном объёме должно проводиться с учётом всех перечисленных выше источников индивидуальной вариативности, для чего необходимы совместные усилия лингвистов, психологов и специалистов в обработке сигнала. Однако на современном этапе развития речевых технологий невозможно охватить все аспекты компьютерного клонирования речи личности. В данном разделе рассматриваются вопросы создания клона речи личности ограниченные следующими условиями:

- личность находится в нейтральном эмоционально-психологическом состоянии,
- удовлетворительном физическом состоянии,
- стиль речи ограничен условиями чтения вслух текста жанра «деловая проза».

При таких ограничениях индивидуальное варьирование будет проявляться, в основном, только на фонетическом, просодическом и акустическом уровнях. В результате исследования индивидуальных фонетических, просодических и акустических характеристик речи личности создаются индивидуализированные базы данных, содержащие эти характеристики. Далее используется такой метод синтеза речи по тексту, который в процессе генерации способен в полной мере

передавать указанные выше индивидуальные характеристики речи конкретной личности.

5.2. Основные этапы компьютерной технологии клонирования фонетико-акустических характеристик речи

К основным этапам реализации компьютерной технологии клонирования речи следует отнести следующие:

– выбор и подготовка текстовых и речевых корпусов для клонирования речи;

– выбор вариантов разметки речевого корпуса на фонетические и просодические сегменты для создания оптимального набора элементов компиляции;

– выбор методов и средств автоматической сегментации и маркировки речевого корпуса на фонетические и просодические сегменты различного уровня;

– выбор методов и средств автоматического создания БД элементов компиляции для синтеза речи конкретного диктора.

5.2.1. Выбор и подготовка текстовых и речевых корпусов

Формирование корпусов должно удовлетворять следующим основным требованиям:

- 1) результирующий корпус должен быть фонетически полным;
- 2) созданный корпус должен быть фонетически сбалансированным;
- 3) объём корпуса должен быть, по возможности, минимизирован.

Одновременное выполнение трёх указанных требований путём использования какого-либо общедоступного корпуса текстов и соответствующих ему речевых фонограмм практически невыполнимо. Нет никакой гарантии, что даже при чтении очень обширного текста в нём встретятся все основные варианты фонем – аллофоны и все варианты интонационных конструкций. Даже если такое случится, этот текст будет слишком длинным, для того чтобы он мог быть произнесён конкретным диктором. Опыт создания речевого корпуса для синтеза английской речи [178] показал, что для удовлетворения первых двух требований необходимо осуществить запись от 10 до 40 часов речи. При этом третье требование явно не удовлетворяется.

Хорошо известно, что обычный человек утомляется даже после 15-ти минут непрерывного чтения, а после 20 минут чтения его голос может вообще сорваться. Даже для профессионального диктора 45 минут непрерывного чтения с сохранением всего комплекса индивидуальных характеристик речи - довольно тяжёлая задача.

В связи со сказанным с целью одновременного выполнения трёх указанных выше требований для создания БД элементов компиляции

разработан минимальный по объёму корпус, содержащий два различных текста: «мини-текст», удовлетворяющий требованию фонетической полноты и «макси-текст», удовлетворяющий требованию фонетической сбалансированности.

Фонетический «мини-текст» создан путём специального подбора минимального количества слов, в которых реализуются все основные аллофоны, требуемые для синтеза речи. Слова мини-текста подобраны таким образом, что присутствующие в них фонемы имеют максимально возможный (в пределах заданных групп левых и правых контекстов) контраст спектральных характеристик. Такой выбор слов упрощает задачу анализа и сегментации речевого сигнала, созданного на основе «мини-текста», и способствует повышению точности сегментации.

Мини-текст состоит из 69 фраз, включающих 318 слов. Содержимое мини-текста приведено в Приложении 1.1.

В качестве «макси-текста» используются «тесты для оценки фразовой разборчивости речи» [173], фонетическая сбалансированность которых декларируется разработчиками ГОСТ 16600-72. Текст состоит из 500 фраз, включающих 1741 слово. Содержимое макси-текста приведено в Приложении 1.2, где позиция ударения помечена знаком «+». Статистический анализ фонетической структуры макси-текста приведен в Приложении 3.

Для создания набора интонационных портретов синтагм русской речи создан просодический «мини-текст» путём специального подбора минимального количества фраз, в которых реализуются основные интонационные конструкции (интономы) из числа требуемых для синтеза речи.

Просодический «мини-текст» состоит из 28-ми предложений, включающих 67 синтагм различного интонационного типа. Содержимое просодического мини-текста приведено в Приложении 1.3.

Кроме мини-текста, для создания расширенного набора интонационных портретов синтагм русской речи использован текстовый корпус (см. Приложение 1.4), включающий тексты различных жанров:

- 1) юмористический рассказ со множеством диалоговых включений,
- 2) сбалансированный текст для анализа просодических характеристик славяноворящих дикторов,
- 3) сбалансированный текст для анализа просодических характеристик неславяноворящих дикторов.

Последние два текста разработаны на филологическом факультете С-пГУ для целей тестирования индивидуальных языковых особенностей речи дикторов.

Создание речевого корпуса – запись естественной речи диктора при прочтении текстового корпуса – осуществляется в студийных условиях. В процессе записи должны соблюдаться следующие технические

требования: расстояние до микрофона в течение всего времени записи было постоянным, длительность сессии записи составляла не более 10 минут при среднем темпе речи дикторов. Кроме того, тексты, предназначенные для клонирования фонетико-акустических характеристик речи, следует читать с минимальными интонационными модификациями голоса.

5.2.2. Разметка речевого корпуса на фонетические и просодические сегменты

В основу классификации речевых сегментов положено понятие аллофона – позиционного и комбинаторного оттенка фонемы.

Аллофонная разметка речевого корпуса может осуществляться как «вручную», так и автоматически в процессе ДП-сопоставления синтезированного и естественного речевых сигналов (см. раздел 5.2.3). На основе аллофонной разметки создаются следующие группы фонетических сегментов: аллофоны – $\{S_a\}$, диаллофоны – $\{S_{da}\}$, аллослоги – $\{S_{as}\}$.

Разметка корпуса на диаллофоны осуществляется посинтагменно, т.е. создаваемые диаллофонные сегменты могут находиться как внутри фонетического слова, так и на границе фонетических слов. Согласно алгоритмам преобразования фонема-аллофон, описанным в разделе 4.3.2, внутрисловное и межсловное различие существенно для диаллофонов типа ГГ и СГ (где Г обозначает гласный, С - согласный). Это различие в местоположении диаллофонов однозначно определяется позиционными индексами аллофонов.

Особенности процедуры разметки речевого корпуса на слоговые комплексы, учитывающие фонетическую и артикуляторно-акустическую природу слога, описаны ранее в разделе 3.2, где дано определение слоговых комплексов трёх типов. Трудность вычленения в речевом потоке (при ручной или автоматической разметке корпуса) того или иного типа слоговых комплексов различна. Этот факт наглядно представлен на примере, приведенном в таблице 5.1. Из приводимого примера видно также, что одновременно с уменьшением трудности сегментации, возрастает количество фонетических элементов в слоге.

Разметка на слоговые комплексы проводится двумя способами: пословно и посинтагменно. В первом случае разметка осуществляется независимо для каждого отдельного фонетического слова, входящего в синтагму. Во втором случае последовательность слов в синтагме рассматривается как единый речевой поток с учётом межсловных фонетико-акустических явлений, исследованных в разделе 3.1. Очевидно, что поскольку на стыках слов могут встретиться любые сочетания фонем, невозможно создать речевой корпус разумного размера, в котором бы реализовались все сочетания. Целесообразно

поэтому при использовании ограниченного по объёму речевого корпуса воспользоваться обоими способами его разметки.

В общем случае каждая речевая синтагма может быть размечена на аллослоги различного типа шестью способами. Пример такого рода разметки синтагмы «Олимпийские чемпионы вернулись на родину» приведен в таблице 5.1, где границы аллослогов помечены значками «<>», «>».

Таблица 5.1

Пример разметки синтагмы на аллослоги

Вид разметки	Тип слогового комплекса	Размеченная на слоговые комплексы синтагма «Олимпийские чемпионы вернулись на родину»
Пословная фонетическая	Трудно сегментируемый (тип 1)	<A ₂₀₃ >, <L' ₀₀₂ I ₂₄₃ >, <M' ₀₀₃ P' ₀₀₁ I ₀₄₃ >, <J' ₀₁₃ S ₀₀₁ K' ₀₀₁ I ₃₄₃ >, <J' ₀₁₂ E ₃₄₃ >, <CH' ₀₀₁ E ₃₄₃ >, <M' ₀₀₃ P' ₀₀₁ I ₂₄₁ >, <O ₀₄₂ >, <N ₀₀₂ Y ₃₂₃ >, <V' ₀₁₂ E ₂₄₂ >, <R ₀₀₁ N ₀₀₂ U ₀₂₃ >, <L' ₀₀₂ I ₃₄₃ S' ₀₀₁ >, <N ₀₀₂ A ₂₂₂ >, <R ₀₀₂ O ₀₂₃ >, <D' ₀₀₂ I ₃₄₂ >, <N ₀₀₂ U ₃₂₀ >
	Умеренной трудности сегментации (тип 2)	<A ₂₀₃ L' ₀₀₂ I ₂₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₀₄₃ J' ₀₁₃ >, <S ₀₀₁ K' ₀₀₁ I ₃₄₃ J' ₀₁₂ E ₃₄₃ >, <CH' ₀₀₁ E ₃₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₂₄₁ O ₀₄₂ >, <N ₀₀₂ Y ₃₂₃ >, <V' ₀₁₂ E ₂₄₂ >, <R ₀₀₁ N ₀₀₂ U ₀₂₃ >, <L' ₀₀₂ I ₃₄₃ S' ₀₀₁ >, <N ₀₀₂ A ₂₂₂ >, <R ₀₀₂ O ₀₂₃ >, <D' ₀₀₂ I ₃₄₂ >, <N ₀₀₂ U ₃₂₀ >
	Сравнительно легко сегментируемый (тип 3)	<A ₂₀₃ L' ₀₀₂ I ₂₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₀₄₃ J' ₀₁₃ >, <S ₀₀₁ K' ₀₀₁ I ₃₄₃ J' ₀₁₂ E ₃₄₃ >, <CH' ₀₀₁ E ₃₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₂₄₁ O ₀₄₂ N ₀₀₂ Y ₃₂₃ >, <V' ₀₁₂ E ₂₄₂ >, <R ₀₀₁ N ₀₀₂ U ₀₂₃ L' ₀₀₂ I ₃₄₃ S' ₀₀₁ >, <N ₀₀₂ A ₂₂₂ >, <R ₀₀₂ O ₀₂₃ >, <D' ₀₀₂ I ₃₄₂ N ₀₀₂ U ₃₂₀ >
Посинтагменная фонетическая	Трудно сегментируемый (тип 1)	<A ₂₀₃ >, <L' ₀₀₂ I ₂₄₃ >, <M' ₀₀₃ P' ₀₀₁ I ₀₄₃ >, <J' ₀₁₃ S ₀₀₁ K' ₀₀₁ I ₃₄₃ >, <J' ₀₁₂ E ₃₄₃ >, <CH' ₀₀₁ E ₃₄₃ >, <M' ₀₀₃ P' ₀₀₁ I ₂₄₁ >, <O ₀₄₂ >, <N ₀₀₂ Y ₃₂₃ >, <V' ₀₁₂ E ₂₄₂ >, <R ₀₀₁ N ₀₀₂ U ₀₂₃ >, <L' ₀₀₂ I ₃₄₃ >, <S' ₀₀₁ N ₀₀₂ A ₂₂₂ >, <R ₀₀₂ O ₀₂₃ >, <D' ₀₀₂ I ₃₄₂ >, <N ₀₀₂ U ₃₂₀ >
	Умеренной трудности сегментации (тип 2)	<A ₂₀₃ L' ₀₀₂ I ₂₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₀₄₃ J' ₀₁₃ >, <S ₀₀₁ K' ₀₀₁ I ₃₄₃ J' ₀₁₂ E ₃₄₃ >, <CH' ₀₀₁ E ₃₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₂₄₁ O ₀₄₂ >, <N ₀₀₂ Y ₃₂₃ >, <V' ₀₁₂ E ₂₄₂ >, <R ₀₀₁ N ₀₀₂ U ₀₂₃ >, <L' ₀₀₂ I ₃₄₃ >, <S' ₀₀₁ N ₀₀₂ A ₂₂₂ >, <R ₀₀₂ O ₀₂₃ >, <D' ₀₀₂ I ₃₄₂ >, <N ₀₀₂ U ₃₂₀ >
	Легко сегментируемый (тип 3)	<A ₂₀₃ L' ₀₀₂ I ₂₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₀₄₃ J' ₀₁₃ >, <S ₀₀₁ K' ₀₀₁ I ₃₄₃ J' ₀₁₂ E ₃₄₃ >, <CH' ₀₀₁ E ₃₄₃ M' ₀₀₃ >, <P' ₀₀₁ I ₂₄₁ O ₀₄₂ N ₀₀₂ Y ₃₂₃ V' ₀₁₂ E ₂₄₂ >, <R ₀₀₁ N ₀₀₂ U ₀₂₃ L' ₀₀₂ I ₃₄₃ S' ₀₀₁ N ₀₀₂ A ₂₂₂ >, <R ₀₀₂ O ₀₂₃ >, <D' ₀₀₂ I ₃₄₂ N ₀₀₂ U ₃₂₀ >

просодическая	Разметка на АЕ	A ₂₀₃ L' ₀₀₂ I ₂₄₃ M' ₀₀₃ P' ₀₀₁ I ₀₄₃ J' ₀₁₃ S ₀₀₁ K' ₀₀₁ I ₃₄₃ J' ₀₁₂ E ₃₄₃ CH' ₀₀₁ E ₃₄₃ M' ₀₀₃ P' ₀₀₁ I ₂₄₁ O ₀₄₂ N ₀₀₂ Y ₃₂₃ V' ₀₁₂ E ₂₄₂ R ₀₀₁ N ₀₀₂ U ₀₂₃ L' ₀₀₂ I ₃₄₃ S' ₀₀₁ N ₀₀₂ A ₂₂₂ R ₀₀₂ O ₀₂₃ D' ₀₀₂ I ₃₄₂ N ₀₀₂ U ₃₂₀
	Разметка на ЭАЕ	/A ₂₀₃ L' ₀₀₂ I ₂₄₃ M' ₀₀₃ P' ₀₀₁ /I ₀₄₃ / J' ₀₁₃ S ₀₀₁ K' ₀₀₁ I ₃₄₃ J' ₀₁₂ E ₃₄₃ / /CH' ₀₀₁ E ₃₄₃ M' ₀₀₃ P' ₀₀₁ I ₂₄₁ /O ₀₄₂ / N ₀₀₂ Y ₃₂₃ / /V' ₀₁₂ E ₂₄₂ R ₀₀₁ N ₀₀₂ /U ₀₂₃ / L' ₀₀₂ I ₃₄₃ S' ₀₀₁ / /N ₀₀₂ A ₂₂₂ R ₀₀₂ /O ₀₂₃ /D' ₀₀₂ I ₃₄₂ N ₀₀₂ U ₃₂₀ /

При создании БД просодических элементов компиляции аллофонные сегменты группируются в сегменты более высокого уровня, чем диаллофоны и аллослоги, а именно: в сегменты, соответствующие АЕ – акцентным единицам и ЭАЕ – элементам АЕ. Правила объединения аллофонов в АЕ и ЭАЕ описаны в разделе 4.4. Примеры разметки синтагмы на просодические сегменты АЕ (ограничены знаками - |) и ЭАЕ (ограничены знаками - /) представлены в двух нижних строках таблицы 5.1.

5.2.3. Автоматическая сегментация и маркировка речевого сигнала

В разделе 2.5 проанализированы существующие методы автоматической сегментации и маркировки речевого сигнала и обоснован выбор метода анализа через синтез с использованием математического аппарата динамического программирования (ДП-метод). В основу ДП-метода положена процедура динамического сопоставления синтезированного и естественного сигналов [146, 147]. Потенциально процедура динамического сопоставления обеспечивает прецизионную разметку речевого корпуса, не требует предварительного обучения системы сегментации и, кроме того, является в значительной степени дикторонезависимой.

Основная идея автоматизации процессов сегментации и аллофонной маркировки заключается в реализации алгоритмов переноса меток начала и конца аллофонов с синтезированного сигнала на естественный речевой сигнал, произнесенный голосом клонируемого диктора. Для автоматического переноса меток выбирается один из синтезированных голосов, наиболее близкий к клонируемому голосу.

Динамическое сопоставление (ДП-сопоставление) синтезированного и естественного сигналов осуществляется путём вычисления матрицы интегральных расстояний между векторами спектральных признаков сигналов по рекуррентной формуле

$$D(n + 1, m + 1) = \max (D(n + 1, m), D(n, m + 1), D(n, m) + \delta_A(n, m)) \quad (5.1)$$

где $\delta_A(n, m)$ – локальные расстояния между векторами спектральных признаков синтезированного и естественного сигналов.

Начальные условия для вычисления матрицы интегральных расстояний следующие: $D(n, 0) = 0$; $D(0, m) = 0$.

Значения матрицы локальных расстояний $d_A(n,m)$ вычисляются по формуле

$$d_A(n,m) = d(S(n), E(m)) = \frac{1}{K} \sum_{k=1}^K |S(n,k) - E(m,k)| \quad (5.2)$$

где $S(n)$ – вектор спектральных признаков синтезированного сигнала в точке n , $E(m)$ – вектор спектральных признаков естественного сигнала в точке m , K – размерность вектора спектральных признаков.

Соответствие между синтезированным и естественным сигналами находится путём отображения оптимального пути на вычисленной матрице интегральных расстояний D . Пример переноса меток границ аллофонов для слова «абракадабра» с синтезированного речевого сигнала (вертикальная ось) на естественный (горизонтальная ось) с использованием найденного оптимального пути соответствия показан на рис. 5.2.

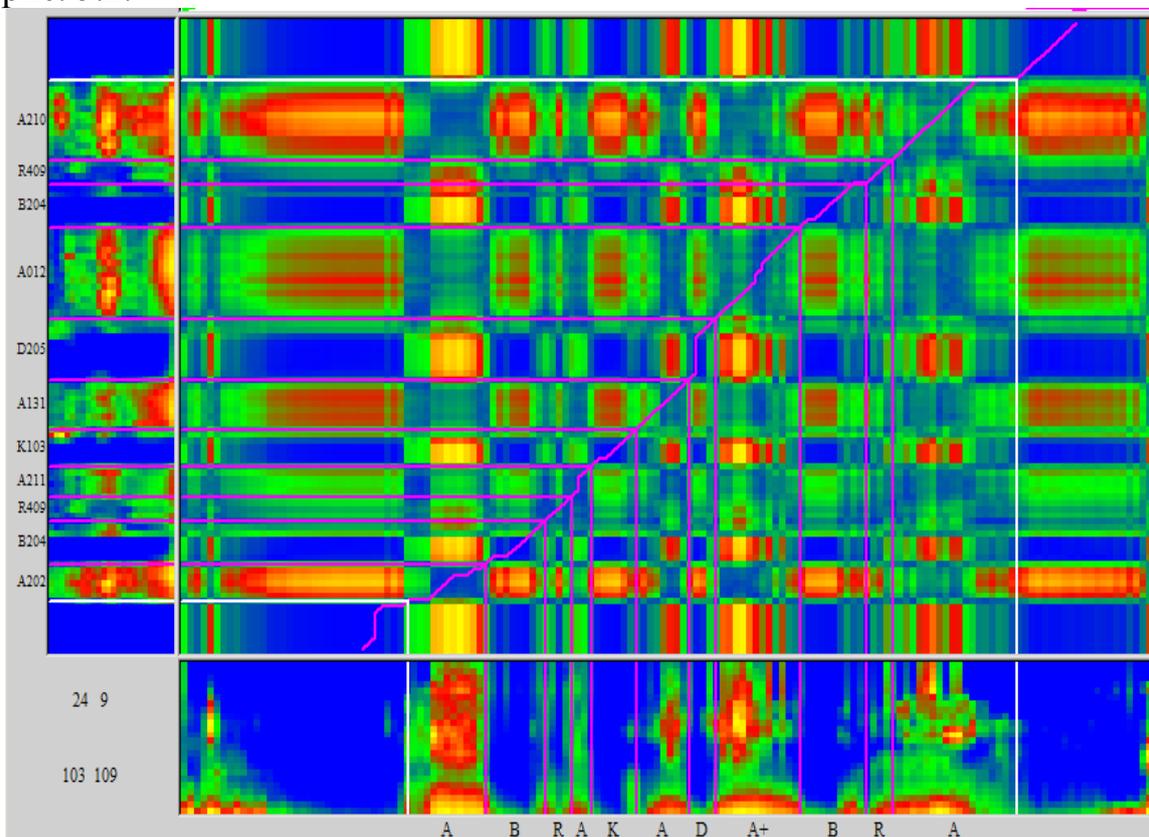


Рис. 5.2. Иллюстрация оптимального пути соответствия между синтезированным и естественным сигналами

На рис. 5.3 приведена структурная схема системы сегментации и аллофонной маркировки. В состав системы входят блоки, выполняющие следующие функции:

- преобразование исходного орфографического текста (эталонный набор русских слов для клонирования) в аллофонный текст;
- синтез речевого сигнала (РС) и его аллофонную разметку;

- выделение спектральных признаков синтезированного и естественного речевых сигналов;
- автоматический перенос меток аллофонов с синтезированных спектральных параметров на естественный речевой сигнал и автоматическую маркировку аллофонных сигналов.

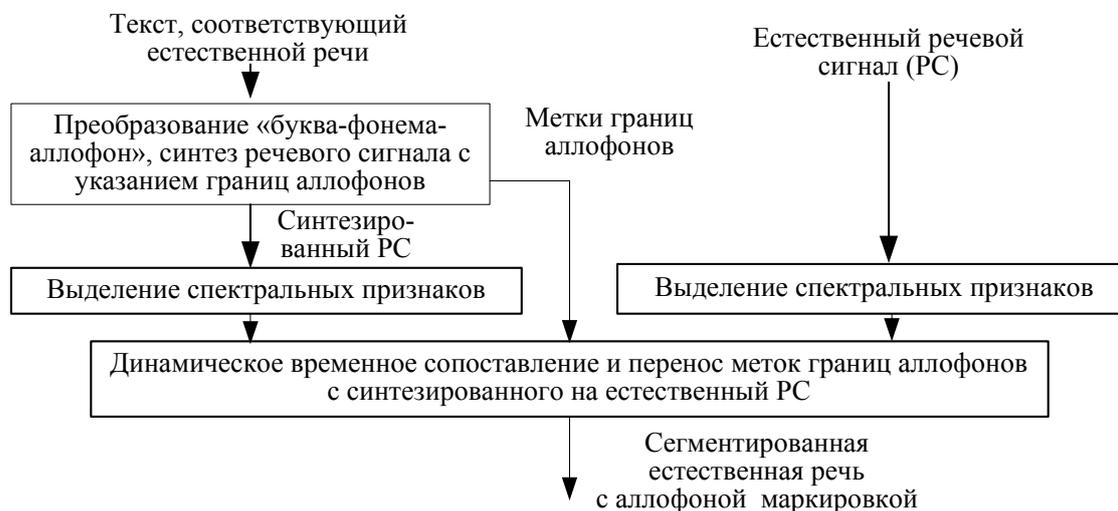


Рис. 5.3. Структурная схема системы сегментации и аллофонной маркировки

Исходя из самого определения ДП-метода автоматической сегментации речевого сигнала, как метода анализа через синтез, вытекает, что уже в самом начале необходимо иметь, по крайней мере, одну готовую мини-БД аллофонов для реализации процедуры синтеза речи по тексту. Если ранее мини-БД аллофонов не была сформирована ни для одного из голосов, то сегментацию и аллофонную разметку мини-корпуса необходимо осуществить «вручную». Процедура «ручной» сегментации более 500 аллофонов для создания мини-БД достаточно трудоёмкая и требует определённых навыков. Этот недостаток ДП-метода с лихвой окупается, однако, указанными ранее его преимуществами перед другими методами.

Общая схема процедуры создания мини-БД аллофонов и макси-БД мультифонов, включающей систему автоматической сегментации речевого сигнала ДП-методом, представлена на рис. 5.4.

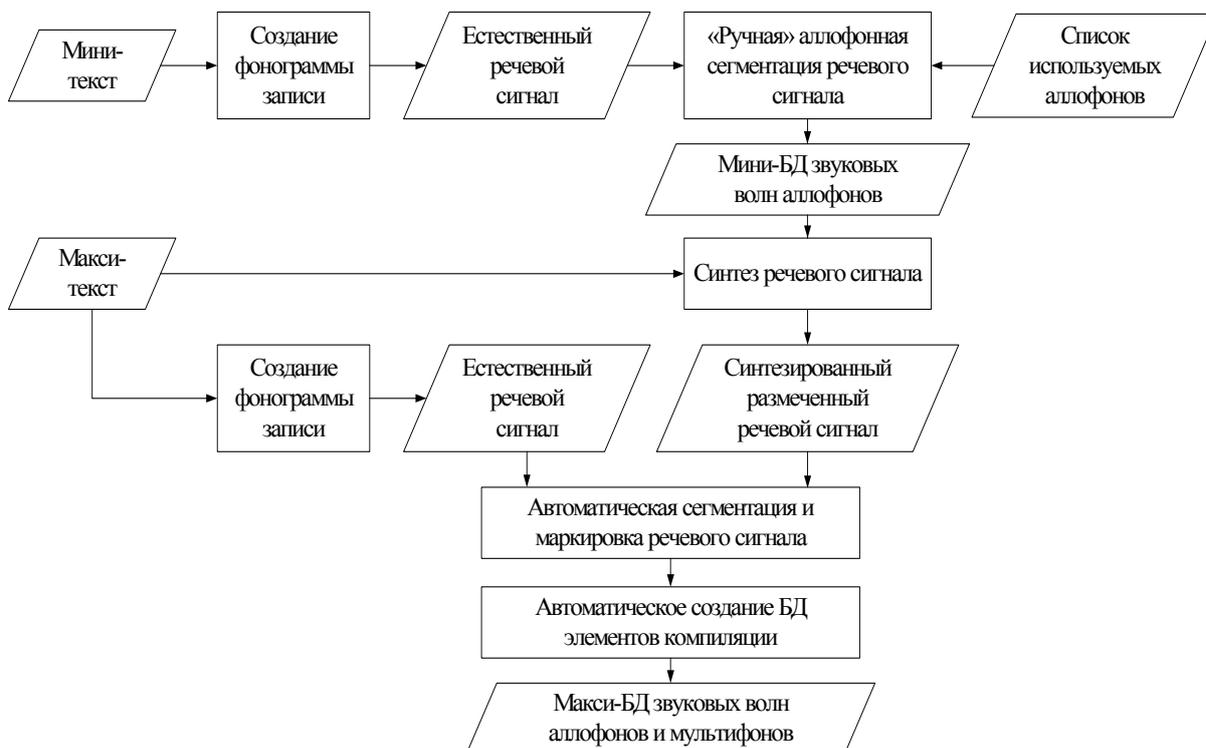


Рис. 5.4. Процедура создания мини- и макси-БД звуковых волн аллофонов и мультифонов

На основе мини-текста осуществляется создание фонограммы записи голоса одного из дикторов. Полученный естественный речевой сигнал анализируется и сегментируется опытным экспертом-фонетистом, в результате чего создаётся мини-БД, содержащая полный набор звуковых волн аллофонов.

Фонограмма записей макси-текста, прочитанного тем же либо новым диктором, сегментируется автоматически с использованием описанного выше ДП-метода «анализа через синтез». При этом для синтеза размеченного речевого сигнала используется созданная ранее мини-БД звуковых волн аллофонов. Аллофонно-размеченный естественный речевой сигнал поступает в блок автоматического создания БД элементов компиляции, осуществляющий выбор фонетических сегментов различного уровня, их анализ и обработку. Результаты обработки помещаются в макси-БД звуковых волн аллофонов и мультифонов [179, 180, 181].

5.2.4 Создание оптимального набора элементов компиляции

В результате работы автоматической системы сегментации и маркировки речевого корпуса генерируются множественные наборы описанных выше фонетических сегментов – аллофонов и мультифонов. Для создания БД элементов компиляции необходимо детально проанализировать полученные наборы для того, чтобы:

– исключить фонетические сегменты, в которых допущены ошибки при прочтении диктором или при автоматической сегментации, и отобрать только лучшие из них (операция «отсекающий отбор»);

– при наличии множественной реализации отобранных сегментов выбрать наилучшие из них (операция «селекция»);

– по определённым критериям провести оценку качества каждого из отобранных сегментов и отметить отклонения от нормы (операция «диагностика»);

– провести, по возможности, корректировку параметров сегментов с замеченными отклонениями от нормы (операция «коррекция»);

– проанализировать состав полученных элементов компиляции и при необходимости создать на их основе недостающие путём их видоизменения (операция «размножение»).

Операция «отсекающий отбор» осуществляется путём сопоставления акустических и временных характеристик естественного и синтезированного речевых сегментов. Если различия между ними будут выше некоторой пороговой величины, это означает, что такой сегмент не сможет обеспечить даже минимально необходимого качества синтезированной речи, достигаемого при использовании только мининабора аллофонов, и должен быть исключён.

Операция «селекция» осуществляется путём выбора наилучшего, по определённому критерию, экземпляра сегмента в случае его множественной реализации. В качестве такого критерия используется расстояние между значениями просодических параметров каждого из экземпляров и медианными значениями просодических параметров в полученной выборке.

Операция «диагностика» осуществляется над полученными в соответствии с указанными выше критериями сегментами. Целью данной операции является выявление возможных незначительных неточностей сегментации.

Найденные сегменты с неточно определёнными границами подвергаются операции «коррекция» путём проведения специальных алгоритмических процедур по удалению некорректных и установке недостающих граничных периодов основного тона.

Операция «размножение» осуществляется над мультифонными сегментами, помещёнными в БД элементов компиляции. Существующий в БД сегмент может дублироваться под другим именем, если по своим характеристикам он способен заменить отсутствующий в созданной БД сегмент.

Алгоритмическая реализация перечисленных выше операций в системе автоматического создания БД элементов компиляции рассматривается в следующем разделе.

5.3. Алгоритмы автоматического создания БД фонетических элементов компиляции

Как уже было отмечено в разделе 5.2.4, процесс создания БД элементов компиляции реализуется с использованием следующих процедур: «отсекающий отбор», «селекцию», «диагностику» и «коррекцию» фонетических сегментов, а также «размножение» элементов компиляции.

5.3.1 Алгоритмы «отсекающего отбора» речевых сегментов

Задача отбора речевых сегментов возникает по следующим причинам:

а) редукции в естественной речи аллофонов в некоторых позициях вплоть до полного исчезновения, в результате чего фонетическое содержимое синтезированной и естественной речевых синтагм не совпадают;

б) неточности разметки естественного сигнала, в результате чего из естественной речи «вычленяется» сегмент, акустическое содержимое которого не соответствует фонетическому содержимому.

С учётом причин возможного некорректного вычленения сегментов алгоритм «отсекающего отбора» основан на вычислении степени сходства временных и акустических характеристик синтезированного и естественного сегментов.

Пусть n – количество аллофонов в сегменте, m – количество экземпляров естественного сегмента. Для определения степени сходства временных характеристик синтезированного и j -го экземпляра естественного сегментов (где $1 \leq j \leq m$) для $\forall i, 1 \leq i \leq n$, вычисляются значения длительностей T_i^S i -го аллофона синтезированного сегмента, а также для $\forall i, 1 \leq i \leq n, \forall j, 1 \leq j \leq m$, вычисляются значения длительностей T_i^j i -го аллофона j -го экземпляра сегмента. Для определения степени сходства акустических характеристик синтезированного и j -го экземпляра естественного сегментов (где $1 \leq j \leq m$) для $\forall i, 1 \leq i \leq n, \forall j, 1 \leq j \leq m$, вычисляются акустические расстояния D_i^j между i -м аллофоном синтезированного сегмента и j -го экземпляра естественного сегмента. Расстояние D_i^j определяется как нормированная в диапазоне 0 – 1 сумма локальных расстояний $d_A(n, m)$, расположенных на траектории оптимального соответствия синтезированного и j -го естественного сигналов, полученной в процессе ДП-сопоставления (см. раздел 5.2.3).

Степень сходства временных характеристик L_T^j и степень сходства акустических характеристик L_D^j j -го естественного и синтезированного сегментов вычисляются как функции принадлежности к эталонной точке метрического пространства, где в качестве эталонной реализации

используется вектор соответствующих значений синтезированного сегмента. Для вычисления расстояния между векторами используется метрика l_∞ [182]:

$$L_T^j = \min_{i=1..n} \frac{\min(T_i^j, T_i^S)}{\max(T_i^j, T_i^S)} \quad (5.3)$$

$$L_D^j = 1 - \max_{i=1..n} D_i^j \quad (5.4)$$

В результате операции «отсекающий отбор» из дальнейшей обработки исключаются все экземпляры p сегментов, для которых $L_T^p \leq \beta_1$ или $L_D^p \leq \beta_2$. Здесь β_1, β_2 – пороги сходства соответственно временных и акустических характеристик синтезированного и естественного сигналов. Как показали эксперименты, оптимальные результаты операции «отсекающий отбор» достигаются при $\beta_1 = 0,5$ и $\beta_2 = 0,6$.

5.3.2. «Селекция» речевых сегментов

Количество экземпляров одного и того же фонетического сегмента, прошедших операцию «отсекающий отбор», зависит от исходного количества таких экземпляров в корпусе, от качества их произношения диктором и от точности аллофонной разметки. В результате «отсекающего отбора» некоторые экземпляры сегментов могут быть исключены из дальнейшей обработки. Для оставшихся экземпляров выполняется операция «селекция», в результате которой выбирается один, наилучший представитель каждого сегмента.

Фонетико-акустические характеристики каждого экземпляра сегмента, прошедшего операцию «отсекающий отбор», должны быть приемлемыми для помещения данного экземпляра в создаваемую БД элементов компиляции. Учитывая, что созданные элементы будут подвергаться просодической модификации в процессе синтеза речи, целесообразно выбрать экземпляр, наиболее типичный по значениям просодических характеристик: частоты основного тона F_0 , амплитуды A и длительности T . В качестве такого экземпляра в процессе операции «селекция» выбирается сегмент, имеющий просодические характеристики, наиболее близкие к медианным.

Пусть n – количество аллофонов в сегменте, m – количество экземпляров сегмента. Тогда для $\forall i, 1 \leq i \leq n, \forall j, 1 \leq j \leq m$, вычисляются значения длительности T_i^j , средние значения амплитуды $A_{ave\ i}^j$ и средние значения частоты основного тона $F_{0\ ave\ i}^j$ i -го аллофона j -го экземпляра сегмента. Для невокализованных аллофонов значение $F_{0\ ave\ i}^j$ принимается равным 0. Формируется вектор медианных значений

длительностей T_i^M , амплитуд $A_{ave\ i}^M$, частот основного тона $F_{0\ ave\ i}^M$ аллофонов сегмента. Размерность вектора равна $3*n$.

Нормированное в диапазоне 0–1 расстояние D_M^j между вектором значений просодических характеристик j -го экземпляра сегмента и вектором медианных значений вычисляется по метрике l_1 [182] в соответствии с формулой:

$$D_M^j = \frac{1}{3} \sum_{i=1}^n \left(\frac{|T_i^j - T_i^M|}{\max_{j=1..m} |T_i^j - T_i^M|} + \frac{|A_{ave\ i}^j - A_{ave\ i}^M|}{\max_{j=1..m} |A_{ave\ i}^j - A_{ave\ i}^M|} + \frac{|F_{0\ ave\ i}^j - F_{0\ ave\ i}^M|}{\max_{j=1..m} |F_{0\ ave\ i}^j - F_{0\ ave\ i}^M|} \right) \quad (5.5)$$

В результате операции «селекция» выбирается экземпляр k сегмента, удовлетворяющий условию:

$$k = \arg \min_{j=1..m} (D_i^M) \quad (5.6)$$

5.3.3 «Диагностика» и «коррекция» сегментов

Использование создаваемых сегментов в качестве элементов компиляции при персонализированном синтезе речи по тексту накладывает требование особой точности, прецизионности разметки. Целью операции «диагностика» является тестирование полученных сегментов на наличие некорректно определённых границ. Обнаруженные неточности разметки «корректируются» путём удаления граничных периодов основного тона и добавления необходимых периодов.

Операция «диагностика» осуществляется путём попарного определения степени сходства временных и акустических характеристик сигнала на первом и втором периодах основного тона сегмента, а также на предпоследнем и последнем периодах. «Диагностика» осуществляется только в случае, когда граничные периоды принадлежат вокализованным аллофонам.

Степень сходства акустических характеристик L_D и степень сходства временных характеристик L_T вычисляются как функции принадлежности к эталонной точке метрического пространства, для чего соответствующие значения на пред-граничных периодах принимаются за эталонные. Для определения расстояния между двумя точками по акустическим характеристикам используется метрика l_1 , расстояние по временным характеристикам определяется как отношение длительностей периодов:

$$L_D = 1 - \frac{\sum_{i=1}^N |s_i^u - s_i^p|}{\sum_{i=1}^N (|s_i^u| + |s_i^p|)} \quad (5.7)$$

$$L_D = 1 - \frac{\min(T^u, T^p)}{\max(T^u, T^p)} \quad (5.8)$$

где s_i^u – значение сигнала на i -том отсчёте граничного периода, s_i^p – значение сигнала на i -том отсчёте пред-граничного периода, T^u , T^p – длительности, соответственно, граничного и пред-граничного периодов.

Суммарная степень сходства периодов вычисляется как

$$L = \alpha L_D + (1 - \alpha)L_T \quad (5.9)$$

где α – весовой коэффициент, $0 \leq \alpha \leq 1$.

Левая или правая граница сегмента считается некорректно определённой, если значение степени сходства, соответственно, первого и второго или последнего и предпоследнего периодов $L \leq \beta$, где β – порог сходства периодов. Значения α и β рассчитаны экспериментально и равны, соответственно, 0,4 и 0,7. Область значений L_D и L_T , при которых границы сегмента считаются корректно определёнными, в соответствии с формулой 5.9 и значениями $\alpha = 0,4$ и $\beta = 0,7$, показана на рис. 5.5.

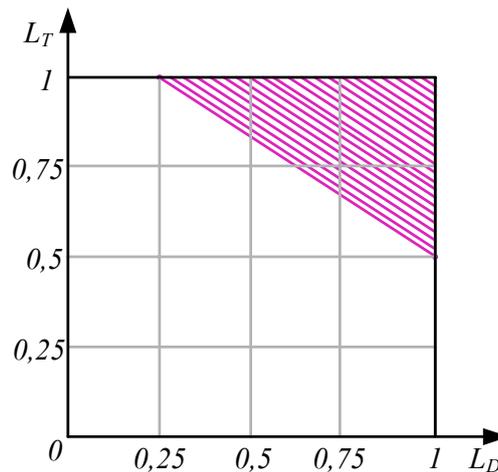


Рис. 5.5. Область допустимых значений степеней сходства акустических L_D и временных L_T характеристик периодов при весовом коэффициенте $\alpha = 0,4$ и пороге сходства $\beta = 0,7$

«Коррекция» осуществляется в случае, когда «неточный» период не является единственным периодом аллофона, и заключается в удалении граничного периода и дублировании пред-граничного. В результате количество периодов основного тона сегмента не изменяется.

Сегменты, прошедшие операции «отсекающий отбор», «селекция», «диагностика» и «коррекция», помещаются в БД элементов компиляции.

5.3.4 «Размножение» элементов компиляции

При «размножении» элементов компиляции учитывается известный факт [183] идентичности слухового восприятия аллофонов некоторых гласных русской речи в безударных позициях, а именно:

– *E* и *Y* первой степени редукции после твёрдых согласных не перед паузой;

– *A*, *E* и *Y* второй степени редукции после твёрдых согласных не перед паузой;

– *A* и *E* первой степени редукции после мягких согласных не перед паузой;

– *A*, *E* и *I* второй степени редукции после мягких согласных не перед паузой.

Таким образом, отсутствующий в БД мультифон может быть заменён другим, содержащим гласные, идентичные по слуховому восприятию с требуемыми для синтеза.

С учётом используемых по разработанной классификации аллофонов гласных первой и второй степеней редукции, описанных в Приложении 2, а также индексов аллофонов гласных, описанных в разделе 4.3, правила замены аллофонов при «размножении» элементов компиляции следующие:

$E_{2jk} \rightarrow Y_{21k}, j = 1, 2; k = 1, 2, 3;$ $E_{3jk} \rightarrow Y_{3jk}, j = 1, 2; k = 1, 2, 3;$

$Y_{2jk} \rightarrow E_{21k}, j = 1, 2; k = 1, 2, 3;$ $Y_{3jk} \rightarrow A_{3jk}, j = 1, 2; k = 1, 2, 3;$

$A_{24k} \rightarrow E_{24k}, k = 1, 2, 3;$ $Y_{3jk} \rightarrow E_{3jk}, j = 1, 2; k = 1, 2, 3;$

$E_{24k} \rightarrow A_{24k}, k = 1, 2, 3;$ $A_{34k} \rightarrow I_{34k}, k = 1, 2, 3;$

$A_{3jk} \rightarrow E_{3jk}, j = 1, 2, 3, 4; k = 1, 2, 3;$ $E_{34k} \rightarrow I_{34k}, k = 1, 2, 3;$

$A_{3jk} \rightarrow Y_{3jk}, j = 1, 2; k = 1, 2, 3;$ $I_{34k} \rightarrow A_{34k}, k = 1, 2, 3;$

$E_{3jk} \rightarrow A_{3jk}, j = 1, 2, 3, 4; k = 1, 2, 3;$ $I_{34k} \rightarrow E_{34k}, k = 1, 2, 3.$

Операции «размножение» подвергаются диаллофоны, а также внутрисловные и внутрисинтагменные слоговые комплексы первого типа.

Необходимо отметить, что при замене аллофонов могут не учитываться фонетические особенности различных диалектов, а также индивидуальные фонетико-акустические особенности произношения. Таким образом, создаются элементы компиляции, которые, возможно, не в полной мере отражают индивидуальные фонетико-акустические особенности голоса диктора, что влечёт в результате некоторое уменьшение естественности синтезируемой речи. Но при операции дублирования количество элементов компиляции в БД увеличивается, таким образом, при синтезе речи вместо отдельных аллофонов используются мультифоны, что влечёт меньшие искажения сигнала в

процессе просодической модификации и, как следствие, повышение разборчивости синтезируемой речи.

Процесс «размножения» сегментов может осуществляться статически – путём добавления сегментов в фонетико-акустическую БД, и динамически – при поиске речевого сегмента в процессе синтеза речи. Во втором случае, если требуемый сегмент отсутствует в БД, формируется список сегментов, которые могут использоваться вместо требуемого, и происходит поиск одного из них.

Статическое размножение сегментов влечёт увеличение объёма фонетико-акустической БД при практически неизменном содержании, но при этом, в отличие от динамического размножения, не увеличивается время поиска сегментов в процессе синтеза речи. Таким образом, в прикладных программах, где критическим ресурсом является объём памяти, целесообразней использовать динамическое размножение, в то время как в прикладных программах, критичных к временному ресурсу, необходимо использовать статическое размножение.

5.4. Компьютерная система клонирования фонетико-акустических характеристик речи – «ФоноКлонатор»

5.4.1. Функциональная схема системы

Функциональная схема, входные и выходные данные, взаимодействие блоков системы представлены на рис 5.6. В системе реализованы описанные выше этапы создания БД элементов компиляции.

Входные данные системы:

– предварительно обработанная фонограмма записи – набор речевых синтагм, каждая из которых хранится в виде оцифрованной звуковой волны в отдельном файле в формате *WAVE PCM*;

– предварительно обработанная стенограмма записи – текстовый файл, содержащий пометы границ синтагм;

– базовая БД звуковых волн аллофонов. В качестве такой БД используется БД элементов компиляции синтезатора, созданная «вручную» или автоматически на основе записей голоса одного из дикторов.

Выходные данные системы:

- БД аллофонов и мультифонов голоса «клонлируемого» диктора.

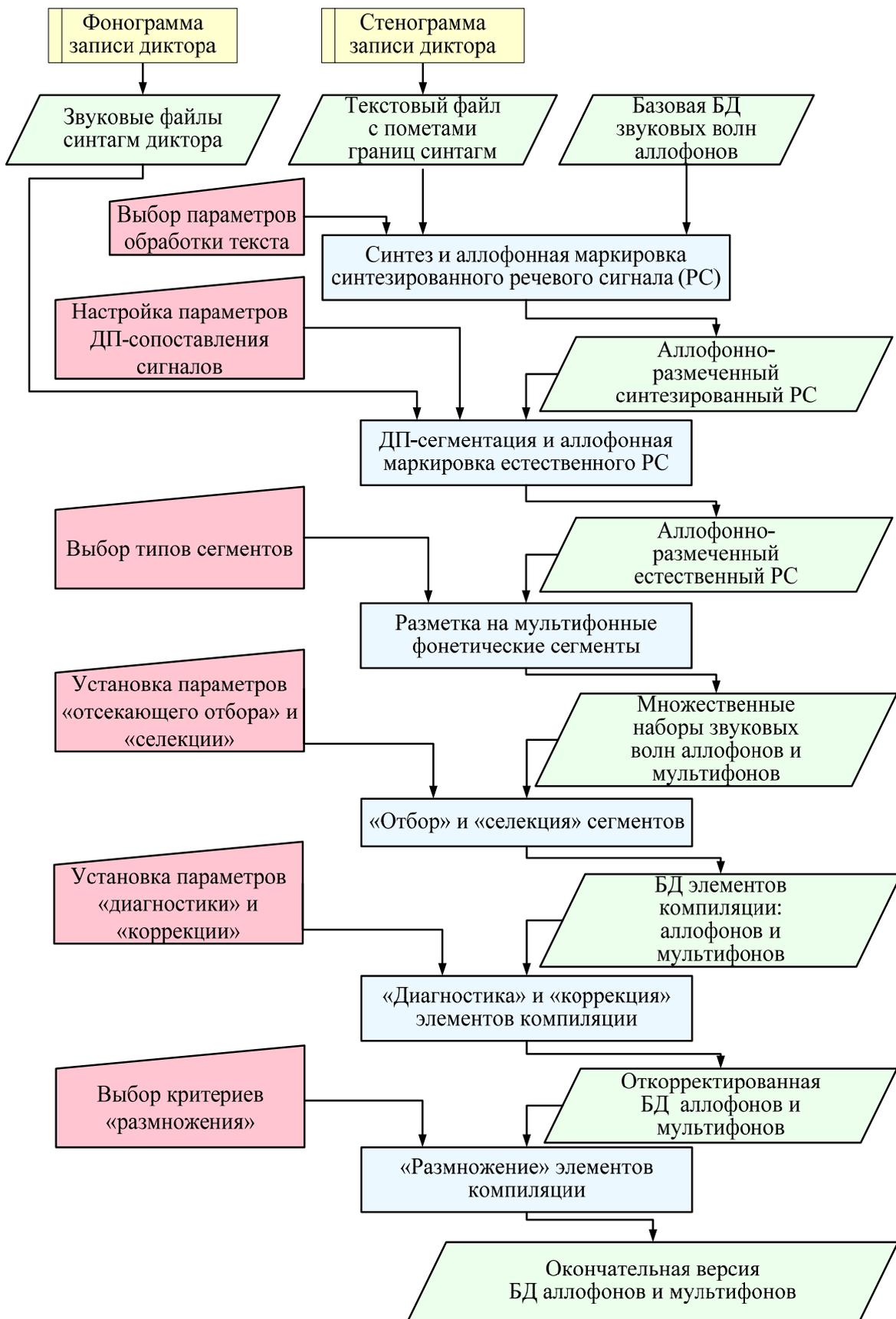


Рис. 5.6. Общая функциональная схема автоматической системы клонирования фонетико-акустических характеристик речи

На первом этапе синтеза осуществляется фонетическая обработка текста, включающая расстановку словесных ударений, преобразования «буква-фонема» и «фонема-аллофон». Результат обработки – последовательность аллофонов – передается на второй этап, где происходит выбор звуковых волн аллофонов из БД, их компиляция и аллофонная маркировка.

Настройка параметров обработки текста включает выбор используемого словаря с пометами позиции ударения, а также указание индикаторов границ фонетических слов и синтагм в тексте.

Каждая пара синтагм: «аллофонно-размеченный синтезированный сигнал – естественный сигнал», поступает в блок ДП-сегментации и аллофонной маркировки естественного РС, где осуществляется анализ спектральных признаков сигналов, их ДП-сопоставление и перенос маркеров границ аллофонов с синтезированного на естественный РС. В системе реализована настройка параметров вычисления спектральных признаков и параметров ДП-сопоставления.

Аллофонно-размеченный естественный РС поступает в блок разметки на фонетические сегменты. Пользователь системы может выбрать типы получаемых сегментов, среди которых аллофоны различного типа: ударные гласные, гласные первой степени редукции, гласные второй степени редукции, согласные; диаллофоны типов ГГ, СГ, СС, ГС, а также внутрисловные и внутрисинтагменные аллослоги первого, второго и третьего типов.

Результат работы данного блока – множественные наборы звуковых волн сегментов указанных типов – подвергается операциям «отсекающий отбор» и «селекция». На этом этапе обработки пользователь может указать параметры «отсекающего отбора»: пороги сходства синтезированного и естественного сегментов по временным и акустическим характеристикам. В результате операций «отсекающий отбор» и «селекция» создается первая версия БД аллофонов и мультифонов, содержащая по одному экземпляру элементов компиляции для каждого фонетического сегмента.

При осуществлении следующего этапа обработки – «диагностики» и «коррекции» – пользователь системы может изменить значения весовых коэффициентов акустических и временных характеристик, а также порог сходства периодов основного тона. Откорректированная БД аллофонов и мультифонов поступает в блок «размножения». Настройка параметров на этом этапе включает выбор критериев размножения: типы «размножаемых» мультифонов, характеристики «заменяемых» аллофонов гласных. Результатом работы системы является окончательная версия БД аллофонов и мультифонов.

Промежуточные данные, получаемые в результате работы каждого из блоков, могут быть сохранены для дополнительного анализа и коррекции опытным экспертом-фонетистом.

Все блоки системы работают независимо и могут использовать как данные, полученные в текущем сеансе в результате работы предыдущих блоков, так и предварительно сохранённые промежуточные данные.

Алгоритмы работы основных блоков системы «ФоноКлонатор» соответствуют положениям, рассмотренным в разделах 5.2, 5.3.

5.4.2. Пользовательский интерфейс системы «ФоноКлонатор»

Пользовательский интерфейс системы «ФоноКлонатор» (рис. 5.7) включает следующие основные блоки:

- блок окон, в котором отображается список синтагм (орфографический текст), последовательность аллофонов, соответствующих каждой синтагме, и последовательность фонетических сегментов заданных типов;
- окна отображения осциллограмм естественного и синтезированного речевых сигналов;
- окно просмотра дерева сегментов с указанием синтагмы, из которой будет выделен соответствующий сегмент и помещен в БД.
- диалоговые окна настроек параметров системы.

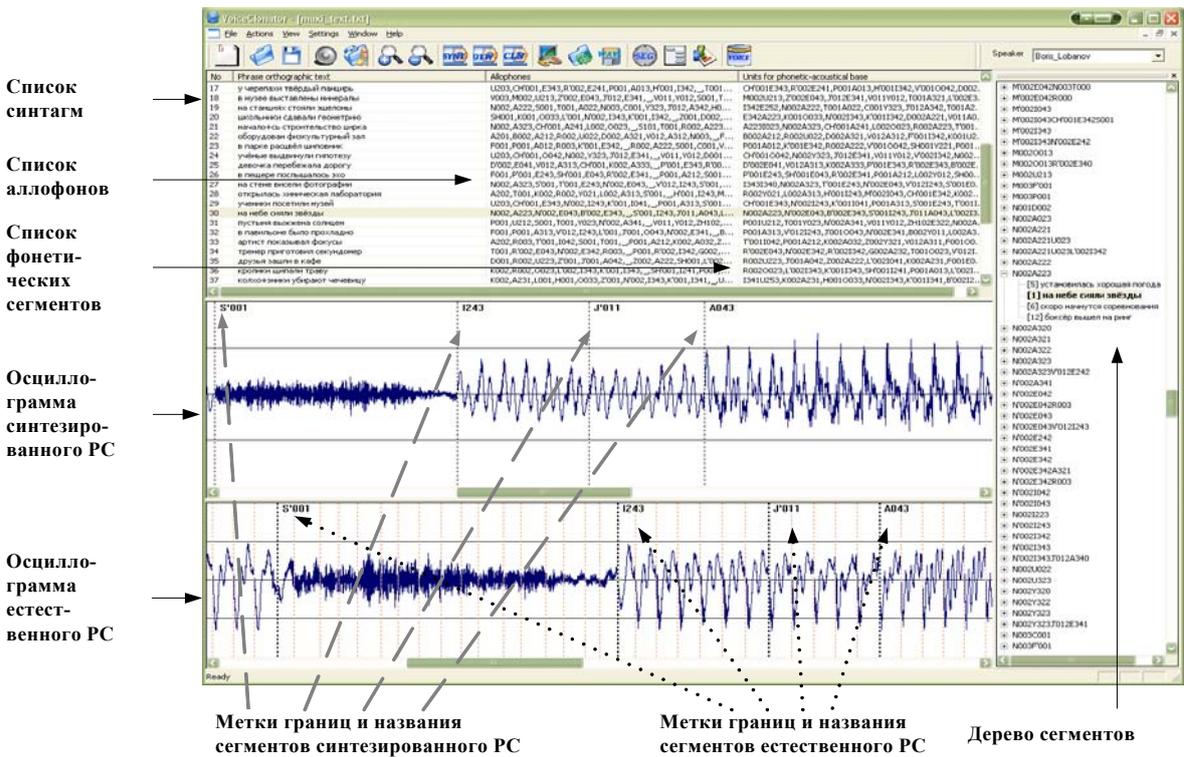


Рис.5.7. Общий вид и основные блоки пользовательского интерфейса системы

На осциллограммах синтезированного и естественного сигналов (рис. 5.8) указаны метки границ сегментов и названия сегментов. На осциллограмме естественного РС указаны, кроме того, границы периодов основного тона (питчи). В системе реализовано масштабирование отображения и прослушивание выделенного фрагмента синтезированного либо естественного сигналов.

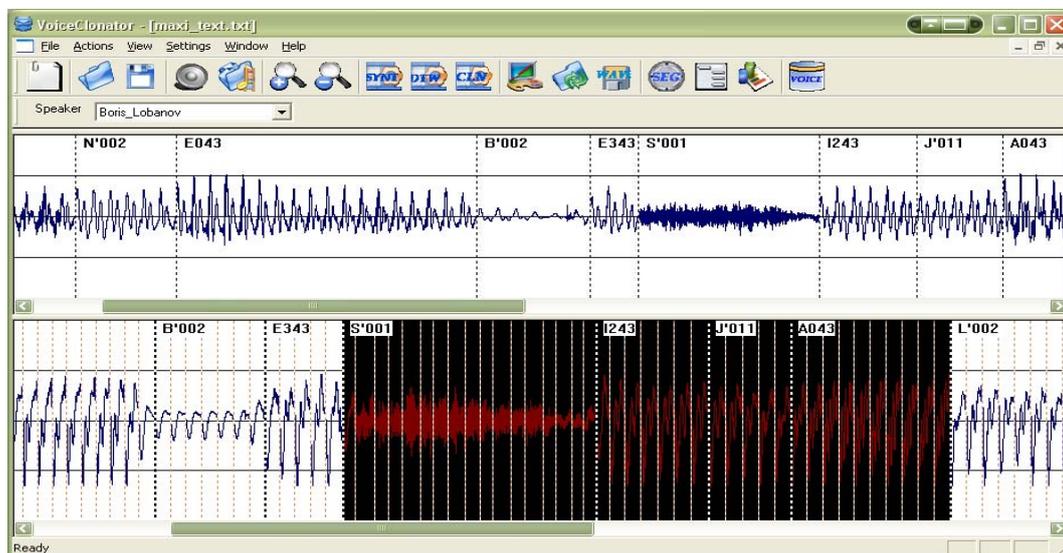


Рис. 5.8. Отображение осциллограмм речевых сигналов, меток границ сегментов и названий сегментов

Дерево сегментов (рис.5.9) формируется на основании текстового содержимого списка синтагм и выбранных пользователем типов сегментов. Верхний уровень дерева – это список указанных фонетических сегментов, содержащихся в тексте. Элементы второго уровня – синтагмы, в которых содержится данный сегмент. Для каждой синтагмы указана позиция начала сегмента, длительность сегмента, степень сходства (в %) с синтезированным сегментом, а также расстояние от данного сегмента до медианного. Синтагма, содержащая сегмент, выбранный в результате операции «селекция», выделяется жирным шрифтом.

Настройки параметров системы реализованы в следующих диалоговых окнах:

- диалог настроек синтеза речевого сигнала;
- диалог настроек сегментации и аллофонной маркировки естественного РС,
- диалог выбора типов сегментов для создаваемой БД;
- диалоги установок параметров «отсекающего отбора», «диагностики», «коррекции» и выбора критериев «размножения».

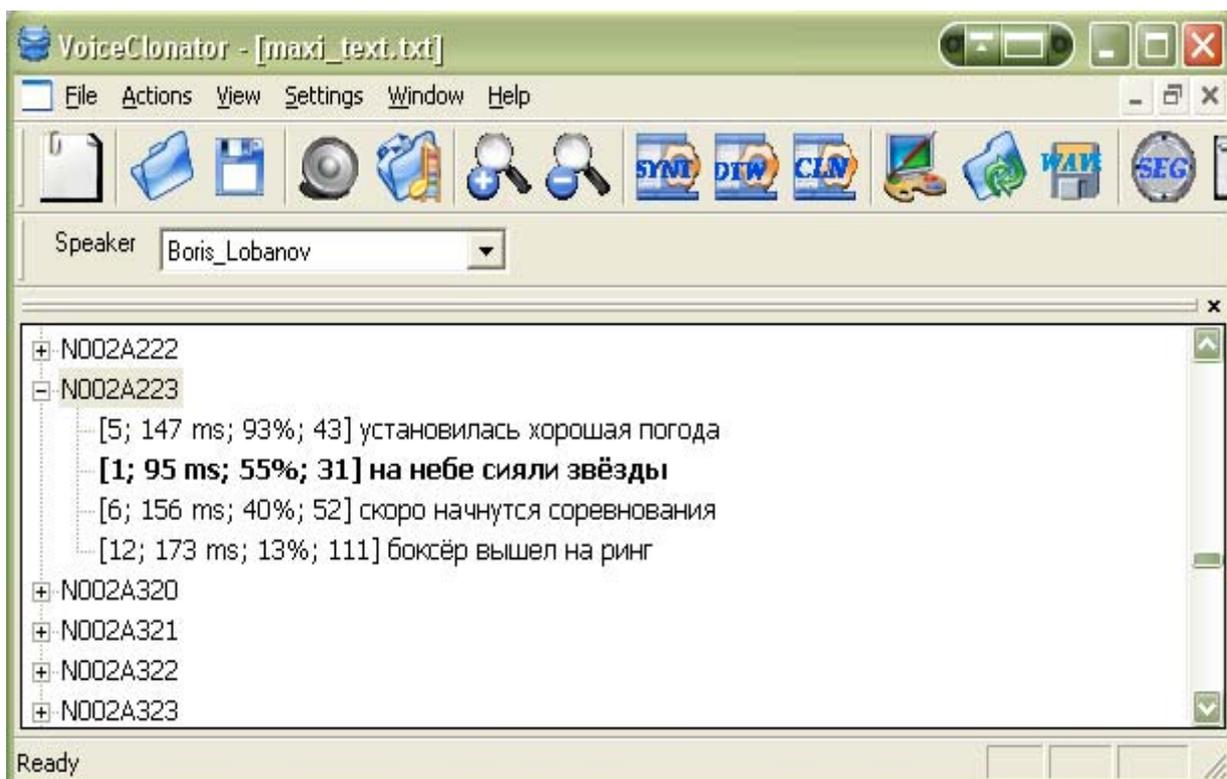


Рис. 5.9. Список выбираемых фонетических сегментов и синтагм, содержащих данные сегменты

Настройки синтеза речевого сигнала. Окно настроек синтеза РС позволяет варьировать используемые лингвистические и акустические ресурсы: грамматический словарь, списки слов – фонетических исключений, БД звуковых волн аллофонов.

Настройки блока сегментации и аллофонной маркировки РС. Окно настроек блока сегментации и аллофонной маркировки позволяет устанавливать параметры вычисления спектральных характеристик сигналов и ДП-сопоставления естественного и синтезированного РС. Установка параметров влияет на точность сегментации естественного РС.

Выбор сегментов для создаваемой БД. Диалог выбора типов сегментов, отображённый на рис. 5.10, даёт возможность пользователю указать типы сегментов для создания и пополнения БД элементов компиляции. Система находит в каждой синтагме входного списка указанные сегменты, которые отображаются затем в окне списка сегментов и в окне дерева сегментов.



Рис. 5.10. Окно выбора типов сегментов

Создание БД элементов компиляции. Диалог настроек данного блока (рис. 5.11) включает настройки операций «отсекающий отбор», «диагностика», «коррекция» и «размножение».

Настройки «отсекающего отбора» включают значения порогов сходства временных и акустических признаков синтезированного и естественного сегментов.

Настройки «диагностики» и «коррекции» включают значения весовых коэффициентов сходства временных и акустических признаков двух соседних периодов основного тона и порог сходства периодов.

Критерии размножения включают типы «размножаемых» мультифонов: гласные первой и (или) второй степени редукции.

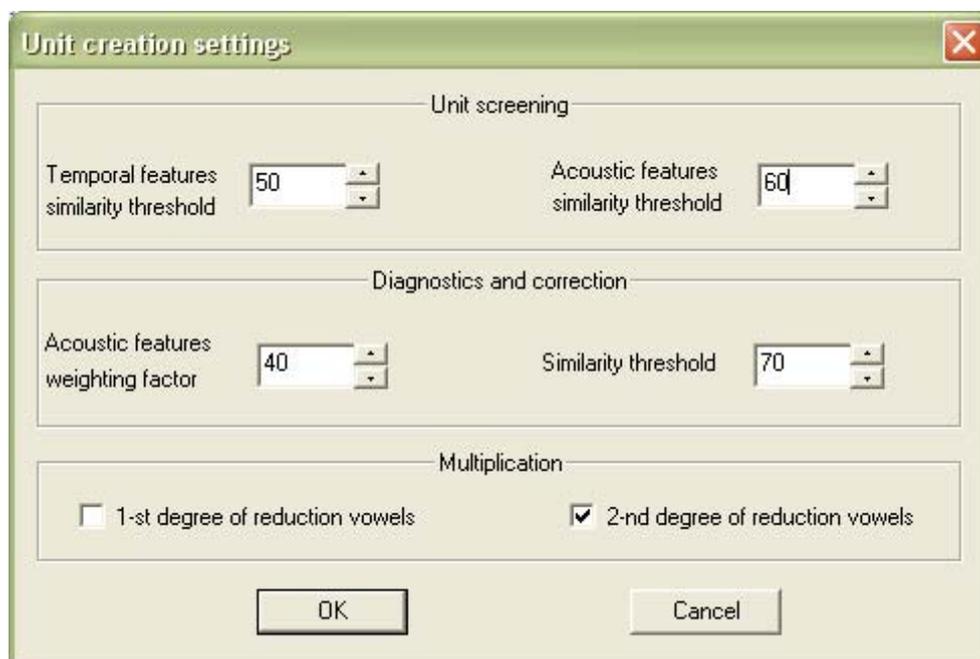


Рис. 5.11. Окно настроек отсекающего отбора, диагностики, коррекции сегментов и размножения элементов компиляции

5.4.3. Результаты практического использования системы «ФоноКлонатор»

Система «ФоноКлонатор» использовалась для получения компьютерных клонов голосов пяти дикторов: трёх женщин и двух мужчин. Двое из них: женщина (ДЖ1) и мужчина (ДМ1) – являлись профессиональными дикторами, трое – не профессиональными: две женщины (ДЖ2, ДЖ3) и один мужчина (ДМ2). Фонограммами для создания каждого из клонов являлись записи мини- и макси-тестов, описанных в приложениях В1, В2. Запись фонограмм осуществлена в акустических условиях профессиональной радиостудии. Результатом работы системы явилось создание пяти БД элементов компиляции: БД_Ж1, БД_Ж2, БД_Ж3, БД_М1, БД_М2.

В качестве иллюстрации акустических особенностей реализации основных аллофонов гласных и согласных фонем русской речи в Приложении 4 представлены их спектральные портреты на примере аллофонов, взятых из БД_М1.

Перед применением процедуры автоматического клонирования фонограммы записей и соответствующие стенограммы были обработаны вручную следующим образом:

- из фонограмм и стенограмм были удалены синтагмы, в записи которых присутствовали различного рода шумы (например, шелест страниц);

- содержимое текстов было приведено в соответствие с фонограммами (поскольку некоторые слова были прочитаны дикторами неверно);

- в текстовых файлах стенограмм записей были проставлены ударения, соответствующие ударениям, сделанным дикторами при чтении текста (которые не всегда являлись каноническими, например, «приняла», «начался»);

- в звуковых и текстовых файлах были установлены пометы границ фонетических синтагм, реализованных в речи дикторов.

Количество синтагм в «мини-тексте» (см. Приложение 1.1) для всех дикторов было одинаковым и равнялось 69. Количество синтагм в «макси-тексте» (см. Приложение 1.2) для каждого из дикторов приведено в таблице 5.2. Отличия в количестве синтагм для разных дикторов связано с их индивидуальными особенностями прочтения отдельных предложений «макси-текста».

Таблица 5.2

Количество фонетических синтагм в «макси-тексте»

Имя диктора	Количество синтагм
ДЖ1	518
ДЖ2	496
ДЖ3	500
ДМ1	635
ДМ2	494

Создание компьютерных голосовых клонов проходило в два этапа. На первом этапе входными данными системы являлись:

- мини-текст с пометами границ синтагм;
- список звуковых файлов синтагм мини-текста;
- имеющиеся БД элементов компиляции женского и мужского голосов. Для обработки естественной речи женских голосов использовалась БД элементов компиляции женского голоса, а для обработки естественной речи мужских голосов – БД элементов компиляции мужского голоса.

При осуществлении операции «отсекающий отбор» использовались: порог сходства временных признаков $\beta_1 = 0,6$ и порог сходства акустических признаков $\beta_2 = 0,7$. При осуществлении операций «диагностика» и «коррекция» использовались: весовой коэффициент $\alpha = 0,4$ и порог сходства периодов $\beta = 0,7$.

Результатом первого этапа явились вновь созданные БД аллофонов для каждого голоса. Время автоматического создания БД на персональном компьютере, имеющем процессор AMD Athlon 2000 с реальной тактовой частотой $1,67 \cdot 10^9$ Гц, составило $7 \pm 1,5$ минуты.

Количество автоматически созданных аллофонов для каждой из БД приведено в таблице 5.3.

Таблица 5.3

Количество элементов в автоматически созданных БД аллофонов

Название БД	Количество элементов
БДЖ1	548
БДЖ2	498
БДЖ3	503
БДМ1	552
БДМ2	514

Аллофоны, присутствующие в мини-тексте, но не прошедшие операцию «отсекающий отбор», добавлялись в БД «вручную».

На втором этапе – пополнения БД мультифонными сегментами – входными данными системы являлись:

- мини- и макси-тексты с пометами границ синтагм;
- список звуковых файлов синтагм мини- и макси-текстов;
- вновь созданные БД аллофонов для каждого из женских и мужских голосов.

При осуществлении операции «отсекающий отбор» использовались менее жёсткие критерии: порог сходства временных признаков $\beta_1 = 0,5$, порог сходства акустических признаков $\beta_2 = 0,6$. При осуществлении операций «диагностика» и «коррекция» использовались: весовой коэффициент $\alpha = 0,3$; порог сходства периодов $\beta = 0,6$.

Количество созданных мультифонов для каждого из дикторов приведено в таблице 5.4. Количество диаллофонов и аллослогов до операции «отсекающий отбор», которое определялось по соответствующим текстовым файлам, не является одинаковым для всех дикторов, поскольку, во-первых, количество фонетических синтагм в макси-тексте было разным для различных дикторов, во-вторых, некоторые слова были произнесены дикторами с фонетическими ошибками или с неканоническим ударением. Общее количество мультифонов не является суммой диаллофонов и аллослогов, так как некоторые аллослоги, полученные по алгоритмам, описанным в разделе 5.2, являются диаллофонами.

Таблица 5.4

Количество различных элементов компиляции в созданных БД мультифонов

Название БД	Количество мультифонов								
	диаллофоны			аллослоги			общее количество		
	после разметки на мультифоны	после «отсекающего отбора»	после «размножения»	после разметки на мультифоны	после «отсекающего отбора»	после «размножения»	после разметки на мультифоны	после «отсекающего отбора»	после «размножения»
БДЖ1	3058	2923	3512	4397	4082	4705	6603	6012	7073
БДЖ2	3052	2875	3425	4405	3979	4593	6595	5918	6772
БДЖ3	3067	2830	3401	4434	3820	4386	6628	5844	6678
БДМ1	3040	2964	3592	4320	3990	4598	6495	5935	6818
БДМ2	3074	2729	3317	4407	3063	3532	6602	4997	5870

Время автоматического создания БД мультифонов на персональном компьютере, имеющем процессор AMD Athlon 2000 с реальной тактовой частотой $1,67 \cdot 10^9$ Гц, составило 42 ± 4 минуты.

5.5. Компьютерная система клонирования просодических характеристик речи – «ИнтоКлонатор»

5.5.1. Функциональная схема и алгоритмы работы системы

Функциональная схема, входные и выходные данные, взаимодействие блоков системы представлены на рис 5.12.

Входные данные системы:

– предварительно обработанная фонограмма записи – набор речевых синтагм, каждая из которых хранится в виде оцифрованной звуковой волны в отдельном файле в формате *WAVE PCM*;

– предварительно обработанная стенограмма записи – набор текстовых файлов синтагм, с указанием интонационного типа и количества акцентных единиц (АЕ) для каждой синтагмы;

– правила просодической маркировки синтезированного речевого сигнала на АЕ и элементы АЕ – предъядро, ядро и заядро (см. разделы 4.3, 4.5).

Выходные данные системы:

БД просодических характеристик речи «клонлируемого» диктора – набор мелодических, энергетических и ритмических портретов акцентных единиц.

Просодическая маркировка естественного РС. Текстовые файлы синтагм является входным данным блока синтеза и просодической маркировки синтезированного речевого сигнала (РС). В блоке синтеза осуществляется фонетическая и просодическая обработка текста, включающая преобразования «буква-фонема» и «фонема-аллофон», выбор звуковых волн аллофонов из акустической БД, их компиляция и маркировка синтезированного речевого сигнала на АЕ и элементы АЕ – предъядро, ядро и заядро (ЭАЕ). Результат обработки – просодически размеченная синтагма синтезированного речевого сигнала.

Каждая пара синтагм «просодически-размеченный синтезированный сигнал – естественный сигнал» поступает в блок ДП-сегментации и просодической маркировки естественного РС, в котором осуществляется разметка естественного сигнала на периоды основного тона (питчи), анализ акустических признаков естественного и синтезированного сигналов, их ДП-сопоставление и перенос маркеров границ аллофонов, АЕ и ЭАЕ с синтезированного на естественный РС. В системе реализована настройка параметров вычисления питчей естественного РС. Результатом работы блока является синтагма естественного РС, в которой расставляются метки питчей, аллофонов, а также предъядра, ядра и заядра для каждой АЕ. Для именования регионов приняты следующие обозначения: предъядро – preN, ядро – N, заядро – postN. По именам этих регионов определяются границы и длительности предъядра, заядра и ядра каждой акцентной единицы.

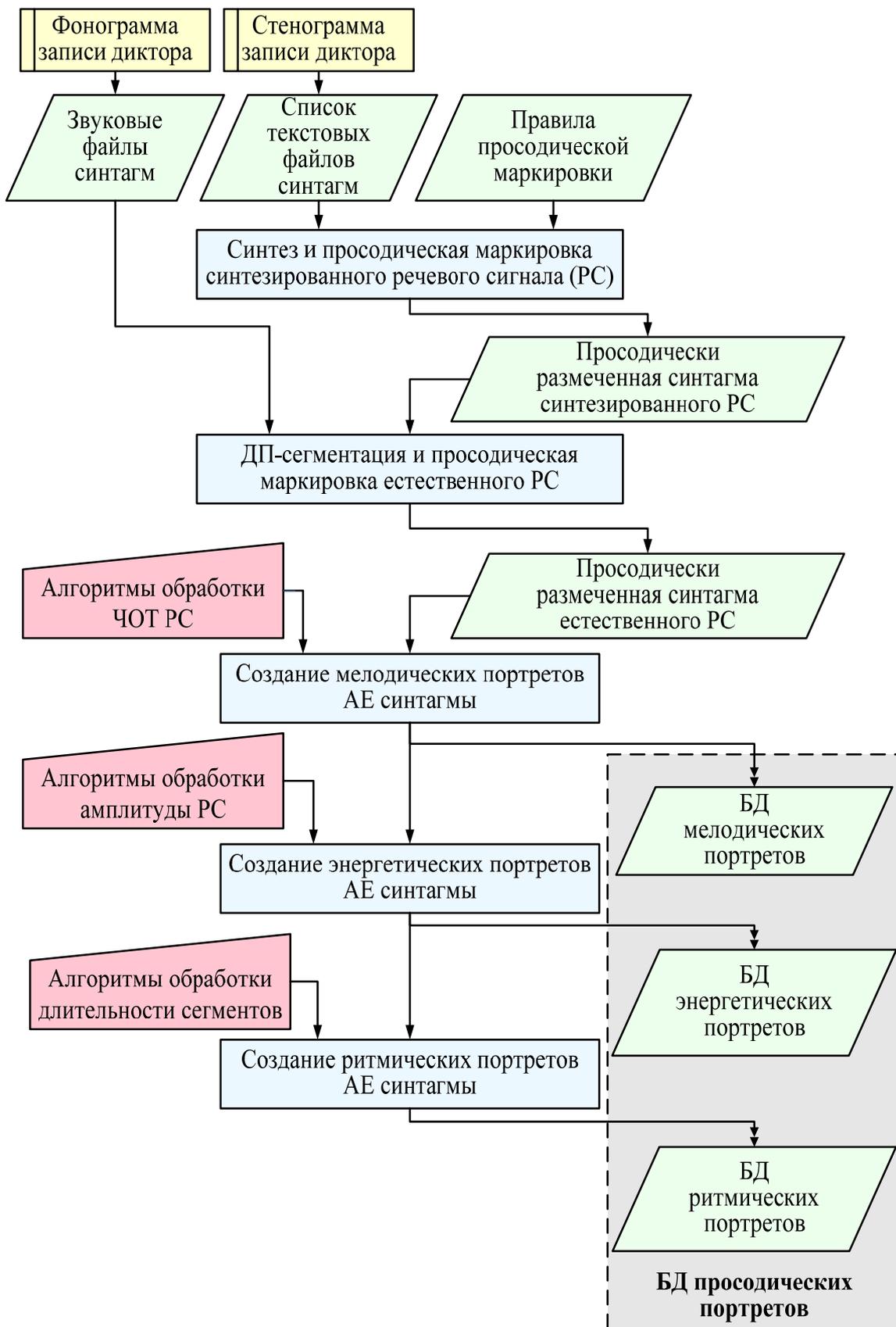


Рис. 5.12. Функциональная схема системы «ИнтоКлонатор»

Пример сигнала синтагмы «Машенька уснула», размеченного на питчи и аллофоны, показан на рис. 5.13. Синтагма состоит из двух АЕ: «Машенька» и «уснула». Ядром первой АЕ является аллофон A_{012} , ядром второй – аллофон U_{022} .

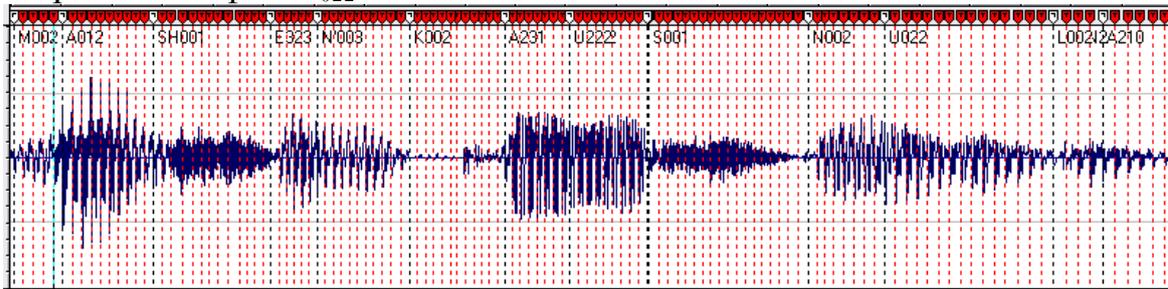


Рис. 5.13. Пример маркировки естественного РС синтагмы

Алгоритмы создания мелодических портретов.

С использованием информации о текущей длительности периодов речевого сигнала, задаваемой метками питчей, вычисляется исходный мелодический контур (ИМК) значений ЧОТ - F_0 с использованием дополнительной процедуры медианного сглаживания. Для каждого элемента АЕ – предъядра, ядра и заядра – равномерно выбирается пять точек ИМК, лежащих во временных пределах каждого элемента АЕ на участках, соответствующих аллофонам гласных и звонких согласных. При этом в ИМК не включаются точки, находящиеся в регионах аллофонов шумных согласных $\{f, f', s, s', sh, sh', c, ch', h, h', p, p', t, t', k, k', b, b', d, d', g, g'\}$. На участках шумных согласных реальные значения ЧОТ заменяются новыми значениями путём вычисления интерполяционной прямой от последней точки предшествующего региона звонкого аллофона к первой точке последующего региона звонкого аллофона. Пример обработки контура ЧОТ для синтагмы «Машенька уснула» показан на рис. 5.14.

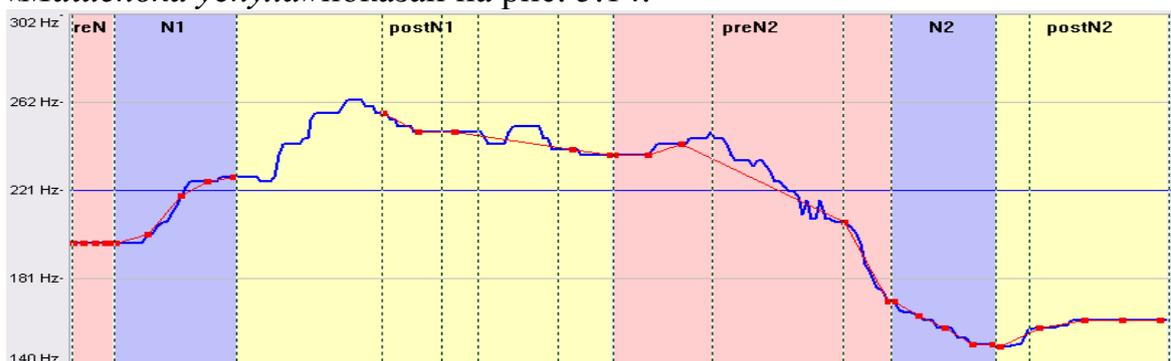


Рис. 5.14. Пример обработки мелодического контура синтагмы

На следующих шагах алгоритма осуществляется нормировка длительности сегментов preN, N, postN путём уравнивания длительности областей предъядра, ядра и заядра каждой АЕ, входящей в синтагму.

Далее осуществляется нормировка контура ЧОТ. Для этого определяются минимальное – $F_{0\ min}$ – и максимальное – $F_{0\ max}$ – значения на всей исследуемой фонограмме. Нормированные значения ЧОТ вычисляются согласно формуле:

$$F_{0\ norm} = \frac{F_0 - F_{0\ min}}{F_{0\ max} - F_{0\ min}} \quad (5.10)$$

Результатом описанных операций является создание последовательности нормированных мелодических портретов АЕ, составляющих синтагму (рис. 5.15).



Рис. 5.15. Нормированный мелодический портрет двухакцентной синтагмы: «Машенька уснула.»

Алгоритмы создания энергетического портрета. Для создания энергетического портрета синтагмы строится контур текущих значений энергии путём усреднения среднеквадратичного значения сигнала на интервале 15 миллисекунд с шагом 5 миллисекунд. На каждом из ядер АЕ синтагмы выбирается максимальное значение текущей энергии – $A_{i\ max}$. Графическое построение контура производится по следующему правилу. От левой границы сигнала до правой границы первого ядра строится горизонталь на уровне значения этого ядра $A_{1\ max}$. Далее от правой границы первого ядра до точки со значением амплитуды второго ядра синтагмы $A_{2\ max}$ строится прямая, а от неё до правой границы этого же ядра строится горизонталь. И так далее, до последнего ядра, от правой границы которого проводится горизонталь до конца сигнала.

Пример обработки энергетического контура для синтагмы «Машенька уснула» показан на рис. 5.16.



Рис. 5.16. Пример обработки энергетического контура синтагмы

Далее производится нормировка энергетического контура контура. Нормировка по длительности сегментов $preN$, N , $postN$ осуществляется, как и в предыдущем случае, путём уравнивания длительности областей предъядра, ядра и заядра каждой АЕ, входящей в синтагму. Нормировка энергетических уровней осуществляется путём деления полученного энергетического контура на величину наибольшего значения $A_{i\ max}$, найденного на всей исследуемой фонограмме.

Результатом описанных операций является создание нормированного энергетического портрета синтагмы (рис. 5.17).

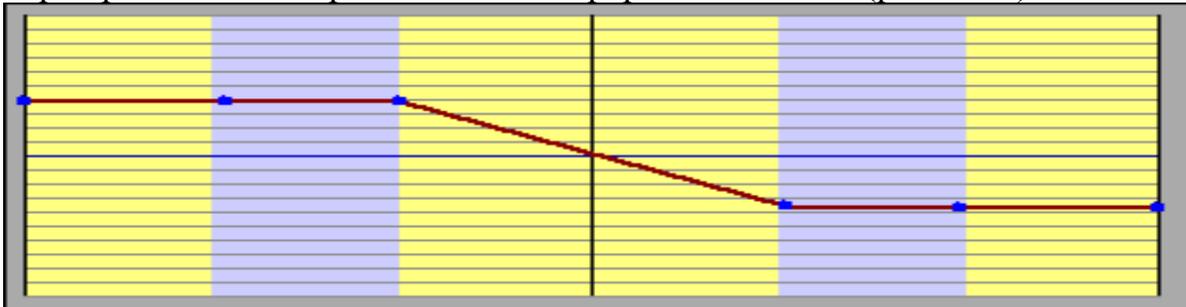


Рис. 5.17. Нормированный энергетический портрет 2-х акцентной синтагмы: «Машенька уснула.»

Алгоритмы создания ритмического портрета. Вычисляются длительности ядер АЕ, входящих в синтагму – T_{N1} , T_{N2} , T_{N3} , Определяется максимальная из длительностей ядер в синтагме и осуществляется вычисление нормированных ритмических коэффициентов изменения длительности ядер в синтагме относительно ядра с максимальной длительностью. Ритмический коэффициент i -ой АЕ R_i вычисляется в соответствии с формулой

$$R_i = \frac{T_{Ni}}{T_{Ni\ max}} \quad (5.11)$$

где T_{Ni} – длительность ядра i -й АЕ синтагмы, $T_{Ni\ max}$ – максимальная из длительностей ядер в синтагме.

Результатом описанных операций является создание нормированного ритмического портрета синтагмы (рис. 5.18). Нижний участок рисунка показывает изменённые под действием ритмического фактора относительные длительности ядер первой и второй АЕ синтагмы.

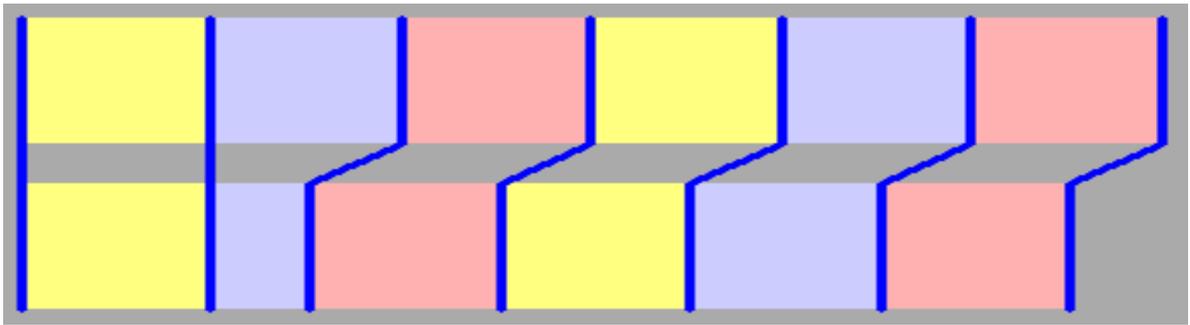


Рис. 5.18. Нормированный ритмический портрет двухакцентной синтагмы: «*Машенька уснула.*»

5.5.2. Пользовательский интерфейс системы «ИнтоКлонатор»

Пользовательский интерфейс системы «ИнтоКлонатор» (рис. 5.19) включает следующие блоки:

- окно отображения осциллограммы речевого сигнала (РС);
- окно отображения графика нормированной амплитуды (A_{norm}) сигнала;
- окно отображения графика нормированной ЧОТ ($F_{0\,norm}$) сигнала;
- диалоговые окна настроек параметров системы.

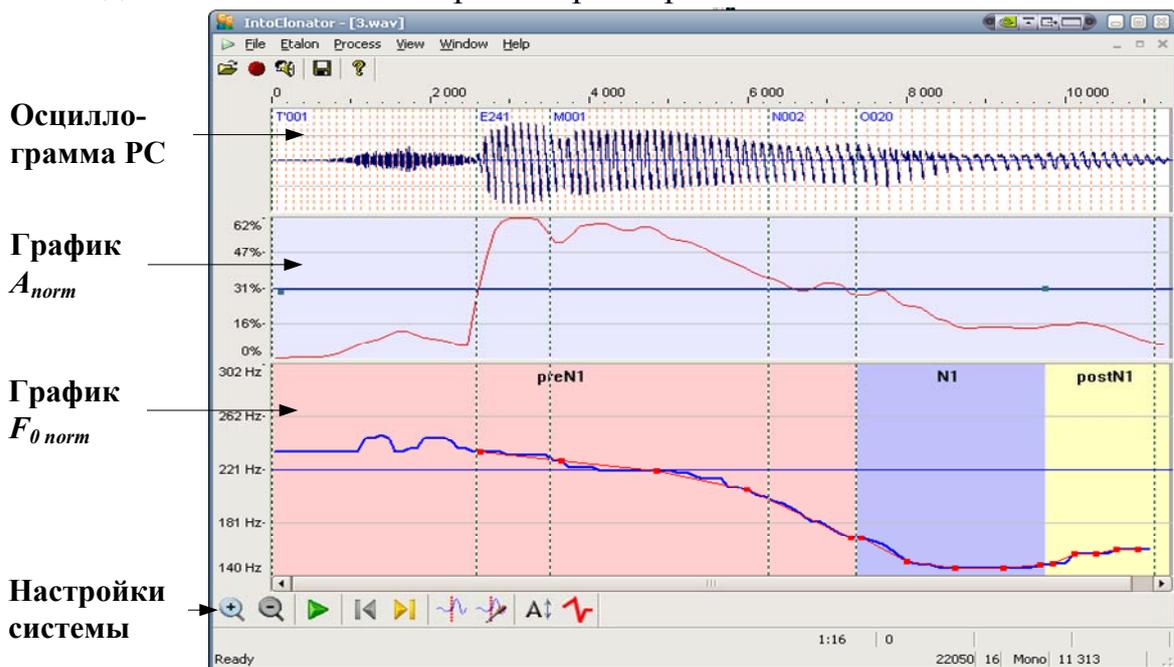


Рис. 5.19. Общий вид пользовательского интерфейса системы «ИнтоКлонатор»

На осциллограмме РС (рис. 5.20) указаны границы периодов основного тона и аллофонов, а также имена аллофонов. В системе реализовано масштабирование отображения и прослушивание выделенного фрагмента РС.

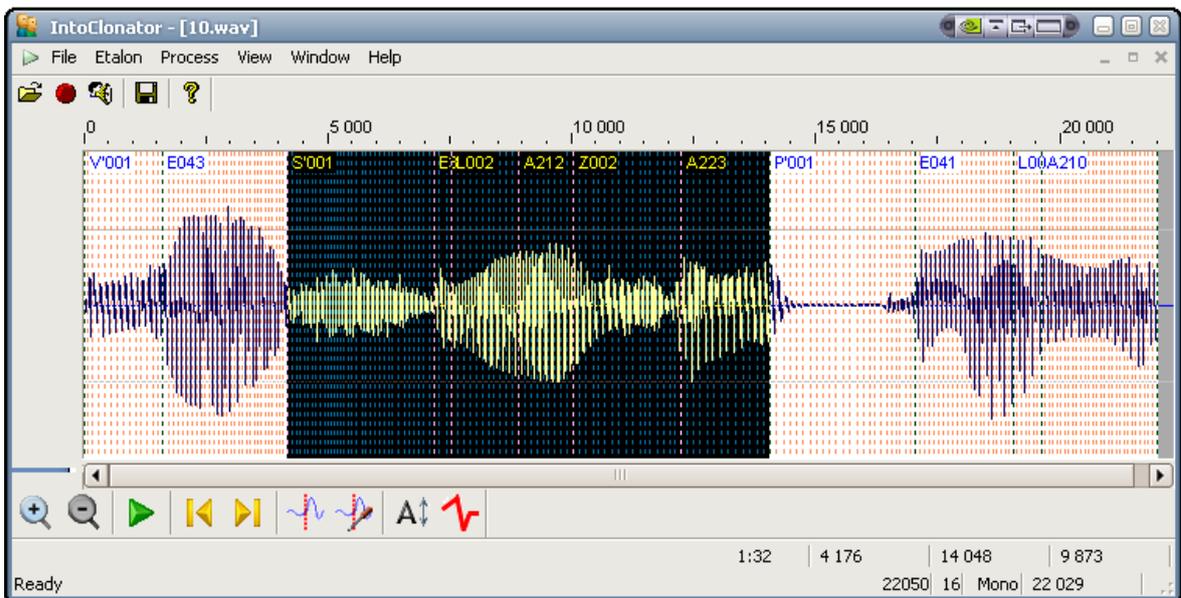


Рис. 5.20. Отображение осциллограммы РС, границ аллофонов и периодов основного тона

Нормированная амплитуда и ЧОТ сигнала (рис 5.21) вычисляются в соответствии с задаваемым диапазоном A_{min} , A_{max} и $F_{0 min}$, $F_{0 max}$. На графиках отображаются границы АЕ синтагмы и предъядра, ядра и заядра каждой АЕ, а также вычисленные динамический и мелодический портреты.

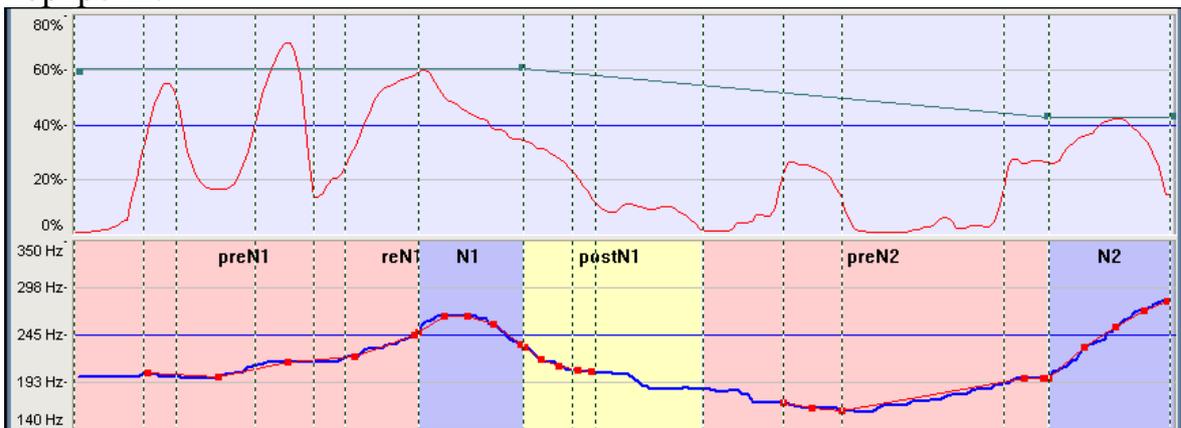


Рис. 5.21. Отображение графиков A_{norm} , $F_{0 norm}$, динамического и мелодического портретов синтагмы

Настройки параметров системы реализованы в следующих диалоговых окнах:

- диалог настроек параметров вычисления ЧОТ;
- диалог настроек параметров сегментации РС,
- диалог установки диапазона амплитуды и ЧОТ;

Настройки блока вычисления ЧОТ позволяют устанавливать параметры вычисления спектральных характеристик и параметры определения вокализованных участков сигнала.

Настройки блока сегментации позволяют устанавливать параметры ДП-сопоставления естественного и синтезированного РС.

Настройки диапазона амплитуды и ЧОТ позволяют указывать значения A_{min} , A_{max} и F_{0min} , F_{0max} , которые должны быть определены заранее для набора речевых синтагм, обрабатываемых системой.

5.5.3. Результаты практического использования системы «ИнтоКлонатор»

Система «ИнтоКлонатор» работает на базе специально разработанного текстового корпуса, включающего «мини-текст» для создания основного набора просодических портретов и «макси-тексты» для создания расширенного набора просодических портретов русской речи (см. Приложение 1).

С использованием системы «ИнтоКлонатор» создана БД просодических портретов для системы синтеза русской речи по тексту «МультиФон», включающая мелодические, динамические и ритмические портреты для следующих интонационных типов.

Для повествовательных предложений.

Синтагмы с интонацией незавершённости, которые образуются в следующих ситуациях:

- 1) С1, если «И»;
- 2) С2, если «ИЛИ»;
- 3) С3, если «,» и не С7 – С11 при условии, что «,» встретилась в тексте впервые или в 4-й, 7-й, ... раз подряд;
- 4) С3_1, если «,» и не С7 – С11 при условии, что «,» встретилась в тексте во 2-й, 5-й, 8-й, ... раз подряд;
- 5) С3_2, если «,» и не С7 – С11 при условии, что «,» встретилась в тексте в 3-й, 6-й, 9-й раз подряд;
- 6) С4, если «-»;
- 7) С5, если «(»;
- 8) С6, если «, - »;
- 9) С7, если «,» и союз сочинительный;
- 10) С8, если «,» и союз вопросительно- подчинительный;
- 11) С9, если «,» и союз подчинительный;
- 12) С10, если «,» и причастие;
- 13) С11, если «,» и деепричастие;
- 14) С01, если первая, третья, пятая и т.д. синтаксическая синтагма;
- 15) С02, если вторая, четвёртая и т.д. синтаксическая синтагма.

Синтагмы с интонацией завершённости, которые образуются в следующих ситуациях:

- 16) Р1, если «:»;
- 17) Р2, если «)»;
- 18) Р3, если «;»;
- 19) Р4, если «,» при условии, что «,» встретилась в тексте в 1-й или 4-й, 7-й и т.д. раз подряд;

20) P4_1, если «.» при условии, что «.» встретилась в тексте во 2-й, 5-й, 8-й и т.д. раз подряд;

21) P4_2, если «.» что «.» встретилась в тексте в 3-й, 6-й, 9-й ... раз подряд;

22) P5, если «...»;

23) P5, если «.» и конец абзаца;

24) P6, если «.» и конец текста;

25) P7, если «.» и в начале союз сочинительный после (,);

26) P8, если «.» и в начале союз вопросительно-подчинительный после (,);

27) P9, если «.» и в начале союз подчинительный после (,);

28) P10, если «.» и в начале причастие после (,);

29) P11, если «.» и в начале деепричастие после (,).

Для вопросительных предложений:

30) Q1, если в составе синтагмы имеется вопросительное слово и если в вопросительном предложении оказалась только одна синтагма;

31) Q2, если в составе синтагмы отсутствует вопросительное слово и если в вопросительном предложении оказалась только одна синтагма;

32) Q1-1, если в составе синтагмы имеется вопросительное слово и если в вопросительном предложении более, чем одна синтагма;

33) Q2-1, если в составе синтагмы отсутствует вопросительное слово и если в вопросительном предложении более, чем одна синтагма.

Для восклицательных предложений.

34) E1, если в составе синтагмы имеется междометие и если в восклицательном предложении оказалась только одна синтагма;

35) E2, если в составе синтагмы отсутствует междометие и если в восклицательном предложении оказалась только одна синтагма;

36) E1_1, если в составе синтагмы имеется междометие и если в восклицательном предложении более, чем одна синтагма;

37) E2_1, если в составе синтагмы отсутствует междометие и если в восклицательном предложении более, чем одна синтагма.

Итого с использованием системы «ИнтоКлонатор» созданы просодические портреты синтагм 37-ми интонационных типов. При этом допускались 4 возможных варианта синтагм, состоящих из одной, двух, трёх и четырёх акцентных единиц. Таким образом, были созданы $37 \cdot 4 = 148$ просодических портретов.

Примеры созданных мелодических портретов основных интонационных типов вопросительной интонации русской речи приведены в Приложении 5.

5.6. Оценка правдоподобия синтезированного речевого клона

Для оценки правдоподобия синтезированного речевого клона решаются следующие задачи:

– разработать адекватную методику оценки степени сходства синтезированного клона с естественной речью (т. е. правдоподобие речевого клона),

– получить численную оценку правдоподобия созданного речевого клона,

– оценить вклад элементов компиляции различного фонетического количества (аллофоны, диаллофоны, аллослоги) в правдоподобие речевого клона,

– оценить вклад фонем различного типа (ударные и безударные гласные, согласные) в правдоподобие речевого клона.

Существует несколько методов оценки качества синтезированной речи [184-187], основанной на расчёте корреляции между естественным и синтезированным речевыми сигналами в пространстве различных параметров сигнала. Однако даже лучшие из них не дают результат, приближающийся к результатам субъективной оценки. Поэтому в экспериментах по определению степени сходства синтезированного клона с естественной речью (т. е. правдоподобия речевого клона) предпочтение было отдано оценке субъективного мнения, так называемой MOS-оценке. Методика проведения эксперимента основывалась на Рекомендации Р.85 ИТУ-Т «Метод субъективной оценки качества речи устройств речевого вывода» [188] и СТБ ГОСТ Р 50840-2000 [176], и была адаптирована для данной задачи [32].

В связи с тем, что оценивалось не качество синтезируемой речи, а правдоподобие речевого клона, были подходящим образом скорректированы форматы стимулов, опросные листы и процедура прослушивания.

БД элементов компиляции для тестирования правдоподобия речевого клона. Для экспериментов использованы записи естественного голоса диктора Д1 и его синтезированного клона, а также клона голоса диктора Д2. Оба выбранных для эксперимента голоса имеют одинаковый диапазон изменения частоты основного тона: 80 – 120 Гц. Клоны голосов двух дикторов – Д1 и Д2 – синтезированы с использованием соответствующих БД элементов компиляции. Каждая из БД содержит полный набор звуковых волн аллофонов. В состав БД диктора Д1 мог быть включён, кроме того, набор звуковых волн диаллофонов и аллослогов в количестве 6818 единиц. Исходным материалом для подготовки БД послужили записи естественной речи, выполненные в студийных условиях.

С целью оценки вклада элементов компиляции различного типа в правдоподобие речевого клона сформированы семь различных БД, две из которых (мини-БД и макси-БД) использовались для синтеза клонов диктора Д1, одна (мини-БД) – для синтеза клонов диктора Д2, и четыре сформированных БД (мини-БД) включали аллофоны различного типа,

взятые из мини-БД дикторов Д1 и Д2 согласно содержанию, описанному в таблице 5.5.

Все элементы баз сохранялись в формате WAVE PCM с частотой дискретизации 22050 Гц и разрядностью 16 бит.

Таблица 5.5

Содержимое фонетико-акустических БД для тестирования правдоподобия речевого клона

Название БД	Содержимое БД
База1	Макси-БД (аллофоны, диаллофоны, аллослоги) диктора Д1
База2	Мини-БД (аллофоны) диктора Д1
База3	Аллофоны гласных диктора Д1+ аллофоны согласных диктора Д2
База4	Аллофоны ударных гласных диктора Д1+ аллофоны безударных гласных и согласных диктора Д2
База5	Аллофоны безударных гласных и согласных диктора Д1+ аллофоны ударных гласных диктора Д2
База6	Аллофоны согласных диктора Д1+ аллофоны гласных диктора Д2
База7	Мини-БД (аллофоны) диктора Д2

Стимулы для тестирования. В качестве сообщений были подобраны 20 фонетически сбалансированных фраз. Каждая фраза состояла из последовательности трёх-четырёх слов.

Все фразы были произнесены диктором Д1 в студийных условиях, идентичных условиям для подготовки фонетико-акустических баз клона. Запись производилась на цифровые носители, и была сохранена в формате WAVE PCM с частотой дискретизации 22050 Гц и разрядностью 16 бит. Фразы были произнесены в среднем темпе с интонацией перечисления. Длительность пауз между словами во фразе была приведена к значению 200 мс. Длительность фраз находилась в диапазоне 3,2 - 4,3 секунды. Частота основного тона во фразах находилась в диапазоне 80-120 Гц.

На основании подготовленных фонетико-акустических БД были синтезированы 7 групп фраз-клонов. Каждая группа состояла из 20 фраз, идентичных по содержанию фразам, произнесённым естественным голосом. Для устранения влияния просодических характеристик на восприятие синтезированной речи во всех синтезированных фразах сохранялись темп речи, длительность пауз между словами, амплитуда сигнала и значения частоты основного тона, характерные для естественных фраз.

Стимулами для тестирования и оценки являлись пары фраз одинакового содержания. При этом первая фраза в паре являлась записью естественной речи диктора Д1, а вторая – записью синтезированного клона, полученного на основе одной из БД, описанных в таблице 5.5, либо записью речи того же диктора с незначительными искажениями, применяемыми для того, чтобы естественный сигнал не был полностью идентичен исходному. В качестве искажений была выбрана частотная модуляция со следующими параметрами: частота модулирующего сигнала 10 Гц, глубина модуляции 4%. Выбранные параметры модуляции не ухудшают степень разборчивости речи и узнаваемости голоса.

Таким образом, было сформировано $20 \cdot 8 = 160$ стимулов, которые представлялись аудиторам в случайном порядке.

Пауза между фразами в паре составляла 700 мс, пауза между стимулами – 5 секунд.

Проведение эксперимента. Для оценки правдоподобия синтезированного речевого клона аудиторам было предложено ответить на вопрос «Похож ли второй из услышанных голосов на первый?», используя шкалу оценки, представленную в таблице 5.6.

Таблица 5.6

Шкала оценки сходства голосов

Оценка	Значение
1	Нет, совсем другой голос
2	Нет, пожалуй, это другой голос
3	Немного похож
4	Да, очень похож
5	Да, практически тот же голос

Каждый стимул аудиторы прослушивали один раз. Для того, чтобы аудиторы сфокусировали внимание на сходстве голосов, а не на разборчивости произносимых фраз (для минимизации напряжения при прослушивании) текстовое содержание каждого из стимулов было записано на опросном листе.

Аудиторами являлись 8 мужчин в возрасте от 21 до 60 лет, носители русского языка, без выявленных дефектов слуха. Тест проходил в тихой комнате, длился около 50 минут и был разбит на 2 сессии по 25 минут.

Статистическая обработка результатов эксперимента. Обобщенная оценка правдоподобия речевых клонов (MOS-оценка) выражается через среднее значение и дисперсию оценок всех аудиторов в соответствии со шкалой сходства голосов (таблица 5.6) по каждому из тестируемых типов БД (таблица 5.5).

Для определения статистической значимости оценок, полученных для различных типов голосов, был осуществлён однофакторный дисперсионный анализ результатов с использованием F -критерия и множественное попарное сравнение [189] с использованием критерия Тьюки достоверно значимой разности.

При однофакторном дисперсионном анализе вычислялась общая (по всем значениям) дисперсия оценок $\tilde{\sigma}^2_{общая}$, внутригрупповая дисперсия $\tilde{\sigma}^2_{внгрупп}$ и межгрупповая дисперсия $\tilde{\sigma}^2_{межгрупп}$ в соответствии с формулами:

$$\tilde{\sigma}^2_{общая} = \sum_{i=1}^{n_1} x_{1i}^2 + \sum_{i=1}^{n_2} x_{2i}^2 + \dots + \sum_{i=1}^{n_r} x_{ri}^2 - \frac{\sum_{i=1}^{n_1} x_{1i} + \sum_{i=1}^{n_2} x_{2i} + \dots + \sum_{i=1}^{n_r} x_{ri}}{N} \quad (5.12)$$

$$\tilde{\sigma}^2_{внгрупп} = \frac{(\sum_{i=1}^{n_1} x_{1i})^2}{n_1} + \frac{(\sum_{i=1}^{n_2} x_{2i})^2}{n_2} + \dots + \frac{(\sum_{i=1}^{n_r} x_{ri})^2}{n_r} - \frac{\sum_{i=1}^{n_1} x_{1i} + \sum_{i=1}^{n_2} x_{2i} + \dots + \sum_{i=1}^{n_r} x_{ri}}{N} \quad (5.13)$$

$$\tilde{\sigma}^2_{межгрупп} = \tilde{\sigma}^2_{общая} - \tilde{\sigma}^2_{внгрупп} \quad (5.14)$$

где N – общее количество наблюдений (в данном тесте равное 1280), x_{ji} – значение i -го наблюдения в j -той группе, n_j – количество наблюдений в группе j (в данном тесте равное 160 для каждой группы), r – количество групп (в данном тесте равное 8).

Затем вычислялось межгрупповое $dF_{межгрупп}$ и внутригрупповое $dF_{внгрупп}$ количество степеней свободы (формулы (5.15), (5.16)) и, на их основе, межгрупповое $MS_{межгрупп}$ и внутригрупповое $MS_{внгрупп}$ среднеквадратичные значения (формулы (5.17), (5.18)):

$$dF_{межгрупп} = r - 1 \quad (5.15)$$

$$dF_{внгрупп} = N - r \quad (5.16)$$

$$MS_{межгрупп} = \frac{\tilde{\sigma}^2_{межгрупп}}{dF_{межгрупп}} \quad (5.17)$$

$$MS_{внгрупп} = \frac{\tilde{\sigma}^2_{внгрупп}}{dF_{внгрупп}} \quad (5.18)$$

Значение F вычислялось как отношение среднеквадратичных значений эффекта и ошибки:

$$F = \frac{MS_{межгрупп}}{MS_{внгрупп}} \quad (5.19)$$

Результаты вычислений представлены в таблице 5.7, где p показывает уровень статистической значимости результатов.

Таблица 5.7

Результаты однофакторного дисперсионного анализа

	$\tilde{\sigma}^2$	dF	MS	F	p
Межгрупповое значение	2034.3	7	290.6	425.61	<0.0001
Внутригрупповое значение	868.6	1272	0.7		

Таким образом, различие средних значений является значимым. Для попарной оценки значимости различий между средними значениями конкретных групп i и j вычислялось значение $MS_{\text{попарное}}$:

$$MS_{\text{попарное}} = \frac{M_i - M_j}{\sqrt{\frac{MS_{\text{ошибки}}}{n}}} \quad (5.20)$$

где M_i , M_j – средние значения оценок для групп i , j соответственно, n – количество наблюдений в группе.

Затем были вычислены границы разности с доверительным интервалом 95% и сделан вывод о значимости/не значимости разности.

Результаты вычислений представлены в таблице 5.8.

Таблица 5.8

Результаты множественного попарного сравнения

Пары групп	$MS_{\text{попарное}}$	95%-ный доверительный интервал		Значимость разности
		Нижняя граница	Верхняя граница	
Естественный – База1	0.5	0.3	0.8	Значима
Естественный – База2	0.8	0.6	1.1	Значима
Естественный – База3	1.3	1.0	1.5	Значима
Естественный – База4	2.7	2.4	2.9	Значима
Естественный – База5	2.8	2.6	3.0	Значима
Естественный – База6	3.4	3.2	3.6	Значима
Естественный – База7	3.3	3.1	3.6	Значима
База1 – База2	0.3	0.5	1.0	Значима
База1 – База3	0.8	0.2	0.7	Значима
База1 – База4	2.2	1.9	2.4	Значима
База1 – База5	2.3	2.0	2.5	Значима
База1 – База6	2.9	2.6	3.1	Значима

База1 – База7	2.8	2.6	3.1	Значима
База2 – База3	0.5	0.2	0.7	Значима
База2 – База4	1.8	1.6	2.1	Значима
База2 – База5	2.0	1.7	2.2	Значима
База2 – База6	2.6	2.3	2.8	Значима
База2 – База7	2.5	2.3	2.7	Значима
База3 – База4	1.4	1.1	1.6	Значима
База3 – База5	1.5	1.3	1.8	Значима
База3 – База6	2.1	1.9	2.4	Значима
База3 – База7	2.0	1.8	2.3	Значима
База4 – База5	0.1	-0.1	0.4	Не значима
База4 – База6	0.7	0.5	1.0	Значима
База4 – База7	0.7	0.4	0.9	Значима
База5 – База6	0.6	0.3	0.9	Значима
База5 – База7	0.5	0.3	0.8	Значима
База6 – База7	-0.1	-0.3	0.2	Не значима

Как видно из таблицы 5.8, не является статистически значимой разность между результатами, полученными на основе пар БД База4 – База5 и База6 – База7.

Численные значения MOS-оценок. Наилучшую численную оценку правдоподобия среди созданных речевых клонов получил клон диктора Д1 (База 1), в котором использована БД звуковых волн аллофонов и мультифонов в количестве 6818 единиц. На рис. 5.22 приведены MOS-оценки правдоподобия речевого клона диктора Д1, полученного с использованием БД База1, клона диктора Д2, полученного с использованием БД База7, а также (для сравнения) MOS-оценка правдоподобия естественного речевого сигнала диктора Д1 с искусственно внесёнными незначительными искажениями. Слева на рисунке представлены значения шкалы оценок, для каждого типа голоса показано количественное значение средней оценки и среднеквадратическое отклонение (СКО) оценки.

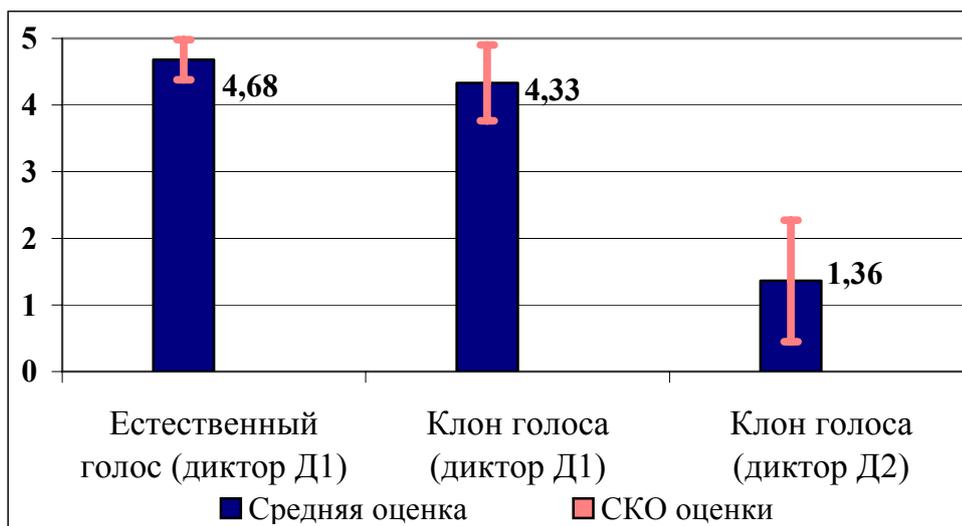


Рис. 5.22. MOS-оценка правдоподобия речевых клонов двух дикторов в сравнении с оценкой естественного голоса диктора Д1

Как видно из диаграммы, достигнутая оценка правдоподобия речевого клона диктора Д1 – 4,33 – близка к оценке естественной речи и существенно отличается от оценки, полученной для клона диктора Д2.

Оценка вклада в правдоподобие речевого клона элементов компиляции различного уровня (аллофонов и мультифонов) иллюстрируется рис. 5.23, где приведены MOS-оценки правдоподобия клонов диктора Д1, полученные с использованием БД База1 (макси-БД) и База2 (мини-БД), а также (для сравнения) MOS-оценка правдоподобия естественного речевого сигнала диктора Д1 с искусственно внесёнными незначительными искажениями. На диаграмме показаны средние значения оценок, а также SKO оценки для каждого типа голоса.

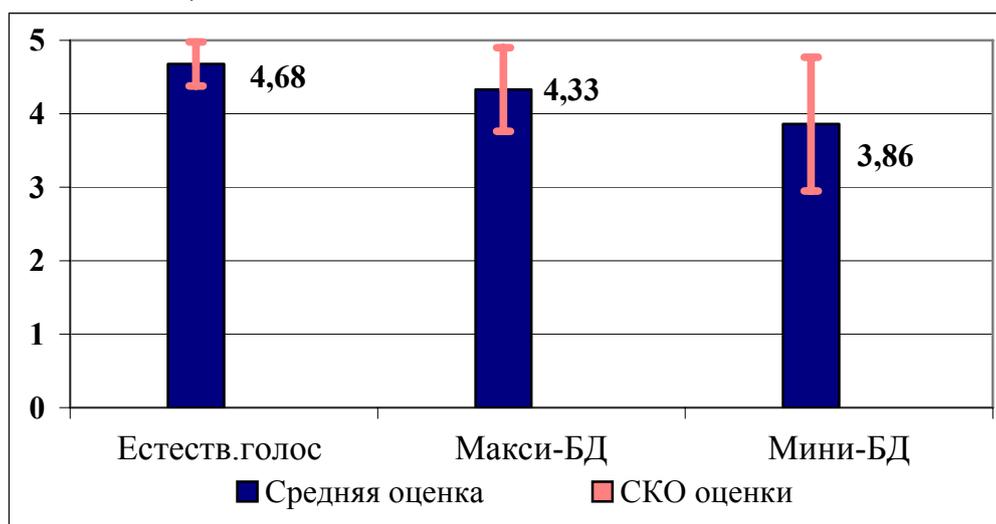


Рис. 5.23. MOS-оценка правдоподобия речевых клонов диктора Д1 на основе двух различных баз в сравнении с оценкой естественного голоса диктора Д1

Как видно из рис. 5.23, добавление в БД диаллофонов и аллослогов даёт ощутимый эффект в восприятии правдоподобия речевого клона, причём разница в оценках БД База1 и База2 является статистически значимой.

Оценка вклада в правдоподобие речевого клона фонем того или иного типа (ударные и безударные гласные, согласные) иллюстрируется на рис. 5.24. Здесь приведены MOS-оценки правдоподобия клонов дикторов Д1 и Д2, полученные с использованием, соответственно, БД База2 (мини-БД диктора Д1) и База7 (мини-БД диктора Д2). Кроме того, приведены MOS-оценки для четырёх БД (База3 – База6), с помощью которых синтезировались «клоны-химеры», обладающие в той или иной степени свойствами голоса дикторов Д1 или Д2.

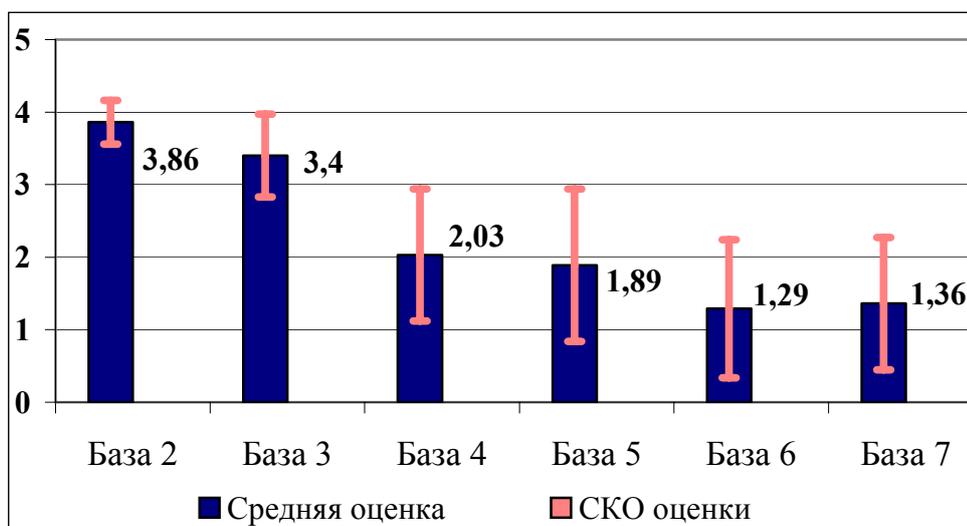


Рис. 5.24. MOS-оценка правдоподобия речевых клонов на основе шести различных баз

Как видно из диаграммы, наибольший вклад, как и ожидалось, в правдоподобие клона вносит комплекс ударных и безударных гласных (База3). Использование в клоне только ударных гласных (База4) или только безударных гласных и согласных (База5), хотя и повышает правдоподобие клона Д1 в сравнении с клоном голоса Д2, однако не столь значительно. Кроме того, как показано в таблице 5.8, разница между оценками Базы4 и Базы5 не является статистически значимой. Можно утверждать, что отсутствие в клоне либо ударных гласных, либо безударных гласных и согласных клонируемого диктора одинаково ощутимо уменьшает правдоподобие создаваемого речевого клона. Замена в базе клона Д1 только согласных на соответствующие согласные из базы клона Д2 не приводит к сколь-нибудь существенному изменению правдоподобия клона.

Приложение 1

Текстовый корпус для создания БД элементов компиляции

П 1.1. Текст для создания мини-набора аллофонов русской речи

1. А.
2. Тяга у тяпки и дяди.
3. Э.
4. Вэф у кэба и посев.
5. О.
6. Кот дочка и капот.
7. У.
8. Гуси у шлюпки и тюльки.
9. И.
- 10.Пытке у выбора быть.
- 11.Атом дата у запада еда.
- 12.Автор Тася у пасхи Уфа.
- 13.Ася и бабка у бати Ока.
- 14.Каска и Катя у капли шутя.
- 15.Этот Тэд тандэм кафэ.
- 16.Эпос мэтр цель не те.
- 17.Эти бэби у Кэтрин каратэ.
- 18.Сети Эдику и тесто Кэ.
- 19.Ода сотка и топка у депо.
- 20.Оба поприща у тёти в пальто.
- 21.Осень почки и косит легко.
- 22.Кофта у тётки Стёпы и плечо.
- 23.Утку и дудку у тубуса кладу.
- 24.Убыла будка и судит марабу.
- 25.Усик бублику и путать бедняку.
- 26.Густо и губы и салют во-всю.
- 27.Искры у сита чипсы и молчи.
- 28.Импульс тихий сытые кроты.
- 29.Изя и зычный сыпали грибы.
- 30.Атака октября у котят облава.
- 31.Опасный пятак гадалки отвага.
- 32.Осилить пятёрку от этапа ограда.
- 33.Этап в эпоху от Эдиты Ася.
- 34.Эпоха к этапу от эпохи о Боге.
- 35.Эдита в этап к Эдите Струвэ.
- 36.Исак к эпохе в Эдиту лучше.
- 37.Упал будильник числом на прачку.
- 38.У тёти в сибери гудели губы.
- 39.Тюльпан и люцерна бывают у шахты.
- 40.В бытовке любой дымок на тётю.
- 41.Иваны в Бутырку хотели.
- 42.Идея бубновая на охоту.
- 43.Удар губастый на бабу.
- 44.Быки и сырки дымят.
- 45.По нашему обыкновению прыгнули.
- 46.Хлопнув дверью по самому высокому.
- 47.Должен из-за эмоциональности извиниться.
- 48.Гордый загромождал Сибирь.
- 49.Обвинял семь тел и семь дел.
- 50.Завис в мозгу комар.
- 51.Объел все наши деньги остолоп.
- 52.Фирму повидал и обвесил Иванов.
- 53.Щука сожгла пожаром степь.
- 54.Художник в борьбе за идею молодец.
- 55.А в хате под горою ночь.
- 56.Одних заело гвоздь вбей.
- 57.Они сказали вбей здесь.
- 58.Отвлекался но отправился на верфь.

59. Вблизи справки спрятали донос.
 60. Невзрачного и скрытного не трожь.
 61. Отхватил сразу мужской носок.
 62. Открутил скважину и вскипел борщ.
 63. Первобытный институт врасплох.
 64. А перстень скользкий выбрось.

65. Компрессор умолкнет как барабан.
 66. Окрепший должник сел на мель.
 67. Бортмеханику скафандр накинь.
 68. Смертный грех зачеркнёт атом.
 69. Рекордных бутербродов ровно семь.

II 1.2. Текст для создания макси-набора аллофонов, диаллофонов и аллослогов русской речи

1. Шту+рман проси+л
продолжа+ть разворо+т.
2. Команди+р корабля+
одо+брил реше+ние.
3. Ско+ро начнё+тся небольшо+й
до+ждик.
4. Лё+тчик наблюда+ет сла+бые
оса+дки.
5. Руководи+тель потре+бовал
прекрати+ть поса+дку.
6. Взво+д пе+рвым форси+ровал
ре+ку.
7. Со+лнце ещё+ нахо+дится в
зени+те.
8. Весё+лые го+сти до+роги
хозя+ину.
9. Спортсме+н до+лжен
пробежа+ть диста+нцию.
10. Хи+мия и фи+зика -
интере+сные нау+ки.
11. Пере+днее колесо+ ре+зко
опу+щено.
12. Фи+льм снима+ют це+лый
го+д.
13. Самолё+т оказа+лся в
возду+шной я+ме.
14. Всю неде+лю иду+т дожди+.
15. Уче+бный истреби+тель
произвё+л взлё+т.
16. Апати+ты неда+вно нашли+ в
Хиби+нах.
17. руководи+тель разреши+л
произвести+ мане+вр.
18. Зи+мняя приро+да о+чень
краси+ва.
19. В шко+лу приезжа+ли геро+и
фрон+та.
20. Экипа+ж та+нка по+нял
зада+чу.
21. Этот бло+к рабо+тает
хорошо+.
22. Дире+ктор сравни+л дохо+д с
расхо+дом.
23. Испо+рченный конта+кт
нару+шил ци+кл.
24. Опера+тор стира+ет ста+рые
за+писи.
25. Кно+пки стоя+т на сигнала+е
"Сто+п".
26. Нала+живаю хо+д мото+ра
по схе+ме.

27. Мно+гие слы+шали
побе+дный во+зглас.
28. Автока+р бы+стро дви+жется
напра+во.
29. Зимо+вщик запусти+л
сло+жный механи+зм.
30. Монтё+р прове+рил но+вый
выключа+тель.
31. Ю+жное направле+ние
давно+ закры+то.
32. Я ви+дел это кре+сло.
33. Геро+и верну+лись домо+й с
побе+дой.
34. Совхо+з получи+л ще+дрый
урожа+й.
35. На реке+ появля+ются
пе+рвые льди+ны.
36. Ле+тний перехо+д был+л
завершё+н.
37. Фи+нны взя+ли но+вую
маши+ну.
38. Меха+ник вы+шел из
заводско+го гаража+.
39. Министе+рство
распределя+ет нау+чные
ка+дры.
40. Хи+мики награждены+
почё+тными ордена+ми.
41. Звоно+к разда+лся
соверше+нно неожи+данно.
42. Больно+й берё+г своё
здоро+вье.
43. О+бщее состоя+ние
генера+ла уху+дшилось.
44. Ма+ть отли+чно игра+ла на
пиани+но.
45. Игру+шки очень обра+довали
цели+нников.
46. Эта бесе+да всех
развесели+ла.
47. В институ+те я изуча+л
фи+зику.
48. Статью+ тру+дно бы+ло
измени+ть.
49. Дежу+рный принё+с
оде+жную щё+тку.
50. Зи+мний ве+чер был+л
ти+хий.
51. На поса+дочной полосе+
зелё+ные огни+.
52. Выполня+ю пра+вый
разворо+т.
53. Переда+йте сро+чно
све+дения.
54. Направле+ние держа+ть
пре+жнее.
55. Набери+те максима+льную
высоту+.
56. Впереди+ ви+дите це+ль.
57. Повтори+те после+дную
фигу+ру.
58. Ви+жу я+ркие ба+кены.
59. Ме+тод счита+йте ве+рным.
60. Фюзеля+ж гидропла+на
прове+рен.
61. За горо+й течё+т река+.
62. Спо+рт тре+бует упо+рства.
63. Спортсме+н смени+л
футбо+лку.
64. Снача+ла сдаю+т
матема+тику.
65. Офице+р о+тдал прика+з.
66. Арти+сты обеща+ют
конце+рт.
67. Сини+цы зиму+ют в
Эсто+нии.
68. Хи+мики прохо+дят
пра+ктику.
69. Гео+логи сдержали сло+во.
70. Лё+тчик ви+дит хорошо+.
71. Футболи+ст заби+л го+л.
72. Гидро+лог смотре+л
плоти+ну.
73. В кишлаке+ весё+лый
пра+здник.

74. Лё+тчицы выполня+ют
фигу+ры.
75. Мотори+ст слы+шит
поме+хи.
76. В степи+ была+ биоста+нция.
77. Гро+мко щё+лкнул
выключа+тель.
78. Дире+ктор поблагодари+л
слу+жащих.
79. Учени+к по+нял реше+ние
зада+чи.
80. Охо+тник уби+л лису+.
81. Ма+льчик слома+л са+нки.
82. Скаме+йка вы+крашена
кра+ской.
83. Гепа+рд бежи+т скачка+ми.
84. Шахтё+ры стуча+ли
молотка+ми.
85. На груди+ виси+т значо+к.
86. Сосе+дка щё+лкнула
две+рью.
87. Они зна+ли хи+мию.
88. Ю+ноша приобрё+л
профе+ссию.
89. Пироте+хник спа+с звере+й.
90. Це+х постро+или зимо+й.
91. Ве+тер поднима+лся
посте+нно.
92. Сте+пи изоби+луют
хи+щниками.
93. Ге+на разби+л стекло+.
94. Студе+нты взя+ли
блокно+ты.
95. Инстру+ктор показа+л
фюзеля+ж.
96. Штукату+р стуча+л в
пере+дней.
97. Узнаю+ кита+йские
орна+менты.
98. Идё+т разгру+зка
тра+нспорта.
99. Та+нк жда+л у доро+ги.
100. Отцу+ рассказа+ли всё+.
101. Маши+ны зажгли+ фа+ры.
102. Эсми+нец борозди+л
океа+н.
103. Ради+ст подгото+вил
приё+м.
104. Хи+мик уточни+л расчё+ты.
105. Раке+та сбива+ет це+ль.
106. Фи+зик прове+рил вво+д.
107. В гора+х поги+б ско+т.
108. Термо+метр пока+зывает
но+ль.
109. Сро+чно выполня+йте
сниже+ние.
110. Пило+т произвё+л взлё+т.
111. За две+рью послы+шался
го+лос.
112. Мя+ч мелька+л над се+ткой.
113. Ребё+нка мы+ли
ежедне+вно.
114. В за+ле звуча+л фокстро+т.
115. Э+дик ста+л гео+логом.
116. Нача+льники вошли+ в
це+х.
117. Ве+хи поста+влены
бы+стро.
118. Же+нщина рабо+тает на
фе+рме.
119. Зи+на получи+ла
назначе+ние.
120. Я+хту мо+жно перегна+ть.
121. Геро+й уча+ствует в бою+.
122. Учени+ца разби+ла
коле+но.
123. В расчё+тах допу+щена
оши+бка.
124. Пионе+ры разложи+ли
костё+р.
125. В магази+не vise+л
костю+м.
126. Тури+стов привлека+ло
Подмоско+вье.
127. Хе+к ме+чет икру+.

128. Фи+зик сде+лал
киноаппара+т.
129. Ё+лка стоя+ла у окна+.
130. Маши+на не вы+держала
перегру+зки.
131. Ма+сло покупа+ют в
магази+нах.
132. Всхо+ды семя+н поги+бли.
133. Химкомбина+т переезжа+ет
за го+род.
134. Оте+ц подари+л мя+ч.
135. Му+зыка греме+ла до утра+.
136. Шофё+р ведё+т маши+ну.
137. Щенки+ откры+ли глаза+.
138. Эта+п бы+л завершё+н.
139. Зада+чу объясни+ть
тру+дно.
140. Де+ло сде+лано за+ново.
141. В феврале+ шё+л сне+г.
142. Трудово+й го+д ко+нчился.
143. Това+рищи уезжа+ли в
Белору+ссию.
144. Жнивьё+ сгоре+ло от
взры+ва.
145. Зда+ния находи+лись в
це+нтре.
146. До+ждь стучи+т по окну+.
147. До+м пол+ностью
перестро+ен.
148. В гора+х цветё+т кизи+л.
149. Материа+л си+льно смя+т.
150. Те+хник погиб+л от ра+н.
151. Дружи+на собрала+сь в
ла+гере.
152. Ста+нция очень
живопи+сна.
153. Электроста+нции даю+т
то+к.
154. Секу+нда прошла+
незаме+тно.
155. Де+вочка оби+делась на
бра+га.
156. Смотре+ли интере+сный
фи+льм.
157. Фона+рь свети+л я+рко.
158. Охо+тник спугну+л за+йца.
159. Кирпи+ч подвози+ли
круглосу+точно.
160. Из музе+я похи+щены
це+нности.
161. Шко+льник разби+л
стекло+.
162. Сте+пь переходи+ла в ле+с.
163. Ма+ть рассказа+ла о
зимо+вке.
164. Я+хта остаё+тся в мо+ре.
165. Живо+тное привезли+
вчера+.
166. Ю+ноши стреля+ли ме+тко.
167. Спо+рт увлека+л офице+ра.
168. Стога+ ги+бнут от дожде+й.
169. В не+бе блесну+ли
звё+здочки.
170. Хи+щник охо+титися за
перепё+лкой.
171. Опера+ция прошла+
успе+шно.
172. Звоно+к разбуди+л щенка+.
173. Зда+ние филармо+нии
краси+во.
174. Здоро+вье отца+
беспоко+ило.
175. Ко+т поцара+пал ребё+нка.
176. Повсю+ду слы+шалось
э+хо.
177. В институ+те изуча+ют
язы+к.
178. Иску+сство возрожда+ется в
Яку+тии.
179. Связи+сты иска+ли
гео+лога.
180. Сапоги+ каза+лись
лё+гкими.
181. Сухове+й уме+ньшил
урожа+й.

182. У ю+нги широ+кие пле+чи.
 183. Конфе+ты подари+л дру+г.
 184. На заво+де появи+лась те+хника.
 185. Тишина+ воцари+лась в за+ле.
 186. Ве+тер помога+л перебе+жке.
 187. Ребё+нок хо+дит в я+сли.
 188. Танцо+вщица устро+илась в ци+рк.
 189. Пла+н утвержда+ют в о+бласти.
 190. Ди+ктора порази+ло сообще+ние.
 191. Трикота+жники вы+полнили пла+н.
 192. Тру+д все+м поле+зен.
 193. Фи+рма перее+хала в Я+лту.
 194. Милиционе+ры смени+ли фо+рму.
 195. Футбо+л привлека+ет боле+льщиков.
 196. Хо+р вы+ступил по телеви+дению.
 197. Худо+жники бегу+т в ателье+.
 198. Этот каранда+ш почи+нен.
 199. Подари+ли дороги+е духи+.
 200. Маля+р разли+л бели+ла.
 201. Велосипеди+сты вмеша+лись в борьбу+.
 202. Шко+льники разгова+ривали дру+жно.
 203. Корабли+ вы+строились у прича+ла.
 204. Эфи+р бы+стро выдыха+ется.
 205. Солда+т всегда+ аккура+тен.
 206. Студе+нты отдыха+ют ле+том.
 207. Ко+шка пойма+ла воробья+.
 208. Ве+село шага+ли гео+логи.
 209. На манеке+н па+дал све+т.
 210. В хи+жине ночева+ли зимо+вщики.
 211. Вдали+ беле+ла я+блоня.
 212. Храбре+ц идё+т в разве+дку.
 213. Эскадро+н пошё+л в ата+ку.
 214. В тайге+ обнару+жили ма+рганец.
 215. Вно+вь откры+лась больни+ца.
 216. Те+хники почини+ли ди+зель.
 217. Пото+к шуме+л впереди+.
 218. Всегда+ рабо+тает буфе+т.
 219. Хиру+рг вы+резал о+пухоль.
 220. Ли+вни освежа+ют ро+щу.
 221. Запреща+ю вхо+д в зо+ну.
 222. Уйди+те на второ+й кру+г.
 223. Выполня+йте пе+рвый полё+т.
 224. Сверя+ю ку+рс по схе+ме.
 225. Экипа+жу разреша+ю взлёт+.
 226. Зафикси+руйте коордina+ты ста+нции.
 227. Переда+йте радиогра+мму зени+тчикам.
 228. Береги+те гла+вный механи+зм.
 229. подгото+вьте инструме+нт бортмеха+нику.
 230. Фи+зику по+няли хорошо+.
 231. Боло+то ско+ро исче+зло.
 232. Зве+рь шёл на ю+г.
 233. Цеме+нт вози+ли на стро+йку.
 234. Хи+щник внима+тельно прислу+шался.
 235. Комсомо+л всегда+ впереди+.

236. Враги+ разби+ли го+род.
 237. Трава+ пробива+лась повсюду+.
 238. Яровы+е всхо+дят хорошо+.
 239. Ребя+та разожгли+ костё+р.
 240. Костю+м подчё+ркивал фигу+ру.
 241. Жи+знь водола+за спасли+.
 242. Ви+димость ста+ла хоро+шая.
 243. За+пуск осуществлё+н уда+чно.
 244. Кора+бль измени+л ку+рс.
 245. Земля+ изры+та я+мами.
 246. Каби+на пило+та освещена+.
 247. Хи+трый мане+вр удался+.
 248. Уточни+ли гра+фики вы+летов.
 249. Це+нтр разрабо+тал ши+фр.
 250. Экипа+ж про+сит взлё+т.
 251. Зелё+ный фона+рь закры+т.
 252. Це+ль ви+дно пло+хо.
 253. В пути+ произошла+ заде+ржка.
 254. Бомбардиро+вщик измени+л ку+рс.
 255. Истреби+тели лете+ли на аэродро+м.
 256. Ги+льзы напо+лнили по+рохом.
 257. капита+н проща+лся с кораблём+.
 258. Офице+р начертил+л схе+му.
 259. Пограни+чники охраня+ли рубежи+.
 260. Шту+рман следи+л за разворо+том.
 261. Ста+вка приняла+ реше+ние.
 262. Свя+зь рабо+тает норма+льно.
263. Прибо+ры зафикси+ровали землетрясе+ние.
 264. Расстоя+ние изме+рили пра+вильно.
 265. Поса+дочная полоса+ расчи+щена.
 266. Полёт+т оказа+лся нелё+гким.
 267. Поступи+ли це+нные све+дения.
 268. Орке+стр гря+нул ги+мн.
 269. Деса+нтники бежа+ли к ле+су.
 270. Майо+р бы+л хитё+р.
 271. Да+н прика+з на взлё+т.
 272. Экипа+ж нахо+дится в зо+не.
 273. Пило+т вы+пустил шасси+ самолё+та.
 274. Отря+д произвёл мане+вр.
 275. Солда+т до+лго це+лился.
 276. Хими+ческое ору+жие запрещено+.
 277. Ло+дку разби+ло прибо+ем.
 278. Начало+сь обледене+ние фюзеля+жа.
 279. Лётчик+ включил+л дви+гатель.
 280. Разда+лся сигнала+ трево+ги.
 281. Лётчику+ разреши+ли взлё+т.
 282. Вы+шел в расчё+тный квадра+т.
 283. Пу+нкт обозна+чен на схе+ме.
 284. Солда+т ждё+т офице+ра.
 285. Со+зданы бето+нные укры+тия.
 286. Приземли+лись в за+данном райо+не.
 287. Погруже+ние проходи+ло норма+льно.

288. Произведё+н за+пуск
раке+ты.
289. На миноно+сце спусти+ли
фла+ги.
290. Пехо+га отдыха+ла на
прива+ле.
291. Адьюта+нт засты+л на
поро+ге.
292. Ефре+йтор растяну+л
брезе+нт.
293. Самолё+т пошё+л на
сниже+ние.
294. Раке+тчики гото+вы к
прове+рке.
295. В ба+ках ко+нчился
бензи+н.
296. Координа+ты
зафикси+рованы на схе+ме.
297. Це+ль поражена+ то+чно.
298. Ку+рс бли+зок к
расчё+тному.
299. Пило+т подбежа+л к
вертолё+ту.
300. В частя+х начала+сь
демобилизация+ция.
301. Пара+д откры+ли
нахи+мовцы.
302. Зна+мя поло+щется на
ветру+.
303. Вы+полнен после+дний
вира+ж.
304. Железнодоро+жник
пропусти+л по+езд.
305. С бе+рега видны+ пози+ции.
306. Гимнастё+рки солда+т
вы+цвели.
307. Поса+дка была+
зафикси+рована.
308. Произведено+ мя+гкое
приземле+ние.
309. Перепра+ва начнё+тся в
ше+сть.
310. Самолё+том управля+л
пило+т.
311. Вертолё+т измени+л ку+рс.
312. В ча+сть прие+хал
журнали+ст.
313. Хи+мик наблюда+л за
прибо+ром.
314. Позывны+е вы+шли в
эфир.
315. Радиогра+мму переда+ли в
це+нтр.
316. Уче+ния начали+сь зимо+й.
317. Поса+дка осуществлена+
уда+чно.
318. На дня+х доста+вили
та+нки.
319. Машину+ запра+вили
бензи+ном.
320. Солда+ты занима+лись
гимна+стикой.
321. Самолё+т вошё+л в облака+.
322. На ре+йде стои+т
дебарка+дер.
323. Матро+с наде+л
бескозы+рку.
324. Гидролока+тор
зафикси+ровал препя+тствие.
325. Пара+д проходи+л на
пло+щади.
326. Эсми+нец патрули+ровал
зали+в.
327. Инжене+рная ча+сть
экипиро+вана.
328. При+был эшелон+ с
гру+зом.
329. Солда+ты взя+ли
винто+вки.
330. Часово+й слы+шал
шо+рохи.
331. Хими+ческое зараже+ние
опа+сно.
332. В полё+те случи+лась
ава+рия.

333. Подло+дка ушла+ в ре+йс.
 334. Бо+цман по+дал сигна+л.
 335. Орби+та близка+ к расчѐ+тной.
 336. В не+бо взмы+л перехва+тчик.
 337. Ради+ст прекрати+л свя+зь.
 338. Привезены+ балло+ны с кислоро+дом.
 339. Включи+ли ше+сть генера+торов.
 340. У самолѐ+та смя+т фюзеля+ж.
 341. Мело+дия понра+вилась все+м.
 342. На столе+ горе+ла свеча+.
 343. Гео+логи откры+ли минера+л.
 344. В хи+жину вошѐ+л лесни+к.
 345. Получи+ли це+нную бандеро+ль.
 346. Те+кст напеча+тали на маши+нке.
 347. Хозя+ин нашѐ+л щенка+.
 348. Пла+тье вы+шито би+сером.
 349. На подоко+ннике стои+т графи+н.
 350. В лесу+ замѐ+рз руче+й.
 351. В ла+гере объяви+ли каранти+н.
 352. В зоопа+рк привезли+ слона+.
 353. Архео+логи вели+ раско+пки.
 354. Ма+стер вошѐ+л в це+х.
 355. Де+ти собира+ли щаве+ль.
 356. В саду+ поспе+ла чере+шня.
 357. На столе+ лежа+ли гра+фики.
 358. Во дворе+ залива+ли каток+к.
 359. Медве+дь живѐ+т в лесу+.
360. Зимо+й ката+ются на лы+жах.
 361. Пионе+ры принесли+ герба+рий.
 362. Стари+к ню+хал таба+к.
 363. Зво+нко пе+л жа+воронок.
 364. На десе+рт пода+ли конфитю+р.
 365. На ме+льнице моло+ли зерно+.
 366. По+вар пригото+вил фрикаде+льки.
 367. На ве+тке сиде+л зя+блик.
 368. Черни+ку собира+ют в лесу+.
 369. О+бувь чи+стят щѐ+ткой.
 370. Парикма+хер сде+лал стри+жку.
 371. На факульте+те начали+сь заня+тия.
 372. Надвига+лся стра+шный цикло+н.
 373. Зако+нчился футбо+льный ма+тч.
 374. Молодѐ+жь прие+хала на фестива+ль.
 375. Появи+лись ма+йские жуки+.
 376. Же+нщина занима+лась хозяй+ством.
 377. Медве+ди спя+т в берло+ге.
 378. Откры+ли хими+ческий элеме+нт.
 379. Ба+нка сде+лана из же+сти.
 380. Отря+д постро+ился по кома+нде.
 381. Ба+бушка расска+зывала ска+зку.
 382. Пока+зывали но+вый фи+льм.
 383. Наступи+ла холо+дная зима+.

384. Разби+ли фарфо+ровую ча+шку.
385. Прилете+ли перелё+тные пти+цы.
386. Коми+ссия подвела+ито+ги.
387. Рыбаки+ вы+шли в мо+ре.
388. Де+ти игра+ли в мя+ч.
389. Инжене+р осма+тривал це+х.
390. Же+нщина поступи+ла хи+тро.
391. Не+бо затя+нуто ту+чами.
392. В лесу+ закрича+л фи+лин.
393. Хозя+йка испекла+ пироги+.
394. В па+рке би+л фонта+н.
395. Маши+на стоя+ла в гараже+.
396. Лаборато+рия сдала+отчё+т.
397. Письмо+ опусти+ли в я+щик.
398. Де+ти гото+вились к пра+зднику.
399. На столе+ лежа+ла конфе+тница.
400. В магази+не продава+лся хе+к.
401. Комсомо+льцы пое+хали на целину+.
402. Де+вочка звони+т по телефо+ну.
403. Брига+ду награди+ли вы+мпелом.
404. В саду+ цветё+т георги+н.
405. Ба+кенщик зажё+т мая+к.
406. В газе+те напеча+тали фельето+н.
407. Ше+фы пришли+ в колхо+з.
408. Биле+ты оказа+лись ста+рыми.
409. Агроно+м изме+рил уча+сток.
410. Студе+нты слу+шали ле+кцию.
411. На пло+щади проходи+л ми+тинг.
412. В саду+ жужжа+ли пчё+лы.
413. Ря+дом послы+шались шо+рохи.
414. Цыпля+та бе+гали по двору+.
415. Продаве+ц взя+л варе+нье.
416. На балко+не цвели+фиа+лки.
417. На плите+ закипе+л ча+йник.
418. Доро+ги вели+ в ро+щу.
419. Вы+пал пе+рвый сне+г.
420. Брю+ки бы+ли из ше+рсти.
421. Спортсме+н дости+г фи+ниша.
422. Этно+граф соста+вил ка+рту.
423. Де+вушки води+ли хорово+ды.
424. На у+лице зажгли+сь фонари+.
425. Под окно+м растё+т берё+за.
426. Ученики+ пойма+ли чижа+.
427. Бра+т изуча+ет физиоло+гию.
428. В тайге+ мно+го хи+щников.
429. Пионе+ры пе+ли пе+сню.
430. Заня+тия подошли+ к концу+.
431. Фотогра+фии получи+лись плохи+е.
432. В шкафу+ вися+т пиджаки+.
433. Метеоро+логи прогнози+руют пого+ду.
434. Рабо+чие забива+ют сва+и.
435. Дирижёр+р взмахну+л па+лочкой.

436. Из норы+ вы+скочила
куни+ца.
437. Бы+л интере+сный
конце+рт.
438. Шко+льники выра+щивали
цветы+.
439. У неё+ краси+вые глаза+.
440. В реке+ мно+го ры+бы.
441. Больно+й при+нял хини+н.
442. Де+ти собира+ли хво+рост.
443. Начался+ зи+мний сезо+н.
444. Профе+ссор привё+л
цита+ту.
445. Шахтё+ры зажгли+
фона+рь.
446. Гидроте+хники оказа+ли
по+мощь.
447. На сце+не стоя+ло
пиани+но.
448. На у+лице метё+т мете+ль.
449. Рабо+тают электро+нные
маши+ны.
450. Установи+лась хоро+шая
пого+да.
451. Цикло+н идё+т с се+вера.
452. В лесу+ щебе+чут пти+цы.
453. Фи+зики провели+
экспериме+нт.
454. На пра+здник прие+хали
го+сти.
455. В го+роде откры+лась
я+рмарка.
456. За кормо+й бушу+ет мо+ре.
457. На веса+х лежа+ли ги+ри.
458. Футболи+сты выи+грали
ма+тч.
459. Ма+льчики собира+ли
оре+хи.
460. Дни+ ста+ли коро+че.
461. Начался+ обе+денный
переры+в.
462. Зажгла+сь я+ркая ла+мпа.
463. В ва+зе лежа+ли фру+кты.
464. Боксё+р вы+шел на ри+нг.
465. Виногра+д уложи+ли в
я+щики.
466. Филатели+ст собира+ет
ма+рки.
467. У черепа+хи твё+рдый
па+нцирь.
468. В музе+е вы+ставлены
минера+лы.
469. На ста+нциях стоя+ли
эшело+ны.
470. Шко+льники сдава+ли
геоме+трию.
471. Начало+сь строи+тельство
ци+рка.
472. Обору+дован
физкульту+рный за+л.
473. В па+рке расцвё+л
шипо+вник.
474. Учё+ные вы+двинули
гипо+тезу.
475. Де+вочка перебежа+ла
доро+гу.
476. В пеще+ре послы+шалось
э+хо.
477. На стене+ висе+ли
фотогра+фии.
478. Откры+лась хими+ческая
лаборато+рия.
479. Ученики+ посети+ли
музе+й.
480. На не+бе сия+ли звё+зды.
481. Пусты+ня вы+жжена
со+лнцем.
482. В павильо+не бы+ло
прохла+дно.
483. Арти+ст пока+зывал
фо+кусы.
484. Тре+нер пригото+вил
секундоме+р.
485. Друзья+ зашли+ в кафе+.
486. Кро+лики щипа+ли траву+.

487. Колхо+зники убира+ют
чечеви+цу.
488. Состоя+лся пионе+рский
сбо+р.
489. Идѐ+т фина+льный
ко+нкурс.
490. В крови+ соде+ржится
гемоглоби+н.
491. Чемода+ны сда+ли в
бага+ж.
492. Из кни+ги вы+рвана
страни+ца.
493. Щенки+ вы+бежали во
дво+р.
494. Спортсме+н пригото+вил
ше+ст.
495. Ско+ро начну+тся
соревнова+ния.
496. На у+лице льѐ+т до+ждь.
497. Ду+ет си+льный ве+тер.
498. Для реа+кции необходи+м
фи+льтр.
499. Слома+лся хими+ческий
каранда+ш.
500. На поля+не росла+
земляника.

П 1.3. Мини-текст для создания основного набора просодических портретов русской речи

Утро первоклашки

Ночь. Темно. Машенька уснула.

Настало утро. Светает.

Машенька открыла глаза, потянулась, повернулась к окну, весело запела и радостно захлопала в ладошки.

- Ура, сегодня школа! Я уже большая, уже не маленькая! Ура! Мама, мама! Где мой портфель? Почему его нет на месте, он же был здесь? Ты его мне собрала? Ты положила в него пенал, тетрадки и букварь? Не забыла ли ты положить карандаши? А моя ручка на месте? Ленточки заплетать красные или синие?

- Красные или синие - это не важно, – сказала мама. Самое важное, что ты идёшь в школу. Запомни свои уроки: русский, арифметика, рисование и физкультура.

- Я вчера всё запомнила, а сегодня я всё забыла. Я всё забыла, потому что ты меня поздно разбудила.

Они пошли (школа была рядом) вдоль цветущего сада, заглядевшись на удивительные кустарники. Вскоре они увидели первоклашек, несущих огромные букеты цветов. Перwokлашки в новеньких платьицах казались окрылёнными чистым чувством впервые за всю свою жизнь быть такими значительными людьми с большой буквы.

И всё-же... Впереди - десять классов школы, наполненных и двойками, и вызовом родителей, и первыми успехами; затем - университет, занятия, первая любовь; затем - выход в самостоятельную жизнь, серьёзные решения, взлёты и падения.

Это было лишь начало их долгого жизненного пути...

II. 1.4. Макси-тексты для создания расширенного набора просодических портретов русской речи

(1) Семён Альтов. Кто там?

Галя еще раз проверила, закрыла ли она окна, спички спрятала и, присев у зеркала, говорила, отделяя слова от губ движениями помады:

- Светочка, мама пошла в парикмахерскую... Позвонит приятный мужской голос, скажешь: "Мама уже вышла". Это парикмахер... Позвонит противный женский голос, спросит: "А где Галина Петровна?" Это с работы. Скажешь: "Она пошла в поликлинику... выписываться!" Не перепутай. Ты девочка умненькая. Тебе шесть лет.

- Будет семь, - поправила Света.

- Будет семь. Помнишь, кому можно открывать дверь?

- Помню, - ответила Света. - Никому.

- Верно! - Галя облизнула накрашенные губы. - А почему нельзя открывать, не забыла?

- Бабушка говорит: "По лестнице нехорошие бандиты с топорами ходят, прикидываются водопроводчиками, тетями, дядями, а сами распиливают непослушных девочек и топят в ванне!" Правильно?

- Правильно, - сказала Галя, прикалывая брошку. - Бабушка хоть и старенькая, руки дрожат, посуду всю перебила, но про бандитов верно долдонит... недавно в одном доме три водопроводчика пришли чинить телевизор. Мальчик открыл...

- А они его топором? - подсказала Света.

- Если бы, - пробормотала Галя, пытаясь застегнуть брошку. - В ванне утопили и все вынесли.

- И ванну?

- Ванну оставили.

- А бабушка придет, ей открыть? - спросила Света, откручивая кукле ногу.

- Бабушка не придет, она на даче. Приедет завтра.

- А если сегодня?

- Я сказала, завтра!

- А если сегодня?

- Если сегодня, это уже не бабушка, а бандит! По домам ходит, деток ворует. Куда я пудру сунула?

- А зачем детей воровать? - Света отвернула кукле ногу и теперь приворачивала ее обратно. - У бандитов своих нету?

- Нету!

- А почему нету?

- Почему, почему? - Галя тушью сделала реснички. - Потому что, в отличие от твоего папочки, хотят что-то в дом принести! Некогда им! Еще есть дурацкие вопросы?

- Нету вопросов! - ответила Света, внимательно глядя, куда мать прячет от нее французские духи.

- Вроде порядок, - Галя цепким глазом таможенника ощупала отражение в зеркале.

- Буду часа через два. Нет, через три!

- Так долго обстригать будут? Ты же не слон!

- Не обстригать, а стричь! Это плохой мастер все делает тяп-ляп, а хороший мастер, - Галин голос потеплел. - Настоящий мастер все делает хорошо, поэтому долго. Никому не открывать!

Мать чмокнула Свету и, хлопнув дверью, ушла. Света достала из тумбочки французские духи, полфлакона опрокинула кукле на голову, приговаривая:

- Вымоем Дашке голову и будем обстригать. Не волнуйтесь, настоящий мастер все делает так долго, пока вам не станет хорошо!

Тут раздался звонок в дверь. Света побежала в прихожую и звонко спросила:

- Кто там?

Хриплый голос ответил:

- Открой! Это я, твоя бабушка.

- Здравствуй, бабушка! А зачем таким страшным голосом говоришь?

- Да простыла, внученька! Уже и молоко с медом пила, а все хриплю.

Открывай! Вишенки вкусеньких привезла!

Света наполовину открыла трудный замок, но вдруг, наморщив лобик, остановилась:

- Бабушка! Мама сказала: ты завтра приедешь! А сегодня еще сегодня.

- А я сегодня и приехала! Открой! Темно на лестнице, и ноги болят!

Света набросила на дверь цепочку.

- Бабушка, - задумчиво сказала она через дверь. - Я открою, а ты бандит?

- Какой еще бандит?! - бабушка закашлялась.

- Обыкновенный. Сама говорила, прикинется бабушкой, а сам распилит и в ванне утопит. Приезжай завтра, будешь бабушкой!

Старушка опустила на ступеньку, заплакала:

- Во, дите воспитали! Родной бабке через дверь не верит! Надо людям верить! Когда мать с отцом будут?

- Папа после работы, - донеслось из-за двери. - А мама пошла выписываться к парикмахеру.

- Куда? - бабушка вскочила. - Все Сереже расскажу! Вертихвостка! И ты вся в нее, вся! Вот возьму и умру тут!

- Бабуль! Бабуль, - пробивался из-за двери детский голосок. - Ты не умрешь! Мама сказала, ты сначала нас всех похоронишь!

- Это Галка про меня такое сказала? Змеюка! Все Сереже расскажу! Про всех парикмахеров! Еще неизвестно, от какого парикмахера дочь!

В это время по лестнице подымался мужчина в сапогах и спецовке. Разглядев в тусклом свете умирающей лампочки, старушку в слезах, он остановился:

- Кого оплакиваем, бабуля?

Признаться постороннему, что тебя не пускает в дом собственная внучка, было так стыдно, что бабушка, проглотив слезу, соврала:

- Давление у меня пониженное, сынок... Вторые сутки с лестницы падаю.

- А мы в квартиру позвоним, валерьяночки хлопнешь! - весело сказал мужчина, нащупывая на двери звонок.

- Кто там? - спросила Света.

- Ребяенок, открой! Тут старуха концы отдаст!

- Дядя! Там темно, вы потрогайте - эта старуха, может быть, старый бандит!

- Мерзавка! - взвыла бабушка. - Перед людьми не позорь!

- А-а, так там внучка твоя оказалась? - сообразил мужчина. - Чья бы внучка ни была, а отпереть будь любезна! Слышь меня, стерва несовершеннолетняя!

- А при детях ругаться нельзя, - сказала Света. - Папа при мне никогда не ругается. Сначала уложит спать, дверь закроет, и потом ругаются с мамой! Понял, сын сукин?

- Во, дает! - одобрительно хмыкнул мужчина.

- А вы там кто такой? Один бабушка, второй дедушка, что ли?

- Я-то? Я дядя Коля водопроводчик...

Бабушка, ладошкой заложив мужчине рот, зашипела:

- Не водопроводчик! Только не водопроводчик! Ей про водопроводчиков такого наговорили! Вы... почтальон!

Дядя Коля пытался оторвать от себя бабушку, бранился шепотом:

- Чтоб вы сгорели! Почему водопроводчиками пугаете? У нас что, почтальон не может стать бандитом? По Конституции...

- Но я прошу вас, - ныла бабушка, - скажите, почтальон, она откроет!

Дядя Коля сплюнул в сердцах:

- Слышь ты там! Открывай! Оказывается, я почтальон!

- А голос, как у водопроводчика!

- Бабусь, внучка воспитана крепко. Граница на замке! Придется ломать дверь.

- Ломайте! - бабушка махнула рукой. - Только аккуратно, как свою.

Водопроводчик достал инструменты и, напевая романс - "Отвори потихоньку калитку" - начал выламывать дверь. Удары кувалды гулко бухали на всю лестницу. За это время шесть человек тихими мышками прошмыгнули по лестнице. Но, во-первых, соседи плохо знали друг друга в лицо. Во-вторых, на площадке был полумрак. А в-третьих, как-то неловко спрашивать у незнакомого человека, в свою квартиру он ломится или в чужую. Видя, как дверь начинает шататься, Света заплакала:

- Мама! Мамочка! Меня утопят!

Она дрожащей рукой задвинула засов старого замка, которым давно не пользовались, но с двери так и не сняли. По лестнице, насвистывая, взбегал Светин папа. Увидев в полутьме стоящих у его двери, Сергей с ходу заехал водопроводчику в ухо.

- Сереженька, не бей! Это свои! - завопила бабушка и кинулась разнимать. Мужчины метили друг в друга, но в темноте в основном все доставалось бабушке. Когда старушка была положена на обе лопатки, мужчины успокоились и начали приводить ее в чувство. Наконец, все очухались, помирились и, потирая ушибленные места, уставились на дверь.

- Света, открой, деточка! - простонал Сергей, держась за скулу.

- Сейчас, папочка, - ответила Света. - А ты, правда, мой папа?

- Другого папы у тебя пока нет!

- А бабушка говорит, я от какого-то парикмахера получилась...

- От какого парикмахера?!

- Сереженька! - бабушка в темноте выразительно посмотрела на сына. - Ты его не знаешь! Я все расскажу, если попадем в квартиру!

В это время на площадке остановился пухлый мужчина. Переводя дух, он сказал:

- Бог в помощь! А что с дверью делаете?

- Видите ли... - замаялся Сергей, - сигнализацию ставим. Мало ли...

- Ага, - ухмыльнулся водопроводчик. - Можно вызвать на дом, а можете сами дверь в милицию отволочь. Дешевле обойдется!

- Понял! - пухлый стал быстро подниматься по лестнице.

Сергей тряхнул дверь:

- Светочка, открой немедленно, гадина! У меня ключи, отопру - выпорю!

- Выпорешь, если отопрешь! - Света вздохнула. - Я на старый замок закрыла, от него ключей ни у кого нету, даже у папы, если он - это вы.

- Ну, что делать? - Сергей закурил. - Деньги мне надо заработать! Деньги! На двух держат цветной телевизор, маленький, за двести рублей.

- Как двести? Он четыреста стоит! - удивился дядя Коля. - Да цельнотянутый! Прямо с завода!

- Что такое "цельнотянутый"? - спросила бабушка.

- Ворованный, утянутый, значит, - объяснил водопроводчик. - Цельнотянутый грех не взять, грех!

- А мой папа говорил: "Воровать нехорошо!" - произнес за дверью детский голос. - Значит, точно там бандиты собрались!

- Нехорошо родному отцу дверь не открывать! - крикнул дядя Коля. - Дура глупая!

В это время наверху что-то гроыхнуло, охнуло и, ругаясь, покатилося вниз. Это были супруги Бирюковы из пятьдесят седьмой. Они волокли дубовую дверь. Тяжелая дверь неслась быстрее Бирюковых, била их о стены, перила, мотая из стороны в сторону.

- Что случилось? - успел крикнуть водопроводчик.

- Сказали: сегодня всем поставить сигнализацию! Завтра бандитов ждут!.. - Дверь уволокла Бирюковых вниз.

- Сумасшедшие! - бабушка покачала гоовой. - Уже слух пустили с вашей легкой руки про сигнализацию! Сейчас все двери посрывают!

И, действительно, наверху что-то грохнуло. Потом еще раз грохнуло в другом месте. Люди рвали двери с петель.

- Света! - Сергей постучал в дверь кулаком. - Слушай внимательно! Мама сказала, чтобы ты никому не открывала дверь, и ты молодец, что слушаешься, дрянь! Но про то, что деньги просовывать под дверь нельзя, мать ничего не говорила! А если человек, не заходя в дом, скажет, где у вас лежат деньги, значит, он кто?

- Вор! - ответила Света.

- Идиотка! Он твой отец!

- Были бы моим папой, знали бы, денег у нас нет! Папа все время маме кричит: "Нет у них, я не ворую!" Никак бандитом не может устроиться!

Сергей стукнул головой в дверь:

- Светочка! У меня в копилке припрятано! Клянусь тебе, в ванной. За ведром. В мыльнице. Под мылом лежат деньги! Вынь, просунь бумажки под дверь!

За дверью было тихо. Наконец, послышались Светины шажки:

- Бандиты, вы здесь?

- Здесь мы, здесь, доченька! Просовывай!

- Там нету мыльницы с деньгами, только папины носки. Сувать?

- Украли! - охнул Сергей. - В кои-то веки в доме появились деньги, сперли! Бандиты пронюхали! А может, они там? Дядя Коля, навалились.

Мужчины прыгнули на дверь и вместе с ней рухнули в квартиру...

Вечером вся семья и дядя Коля ужинали, смотря новенький телевизор. Дверь уже поставили на место, с водопроводчиком расплатились, и он, возбужденный червонцем, хвалил хозяйку.

- Что ж ты папочку обманула, доченька, - перебила его Галя, - сказала "мыльницы нет"?

- Испугалась! Бандиты узнают, деньги есть, и дверь выломают. А они все равно разломали!

Все дружно рассмеялись.

- Соображаешь! Молодец! - сказал водопроводчик, укладывая на хлеб девятый кружок колбасы. - Но запомни: обманывать старших нехорошо! Сначала надо вырасти, стать человеком!..

За стеной временами слышались стоны и треск. Это соседи волокли двери на установку сигнализации.

- Ничего не понимаю, - сказала Галя. - Тащат и тащат! Может, правда? Весь дом засигнализируется, а мы опять, как дураки!

- На какие, позвольте, шиши?! - вскинулся Сергей. - Я и так весь в долгах!

- Папочка, заплати, а то меня в ванне утопят!
- Да где ж я деньги возьму, доченька?
- Я знаю, где, - сказала Света, - у бабушки на антресолях полваленка денег набито!

- Врунья бессовестная! - завопила бабушка. - Какие полваленка? Там еле подошва прикрыта! При моей-то пенсии...

- Извини, бабушка, я во втором валенке не глядела. Выходит, это у мамочки полваленка денег... А у папы в мыльнице кусок мыла остался - долги раздать.

Наступила тишина. Разряжая обстановку, водопроводчик дипломатично высморкался и сказал:

- Я извиняюсь. Домой надо. Девять часов... Пора сына пороть. Ведь кроме меня у него никого нет. Это у вас нормальная семья. Счастливо оставаться!

(2) Текст для славяноговорящих

Был тихий серый вечер. Дул ветер, слабый и тёплый. Небо было покрыто тучами, сквозь которые прорывались лучи заходящего солнца. Наш автобус номер 7 шёл на запад. Мы все ехали в великий старый город, электрические огни которого были видны далеко впереди.

По центральному радио передавали лёгкую музыку. Хор ребят исполнял песенку "Золото осени". Я невольно прислушивался к разговорам в автобусе. Сосед-моряк, не раз бывавший за границей, рассказывал о различных странах, о музеях, о газетах. В них защищают всеобщее равенство, говорят о нужде и судьбе обыкновенных рабочих.

Тема была мне знакома: у нас достаточно много пишут о внешних связях, необходимых для механизации в объединениях заводов.

Впереди меня двое учёных говорили о предстоящей конференции: "Наша лаборатория семь лет отдала эксперименту по проверке этой формулы. Нам было неизвестно значение функции сопротивления, особенно при высоком давлении над поверхностью жидкости. Опыты показали, что оно равно десяти в пятой степени условных единиц."

"Странно, -ответил его товарищ,- а наш преподаватель Тревогин в лекции называл другую цифру. Очевидно, это вызывается неравномерным размещением плотности составляющих. Я думаю,

теперь надо изучить соотношение длины и объёма."

Их разговор был совсем непонятным для меня. Я обратил внимание на женщину, которая уже не в первый раз сажала на сиденье своего малыша. Рыжий мальчик, которому было года три, отворачивался, вырывался, показывал маме розовый язык. Я решил, что мама сейчас рассердится. Но она, стараясь сохранять строгий вид, не могла не улыбаться. Было видно, что она очень любит своего сына, считает, что лучше нет детей на свете: он самый умный, самый милый, самый весёлый. Юный шалун мог позволить себе многое.

Я отыскал в дорожной сумке книгу о животных, которую купил для дочки, вынул её, открыл. На первой странице по голубому небу летели голуби.

"Дядя, это кто?" - спросил малыш, заглядывая на следующую страницу.

"Это кит."

"А ты знаешь, сколько килограммов он весит? Сто? Сколько сантиметров в секунду он проплывает? Двадцать? Кит сильный? Он ест мясо? А мёд? Кто его нарисовал? Художник?"

Вопросы следовали один за другим. Мы привлекли внимание пассажиров. Все с улыбкой следили за нами. Наконец, мама сказала:

"Шурочка, душа моя, уже поздно, скоро ночь. Здесь нельзя шуметь. Пойдём спать. Сядь сюда."

Сначала Шурочка сопротивлялся, тащил маму в сторону. Потом затих, держа в руке жёлтого пёсика. Его дыхание было чуть слышно.

Я выключил лампочку дневного света. Вид спящего ребёнка вызвал во мне воспоминания о моём детстве. Тогда беда ходила рядом, и даже необходимый для жизни хлеб выдавали граммами. Потом вспомнил об учёбе в вузе, о первых годах на фабрике. Там я понял о себе главное:

кроме интересной работы, мне мало что нужно в жизни. Так, задумавшись, я не заметил, как заснул. Город - цель нашего пути - был ещё не близко...

(3) Текст для неславяноговорящих

Я живу в самом красивом местечке на свете – таково мое мнение. Ищи, лучше не найдешь! Посмотреть наш городок приезжали три моих друга. Один, приехавший вместе с женой – красивой Анной с голубыми глазами, шестой год живет на целине. Другого затянули научные конференции, бюро, и только недавняя болезнь позволила ему пообщаться с нами пять-шесть дней. Слабость его еще чувствовалась:

он часто ругал пески на набережной, был недоволен жарой и привычно уселся на кочке, когда мы остановились.

Я решил показать гостям море. Город наш расположен в горах. Ехать на автобусе до моря часа четыре. Но достаточно увидеть вдали скользящие по поверхности белые, розовые, зеленые катера, эту водную ширь без пределов – и сразу легка жизнь!

Ранним утром было особенно тихо вблизи воды. Только невидимая нам сверху Ольга кричала невидимой Даше что-то про форму и объем.

Когда мы подошли поближе, их разговор стал слышнее. Можно было разобрать отдельные фразы.

- Не знаю, перевыполнит ли завод план в этом квартале.

- А что позавчера на совещании мастеров говорили? Какой был ответ?

- Что уже недалеко до ста процентов. Осталось только ваш цех подтянуть. Поторопитесь, не тяните время.

- Постараемся. Вчера уже железо привезли, ямы на центральном дворе вырыли. Тяжело пришлось.

- Ладно, Ольга, будни у всех такие.

- Да, а ты ведь квартиру получила! Довольна?

- Довольна.

- Когда переехали? Давно?

- Во вторник и переехали. Всем хороша квартира, только на кухне негде повернуться. И над дверью щель очень большая, заделать надо, а я сама и гвоздь вбить не могу.

Женщины еще говорили о том, какая цена на новый гарнитур и сколько стоил старенький. Их лодка тихо отплыла от берега, только легкая рябь прошла по волнам.

В знойный полдень лучи солнца доставали отдыхающих всюду. Единственным прохладным уголком (условно прохладным) была тень от яблони, непонятно как выросшей здесь. Поверхность ее ствола, отшлифованная постоянными ветрами, была ровной и гладкой, как карандаш.

Мы медленно шли по берегу и вдруг увидели разошедшийся плот. Громадина его темнела на желтоватом песке и привлекала внимание еще издали. Сразу представилось, как рыбак, когда нет клева, ляжет на свой тулуп, сдвинет на бровь шапку, закинет руки за голову, прищурится на солнце и постепенно засыпает... Солнце согревает его, поверхность плота становится обжигающей, но ничего не чувствует наш рыбак – любит поспать в жаркое время, вытянувшись во весь рост. Увидишь издали и подумаешь: "Жив ли?"

Долго могли бы служить крепкие доски плота, но эпоха уже не та. Теперь в море на лов выходят устойчивые, надежные суда.

А море перестает быть спокойным. Переменный ветер гонит по волнам солнечные шары, похожие на яркие огни рампы. Набежавшая волна, как изящная плясунья, подлетает к кому-нибудь из купальщиков и, едва коснувшись его ног, устремляется опять в море.

Мы уезжаем домой, а вечером идем к моему дяде, человеку очень ровному, у которого редко падает настроение. У него еще с деревенских времен живет кот, да не один. Хоть и не злые животные, а часто царапаются. Как цари ходят они по квартире, занимают уголок у теплой стены и едят исключительно мясо.

Сейчас память у дяди как сито, но рассказы его всегда интересны – не держит он свои знания за семью печатями. Мы входим в комнату и слышим приятный бас:

"Здравствуйте! Вы меня помните? Давно не встречались! Лет девять прошло, да?"

Я сразу узнаю этого человека – сущность его в любом случае не меняется. Основное его качество – страстная любовь к жизни. С зарядом его энергии это и не удивительно!

- Жизнь прибавила тебе разума? К счастью как относишься? О чем-нибудь жалеешь? Где бываешь? Какие книги прочел?

Меня засыпают вопросами. Щурится кот. Начинается долгий разговор-воспоминание...

Приложение 2

Таблицы генерации мини-набора аллофонов русской речи

Каждый из генерируемых аллофонов обозначается именем фонемы и следующими за ним тремя целочисленными индексами: *i, j, k*. Индекс *i* для гласной указывает её позицию по отношению к словесному ударению и может принимать одно из следующих значений: 0 – полноударный, 1 – частично-ударный, 2 – гласный первой степени редукции, 3 – гласный второй степени редукции. Индекс *i* для согласной указывает удвоение фонемы и может принимать одно из следующих значений: 0 – обычная фонема, 1 – удвоенная фонема. Индекс *j* для гласных и согласных указывает группу левого контекста; индекс *k* – группу правого контекста.

П 2.1. Аллофоны гласных фонем

Возможные левые и правые контексты (с указанием соответствующих индексов) для гласных различной степени редукции приведены в таблицах П 2.1 – П 2.4. Поскольку группы левых контекстов гласных не всегда совпадают для ситуаций внутри слова и на стыках слов, в таблицах помечен тип левого контекста: внутрисловный и/или межсловный.

Таблица П 2.1

Аллофоны полноударных гласных

Тип левого контекста		Левый контекст	A ₀ , E ₀ , O ₀ , U ₀				I ₀				Y ₀			
			Правый контекст				Правый контекст				Правый контекст			
Межсловный	Внутрисловный		Пауза	Губные, U, O	Зубн., з/язычн., A, E	Мягкие, I	Пауза	Губные, U, O	Зубн., з/язычн., A, E	Мягкие, I	Пауза	Губные, U, O	Зубн., з/язычн., A, E	Мягкие, I
-	+	Пауза	00	01	02	03	00	01	02	03	-	-	-	-
+	+	Губные, U, O	10	11	12	13	10	11	12	13	10	11	12	13
+	+	Зубные, A, E	20	21	22	23	20	21	22	23	20	21	22	23
+	+	З/язычн.	30	31	32	33	-	31	32	33	-	-	-	-
-	+	Мягкие, I	40	41	42	43	40	41	42	43	-	-	-	-
+	-	Мягкие	-	51	52	53	-	51	52	53	-	-	-	-

Таблица П 2.2

Аллофоны частичноударных гласных

Тип левого контекста		Левый контекст	А ₁ , Е ₁ , О ₁ , U ₁				I ₁				Y ₁			
Межсловный	Внутрисловный		Правый контекст				Правый контекст				Правый контекст			
			Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, I	Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, I	Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, I
-	+	Пауза	-	01	02	03	-	01	02	03	-	-	-	-
+	+	Губные, U, O	10	11	12	13	10	11	12	13	10	11	12	13
+	+	Зубные, А, Е	20	21	22	23	20	21	22	23	20	21	22	23
+	+	З/язычн.	30	31	32	33	-	31	32	33	-	-	-	-
-	+	Мягкие, I	40	41	42	43	40	41	42	43	-	-	-	-
+	-	Мягкие	-	51	52	53	-	51	52	53	-	-	-	-

Таблица П 2.3

Аллофоны гласных первой степени редуции

Тип левого контекста		Левый контекст	А ₂ , Е ₂ , U ₂				I ₂				Y ₂			
Межсловный	Внутрисловный		Правый контекст				Правый контекст				Правый контекст			
			Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, I	Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, I	Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, I
-	+	Пауза	-	01	02	03	-	01	02	03	-	-	-	-
+	+	Губные, U, O	10	11	12	13	10	11	12	13	10	11	12	13
+	+	Зубные, А, Е	20	21	22	23	20	21	22	23	20	21	22	23
+	+	З/язычн.	30	31	32	33	-	31	32	33	-	-	-	-
-	+	Мягкие, I	40	41	42	43	40	41	42	43	-	-	-	-
+	-	Мягкие	-	51	52	53	-	51	52	53	-	-	-	-

Таблица П 2.4

Аллофоны гласных второй степени редукции

Тип левого контекста			Аз, Ез, Уз				Із				Уз			
			Правый контекст				Правый контекст				Правый контекст			
Межсловный	Внутрисловный	Левый контекст	Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, І	Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, І	Пауза	Губные, U, O	Зубн., з/язычн., А, Е	Мягкие, І
+	+	Губные, U, O	-	11	12	13	-	-	-	-	-	11	12	13
+	+	Зубные, А, Е	-	21	22	23	-	-	-	-	-	21	22	23
+	+	З/язычн.	-	31	32	33	-	-	-	-	-	-	-	-
-	+	Мягкие, І	-	41	42	43	-	41	42	43	-	-	-	-

П 2.2. Аллофоны согласных фонем

Списки возможных внутрисловных и межсловных аллофонов (с указанием соответствующих индексов) согласных, различных по способу образования, приведены в таблицах П 2.5 – П 2.8.

Таблица П 2.5

Аллофоны глухих согласных

Правый контекст	Пауза		Любая фонема или вспомогательный символ, кроме паузы	
	Одиночные аллофоны	Удвоенные аллофоны	Одиночные аллофоны	Удвоенные аллофоны
Левый контекст				
Любая фонема или вспомогательный символ	P ₀₀₀	-	P ₀₀₁	P ₁₀₁
	P' ₀₀₀	-	P' ₀₀₁	P' ₁₀₁
	T ₀₀₀	T ₁₀₀	T ₀₀₁	T ₁₀₁
	T' ₀₀₀	-	T' ₀₀₁	T' ₁₀₁
	-	-	K' ₀₀₁	K' ₁₀₁

Продолжение таблицы П 2.5

Любая фонема или вспомогательный символ	C ₀₀₀	-	C ₀₀₁	C ₁₀₁
	CH' ₀₀₀	CH' ₁₀₀	CH' ₀₀₁	CH' ₁₀₁
	F ₀₀₀	F ₁₀₀	F ₀₀₁	F ₁₀₁
	F' ₀₀₀	-	F' ₀₀₁	F' ₁₀₁
	S ₀₀₀	S ₁₀₀	S ₀₀₁	S ₁₀₁
	S' ₀₀₀	-	S' ₀₀₁	S' ₁₀₁
	SH ₀₀₀	-	SH ₀₀₁	SH ₁₀₁
	SH' ₀₀₀	-	SH' ₀₀₁	SH' ₁₀₁
	-	-	H' ₀₀₁	H' ₁₀₁

Таблица П 2.6

Аллофоны твёрдых аспиративных согласных

Правый контекст	Пауза		У, О		Любая фонема или вспомогательный символ, кроме паузы, У, О	
	Одиноч аллоф.	Удвоен аллоф.	Одиноч аллоф.	Удвоен аллоф.	Одиноч аллоф.	Удвоен аллоф.
Левый контекст						
Любая фонема или вспомогательный символ	K ₀₀₀	-	K ₀₀₁	K ₁₀₁	K ₀₀₂	K ₁₀₂
	-	-	G ₀₀₁	G ₁₀₁	G ₀₀₂	G ₁₀₂
	H ₀₀₀	-	H ₀₀₁	H ₁₀₁	H ₀₀₂	H ₁₀₂

Таблица П 2.7

Аллофоны звонких взрывных, щелевых и сонорных согласных

Правый контекст	Пауза		Звонкие согласные		Гласные		Глухие согласные	
	Одиночные аллофоны	Удвоенные аллофоны						
Левый контекст								
Любая фонема или вспомогательный символ	-	-	B ₀₀₁	B ₁₀₁	B ₀₀₂	B ₁₀₂	-	-
	-	-	B' ₀₀₁	B' ₁₀₁	B' ₀₀₂	B' ₁₀₂	-	-
	-	-	D ₀₀₁	D ₁₀₁	D ₀₀₂	D ₁₀₂	-	-
	-	-	D' ₀₀₁	D' ₁₀₁	D' ₀₀₂	D' ₁₀₂	-	-
	-	-	G' ₀₀₁	G' ₁₀₁	G' ₀₀₂	G' ₁₀₂	-	-
	-	-	Z ₀₀₁	Z ₁₀₁	Z ₀₀₂	Z ₁₀₂	-	-
	-	-	Z' ₀₀₁	Z' ₁₀₁	Z' ₀₀₂	Z' ₁₀₂	-	-
	-	-	ZH ₀₀₁	ZH ₁₀₁	ZH ₀₀₂	ZH ₁₀₂	-	-
	L ₀₀₀	L ₁₀₀	L ₀₀₁	L ₁₀₁	L ₀₀₂	L ₁₀₂	L ₀₀₃	L ₁₀₃
	L' ₀₀₀	-	L' ₀₀₁	L' ₁₀₁	L' ₀₀₂	L' ₁₀₂	L' ₀₀₃	L' ₁₀₃
	M ₀₀₀	M ₁₀₀	M ₀₀₁	M ₁₀₁	M ₀₀₂	M ₁₀₂	M ₀₀₃	M ₁₀₃
	M' ₀₀₀	-	M' ₀₀₁	M' ₁₀₁	M' ₀₀₂	M' ₁₀₂	M' ₀₀₃	M' ₁₀₃
	N ₀₀₀	N ₁₀₀	N ₀₀₁	N ₁₀₁	N ₀₀₂	N ₁₀₂	N ₀₀₃	N ₁₀₃
	N' ₀₀₀	-	N' ₀₀₁	N' ₁₀₁	N' ₀₀₂	N' ₁₀₂	N' ₀₀₃	N' ₁₀₃
	R ₀₀₀	-	R ₀₀₁	R ₁₀₁	R ₀₀₂	R ₁₀₂	R ₀₀₃	R ₁₀₃
	R' ₀₀₀	-	R' ₀₀₁	R' ₁₀₁	R' ₀₀₂	R' ₁₀₂	R' ₀₀₃	R' ₁₀₃

Таблица П 2.8

Аллофоны плавных сонорных согласных

Правый контекст	Пауза		Ударные гласные		Безударные гласные		Согласные	
	Одиночные аллофоны	Удвоенные аллофоны						
Левый контекст								
Согласные и пауза	-	-	V ₀₀₁	V ₁₀₁	V ₀₀₂	V ₁₀₂	V ₀₀₃	V ₁₀₃
	-	-	V' ₀₀₁	V' ₁₀₁	V' ₀₀₂	V' ₁₀₂	V' ₀₀₃	V' ₁₀₃
	-	-	J' ₀₀₁	-	J' ₀₀₂	-	-	-
Гласные	-	-	V ₀₁₁	V ₁₁₁	V ₀₁₂	V ₁₁₂	V ₀₁₃	V ₁₁₃
	-	-	V' ₀₁₁	V' ₁₁₁	V' ₀₁₂	V' ₁₁₂	V' ₀₁₃	V' ₁₁₃
	J' ₀₁₀	-	J' ₀₁₁	J' ₁₁₁	J' ₀₁₂	J' ₁₁₂	J' ₀₁₃	-

Приложение 3

Статистика фонем, дифонем, фонослогов и их аллофонических вариантов

В приложении приведены таблицы нормированных частот встречаемости речевых элементов различного фонетического качества в макси-тексте (см. Приложение 1), используемого для создания расширенного набора аллофонов, диаллофонов и аллослогов русской речи. Для всех типов элементов, кроме фонем, в соответствующие списки включены элементы, нормированная частота встречаемости которых – не ниже 10%.

П 3.1. Фонемы, дифонемы, фонослоги

Нормированная частота встречаемости фонем, дифонем (последовательности двух фонем) и фонослогов (фонетических слогов) приведена в таблицах П 3.1 – П 3.3.

Таблица П 3.1

Нормированная частота встречаемости фонем

№ пп	Имя фонемы	Частота, %	№ пп	Имя фонемы	Частота, %	№ пп	Имя фонемы	Частота, %
1	a	100	15	y	14	29	c	5
2	i	50	16	v	13	30	g	5
3	e	46	17	n'	11	31	f	5
4	r	26	18	z	10	32	h	5
5	t	22	19	d	9	33	Zh	5
6	n	22	20	r'	9	34	m'	4
7	o	21	21	m	9	35	k'	4
8	l	19	22	s'	8	36	z'	4
9	u	19	23	t'	8	37	p'	4
10	k	17	24	Ch'	7	38	b'	3
11	l'	17	25	Sh	6	39	f'	3
12	s	16	26	d'	6	40	Sh'	3
13	j'	16	27	v'	6	41	g'	3
14	p	15	28	b	5	42	h'	2

Таблица П 3.2

Нормированная частота встречаемости дифоном

№ пп	Имя ди- фонемы	Частота, %	№ пп	Имя ди- фонемы	Частота, %	№ пп	Имя ди- фонемы	Частота, %
1	na	100	45	en'	20	89	ir	13
2	l'i	87	46	er'	20	90	on	13
3	pa	68	47	in	20	91	sk	13
4	ra	58	48	m'e	20	92	z'i	13
5	ka	58	49	ach'	20	93	Ch'e	13
6	na	57	50	d'e	20	94	im	13
7	al	55	51	d'i	20	95	iz	13
8	ar	50	52	sa	20	96	lo	13
9	ta	50	53	es	19	97	ry	13
10	n'i	47	54	il'	19	98	ut	13
11	al'	43	55	ro	19	99	uCh'	13
12	aj'	42	56	t'i	19	100	ar'	12
13	j'e	40	57	an'	19	101	ch'i	12
14	st	40	58	vy	19	102	kr	12
15	da	37	59	ny	18	103	l'o	12
16	r'e	37	60	p'e	18	104	m'i	12
17	at	37	61	pr	18	105	ic	12
18	va	37	62	et	18	106	tu	12
19	as	37	63	s't'	18	107	at'	11
20	ma	34	64	am	17	108	ad'	11
21	an	34	65	ash	17	109	av'	11
22	za	33	66	fi	17	110	ek	11
23	il	31	67	pr'	17	111	nu	11
24	ik	30	68	is	17	112	s'i	11
25	er	29	69	ab	16	113	zha	11
26	r'i	29	70	ku	16	114	ab'	11
27	k'i	28	71	ol	16	115	cy	11
28	ot	28	72	ha	16	116	ls'	11
29	t'e	27	73	ru	16	117	p'i	11
30	en	27	74	tr	16	118	eh	10
31	j'a	27	75	v'i	16	119	ig	10
32	ap	26	76	ba	15	120	s'e	10
33	az	26	77	ga	15	121	tv	10
34	el'	26	78	iv	15	122	azh	10
35	ij'	26	79	or	15	123	ai	10
36	v'e	26	80	g'i	15	124	az'	10
37	av	25	81	h'i	15	125	ca	10
38	ad	24	82	ah	14	126	ip	10
39	j'u	23	83	ce	14	127	ik'	10

40	l'e	23	84	b'e	13	128	ur	10
41	as'	22	85	ag	13	129	ed'	10
42	n'e	22	86	b'i	13	130	gr	10
43	ak	21	87	ch'a	13	131	s'a	10
44	yj'	21	88	el	13	132	z'e	10

Таблица П 3.3

Нормированная частота встречаемости фонослогов

№ пп	Имя фонослога	Частота, %	№ пп	Имя фонослога	Частота, %	№ пп	Имя фонослога	Частота, %
1	na	100	27	ba	22	53	p'i	14
2	l'i	95	28	f'i	22	54	t'i	14
3	pa	87	29	g'i	22	55	Shy	14
4	ra	85	30	h'i	22	56	g'e	13
5	la	70	31	l'e	22	57	j'e	13
6	ka	60	32	m'e	22	58	ny	13
7	za	52	33	b'e	21	59	s'i	13
8	da	46	34	ga	21	60	s't'i	13
9	va	46	35	n'i	20	61	lo	12
10	ma	44	36	z'i	20	62	s'e	12
11	ta	44	37	Ch'a	20	63	tu	12
12	r'e	41	38	ku	18	64	z'e	12
13	t'e	35	39	m'i	18	65	las'	11
14	k'i	32	40	ca	17	66	Sha	11
15	sa	31	41	Sh'i	17	67	zhe	11
16	d'e	29	42	s'a	17	68	gra	11
17	d'i	29	43	b'i	16	69	k'e	11
18	vy	27	44	j'a	16	70	nu	11
19	n'e	26	45	Sh'e	16	71	r'e	11
20	pra	24	46	ce	15	72	Sho	11
21	sta	24	47	Ch'e	15	73	to	11
22	ha	23	48	r'i	15	74	ty	11
23	p'e	23	49	ro	15	75	Zhy	11
24	v'i	23	50	Zha	15	76	kra	10
25	pr'i	23	51	l'o	14	77	ru	10
26	v'e	23	52	ls'a	14	78	ska	10

П 3.2. Позиционные диаллофоны и аллослоги

Нормированная частота встречаемости позиционных диаллофонов и позиционных аллослогов приведена в таблицах П 3.4 и П 3.5 соответственно. Индексы гласных имеют следующие значения: 0 – ударный, 2 – гласный первой степени редукции, 3 – гласный второй степени редукции. Индексы согласных означают: 0 – одинарный согласный, 1 – удвоенный согласный.

Таблица П 3.4

Нормированная частота встречаемости позиционных диаллофонов

№ пп	Имя позици- онного диалло- фона	Частота, %	№ пп	Имя позици- онного диалло- фона	Частота, %	№ пп	Имя позици- онного диалло- фона	Частота, %
1	r ₀ a ₂	100	85	l ₀ s' ₀	23	169	s ₀ p ₀	15
2	l' ₀ i ₂	90	86	v ₀ y ₀	23	170	sh ₀ o ₀	15
3	n ₀ a ₂	88	87	z ₀ a ₃	23	171	t ₀ p ₀	15
4	s ₀ t ₀	77	88	a ₂ j' ₀	22	172	t ₀ r' ₀	15
5	k ₀ a ₂	62	89	d' ₀ i ₀	22	173	v ₀ l' ₀	15
6	l ₀ a ₂	60	90	e ₀ r ₀	22	174	zh ₀ a ₀	15
7	n ₀ a ₃	58	91	e ₃ r' ₀	22	175	a ₂ g ₀	14
8	o ₀ t ₀	58	92	t ₀ v ₀	22	176	a ₃ z ₀	14
9	p ₀ a ₂	58	93	a ₂ b ₀	21	177	e ₂ n' ₀	14
10	j' ₀ e ₂	57	94	k ₀ a ₀	21	178	e ₃ l' ₀	14
11	n' ₀ i ₃	54	95	v' ₀ e ₀	21	179	i ₀ n ₀	14
12	r ₀ a ₃	53	96	a ₃ l' ₀	20	180	p ₀ a ₀	14
13	a ₂ s ₀	52	97	c ₀ e ₀	20	181	p ₀ o ₀	14
14	a ₂ r ₀	51	98	g ₀ r ₀	20	182	u ₀ r ₀	14
15	t ₀ a ₂	50	99	r' ₀ i ₂	20	183	v ₀ a ₀	14
16	i ₃ j' ₀	46	100	a ₃ d ₀	19	184	v ₀ y ₃	14
17	a ₀ l ₀	46	101	d' ₀ e ₀	19	185	v ₀ z ₀	14
18	p ₀ a ₃	46	102	e ₀ n' ₀	19	186	z' ₀ i ₀	14
19	r ₀ a ₀	46	103	g ₀ a ₂	19	187	a ₃ v ₀	13
20	a ₀ l' ₀	42	104	i ₂ v ₀	19	188	a ₃ n ₀	13
21	r' ₀ e ₂	41	105	i ₃ l ₀	19	189	a ₃ ch' ₀	13
22	y ₃ j' ₀	41	106	i ₃ k' ₀	19	190	a ₃ p ₀	13
23	k' ₀ i ₂	39	107	j' ₀ e ₃	19	191	b ₀ o ₀	13
24	r ₀ o ₀	39	108	m' ₀ e ₀	19	192	d' ₀ e ₂	13
25	i ₀ l ₀	38	109	n ₀ o ₀	19	193	f' ₀ i ₂	13
26	i ₃ k ₀	38	110	t ₀ k ₀	19	194	h' ₀ i ₂	13
27	p ₀ r ₀	38	111	a ₂ d ₀	18	195	i ₂ r ₀	13

28	a_3l_0	37	112	ch'_0a_0	18	196	n_0t_0	13
29	$s'_0t'_0$	37	113	k_0o_0	18	197	r'_0i_0	13
30	a_2p_0	36	114	l'_0i_3	18	198	r_0y_0	13
31	$a_0j'_0$	35	115	m'_0e_2	18	199	t'_0e_0	13
32	$p_0r'_0$	35	116	n_0y_3	18	200	t_0n_0	13
33	a_3r_0	34	117	s'_0a_2	18	201	v'_0i_0	13
34	l_0a_3	34	118	s_0l_0	18	202	a_2h_0	12
35	a_2t_0	33	119	z_0a_0	18	203	$a_2b'_0$	12
36	o_0l_0	33	120	$a_2z'_0$	17	204	$a_2ch'_0$	12
37	o_0r_0	31	121	b_0a_2	17	205	$a_2v'_0$	12
38	a_0t_0	30	122	b'_0e_2	17	206	a_3g_0	12
39	d_0a_2	30	123	i_2z_0	17	207	c_1a_2	12
40	k_0a_3	30	124	i_2p_0	17	208	e_2zh_0	12
41	t_0r_0	30	125	l'_0e_0	17	209	e_3r_0	12
42	v_0a_2	30	126	m_0a_3	17	210	i_0r_0	12
43	a_2v_0	30	127	n'_0e_2	17	211	k'_0i_3	12
44	j'_0a_2	30	128	p'_0e_3	17	212	s'_0e_0	12
45	a_2n_0	29	129	r_0u_0	17	213	t'_0i_2	12
46	a_2z_0	29	130	t'_0i_0	17	214	z_0d_0	12
47	l'_0i_0	29	131	e_0t_0	16	215	z'_0e_2	12
48	n'_0i_0	29	132	e_2n_0	16	216	$a_0t'_0$	11
49	v_0a_3	29	133	h_0o_0	16	217	$a_0ch'_0$	11
50	$a_3j'_0$	28	134	h_0a_2	16	218	a_3s_0	11
51	d_0a_0	28	135	i_2s_0	16	219	a_3t_0	11
52	m_0a_2	28	136	i_2m_0	16	220	c_0y_3	11
53	a_2l_0	27	137	n'_0e_0	16	221	d_0r_0	11
54	a_3m_0	26	138	t'_0e_3	16	222	e_2k_0	11
55	r'_0i_3	26	139	t_0o_0	16	223	h_0a_3	11
56	t'_0e_2	26	140	t_0y_2	16	224	$i_0t'_0$	11
57	$a_2l'_0$	25	141	$a_2d'_0$	15	225	l'_0a_0	11
58	e_2r_0	25	142	b'_0i_0	15	226	n'_0e_3	11
59	l_0o_0	25	143	e_0l_0	15	227	n_0c_0	11
60	l'_0o_0	25	144	e_0n_0	15	228	s'_0i_0	11
61	o_0n_0	25	145	$e_0l'_0$	15	229	sh'_0i_3	11
62	t_0a_0	25	146	e_3n_0	15	230	v'_0i_3	11
63	t_0a_3	25	147	e_3t_0	15	231	y_0l_0	11
64	a_0n_0	24	148	h'_0i_0	15	232	y_0sh_0	11
65	$a_2n'_0$	24	149	i_2g_0	15	233	$a_2p'_0$	10
66	a_2k_0	24	150	i_2n_0	15	234	$b_0l'_0$	10

67	i ₀ l' ₀	24	151	s ₀ a ₂	15	235	d' ₀ i ₃	10
68	k ₀ r ₀	24	152	s ₀ a ₃	15	236	e ₀ r' ₀	10
69	m ₀ a ₀	24	153	t ₀ u ₂	15	237	e ₂ p ₀	10
70	r' ₀ e ₀	24	154	u ₃ t ₀	15	238	g' ₀ e ₂	10
71	s ₀ k ₀	24	155	a ₀ r ₀	15	239	g' ₀ i ₀	10
72	u ₂ ch' ₀	24	156	a ₂ zh ₀	15	240	i ₃ n ₀	10
73	z ₀ a ₂	24	157	a ₂ r' ₀	15	241	i ₃ l' ₀	10
74	a ₂ sh ₀	23	158	a ₂ s' ₀	15	242	l ₀ k ₀	10
75	e ₂ l' ₀	23	159	ch' ₀ e ₃	15	243	l ₀ p ₀	10
76	j' ₀ u ₂	23	160	d ₀ a ₃	15	244	n' ₀ i ₂	10
77	l ₀ a ₀	23	161	e ₀ h ₀	15	245	n ₀ y ₂	10
78	n ₀ a ₀	23	162	f' ₀ i ₀	15	246	o ₀ d ₀	10
79	v' ₀ e ₂	23	163	g' ₀ i ₂	15	247	p' ₀ e ₀	10
80	a ₃ s' ₀	23	164	i ₀ s ₀	15	248	r' ₀ e ₃	10
81	e ₂ s ₀	23	165	i ₂ k ₀	15	249	sh ₀ y ₀	10
82	j' ₀ a ₀	23	166	j' ₀ u ₃	15	250	v ₀ o ₀	10
83	k ₀ u ₂	23	167	m ₀ o ₀	15			
84	l' ₀ e ₂	23	168	o ₀ j' ₀	15			

Таблица П 3.5

Нормированная частота встречаемости позиционных аллослогов

№ пп	Имя позици- онного алло- слога	Часто- та, %	№ пп	Имя позици- онного алло- слога	Часто- та, %	№ пп	Имя позици- онного алло-слога	Часто- та, %
1	l' ₀ i ₂	100	39	d' ₀ e ₀	17	77	l' ₀ e ₀	12
2	r ₀ a ₂	66	40	g ₀ a ₂	17	78	m ₀ a ₃	12
3	n ₀ a ₂	65	41	g' ₀ i ₂	17	79	m' ₀ e ₂	12
4	p ₀ a ₂	61	42	r ₀ a ₀	17	80	n' ₀ e ₀	12
5	l ₀ a ₂	54	43	r' ₀ e ₀	17	81	n ₀ y ₂	12
6	k ₀ a ₂	51	44	z ₀ a ₀	17	82	p ₀ a ₀	12
7	n ₀ a ₃	51	45	k ₀ a ₀	16	83	r' ₀ e ₂	12
8	p ₀ a ₃	48	46	l ₀ o ₀	16	84	p ₀ o ₀	12
9	t ₀ a ₂	40	47	n ₀ a ₀	16	85	t' ₀ e ₃	12
10	k' ₀ i ₂	34	48	p' ₀ e ₃	16	86	b' ₀ i ₀	11
11	r' ₀ e ₂	34	49	p ₀ r' ₀ i ₂	16	87	d ₀ a ₃	11
12	l ₀ a ₃	33	50	s ₀ a ₃	16	88	f' ₀ i ₀	11
13	r ₀ a ₃	32	51	s ₀ t ₀ a ₂	16	89	h ₀ o ₀	11

14	d ₀ a ₂	30	52	h' ₀ i ₀	15	90	h ₀ a ₃	11
15	v ₀ a ₃	30	53	k ₀ a ₃	15	91	p ₀ r ₀ a ₃	11
16	z ₀ a ₂	29	54	l ₀ a ₃ s' ₀	15	92	r' ₀ i ₀	11
17	m ₀ a ₂	28	55	l' ₀ e ₂	15	93	sh ₀ y ₀	11
18	t' ₀ e ₂	26	56	s ₀ a ₂	15	94	t ₀ a ₀	11
19	v ₀ y ₀	25	57	t ₀ o ₀	15	95	t ₀ a ₃	11
20	z ₀ a ₃	25	58	c ₁ a ₂	14	96	t ₀ u ₂	11
21	d' ₀ i ₀	23	59	ch' ₀ e ₃	14	97	v' ₀ i ₃	11
22	m ₀ a ₀	22	60	h' ₀ i ₂	14	98	v ₀ y ₃	11
23	r ₀ o ₀	22	61	p ₀ r ₀ a ₂	14	99	v ₀ a ₀	11
24	s' ₀ a ₂	22	62	sh ₀ o ₀	14	100	b' ₀ i ₂	10
25	l' ₀ o ₀	21	63	t ₀ y ₂	14	101	b ₀ o ₀	10
26	v ₀ a ₂	21	64	v' ₀ e ₂	14	102	c ₀ y ₂	10
27	d ₀ a ₀	20	65	z' ₀ e ₂	14	103	ch' ₀ i ₀	10
28	j' ₀ a ₀	20	66	z' ₀ i ₀	14	104	g' ₀ e ₂	10
29	l' ₀ i ₃	20	67	zh ₀ a ₀	14	105	k ₀ r ₀ a ₂	10
30	s' ₀ a ₂	20	68	c ₀ e ₀	13	106	l ₀ a ₀	10
31	n' ₀ e ₂	20	69	d' ₀ e ₂	13	107	n ₀ u ₂	10
32	b' ₀ e ₂	18	70	f' ₀ i ₂	13	108	p' ₀ e ₀	10
33	ch' ₀ a ₀	18	71	l' ₀ i ₀	13	109	s ₀ a ₀	10
34	h ₀ a ₂	18	72	m' ₀ e ₀	13	110	s' ₀ t' ₀ i ₂	10
35	k ₀ u ₂	18	73	p ₀ r' ₀ i ₃	13	111	sh' ₀ o ₀	10
36	n' ₀ i ₀	18	74	s ₀ t ₀ a ₀	13	112	t' ₀ e ₀	10
37	j' ₀ e ₂	18	75	v' ₀ e ₀	13	113	t' ₀ i ₀	10
38	b ₀ a ₂	17	76	v' ₀ i ₀	13			

П 3.3. Позиционно-комбинаторные аллофоны, диаллофоны, аллослоги

Нормированная частота встречаемости позиционно-комбинаторных аллофонов, диаллофонов и аллослогов приведена в таблицах П 3.6 – П 3.8 соответственно. Для обозначения позиционных и комбинаторных значений используются индексы, описанные в Приложении 2.

Таблица П 3.6

Нормированная частота встречаемости аллофонов

№ пп	Имя алло- фона	Часто- та, %	№ пп	Имя алло- фона	Часто- та, %	№ пп	Имя алло- фона	Часто- та, %
1	t ₀₀₁	100	31	z ₀₀₂	24	61	j' ₀₁₃	16
2	r ₀₀₂	91	32	e ₀₄₃	24	62	z' ₀₀₂	16

3	s ₀₀₁	77	33	z ₀₀₁	24	63	zh ₀₀₂	16
4	n ₀₀₂	75	34	a ₂₂₁	23	64	d ₀₀₁	16
5	p ₀₀₁	73	35	a ₃₂₃	23	65	f ₀₀₁	16
6	l' ₀₀₂	70	36	c ₀₀₁	23	66	k ₀₀₁	15
7	k ₀₀₂	63	37	o ₀₄₂	23	67	sh' ₀₀₁	15
8	l ₀₀₂	52	38	e ₃₄₂	23	68	g' ₀₀₂	15
9	n' ₀₀₂	49	39	i ₀₄₂	23	69	v ₀₁₂	15
10	r' ₀₀₂	44	40	m' ₀₀₂	23	70	a ₃₁₃	14
11	a ₂₂₂	42	41	a ₃₂₂	23	71	o ₀₂₃	14
12	s' ₀₀₁	41	42	g ₀₀₂	23	72	r ₀₀₁	14
13	j' ₀₁₂	40	43	k' ₀₀₁	23	73	a ₀₁₃	13
14	e ₂₄₂	40	44	i ₂₄₁	22	74	a ₂₃₂	13
15	t' ₀₀₁	38	45	o ₀₂₂	21	75	i ₂₄₀	12
16	a ₂₂₃	37	46	a ₀₂₃	21	76	h' ₀₀₁	12
17	a ₂₁₂	33	47	i ₀₄₁	20	77	a ₀₄₃	12
18	i ₃₄₃	32	48	i ₂₄₃	20	78	y ₃₂₃	12
19	i ₂₄₂	31	49	b ₀₀₂	19	79	a ₂₂₀	11
20	ch' ₀₀₁	30	50	a ₀₂₂	19	80	v ₀₀₃	11
21	e ₂₄₃	30	51	p' ₀₀₁	19	81	v ₀₁₃	11
22	sh ₀₀₁	30	52	a ₂₁₃	19	82	e ₂₄₁	10
23	m ₀₀₂	29	53	a ₃₂₁	19	83	e ₂₄₀	10
24	d ₀₀₂	29	54	n ₀₀₃	19	84	v ₀₀₂	10
25	d' ₀₀₂	27	55	a ₃₁₂	18	85	v ₀₁₁	10
26	e ₀₄₂	26	56	b' ₀₀₂	18	86	a ₀₂₁	10
27	i ₀₄₃	26	57	r ₀₀₃	17	87	a ₀₄₂	10
28	i ₃₄₂	25	58	e ₃₄₃	17	88	a ₂₃₃	10
29	l ₀₀₃	25	59	l ₀₀₁	17	89	u ₀₂₂	10
30	f ₀₀₁	24	60	h ₀₀₂	17			

Таблица П 3.7

Нормированная частота встречаемости диаллофонов

№ пп	Имя диалло- фона	Часто- та, %	№ пп	Имя диалло- фона	Часто- та, %	№ пп	Имя диалло- фона	Часто- та, %
1	s ₀₀₁ t ₀₀₁	100	94	e ₂₄₂ r ₀₀₂	17	187	v ₀₁₁ a ₀₁₃	13
2	i ₃₄₃ j' ₀₁₂	53	95	i ₀₄₁ l ₀₀₃	17	188	zh ₀₀₂ a ₀₂₃	13
3	r ₀₀₂ a ₂₂₃	52	96	k' ₀₀₁ i ₂₄₁	17	189	a ₀₂₃ ch' ₀₀₁	11
4	p ₀₀₁ r ₀₀₂	51	97	l ₀₀₂ a ₂₁₃	17	190	a ₀₄₁ l ₀₀₃	11
5	n' ₀₀₂ i ₃₄₃	48	98	m ₀₀₂ a ₂₁₃	17	191	a ₀₄₂ t ₀₀₁	11
6	s' ₀₀₁ t' ₀₀₁	48	99	t ₀₀₁ n ₀₀₂	17	192	a ₂₁₂ t ₀₀₁	11

7	l' ₀₀₂ i ₂₄₂	47	100	a ₃₃₂ r ₀₀₂	16	193	a ₂₂₁ b ₀₀₂	11
8	p ₀₀₁ r' ₀₀₂	47	101	d ₀₀₂ a ₀₂₃	16	194	a ₂₂₁ v ₀₁₃	11
9	r ₀₀₂ a ₂₂₂	47	102	e ₂₄₂ n ₀₀₂	16	195	a ₂₂₂ k ₀₀₂	11
10	p ₀₀₁ a ₂₁₂	45	103	e ₂₄₂ zh ₀₀₂	16	196	a ₂₂₃ n' ₀₀₂	11
11	l ₀₀₂ a ₂₁₂	41	104	f' ₀₀₁ i ₂₄₂	16	197	a ₃₂₁ m ₀₀₂	11
12	k ₀₀₂ a ₂₃₂	40	105	i ₂₄₂ r ₀₀₂	16	198	a ₃₂₃ l' ₀₀₂	11
13	t ₀₀₁ r ₀₀₂	36	106	k' ₀₀₁ i ₂₄₀	16	199	b ₀₀₂ a ₂₁₃	11
14	l' ₀₀₂ i ₂₄₁	35	107	l' ₀₀₂ i ₀₄₁	16	200	b' ₀₀₂ e ₂₄₃	11
15	r ₀₀₂ a ₃₂₂	35	108	l' ₀₀₂ i ₃₄₃	16	201	d ₀₀₂ a ₂₂₃	11
16	t ₀₀₁ a ₂₂₃	34	109	t ₀₀₁ a ₃₂₂	16	202	d' ₀₀₂ e ₂₄₂	11
17	a ₂₂₂ s ₀₀₁	33	110	z ₀₀₁ d ₀₀₂	16	203	e ₀₄₁ l ₀₀₂	11
18	i ₃₄₂ k ₀₀₂	33	111	a ₀₂₂ n ₀₀₃	15	204	e ₀₄₂ n ₀₀₃	11
19	k ₀₀₂ r ₀₀₂	33	112	d ₀₀₁ r ₀₀₂	15	205	e ₂₄₂ r ₀₀₁	11
20	l' ₀₀₂ i ₂₄₃	33	113	d ₀₀₂ a ₂₂₁	15	206	h ₀₀₁ o ₀₃₂	11
21	n ₀₀₂ a ₂₂₂	33	114	e ₀₄₂ t ₀₀₁	15	207	i ₀₄₂ r ₀₀₂	11
22	n ₀₀₂ a ₂₂₃	33	115	h ₀₀₂ a ₂₃₃	15	208	i ₂₄₁ v ₀₁₁	11
23	o ₀₄₂ t ₀₀₁	33	116	i ₂₄₂ n ₀₀₂	15	209	i ₂₄₂ g ₀₀₂	11
24	j' ₀₁₂ e ₂₄₂	32	117	l ₀₀₂ a ₀₁₃	15	210	j' ₀₁₂ a ₂₄₀	11
25	n ₀₀₂ a ₃₂₃	32	118	m' ₀₀₂ e ₀₄₂	15	211	k ₀₀₂ a ₃₃₂	11
26	p ₀₀₁ a ₃₁₂	32	119	n ₀₀₃ c ₀₀₁	15	212	m ₀₀₂ a ₃₁₂	11
27	l ₀₀₃ s' ₀₀₁	30	120	p ₀₀₁ a ₃₁₁	15	213	n' ₀₀₂ e ₂₄₂	11
28	e ₃₄₃ r' ₀₀₂	29	121	r ₀₀₂ a ₃₂₁	15	214	n ₀₀₃ p ₀₀₁	11
29	l ₀₀₂ a ₃₁₃	29	122	r' ₀₀₂ e ₀₄₂	15	215	r ₀₀₁ m ₀₀₂	11
30	r ₀₀₂ a ₂₂₁	28	123	t' ₀₀₁ e ₂₄₃	15	216	r' ₀₀₂ e ₀₄₃	11
31	r ₀₀₂ o ₀₂₃	28	124	u ₂₂₃ ch' ₀₀₁	15	217	r ₀₀₃ p ₀₀₁	11
32	a ₀₂₃ l' ₀₀₂	27	125	v' ₀₀₁ e ₀₄₃	15	218	s ₀₀₁ a ₀₂₂	11
33	a ₂₁₂ s ₀₀₁	27	126	y ₀₁₂ sh ₀₀₁	15	219	s ₀₀₁ a ₂₂₁	11
34	g ₀₀₂ r ₀₀₂	27	127	z ₀₀₂ a ₃₂₃	15	220	s ₀₀₁ k ₀₀₁	11
35	i ₃₄₃ k' ₀₀₁	26	128	a ₂₁₂ n ₀₀₂	14	221	sh' ₀₀₁ i ₃₄₂	11
36	n ₀₀₂ a ₂₂₀	26	129	a ₂₁₂ r ₀₀₂	14	222	sh' ₀₀₁ o ₀₄₂	11
37	n ₀₀₂ a ₂₂₁	26	130	a ₂₂₂ sh ₀₀₁	14	223	t ₀₀₁ u ₂₂₃	11
38	r ₀₀₂ a ₀₂₂	26	131	a ₂₂₂ z ₀₀₁	14	224	t ₀₀₁ v ₀₀₂	11
39	a ₂₂₁ p ₀₀₁	25	132	a ₂₃₂ t ₀₀₁	14	225	v ₀₁₁ y ₀₁₂	11
40	e ₂₄₂ s ₀₀₁	25	133	a ₃₂₂ z ₀₀₁	14	226	v ₀₁₂ a ₃₁₃	11
41	i ₀₄₃ l' ₀₀₂	25	134	a ₃₂₃ j' ₀₁₂	14	227	v ₀₁₃ l' ₀₀₂	11
42	j' ₀₁₂ e ₃₄₂	25	135	b ₀₀₁ l' ₀₀₂	14	228	a ₀₄₃ j' ₀₁₂	10
43	n ₀₀₂ a ₃₂₂	25	136	b ₀₀₂ o ₀₁₂	14	229	a ₂₂₃ ch' ₀₀₁	10
44	n' ₀₀₂ i ₃₄₂	25	137	c ₀₀₁ y ₃₂₃	14	230	a ₂₂₃ l' ₀₀₂	10
45	r' ₀₀₂ e ₂₄₃	25	138	d ₀₀₂ a ₀₂₂	14	231	a ₂₃₂ r ₀₀₂	10

46	s ₀₀₁ l ₀₀₂	25	139	e ₂₄₁ p ₀₀₁	14	232	a ₂₃₂ z ₀₀₂	10
47	a ₀₂₂ t ₀₀₁	23	140	e ₂₄₃ n' ₀₀₂	14	233	a ₃₁₁ l ₀₀₂	10
48	i ₂₄₁ p ₀₀₁	23	141	f' ₀₀₁ i ₀₄₃	14	234	a ₃₁₂ g ₀₀₂	10
49	l' ₀₀₂ o ₀₄₂	23	142	g ₀₀₂ a ₂₃₂	14	235	a ₃₂₃ b' ₀₀₂	10
50	n ₀₀₂ y ₃₂₃	23	143	h' ₀₀₁ i ₀₄₃	14	236	a ₃₂₃ ch' ₀₀₁	10
51	p ₀₀₁ a ₂₁₃	23	144	j' ₀₁₂ a ₂₄₁	14	237	b' ₀₀₂ e ₂₄₂	10
52	a ₃₁₃ s' ₀₀₁	22	145	j' ₀₁₂ e ₂₄₁	14	238	ch' ₀₀₁ a ₀₄₃	10
53	p' ₀₀₁ e ₃₄₃	22	146	k ₀₀₂ a ₀₃₂	14	239	ch' ₀₀₁ e ₃₄₂	10
54	t ₀₀₁ a ₂₂₂	22	147	k ₀₀₂ a ₂₃₁	14	240	d ₀₀₁ b ₀₀₂	10
55	y ₃₂₃ j' ₀₁₂	22	148	l ₀₀₂ a ₂₁₁	14	241	d' ₀₀₂ i ₀₄₁	10
56	e ₂₄₃ l' ₀₀₂	21	149	l ₀₀₃ p ₀₀₁	14	242	d' ₀₀₂ o ₀₄₂	10
57	e ₃₄₂ t ₀₀₁	21	150	n ₀₀₃ t ₀₀₁	14	243	e ₀₄₃ n' ₀₀₂	10
58	i ₂₄₁ m ₀₀₂	21	151	p ₀₀₁ a ₀₁₂	14	244	e ₂₄₃ b' ₀₀₂	10
59	l' ₀₀₂ e ₂₄₂	21	152	p ₀₀₁ a ₃₁₃	14	245	e ₂₄₃ d' ₀₀₂	10
60	m ₀₀₂ a ₀₁₃	21	153	sh ₀₀₁ o ₀₂₂	14	246	e ₂₄₃ s' ₀₀₁	10
61	n ₀₀₂ a ₃₂₁	21	154	t' ₀₀₁ e ₂₄₂	14	247	e ₃₄₂ s ₀₀₁	10
62	r ₀₀₂ a ₃₂₃	21	155	t' ₀₀₁ e ₃₄₃	14	248	h ₀₀₁ o ₀₃₃	10
63	s ₀₀₁ k ₀₀₂	21	156	t ₀₀₁ v ₀₀₃	14	249	i ₀₄₂ c ₀₀₁	10
64	t ₀₀₁ a ₀₂₂	21	157	a ₀₂₃ j' ₀₁₂	13	250	i ₀₄₂ n ₀₀₂	10
65	a ₂₂₂ r ₀₀₂	20	158	a ₂₁₂ k ₀₀₂	13	251	i ₃₄₂ k ₀₀₁	10
66	i ₀₄₁ l ₀₀₂	20	159	a ₂₂₁ v ₀₁₁	13	252	i ₃₄₃ l' ₀₀₂	10
67	i ₀₄₂ s ₀₀₁	20	160	a ₂₂₃ r' ₀₀₂	13	253	i ₃₄₃ s' ₀₀₁	10
68	j' ₀₁₂ u ₃₄₂	20	161	a ₂₄₁ p ₀₀₁	13	254	k' ₀₀₁ i ₃₄₃	10
69	n' ₀₀₂ i ₀₄₂	20	162	a ₃₂₁ v ₀₁₂	13	255	k ₀₀₂ a ₂₃₀	10
70	r ₀₀₂ a ₀₂₃	20	163	ch' ₀₀₁ n ₀₀₂	13	256	k ₀₀₂ a ₃₃₁	10
71	r' ₀₀₂ e ₂₄₂	20	164	d ₀₀₂ a ₂₂₂	13	257	k ₀₀₂ n ₀₀₂	10
72	r ₀₀₂ o ₀₂₂	20	165	d' ₀₀₂ e ₀₄₁	13	258	k ₀₀₂ t ₀₀₁	10
73	r ₀₀₂ u ₀₂₂	20	166	e ₀₄₂ h ₀₀₂	13	259	l ₀₀₂ a ₀₁₂	10
74	s ₀₀₁ p ₀₀₁	20	167	e ₀₄₃ t' ₀₀₁	13	260	l ₀₀₂ a ₃₁₂	10
75	t ₀₀₁ k ₀₀₂	20	168	e ₂₄₂ k ₀₀₂	13	261	l ₀₀₂ o ₀₂₃	10
76	t ₀₀₁ p ₀₀₁	20	169	e ₂₄₃ r' ₀₀₂	13	262	m' ₀₀₂ e ₂₄₃	10
77	t ₀₀₁ r' ₀₀₂	20	170	f ₀₀₁ s' ₀₀₁	13	263	m' ₀₀₂ i ₀₄₃	10
78	y ₃₂₃ j' ₀₁₃	20	171	i ₀₄₁ l ₀₀₁	13	264	n' ₀₀₂ e ₀₄₂	10
79	z ₀₀₂ a ₂₂₂	20	172	i ₀₄₃ t' ₀₀₁	13	265	n' ₀₀₂ e ₃₄₂	10
80	c ₀₀₁ e ₀₂₂	19	173	i ₂₄₂ s ₀₀₁	13	266	n' ₀₀₂ i ₀₄₃	10
81	i ₂₄₂ k ₀₀₂	19	174	i ₃₄₁ l ₀₀₃	13	267	o ₀₂₂ t ₀₀₁	10
82	k ₀₀₁ u ₂₃₀	19	175	j' ₀₁₂ e ₂₄₀	13	268	p' ₀₀₁ i ₃₄₂	10
83	k ₀₀₂ a ₂₃₃	19	176	k' ₀₀₁ i ₂₄₂	13	269	p ₀₀₁ l ₀₀₂	10
84	k ₀₀₂ a ₃₃₃	19	177	k ₀₀₁ o ₀₃₂	13	270	r ₀₀₁ n ₀₀₂	10

85	l ₀₀₂ o ₀₂₂	19	178	l' ₀₀₁ n ₀₀₂	13	271	r' ₀₀₂ i ₂₄₂	10
86	m ₀₀₂ a ₂₁₂	19	179	l' ₀₀₂ e ₀₄₃	13	272	r' ₀₀₂ i ₂₄₃	10
87	n ₀₀₂ a ₀₂₃	19	180	l' ₀₀₂ i ₀₄₂	13	273	s ₀₀₁ n ₀₀₂	10
88	r' ₀₀₂ i ₃₄₃	19	181	m ₀₀₂ o ₀₁₃	13	274	t' ₀₀₁ e ₀₄₂	10
89	u ₃₄₂ t ₀₀₁	19	182	r ₀₀₂ a ₀₂₁	13	275	t ₀₀₁ s ₀₀₁	10
90	a ₂₁₂ sh ₀₀₁	17	183	t ₀₀₁ o ₀₂₃	13	276	v ₀₁₂ a ₃₁₂	10
91	a ₂₁₃ l' ₀₀₂	17	184	t' ₀₀₁ o ₀₄₂	13	277	z ₀₀₁ l' ₀₀₂	10
92	a ₂₂₃ j' ₀₁₁	17	185	v ₀₀₂ a ₂₁₂	13	278	z' ₀₀₂ e ₂₄₃	10
93	a ₃₂₁ p ₀₀₁	17	186	v' ₀₀₂ e ₂₄₃	13			

Таблица П 3.8

Нормированная частота встречаемости аллослогов

№ пп	Имя аллослога	Частота, %	№ пп	Имя аллослога	Частота, %	№ пп	Имя аллослога	Частота, %
1	l' ₀₀₂ i ₂₄₂	100	108	p ₀₀₁ a ₂₁₁	17	215	v' ₀₁₂ i ₃₄₁ l ₀₀₃	12
2	p ₀₀₁ a ₂₁₂	84	109	p ₀₀₁ f ₀₀₂ a ₂₂₃	17	216	z ₀₀₁ v' ₀₀₁ o ₀₄₂	12
3	r ₀₀₂ a ₂₂₂	76	110	p ₀₀₁ f ₀₀₂ a ₃₂₂	17	217	z ₀₀₁ v ₀₀₂ a ₂₁₂	12
4	l' ₀₀₂ i ₂₄₁	71	111	p ₀₀₁ f' ₀₀₂ i ₂₄₃	17	218	z' ₀₀₂ a ₀₄₃	12
5	l ₀₀₂ a ₂₁₂	69	112	r ₀₀₂ a ₀₂₂	17	219	z ₀₀₂ a ₃₂₁	12
6	l' ₀₀₂ i ₂₄₃	69	113	r ₀₀₂ a ₃₂₁	17	220	f' ₀₀₁ i ₂₄₂	12
7	k ₀₀₂ a ₂₃₂	66	114	s' ₀₀₁ a ₂₄₃	17	221	z' ₀₀₂ i ₃₄₂	12
8	p ₀₀₁ a ₃₁₂	64	115	s ₀₀₁ a ₃₂₁	17	222	b ₀₀₂ a ₂₁₁	10
9	n ₀₀₂ a ₂₂₂	58	116	s ₀₀₁ a ₃₂₃	17	223	b ₀₀₂ a ₃₁₂	10
10	l ₀₀₂ a ₃₁₃	56	117	s ₀₀₁ l ₀₀₂ y ₀₁₂	17	224	b' ₀₀₂ e ₀₄₂	10
11	n ₀₀₂ a ₂₂₃	48	118	s ₀₀₁ p ₀₀₁ a ₂₁₂	17	225	b' ₀₀₂ i ₀₄₁	10
12	n ₀₀₂ a ₃₂₃	46	119	s ₀₀₁ t ₀₀₁ a ₀₂₂	17	226	b' ₀₀₂ i ₀₄₃	10
13	r ₀₀₂ a ₃₂₂	46	120	sh' ₀₀₁ e ₂₄₂	17	227	ch' ₀₀₁ a ₀₄₁ l ₀₀₃	10
14	p ₀₀₁ a ₂₁₃	43	121	t' ₀₀₁ e ₀₄₂	17	228	ch' ₀₀₁ e ₃₄₃	10
15	r ₀₀₂ a ₂₂₃	43	122	z' ₀₀₂ i ₂₄₁	17	229	ch' ₀₀₁ i ₀₄₁	10
16	z ₀₀₂ a ₂₂₂	43	123	zh ₀₀₂ e ₀₂₃	17	230	ch' ₀₀₁ n ₀₀₂ a ₂₂₀	10
17	n ₀₀₂ a ₃₂₁	41	124	ch' ₀₀₁ a ₀₄₁	15	231	d ₀₀₁ v' ₀₀₁ e ₀₄₃	10
18	n ₀₀₂ a ₃₂₂	38	125	ch' ₀₀₁ a ₀₄₃	15	232	d ₀₀₂ a ₃₂₁	10
19	p' ₀₀₁ e ₃₄₃	38	126	ch' ₁₀₁ i ₃₄₂	15	233	d ₀₀₂ a ₃₂₃	10
20	s ₀₀₁ t ₀₀₁ a ₂₂₃	38	127	d' ₀₀₂ o ₀₄₂	15	234	d' ₀₀₂ i ₀₄₂	10
21	t ₀₀₁ a ₂₂₂	38	128	g ₀₀₂ a ₂₃₃	15	235	d' ₀₀₂ i ₂₄₂	10
22	t ₀₀₁ a ₂₂₃	38	129	h ₀₀₁ o ₀₃₃	15	236	d' ₀₀₂ i ₃₄₂	10

23	$l_{002}a_{313}s'_{001}$	35	130	$j'_{011}a_{041}$	15	237	$e_{203}k'_{001}i_{241}$	10
24	$l'_{002}i_{343}$	35	131	$k_{001}o_{032}$	15	238	$f_{001}a_{312}$	10
25	$m_{002}a_{213}$	35	132	$k_{002}a_{230}$	15	239	$f_{001}o_{012}$	10
26	$p_{001}a_{311}$	33	133	$l_{001}d_{002}a_{022}$	15	240	$f_{001}s_{001}a_{222}$	10
27	$r_{002}a_{221}$	33	134	$l_{002}a_{312}$	15	241	$f'_{001}u_{343}$	10
28	$r'_{002}e_{242}$	33	135	$l'_{002}e_{042}$	15	242	$g_{001}u_{032}$	10
29	$r'_{002}e_{243}$	33	136	$l'_{002}i_{042}$	15	243	$g_{002}a_{332}$	10
30	$z_{002}a_{323}$	33	137	$l_{002}o_{023}$	15	244	$g'_{002}e_{241}$	10
31	$f'_{001}i_{242}$	30	138	$m_{002}a_{313}$	15	245	$g'_{002}e_{242}$	10
32	$h_{002}a_{233}$	30	139	$m'_{002}e_{042}$	15	246	$g'_{002}e_{343}$	10
33	$k_{002}a_{233}$	30	140	$m'_{002}i_{343}$	15	247	$g'_{002}i_{242}$	10
34	$l_{002}a_{213}$	30	141	$n_{002}u_{220}$	15	248	$h'_{001}i_{240}$	10
35	$l'_{002}o_{042}$	30	142	$p'_{001}i_{342}a_{243}$	15	249	$h_{002}a_{331}$	10
36	$m_{002}a_{013}$	30	143	$r_{002}a_{323}$	15	250	$r_{002}a_{222}$	10
37	$m_{002}a_{212}$	30	144	$s_{001}a_{221}l_{001}$	15	251	$j'_{011}o_{042}$	10
38	$n_{002}a_{221}$	30	145	$t_{001}a_{022}n_{003}$	15	252	$j'_{013}t'_{001}e_{243}$	10
39	$d_{002}a_{221}$	28	146	$sh'_{001}e_{342}$	15	253	$k'_{001}e_{042}$	10
40	$h'_{001}i_{043}$	28	147	$t_{001}a_{322}$	15	254	$k'_{001}i_{041}$	10
41	$t'_{001}e_{243}$	28	148	$t_{001}y_{222}$	15	255	$k_{001}o_{032}n_{003}$	10
42	$b'_{002}e_{243}$	25	149	$v'_{001}e_{043}$	15	256	$k_{002}a_{030}$	10
43	$k'_{001}i_{242}$	25	150	$v_{012}a_{210}$	15	257	$k_{002}t_{001}a_{322}$	10
44	$k_{001}u_{230}$	25	151	$v_{012}a_{212}$	15	258	$l'_{001}n'_{002}i_{343}$	10
45	$n'_{002}e_{242}$	25	152	$v'_{012}i_{341}$	15	259	$l_{002}a_{012}$	10
46	$r'_{002}e_{042}$	25	153	$z_{002}a_{021}$	15	260	$l_{002}a_{210}$	10
47	$r_{002}o_{023}$	25	154	$z_{002}a_{023}$	15	261	$l_{002}a_{311}$	10
48	$s_{001}a_{022}$	25	155	$z_{002}a_{322}$	15	262	$l'_{002}e_{243}$	10
49	$s_{001}a_{221}$	25	156	$z'_{002}i_{042}$	15	263	$l'_{002}i_{041}$	10
50	$sh_{001}o_{022}$	25	157	$b_{002}a_{212}$	12	264	$l'_{002}i_{342}$	10
51	$sh'_{001}o_{042}$	25	158	$c_{001}a_{220}$	12	265	$l_{003}s'_{001}a_{241}$	10
52	$t'_{001}e_{242}$	25	159	$c_{001}e_{223}$	12	266	$m'_{002}e_{342}$	10
53	$t_{001}o_{023}$	25	160	$c_{101}a_{221}$	12	267	$m_{002}o_{013}$	10
54	$v_{011}y_{012}$	25	161	$c_{101}a_{222}$	12	268	$n_{002}a_{022}$	10
55	$b_{002}o_{012}$	23	162	$ch'_{001}a_{042}$	12	269	$n_{002}a_{322}s_{001}$	10
56	$c_{001}e_{022}$	23	163	$ch'_{001}i_{043}$	12	270	$n'_{002}e_{243}$	10
57	$ch'_{001}e_{342}$	23	164	$d'_{002}e_{042}$	12	271	$n'_{002}e_{342}$	10
58	$d_{002}a_{022}$	23	165	$d'_{002}e_{043}$	12	272	$n'_{002}i_{041}$	10
59	$d_{002}a_{222}$	23	166	$d_{002}o_{022}$	12	273	$n'_{002}i_{041}l_{003}$	10
60	$f'_{001}i_{043}$	23	167	$g_{001}o_{032}$	12	274	$j'_{012}e_{242}$	10
61	$g_{002}a_{232}$	23	168	$g'_{002}e_{241}o_{022}$	12	275	$n_{002}o_{021}$	10

62	k'001i240	23	169	g'002i240	12	276	n002y220	10
63	k'001i241	23	170	g'002i241	12	277	l002u223	10
64	l002a211	23	171	g002r002a023	12	278	p'001e243	10
65	l002o022	23	172	h001o032	12	279	p001o013	10
66	n'002i042	23	173	h002a332	12	280	p001r002a222	10
67	p001a012	23	174	j'001a042	12	281	p001r002a323i322	10
68	r002o022	23	175	j'011a043	12	282	p001r'002i241	10
69	s001t001u223	23	176	j'011e042	12	283	r002a021	10
70	t001a022	23	177	k'001i243	12	284	r'002e240	10
71	t'001e343	23	178	k002a231	12	285	r'002e342	10
72	v012a312	23	179	k002a232n003	12	286	r'002i041	10
73	z'002e243	23	180	k002a331	12	287	r'002i341	10
74	b002a213	20	181	k002s'001i042	12	288	c001m'002e042	10
75	d002a223	20	182	l002a013	12	289	s'001a241	10
76	d'002e242	20	183	l'002e043	12	290	s001h'001e043	10
77	d'002i041	20	184	l'002o042t001	12	291	s001h'001e043	10
78	d'002i043	20	185	l002u223	12	292	s'001i041	10
79	k002a032	20	186	l003s'001a242	12	293	s001k002a232	10
80	l'002e242	20	187	m002a011	12	294	s001p001a212r003	10
81	l'002i343s'001	20	188	m'002e043	12	295	s'001t'001i242	10
82	l'002o043	20	189	m'002e240	12	296	sh001e023	10
83	m'002i240	20	190	ch'001e342s001	12	297	sh'001e242n003	10
84	n002a023	20	191	n002a221u023	12	298	sh001k001o033	10
85	n002a220	20	192	n'002i043	12	299	sh001l'002i041	10
86	j'012e240	20	193	j'012e243	12	300	sh'001n'002i342	10
87	p001a313	20	194	n002y222	12	301	sh001o020	10
88	r'002e242	20	195	p'001e042	12	302	sh001o022l001	10
89	p001r'002i343	20	196	p'001i241	12	303	j'012e342	10
90	sh001y022	20	197	p001o012	12	304	t'001i042	10
91	v011a013	20	198	p001r'002i242	12	305	v'012i341l003	10
92	v012a311	20	199	r002a023	12	306	t001u220	10
93	v012a313	20	200	r'002e043	12	307	t001u223	10
94	v'012e242	20	201	r002u022	12	308	t001y220	10
95	z002a223	20	202	r002y222	12	309	v'011e042	10
96	zh002a023	20	203	s'001a242	12	310	v011y013	10
97	b'002e242	17	204	m002a213	12	311	v012a213	10
98	b'002i242	17	205	s'001t'001i240	12	312	v'012e243	10
99	d002a023	17	206	sh001a321	12	313	v'012i343	10
100	d'002e041	17	207	sh'001i342	12	314	z001b'002i041	10

101	$f_{001}a_{212}$	17	208	$t_{001}a_{220}$	12	315	$z'_{001}d'_{002}e_{041}$	10
102	$h'_{001}i_{243}$	17	209	$t'_{001}i_{242}$	12	316	$z_{001}l'_{002}o_{042}t_{000}$	10
103	$k_{002}a_{332}$	17	210	$v_{001}y_{012}$	12	317	$z_{001}m'_{002}e_{243}$	10
104	$k_{002}r_{002}a_{223}$	17	211	$v_{003}l'_{002}e_{242}$	12	318	$z'_{002}i_{041}$	10
105	$l_{003}s'_{001}a_{243}$	17	212	$v'_{011}i_{043}$	12	319	$z'_{002}i_{043}$	10
106	$m'_{002}i_{043}$	17	213	$v_{012}a_{311}l_{003}$	12	320	$z_{002}y_{321}$	10
107	$n'_{002}e_{042}$	17	214	$v'_{012}i_{243}$	12	321	$zh_{002}a_{021}$	10

Приложение 4

Спектральные портреты основных аллофонов русской речи

В приложении показаны спектры основных аллофонов гласных и согласных русской речи. Для обозначения аллофонов гласных используются индексы, описанные в приложении 2.

П 4.1. Основные аллофоны гласных фонем: *A, E, U, Y, I, O*

Спектральные портреты аллофонов гласных представлены на рис. П 4.1 – П 4.13.

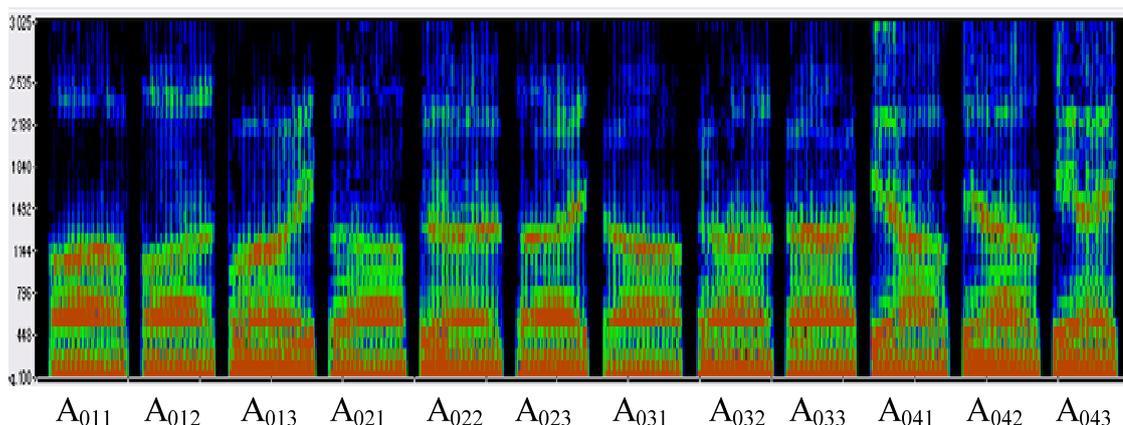


Рис. П 4.1. Спектральные портреты аллофонов ударной гласной *A*

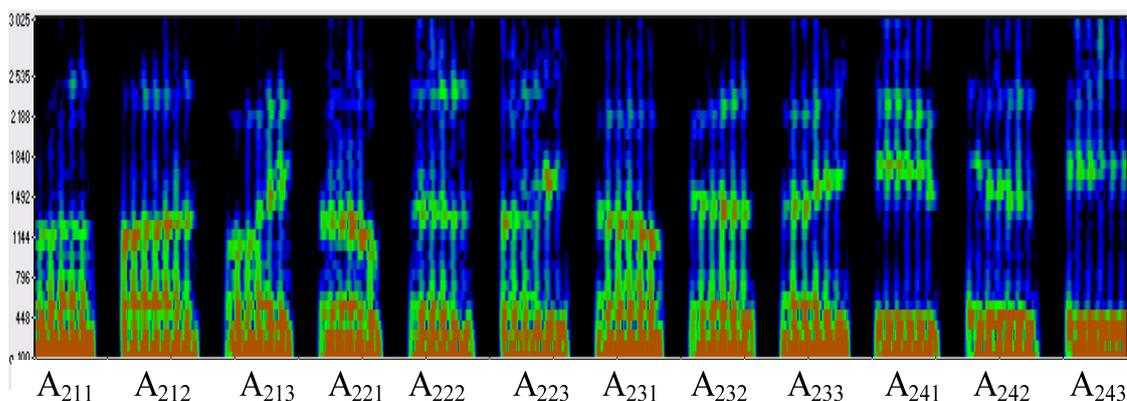


Рис. П 4.2. Спектральные портреты аллофонов гласной *A* первой степени редуции

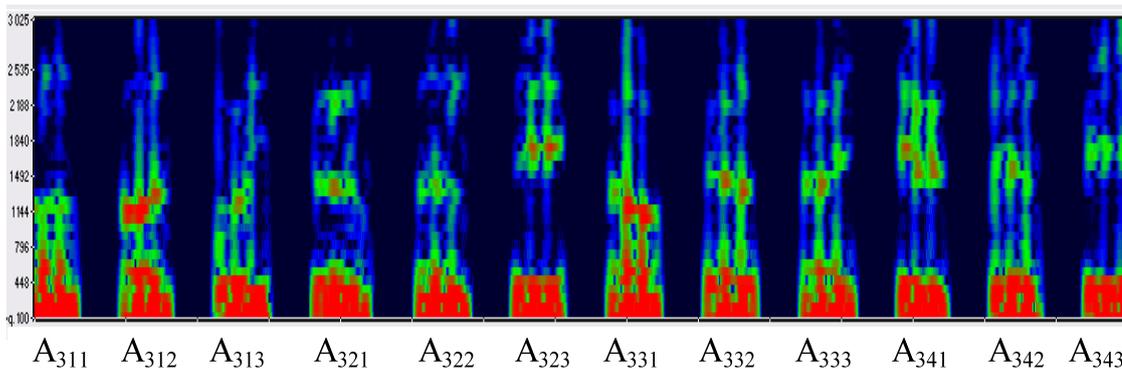


Рис. П 4.3. Спектральные портреты аллофонов гласной *A* второй степени редукции

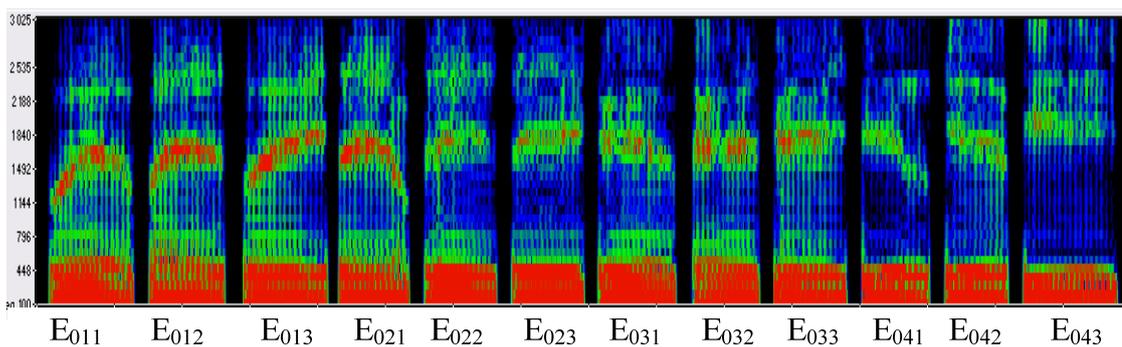


Рис. П 4.4. Спектральные портреты аллофонов ударной гласной *E*

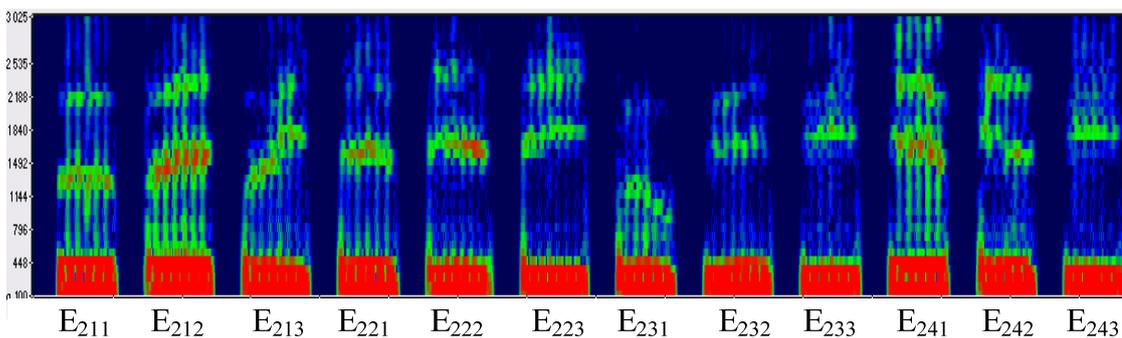


Рис. П 4.5. Спектральные портреты аллофонов гласной *E* первой степени редукции

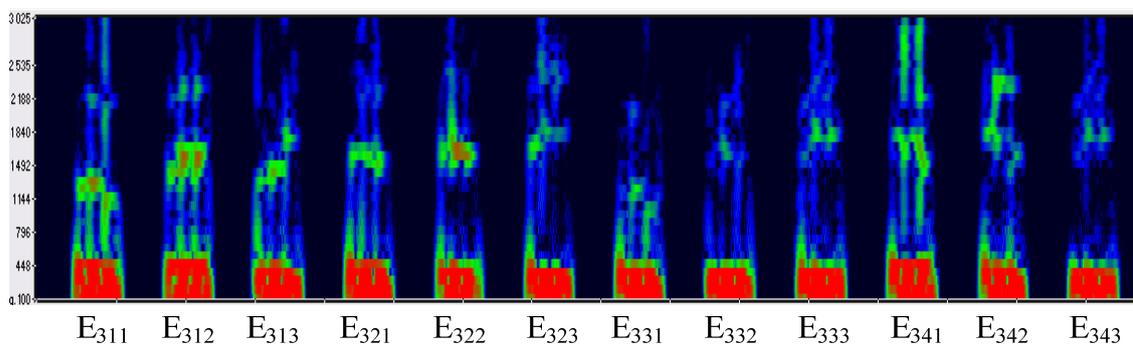


Рис. П 4.6. Спектральные портреты аллофонов гласной *E* второй степени редукции

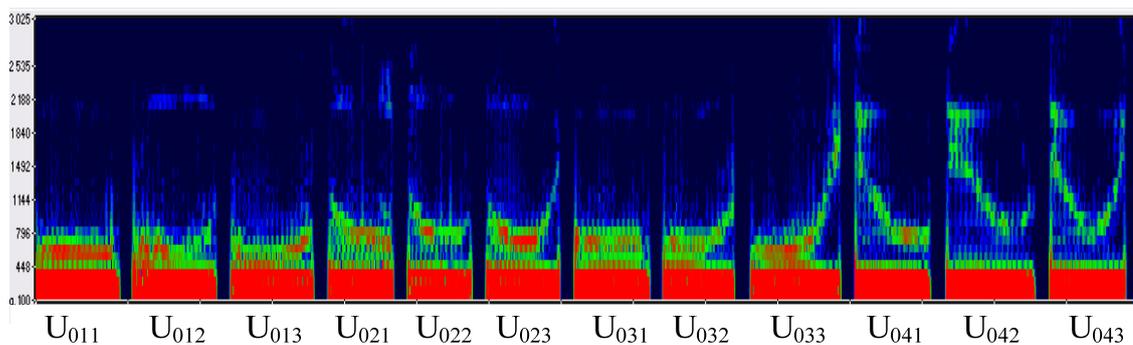


Рис. П 4.7. Спектральные портреты аллофонов ударной гласной U

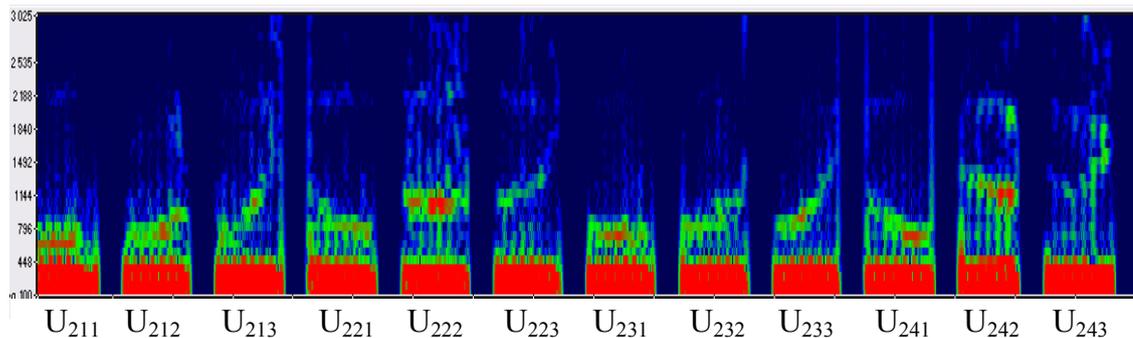


Рис. П 4.8. Спектральные портреты гласной U первой степени редукции

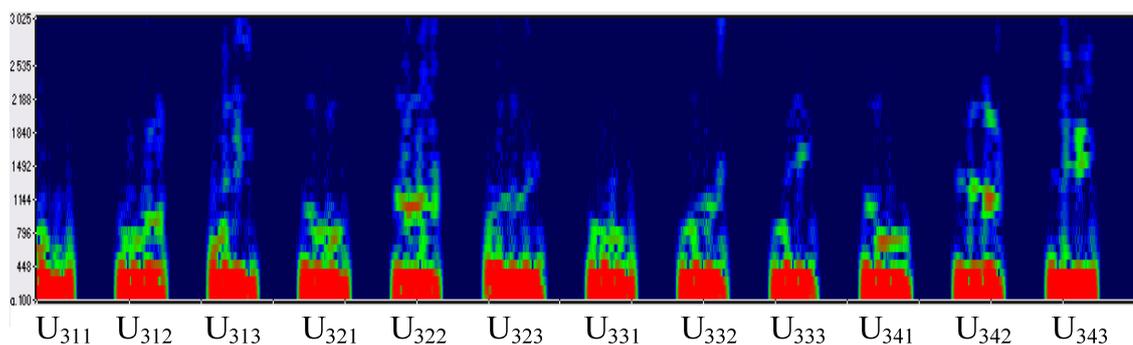


Рис. П 4.9. Спектральные портреты гласной U второй степени редукции

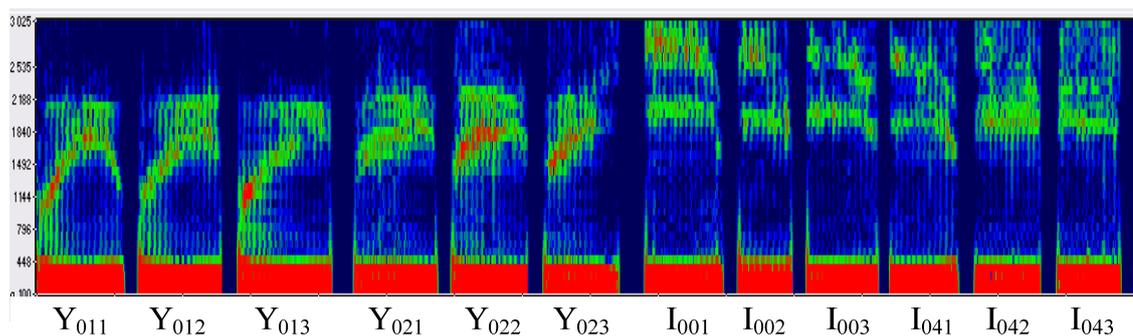


Рис. П 4.10. Спектральные портреты аллофонов ударных гласных Y, I

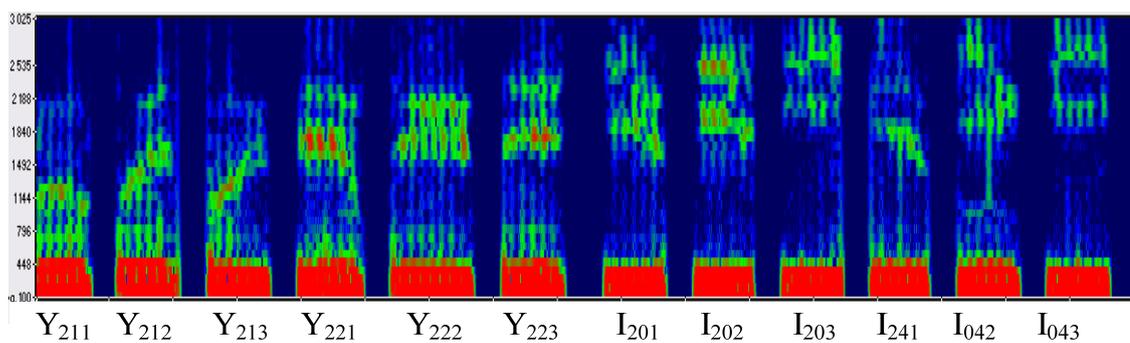


Рис. П 4.11. Спектральные портреты гласных *Y, I* первой степени редукции

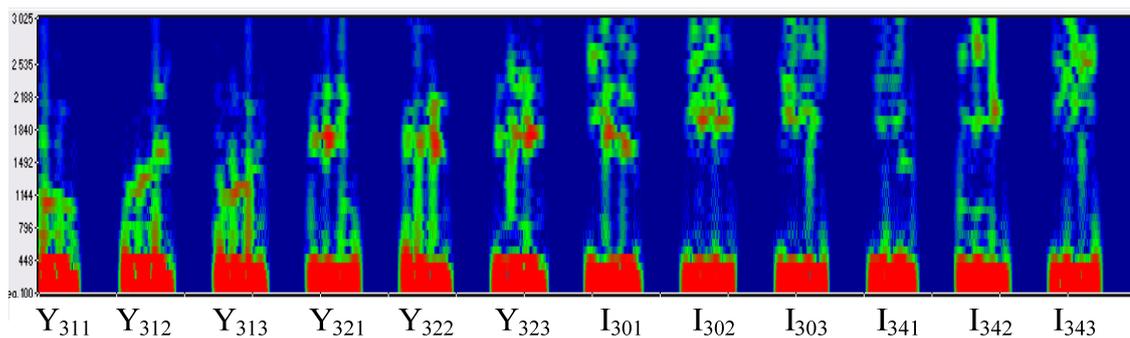


Рис. П 4.12. Спектральные портреты гласных *Y, I* второй степени редукции

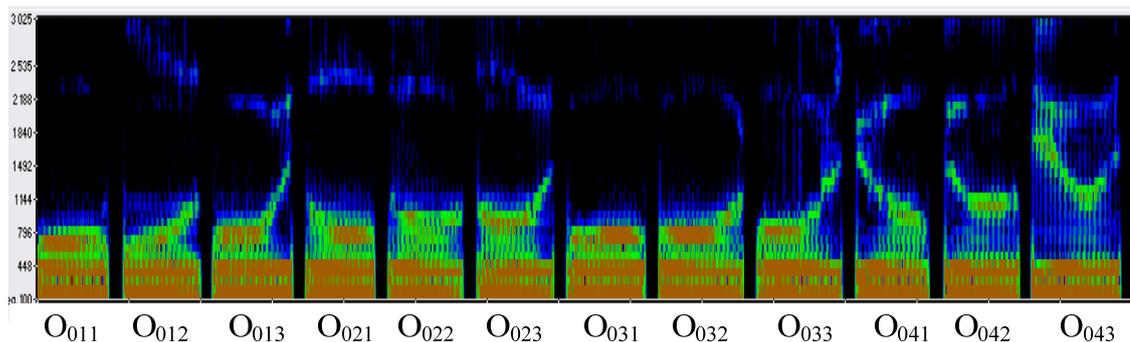
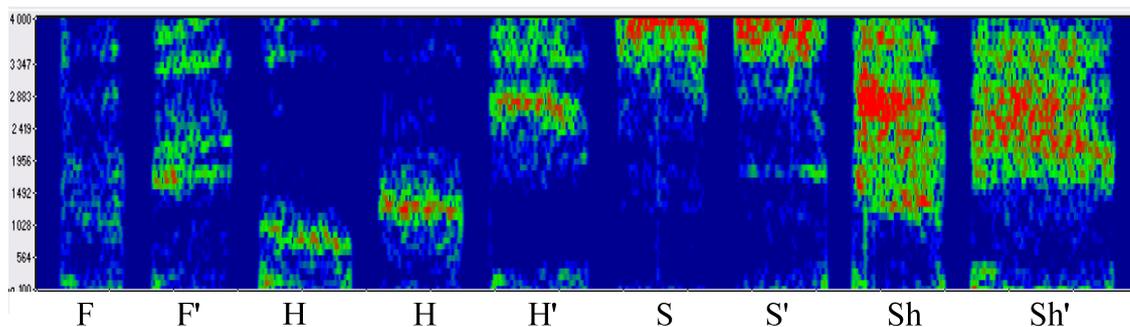


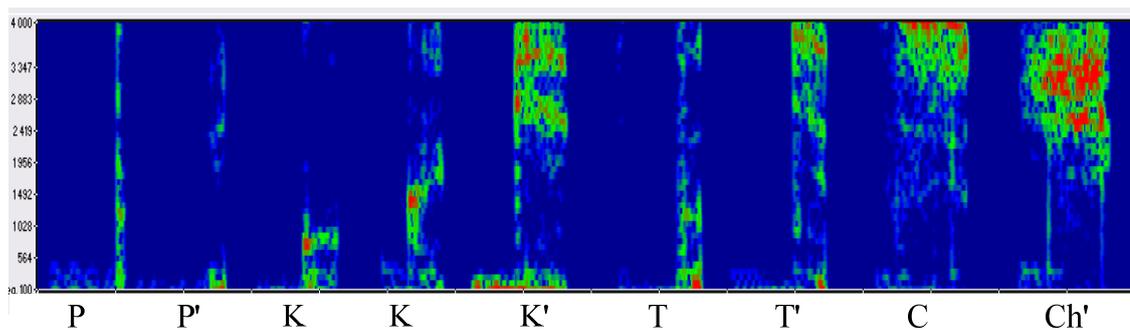
Рис. П 4.13. Спектральные портреты аллофонов ударной гласной *O*

П 4.2. Согласные фонемы

Спектральные портреты согласных представлены на рис. П 4.14 – П 4.15.

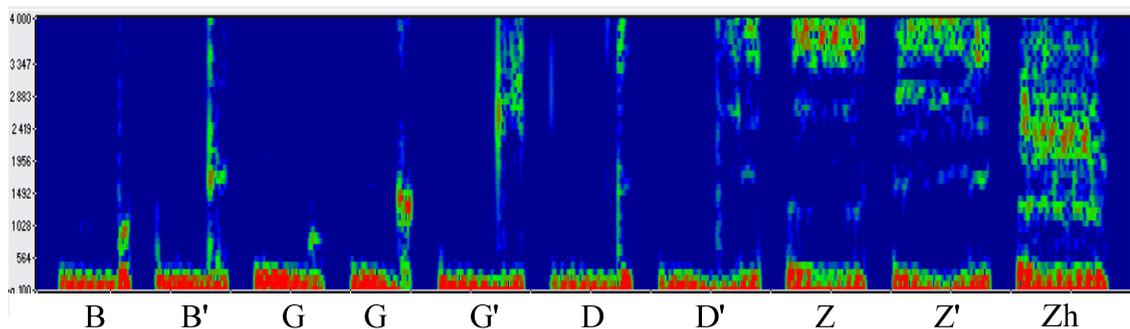


a)

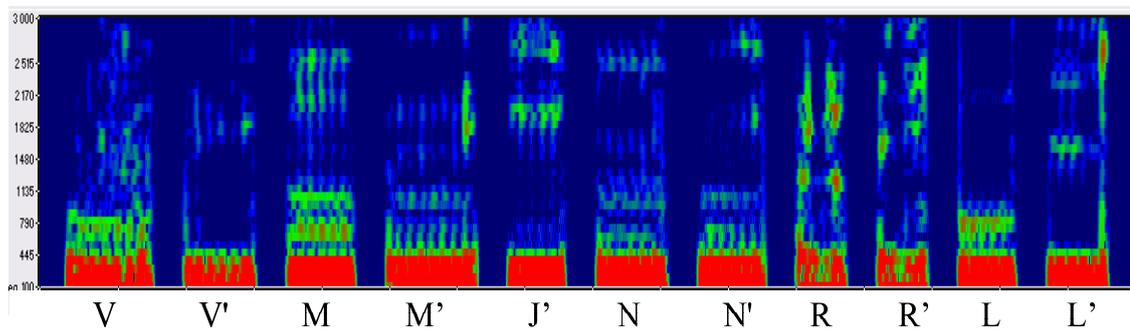


б)

Рис. П 4.14. Спектральные портреты основных аллофонов глухих согласных фонем



а)



б)

Рис. П 4.15. Спектральные портреты основных аллофонов звонких согласных фонем

Приложение 5

Мелодические портреты основных интонационных типов вопросительной интонации русской речи

На графиках ось абсцисс показывает нормированное время: каждая синтагма разделена на акцентные единицы (АЕ), которые, в свою очередь, разделены на три участка: предъядро (все аллофоны, предшествующие полноударному гласному), ядро (полноударный гласный), заядро (все аллофоны, следующие за полноударным гласным). В подрисуночных подписях граница АЕ обозначается символом «/», позиция ударения показана символом «+», следующим за полноударным гласным.

Ось ординат на графиках представляет нормированную частоту основного тона (ЧОТ) для мужского голоса (диапазон ЧОТ: 70-180 Гц).

П 5.1. Мелодические портреты интонации частного и общего вопросов с различным количеством АЕ

Мелодические портреты интонации частного и общего вопросов с количеством АЕ от одной до четырёх представлены на рис. П 5.1 – П 5.4.

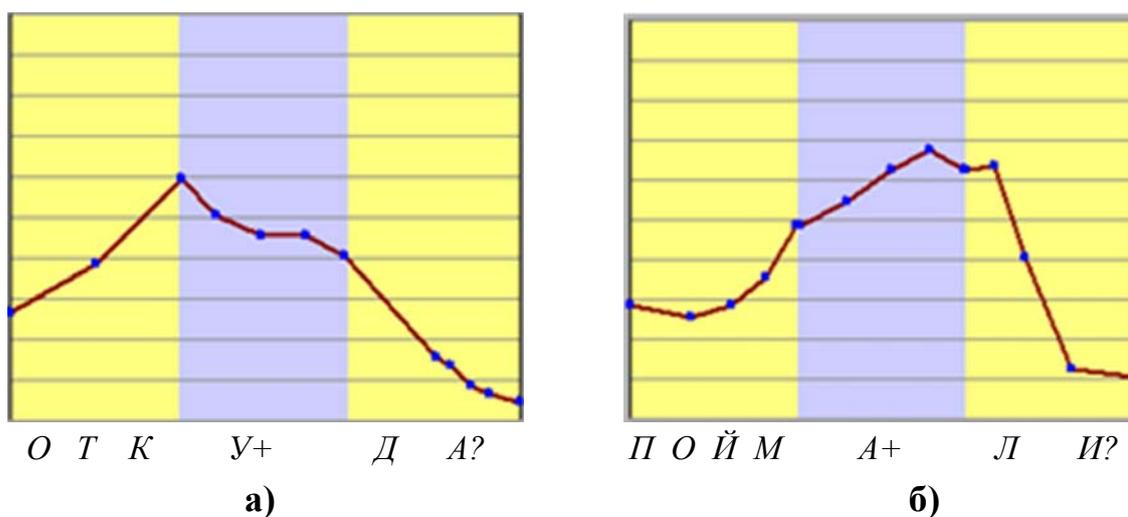
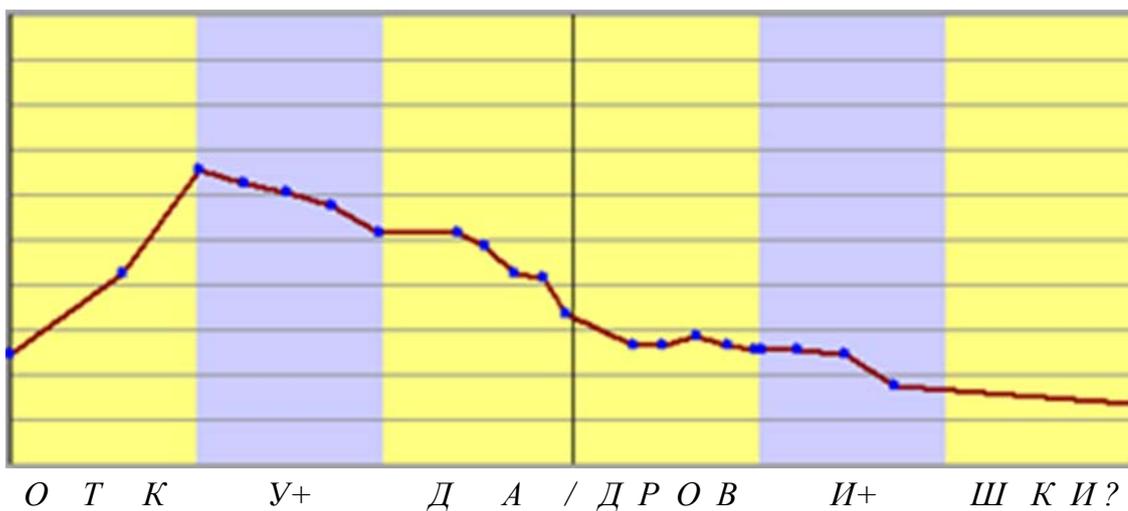


Рис. П 5.1. Мелодические портреты одноакцентной синтагмы
вопросительного типа

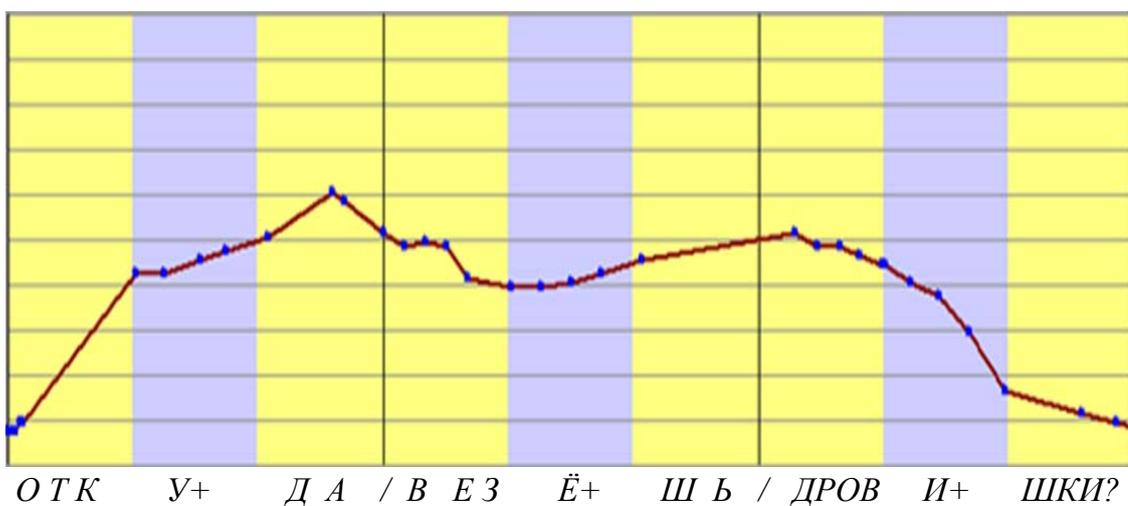


а)



б)

**Рис. П 5.2. Мелодические портреты двухакцентной синтагмы
вопросительного типа**



а)



МИТ Я+ Х У / ВОР И+ ШКУ / ПОЙМ А+ ЛИ?

б)

**Рис. П 5.3. Мелодические портреты трёхакцентной синтагмы
вопросительного типа**



ОТК У+ ДА / ВЕЗ Ё+ ШЬ / ДРОВ И+ ШКИ / ДОМ О+ Й?

а)



ПРОП О+ ЙЦУ / МИТ Я+ ХУ / ВОР И+ ШКУ / ПОЙМ А+ ЛИ?

б)

**Рис. П 5.4. Мелодические портреты четырёхакцентной синтагмы
вопросительного типа**

П 5.2. Мелодические портреты интонации частного вопроса с вопросительным местоимением без частицы и с частицами *бы, же, -нибудь, ну, а, и, да*.



Какую вы хотите?

а)

Когда вы хотите?

б)

Рис. П 5.5. Мелодические портреты интонации частного вопроса без частицы



Какую бы вы хотели?

а)

Когда бы вы хотели?

б)

Рис. П 5.6. Мелодические портреты интонации частного вопроса с частицей *бы*



Какую же вы хотите?

а)

Когда же вы хотите?

б)

Рис. П 5.7. Мелодические портреты интонации частного вопроса с частицей *же*



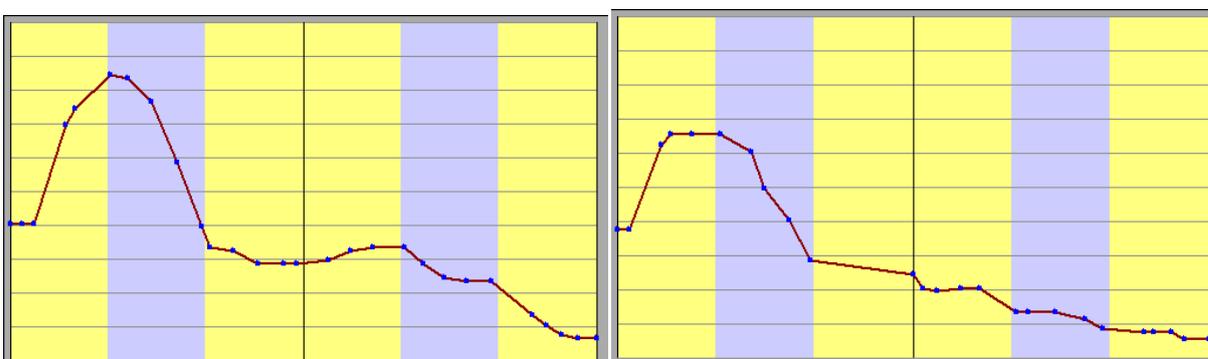
Какую-нибудь хотите?

а)

Когда-нибудь хотите?

б)

Рис. П 5.8. Мелодические портреты интонации частного вопроса с частицей *нибудь*



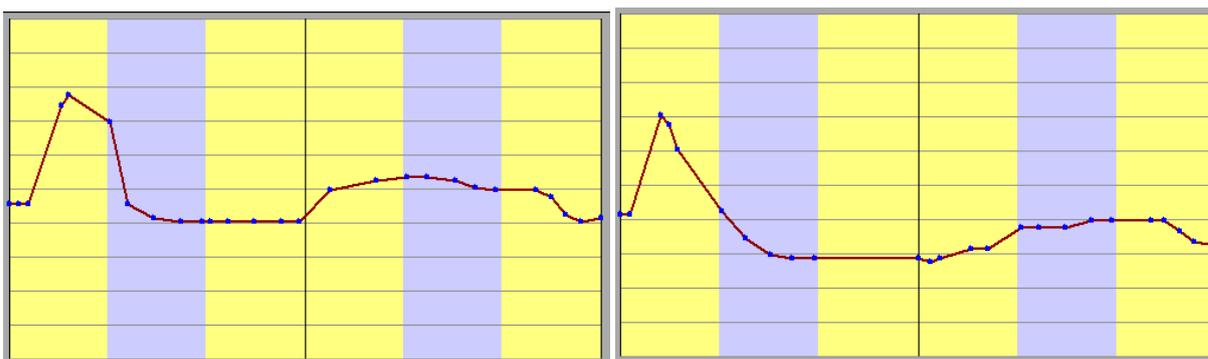
Ну какую вы хотите?

а)

Ну когда вы хотите?

б)

Рис. П 5.9. Мелодические портреты интонации частного вопроса с частицей *ну*



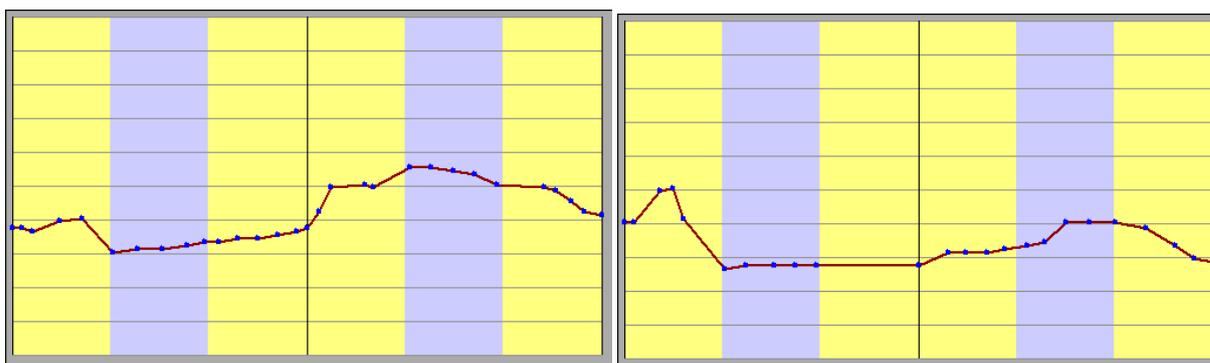
А какую вы хотите?

а)

А когда вы хотите?

б)

Рис. П 5.10. Мелодические портреты интонации частного вопроса с частицей *а*



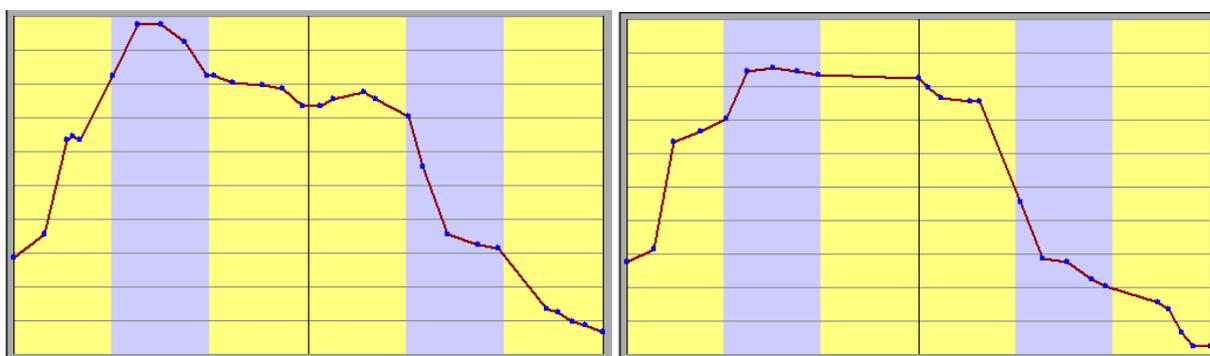
И какую вы хотите?

а)

И когда вы хотите?

б)

Рис. П 5.11. Мелодические портреты интонации частного вопроса с частицей *и*



Да какую вы хотите?

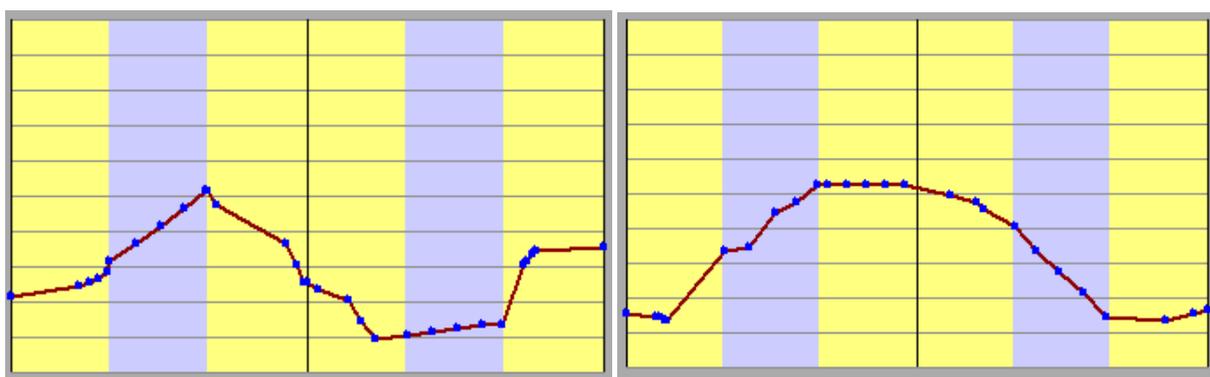
а)

Да когда вы хотите?

б)

Рис. П 5.12. Мелодические портреты интонации частного вопроса с частицей *да*

П 5.3. Мелодические портреты интонации частного вопроса с вопросительными частицами *ли, разве, правда, никак, а, и*.



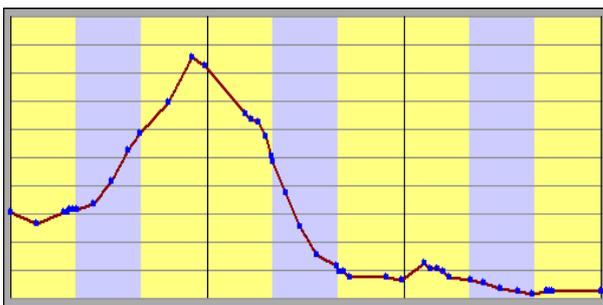
Желаете ли пообедать?

а)

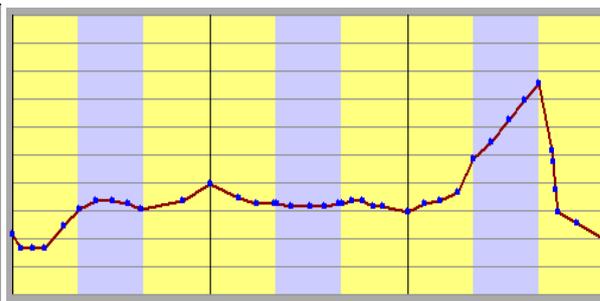
Разве не хочется пообедать?

б)

Рис. П 5.13. Мелодические портреты интонации частного вопроса с вопросительными частицами *ли* (а) и *разве* (б)

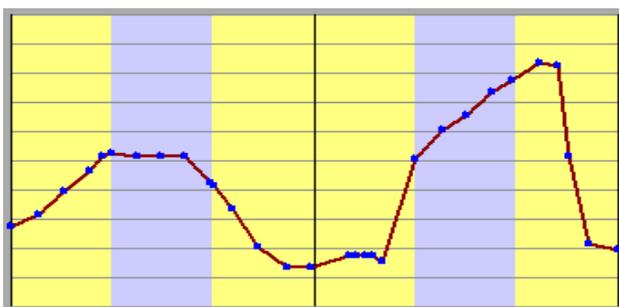


Правда желаете пообедать?



Никак собрались пообедать?

Рис. П 5.14. Мелодические портреты интонации частного вопроса с вопросительными частицами *правда* (а) и *никак* (б)



А Марина пообедала?



И Марина пообедала?

Рис. П 5.15. Мелодические портреты интонации частного вопроса с вопросительными частицами *а* (а) и *и* (б)

Приложение 6

Списки служебных частей речи (предлогов, частиц, междометий, союзов), местоимений и слов-фонетических исключений

II 6.1. Предлоги

Предлоги - служебные слова, выражающие в сочетании с косвенными падежами существительных (а также субстантивированных прилагательных, местоимений и числительных) различные отношения между формами имени и другими словами в предложении.

В первом столбце таблицы в алфавитном порядке представлены простые (несоставные) предлоги с указанием местоположения частичного ударения (=). Отсутствие знака ударения означает, что данный предлог присоединяется к знаменательному слову, образуя с ним одно фонетическое слово.

Во втором столбце указан вид падежного управления (иногда не единственный) следующим за предлогом существительного, местоимения или прилагательного. Указание вида падежного управления полезно в тех случаях, когда в базовом словаре отсутствует информация о падежной форме слова.

<i>Предлог</i>	<i>Управление</i>	<i>во</i>	<i>предл.п.</i>
а-ля=	им.п.	<i>Предлог</i>	<i>Управление</i>
без	род.п.	Вовну=брь	род.п.
Бе=зо	род.п.	Во=зле	род.п.
Благодаря=	дат.п.	Вокру=г	род.п.
близ	род.п.	Вопреки=	дат.п.
в	вин.п.	Восле=д	дат.п.
в	предл.п.	Вперё=д	род.п.
Вблизии=	род.п.	Впередии=	род.п.
Ввиду=	род.п.	Вро=де	род.п.
Вглу=брь	род.п.	Всле=д	дат.п.
Вдо=ль	род.п.	Всле=дствие	род.п.
Взамен=	род.п.	Выключа=я	вин.п.
Включа=я	вин.п.	Вы=ше	род.п.
Вко=сь	род.п.	для	род.п.
Вкру=г	род.п.	до	род.п.
Вме=сто	род.п.	за	вин.п.
вне	род.п.	за	тв.п.
Внизу=	род.п.	из	род.п.
Внутри=	род.п.	И=з-за	род.п.
Вну=брь	род.п.	Изнутри=	род.п.
во	вин.п.	изо	род.п.

И=з-под	род.п.	по	вин.п.
Исключю=чая	вин.п.	по	предл.п.
к	дат.п.	Пове=рх	род.п.
Касса=емо	род.п.	под	вин.п.
Касса=тельно	род.п.	под	тв.п.
ко	дат.п.	По=дле	род.п.
Кро=ме	род.п.	подо	вин.п.
Круго=м	род.п.	подо	тв.п.
меж	род.п.	Подо=бно	дат.п.
меж	тв.п.	Позади=	род.п.
Ме=жду	род.п.	Поми=мо	род.п.
Ме=жду	тв.п.	Поперё=к	род.п.
Ми=мо	род.п.	Посереди=	род.п.
Ми=нус	им.п.	Посереди=не	род.п.
на	вин.п.	Посере=дь	род.п.
на	предл.п.	По=сле	род.п.
Наверху=	род.п.	Посре=дством	род.п.
Навстре=чу	дат.п.	Превы=ше	род.п.
над	тв.п.	пред	тв.п.
надо	тв.п.	предо	тв.п.
Накану=не	род.п.	Пре=жде	род.п.
Наме=сто	род.п.	при	предл.п.
Напереко=р	дат.п.	про	вин.п.
Наподо=бие	род.п.	Проме=ж	род.п.
Напро=тив	род.п.	Проме=жду	род.п.
Насупро=тив	род.п.	Про=тив	род.п.
Насчё=т	род.п.	Путё=м	род.п.
Несмотря=	вин.п.	Ра=ди	род.п.
Ни=же	род.п.	с	вин.п.
о	вин.п.	с	тв.п.
о	предл.п.	с	род.п.
об	вин.п.	Сбо=ку	род.п.
об	предл.п.	Све=рх	род.п.
обо	вин.п.	Све=рху	род.п.
обо	предл.п.	Свы=ше	род.п.
О=коло	род.п.	Се=редь	род.п.
Окре=ст	род.п.	Сза=ди	род.п.
от	род.п.	Скво=зь	вин.п.
Относи=тельно	род.п.	Сни=зу	род.п.
ото	род.п.	со	род.п.
Пе=ред	тв.п.	со	вин.п.
Пе=редо	тв.п.	со	тв.п.
Плю=с	им.п.	Согла=сно	дат.п.
по	дат.п.	Сообра=зно	дат.п.

Соотве=тственно	дат.п.	Су=дя	вин.п.
Соразме=рно	дат.п.	Супроти=в	род.п.
Спе=реди	род.п.	у	род.п.
Спустя=	вин.п.	Че=рез	вин.п.
Среди=	род.п.	Че=резо	вин.п.
Сре=дь	род.п.	Чре=з	вин.п.

II 6.2. Частицы

Частицы - служебные слова, выражающие дополнительные смысловые оттенки предложений, словосочетаний и отдельных слов.

В первом столбце таблицы в алфавитном порядке представлены простые (несоставные) частицы с указанием местоположения полного (+) или частичного ударения (=). Отсутствие знака ударения означает, что данная частица присоединяется к последующему или к предшествующему знаменательному слову, образуя с ним одно фонетическое слово. Частицы, присоединяемые к предшествующему слову помечены звёздочкой (*).

Во втором столбце указана функциональная характеристика частицы (иногда не единственная). Указание функциональной характеристики частицы полезно во многих случаях для адекватного интонационного оформления синтезированной речи.

<i>Частица</i>	<i>Характеристика</i>	<i>Частица</i>	<i>Характеристика</i>
а	Вопросительная	не	Отрицательная
а	Выделительно-ограничительная	небось	Реплика
абы	Реплика	невесть	Реплика
авось	Реплика	нет	Отрицательная
ага	Реплика	неужели	Вопросительная
аж	Реплика	неужели	Отрицательная
али	Реплика	неужели	Утвердительная
аминь	Реплика	неужто	Вопросительная
ан	Реплика	нехай	Реплика
*б	Сослагательное наклонение	нечего	Отрицательная
*бишь	Реплика	нешто	Вопросительная
благо	Сравнительная	ни	Отрицательная
более	Определительно-уточнительная	ни	Усилительная
больше	Определительно-уточнительная	*-нибудь	Определительно-уточнительная
будто	Сравнительная	никак	Вопросительная
*бы	Сослагательное наклонение	никак	Вопросительная

бывает	Признак протекания	ничего	Отрицательная
бывало	Признак протекания	но	Сравнительная
было	Признак протекания	ну	Отрицательная
ведь	Усилительная	ну	Признак протекания
вишь	Эмоционально-восклицательная	ну	Утвердительная
во	Эмоционально-восклицательная	ну-ка	Побудительная
вовсе		однако	Определительно-уточнительная
вон	Указательная	оно	Выделительно-ограничительная
вот	Указательная	поди	Сравнительная
вот	Утвердительная	пожалуй	Выражение сомнения
вот-то	Утвердительная	пожалуйста	Побудительная
вроде	Сравнительная	пока	Определительно-уточнительная
всё	Усилительная	почти	Выделительно-ограничительная
всего	Выделительно-ограничительная	почти	Определительно-уточнительная
где	Выражение сомнения	просто	Усилительная
где	Отрицательная	прямо	Отрицательная
гляди	Побудительная	прямо	Усилительная
глядь	Эмоционально-восклицательная	пускай	Побудительная
да	Повелительное наклонение	пускай	Повелительное наклонение
да	Утвердительная	пусть	Побудительная
давай	Побудительная	пусть	Повелительное наклонение
давай	Повелительное наклонение	разве	Вопросительная
давай	Признак протекания	разве	Вопросительная
давайте	побудительная	ровно	Определительно-уточнительная
давайте	Повелительное наклонение	*-с	Определительно-уточнительная
даже	Усилительная	самое	Сравнительная
дай	Побудительная	себе	Указательная
дай-ка	Побудительная	скорее	Сравнительная
дайте	Побудительная	словно	Сравнительная
*де	Реплика	смотри	Побудительная

*-де	Реплика	совсем	Определительно- уточнительная
дескать	реплика	так	Утвердительная
добро	Выражение сомнения	*-таки	Определительно- уточнительная
если	Определительно- уточнительная	так-то	Утвердительная
есть	Утвердительная	там	Указательная
ещё	Выделительно- ограничительная	*те	Выделительно- ограничительная
*ж	Усилительная	тебе	Выделительно- ограничительная
*же	Усилительная	*-то	Определительно- уточнительная
знай	Утвердительная	того	Указательная
и	Признак протекания	тоже	Отрицательная
и	Усилительная	тоже	Сравнительная
идет	Утвердительная	только	Выделительно- ограничительная
или	Выделительно- Ограничительная	только	Выделительно- ограничительная
именно	Утвердительная	точно	Определительно- уточнительная
инда	Реплика	точно	Сравнительная
ишь	Эмоционально- восклицательная	точно	Утвердительная
*-ка	Выражение смягчения	уж	Усилительная
кабы	Вопросительно- восклицательная	уже	Усилительная
как	вопросительная	ужели	Вопросительная
как	Признак протекания	ужели	Вопросительная
как	сравнительная	ужель	Вопросительная
как	Эмоционально- восклицательная	хорошо	Утвердительная
какое	Отрицательная	хоть	Выделительно- ограничительная
конечно	Утвердительная	хотя	Выделительно- ограничительная
куда	Отрицательная	чай	Реплика
куда	Эмоционально- восклицательная	чего	Эмоционально- восклицательная
ладно	Утвердительная	чисто	Сравнительная
*ли	Вопросительная	что	Вопросительная
*-либо	Определительно-	чтоб	Повелительное

*лишь	уточнительная Выделительно-ограничительная	чтобы	наклонение Усилительная
лучше	Сравнительная	эва	Реплика
*ль	Вопросительная	эвон	Реплика
мерси	Реплика	эк	Реплика
мол	Реплика	эка	Реплика
на	Эмоционально-восклицательная	это	Указательная
на-ка	Эмоционально-восклицательная	якобы	Реплика
нате	Эмоционально-восклицательная		
нате-ка	Эмоционально-восклицательная		

II 6.3. Междометия

Междометия - это слова, которые непосредственно выражают чувства, переживания и волеизъявления, не называя их.

В первом столбце таблицы в алфавитном порядке представлены междометия с указанием местоположения полного (+) или частичного ударения (=). Во втором столбце указана функциональная характеристика междометия (иногда не единственная). Указание функциональной характеристики частицы полезно во многих случаях для адекватного интонационного оформления синтезированной речи.

<i>Междометие</i>	<i>Характеристика</i>	<i>Междометие</i>	<i>Характеристика</i>
А+!	Удовлетворение	Одна+ко!	Удивление
А+!	Удивление	Па+с!	Призыв
А+!	Досада	Па+ф!	Звукоподражание.
А+й!	Удивление	Пардо+н!	Призыв
А+й-а=й-а=й!	Удивление	Пи+ф-па+ф!	Звукоподражание
А+мба!	Досада	Пли+!	Призыв
А+х ты=, го+споди!	Досада	Плюю+х!	Звукоподражание
А+х!	Удовлетворение	Полу+ндра!	Призыв
А+х!	Сожаление	Пра+во!	Эмоциональная оценка
А+х!	Досада	Пши+к!	Звукоподражание
Ага+!	Удовлетворение	Пы+х!	Звукоподражание
Агу+!	Призыв	Скри+п!	Звукоподражание
Айда+!	Призыв	Сто+п!	Призыв
Алло+!	Призыв	Сту+к!	Звукоподражание

Апо+рт!	Призыв	Тарара+х!	Звукоподражание
Атанде+!	Призыв	Ти+к-та=к!	Звукоподражание
Ату+!	Призыв	То+п!	Звукоподражание
Ау+!	Призыв	То+-то!	Удовлетворение
Ахти+!	Удивление	Тпру+!	Призыв
Ба+!	Удивление	Тра+-та=-та=!	Звукоподражание
Ба+иньки-баю=!	Успокоение	Тра+х!	Звукоподражание
	Эмоциональная		
Ба+ста!	оценка	Тра+х-тарара+х!	Звукоподражание
Ба+тюшки!	Удивление	Тре+нь-бре+нь!	Звукоподражание
Ба+х!	Звукоподражание	Трю+х-трю+х!	Звукоподражание
Ба+ц!	Звукоподражаниее	Тс!	Призыв
Ба+юшки-баю=!	Успокоение	Тсс!	Призыв
Ба=й-ба+й!	Успокоение	Ту+к!	Звукоподражание
Ба=ю-ба+й!	Успокоение	Ту+к-ту+к!	Звукоподражание
Баба+х!	Звукоподражание	Тьфу+!	Презрение
	Эмоциональная		
Би+с!	оценка	Тю+к!	Звукоподражание
	Эмоциональная		
Бо+же!	оценка	Тю=-тю+!	Сожаление
Бо+м!	Звукоподражание	Тя+п!	Звукоподражание
Бра+во!	Удовлетворение	У+!	Удовлетворение
	Эмоциональная		
Брави+ссимо!	оценка	У+!	Презрение
	Эмоциональная		
Брр!	оценка	У+-у=-у=!	Звукоподражание
Бры+сь!	Призыв	У+ф!	Сожаление
Бря+к!	Звукоподражание	У+х!	Сожаление
Бу+льк!	Звукоподражание	Уа+!	Звукоподражание
Бу+м!	Звукоподражание	Увы+!	Сожаление
Бу+х!	Звукоподражание	Угу+!	Согласие
Бу=ль-бу+ль!	Звукоподражание	Улюлю+!	Звукоподражание
Булты+х!	Звукоподражание	Ура+!	Удовлетворение
Ве+рть!	Звукоподражание	Фи+!	Презрение
		Физку+льт- ура+!	Призыв
Ви+Ра!	Призыв		
	Эмоциональная		
Вива+т!	оценка	Фу+!	Презрение
	Эмоциональная		
Во+!	оценка	Фу+й!	Презрение
	Эмоциональная		
Во+т еще=!	оценка	Фу+к!	Звукоподражание
Га+в-га+в!	Звукоподражание	Фю+йть!	Звукоподражание
Ге+й!	Призыв	Ха+!	Ирония

Гля+дь!	Призыв	Ха=-ха+!	Ирония
Гм!	Досада	Хва+ть!	Звукоподражание
Го+п!	Звукоподражание	Хе=-хе+!	Ирония
Го+пля!	Звукоподражание	Хи=-хи+!	Ирония
Го+споди!	Удивление	Хле+сть!	Звукоподражание
Гро+х!	Звукоподражание	Хло+п!	Звукоподражание
Гу=ль-гу+ль!	Звукоподражание	Хлы+сть!	Звукоподражание
Дзи+нь!	Звукоподражание	Хлю+п!	Звукоподражание
Ди+нь-ди+нь!	Звукоподражание	Хля+сть!	Звукоподражание
Ду+дки!	Эмоциональная оценка	Хм!	Сомнение
Е+й-бо=гу!	Эмоциональная оценка	Хо+п!	Звукоподражание
Е+й-е=й!	Эмоциональная оценка	Хо=-хо+!	Сомнение
Е+сть!	Эмоциональная оценка	Хру+п!	Звукоподражание
Е=й-бо+гу!	Эмоциональная оценка	Хру+сть!	Звукоподражание
Е=й-е+й!	Эмоциональная оценка	Хрю+к!	Звукоподражание
Звя+к!	Звукоподражание	Хрю=-хрю+!	Звукоподражание
Здо+рово!	Эмоциональная оценка	Хря+сть ¹	Звукоподражание
И+шь!	Эмоциональная оценка	Цо+к!	Звукоподражание
Карау+л!	Призыв	Цы+ц!	Призыв
Ки=с-ки+с!	Призыв	Цы=п-цы+п!	Звукоподражание
Кря+к!	Звукоподражание	Че+рт возьми+!	Досада
Ку=-ку+!	Звукоподражание	Че+рта с два+!	Досада
Кукареку+!	Звукоподражание	Чебура+х!	Звукоподражание
Куку+!	Звукоподражание	Чи+к!	Звукоподражание
Кш!	Призыв	Чи+рк!	Звукоподражание
Кы+ш!	Призыв	Чи+х!	Звукоподражание
Ма+йна!	Призыв	Чи=к-чири+к!	Звукоподражание
Ма+рш!	Призыв	Чмо+к!	Звукоподражание
Ма+тушки!	Удивление	Чу+!	Призыв
М-да+!	Удивление	Чу+р!	Призыв
Мя+у!	Звукоподражание	Чу+р-чура+!	Призыв
Мя+у-мя=у!	Звукоподражание	Чш!	Призыв
Но+!	Презрение	Ша+!	Призыв
Ну+!	Призыв	Ша+сть!	Звукоподражание
Ну+, вали+!	Призыв	Шара+х!	Звукоподражание
Ну+-ка!	Призыв	Шва+рк!	Звукоподражание

Ну= и ну+!	Удивление	Шлее+п!	Звукоподражание
Ну=-ну+!	Удивление	Щё+лк!	Звукоподражание
О+!	Удивление	Э+врика!	Призыв
О+й!	Удивление	Э+й!	Призыв
О+й-о=й-о=й	Удивление	Э+х!	Сожаление
О+п!	Звукоподражание	Э+хма!	Сожаление
О+пля!	Звукоподражание	Э+-э=х!	Сожаление
О+х!	Удивление	Э=-ге=-ге+!	Сожаление
О=-го=-го+!	Удивление	Э=-хе=-хе+!	Сожаление
Ого+!	Удивление	Эге+!	Сожаление

II 6.4. Союзы

Союзы - служебные слова, служащие для связи однородных членов простого предложения и частей сложного предложения - сложносочиненного и сложноподчиненного.

В первом столбце таблицы в алфавитном порядке представлены простые (несоставные) союзы с указанием местоположения частичного ударения (=). Во втором столбце указан тип союза – сочинительный или подчинительный. В третьем столбце указана функциональная характеристика союза (иногда не единственная). Указание функциональной характеристики союза полезно во многих случаях для адекватного интонационного оформления синтезированной речи.

Союзы	Тип союза	Характеристика
а	Сочинительный	Противительный
абы	Подчинительный	Целевой
аж	Подчинительный	Сравнительный
али	Сочинительный	Разделительный
аль	Сочинительный	Разделительный
ан	Сочинительный	Неопределённый
благо	Подчинительный	Уступительный
буде	Подчинительный	Условный
будто	Подчинительный	Сравнительный
ведь	Подчинительный	Причинный
впрочем	Подчинительный	Уступительный
всё-таки	Подчинительный	Уступительный
да	Сочинительный	Соединительный
да	Сочинительный	Противительный
дабы	Подчинительный	Целевой
даже	Сочинительный	Присоединительный
доколе	Подчинительный	Временной
едва	Подчинительный	Временной
ежели	Подчинительный	Условный
если	Подчинительный	Условный

зато	Сочинительный	Противительный
и	Сочинительный	Соединительный
и	Сочинительный	Присоединительный
ибо	Подчинительный	Причинный
или	Сочинительный	Разделительный
или	Сочинительный	Пояснительный
иль	Сочинительный	Разделительный
иначе	Сочинительный	Пояснительный
итак	Подчинительный	Следствия
кабы	Подчинительный	Условный
как	Подчинительный	Изъяснительный
как	Подчинительный	Временной
как	Подчинительный	Условный
как	Подчинительный	Сравнительный
когда	Подчинительный	Временной
когда	Подчинительный	Условный
коли	Подчинительный	Условный
коль	Подчинительный	Условный
либо	Сочинительный	Разделительный
лишь	Подчинительный	Временной
нежели	Подчинительный	Сравнительный
нешто	Подчинительный	Сравнительный
ни	Сочинительный	Соединительный
ни ... ни	Сочинительный	Соединительный
но	Сочинительный	Противительный
однако	Сочинительный	Противительный
отчего	Подчинительный	Причинный
плюс	Сочинительный	Соединительный
пока	Подчинительный	Временной
покамест	Подчинительный	Временной
поколе	Подчинительный	Временной
поколь	Подчинительный	Временной
покуда	Подчинительный	Временной
поскольку	Подчинительный	Причинный
постольку	Подчинительный	Следствия
поэтому	Подчинительный	Следствия
правда	Подчинительный	Уступительный
притом	Сочинительный	Пояснительный
причём	Сочинительный	Пояснительный
пускай	Подчинительный	Уступительный
пусть	Подчинительный	Уступительный
равно	Подчинительный	Сравнительный
раз	Подчинительный	Причинный
разве	Подчинительный	Условный

ровно	Подчинительный	Сравнительный
следовательно	Подчинительный	Следствия
следственно	Подчинительный	Следствия
словно	Подчинительный	Сравнительный
соответственно	Подчинительный	Сравнительный
так	Сочинительный	Соединительный
также	Сочинительный	Присоединительный
тем	Подчинительный	Сравнительный
то	Сочинительный	Разделительный
то ... то	Сочинительный	Разделительный
тоже	Сочинительный	Присоединительный
только	Подчинительный	Временной
только-только	Подчинительный	Временной
точно	Подчинительный	Сравнительный
хоть	Сочинительный	Уступительный
хотя	Подчинительный	Уступительный
чем	Подчинительный	Целевой
что	Подчинительный	Изъяснительный
чтоб	Подчинительный	Изъяснительный
чтоб	Подчинительный	Целевой
чтобы	Подчинительный	Целевой
чуть	Подчинительный	Временной
яко	Подчинительный	Изъяснительный
якобы	Подчинительный	Изъяснительный

II 6.5. Местоимения

К местоимениям относятся слова, которые, не называя предметов или признаков, указывают на них. В таблице в первом столбце указана только основная падежная форма местоимения – именительный падеж. Во втором столбце указан грамматический тип местоимения.

<i>Местоимение</i>	<i>Тип</i>	<i>Местоимение</i>	<i>Тип</i>
я	Личное	что	Вопросительное
ты	Личное	какой	Вопросительное
мы	Личное	каков	Вопросительное
вы	Личное	который	Вопросительное
он	Личное	чей	Вопросительное
она	Личное	сколько	Вопросительное
оно	Личное	никто	Отрицательное
они	Личное	ничто	Отрицательное
себя	Возвратное	никакой	Отрицательное
мой	Притяжательное	ничей	Отрицательное
твой	Притяжательное	некого	Отрицательное
наш	Притяжательное	нечего	Отрицательное

ваш	Притяжательное	нисколько	Отрицательное
его	Притяжательное	некто	Неопределенное
ее	Притяжательное	нечто	Неопределенное
их	Притяжательное	некоторый	Неопределенное
свой	Притяжательное	некий	Неопределенное
тот	Указательное	кто-то	Неопределенное
этот	Указательное	что-то	Неопределенное
такой	Указательное	кто-либо	Неопределенное
таков	Указательное	что-либо	Неопределенное
оба	Указательное	какой-либо	Неопределенное
столько	Указательное	чей-либо	Неопределенное
сей	Указательное	кое-кто	Неопределенное
онный	Указательное	кое-что	Неопределенное
весь	Определительное	какой-то	Неопределенное
сам	Определительное	чей-то	Неопределенное
самый	Определительное	кто-нибудь	Неопределенное
всякий	Определительное	что-нибудь	Неопределенное
каждый	Определительное	какой-нибудь	Неопределенное
иной	Определительное	чей-нибудь	Неопределенное
кто	Вопросительное	кое-какой	Неопределенное

II 6.6. Фонетические исключения

Фонетические исключения – слова, произношение которых не подчиняется стандартным правилам произношения для русского языка. В первом столбце таблицы в алфавитном порядке представлены слова-фонетические исключения, во втором – их эквиваленты, подчиняющиеся стандартным правилам произношения.

Словоформы слов-фонетических исключений, приведенных в таблице, а также слова, производные от приведенных, тоже являются фонетическими исключениями. Правила формирования эквивалентов для них аналогично формированию эквивалента для исходного слова.

<i>Слово – фонетическое исключение</i>	<i>Эквивалент</i>
Альтернатива	Альтэрнатива
Амстердам	Амстэрдам
Ателье	Атэлье
Баден-баден	Бадэн-бадэн
Бизнес	Бизнэс
Бог	Бох
Бутерброд	Бутэрброд
Варьете	Варьетэ

Де	Дэ
Декольте	Дэкольтэ
Депо	Дэпо
Дорого	Дорога
Душанбе	Душанбэ
Заботьяся	Забоцяся
Здравствуй	Здраствуй
Интервью	Интэrvью
Кабаре	Кабарэ
Кафе	Кафэ
Кашне	Кашнэ
Компьютер	Компьютэр
Конечно	Конешна
конечно	конешна
Марсель	Марсэль
Много	Многа
Модель	Модэль
Моделировать	Модэлировать
Мужчина	Мущина
Нависшие	Навишие
Нарочно	Нарошно
Недорого	Недорога
Немного	Немнога
Нестрого	Нестрога
Нью-дели	Нью-дэли
Охотьяся	Охоцяся
Пожалуйста	Пожалуста
Позаботьяся	Позабоцяся
Поохотьяся	Поохоцяся
Пятьдесят	Пядесят
Сегодня	Севодня
Синтез	Синтэз
Солнце	Сонце
Сросшихся	Срошихся
Стенд	Стэнд
Строго	Строга
Тельавив	Тэльавив
Теннис	Тэннис
Тест	Тэст
Филателист	Филатэлист
Фортепьяно	Фортэпиано

Фрикаделька	Фрикадэлька
Цюрих	Цурих
Что	Што
Чтобы	Штобы
Чувства	Чуства
Шестьдесят	Шездесят
Шестьсот	Шессот
Шоссе	Шоссэ
Юнеско	Юнэско

Литература

1. <http://www.sintezator.narod.ru>
2. Лобанов Б.М. и др. Синтезатор персонализированной речи по тексту «ЛобаноФон-2000» // Тр. Междунар. конф., посвященной 100-летию российской экспериментальной фонетики, СПб., 1 – 4 февраля 2001 г. – С. 101 – 104.
3. Gordos, G. General Chairman's Message // Speech Communication and Technology: book of abstracts of the 6-th European conference EUROSPEECH'99, Budapest, Hungary, 5-9 September 1999. – Budapest, 1999. – P. VI–VIII.
4. Trancoso, I. A Welcome Message from the General Chair // European Conference on Speech Communication and Technology: book of abstracts of 9th conference INTERSPEECH'2005, September 4-8, 2005, Lisboa, Portugal. – P. III.
5. Лобанов, Б.М. Компьютерное клонирование персонального голоса и речи / Б.М. Лобанов // Новости искусственного интеллекта. – 2002. – №5. – С. 35–39.
6. Lobanov, B, Karnevskaya, N. TTS-Synthesizer as a Computer Means for Personal Voice “Cloning” // Phonetics and its Applications. Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday. – Stuttgart: Franz Steiner Verlag, 2002. – P. 445–452.
7. Валгина, Н.С., Розенталь, Д.Э., Фомина М.И. Современный русский язык. М., Логос, 2002.
8. Жинкин Н.И. Механизмы речи. М., 1958.
9. Артемов В.А. Экспериментальная фонетика. М., 1956.
10. Скалзуб Л.Г. Палатограммы и рентгенограммы согласных фонем русского литературного языка. Киев, 1963.
11. Матусевич М.И., Любимова Н.И. Альбом артикуляции звуков русского языка. М., 1963.
12. Чистович Л.А. и др. Речь. Артикуляция и восприятие. Л., "Наука", 1965.
13. Сапожков М.А. Речевой сигнал в кибернетике и связи. М., Связьиздат, 1963.
14. Фланаган Дж. Анализ, синтез и восприятие речи. – М., Связь, 1968. – 396 с.
15. Фант Г. Акустическая теория речеобразования. – М., Наука, 1964. – 284 с.
16. Аванесов Р.И. Фонетика современного русского литературного

языка. М., 1956.

17. Панов М.В. Русская фонетика. М., 1967.

18. Зиндер Л.Р. Общая фонетика. – Изд-во ЛГУ, 1960. – 336 с.

19. Матусевич М.И. Современный русский язык. Фонетика. М., 1976.

20. Бондарко Л.В. Звуковой строй современного русского языка. – Л., 1977. – 176 с.

21. Артемов В.А. Метод структурного анализа речевой интонации. М., 1962.

22. Златоустова Л.В. Фонетическая структура слова в потоке речи. Казань, 1962.

23. Брызгунова Е.А. Звуки и интонация русской речи. М., 1977.

24. Светозарова Н.Д. Интонационная система русского языка. Л., 1982.

25. Лобанов, Б.М. Об автоматическом синтезе некоторых просодических характеристик русской речи // Труды всесоюзной конференции «Анализ и синтез речи», Минск, 15-18 июня 1973 г. – Минск, 1973. – С. 196-202.

26. Lobanov, B., Tsirolnik, L. Statistical study of speaker's peculiarities of utterances into phrases segmentation // Speech Prosody: proceedings of the 3rd International conference, Dresden, Germany, May 2–5, 2006. – Dresden, 2006. – V. 2. – P. 557–560.

27. Lobanov, B., Tsirolnik, L., Sizonov, O. AUP's Modeling of Speaker Specific Intonation Contour Peculiarities // Speech and Computer: proceedings of the 12-th International conference SPECOM'2007, Moscow, Russia, 15-18 October, 2007 / Moscow State Linguistic University. – Moscow, 2007. – V.1. – P. 312-317.

28. Логинова, И.М. Взаимодействие ритмической и звуковой структуры русского языка // Лингвистическая полифония: сб. статей в честь юбилея профессора Р.К. Потаповой. – М.: Языки славянских культур, 2007. – С. 445-461.

29. Потебня, А. А. О звуковых особенностях русских наречий // Филологические записки. 1965 г. Вып. 1.

30. Лобанов, Б.М., Пьорковска, Б., Рафалко, Я., Цирульник, Л.И., Шпилевский, Э. Фонетико-акустическая база данных для многоязычного синтеза речи по тексту на славянских языках // Компьютерная лингвистика и интеллектуальные технологии: труды междунар. конф. Диалог'2006, Бекасово, 31мая – 4 июня 2006 г./ Институт проблем информатики РАН; отв. ред.: Н.И. Лауфер [и др.]. – М.: Наука, 2006. – С. 357–363.

31. Lobanov, B., Tsirulnik, L. Development of multi-voice and multi-language TTS synthesizer (languages: Belarussian, Polish, Russian) // *Speech and Computer: proceedings of the 11-th International conference SPECOM'2006, St. Petersburg, Russia, 25-29 June, 2006* / Institute of Informatics and Automation of RAS, Speech Informatics Group. – St.-Petersburg: Anatolia, 2006. – P. 274-283.

32. Цирульник, Л.И., Лобанов, Б.М. Экспериментальная оценка вклада элементов компиляции в правдоподобие синтезированного речевого клона // *Компьютерная лингвистика и интеллектуальные технологии: труды междунар. конф. Диалог'2006, Бекасово, 31мая – 4 июня 2006 г.* / Институт проблем информатики РАН; отв. ред.: Н.И. Лауфер [и др.]. – М.: Наука, 2006. – С. 545–551.

33. Златоустова Л.В. и др. Алгоритм преобразования русских орфографических текстов в фонемную запись. М.: МГУ, 1970.

34. Лобанов Б.М., Панченко Б.В. Преобразователь графема-фонема для синтеза речи по орфографическому тексту // *Тр. Междунар. конф. «Автоматическое распознавание слуховых образов АРСО-8».* – Ч. 4. – Львов, 1974. – С. 15–18.

35. Цирульник Л.И. Алгоритм генерации фонемной последовательности по орфографическому тексту в системе синтеза речи // *Информатика.* – 2006. – № 4. – С.61–70.

36. Лобанов Б.М. Вопросы автоматического анализа и синтеза фразовой интонации. Тезисы докладов Всесоюзной конференции "Анализ и синтез речи". Минск, 1972. – С. 311-312.

37. Лобанов Б.М. Принципы автоматического синтеза интонационных структур. В кн.: *Автоматическое распознавание слуховых образов.* Тбилиси, 1978. – С. 158-160.

38. Dutoit, T. *An Introduction to text-to-speech synthesis.* – Kluwer Academic Publishers, 1997. – 286 p.

39. d'Alessandro C., Liénard J.-S. Synthetic speech generation // *Survey of the State of the Art in Human Language Technology.* – Cambridge University Press, 1996. – Chapter 5. – P. 155–185.

40. *Speech synthesis / editors: J.L. Flanagan, L.R. Rabiner.* – Dowden, Hutchinson & Ross, 1973. – 511 p.

41. Parthasarathy S., Coker C. H. Automatic estimation of articulatory parameters // *Computer Speech and Language.* – 1992. – № 6. – P. 37–75.

42. Лобанов Б.М., Панченко Б.В. Модель артикуляторного синтеза речи по печатному тексту // *Автоматическое распознавание и синтез речи: тезисы докл. Всесоюзн. научной школы-семинара АРСО-9, Минск, 18-26 сентября 1976г.* – Мн., 1976. – С. 72.

43. Itakura F., Saito S. Digital filtering techniques for speech analysis and synthesis // ICA: proceedings of the 7-th International conference, Budapest, May 7-12, 1971. – Budapest, 1971. – P. 261–264.
44. Галунов В.И. Структура речевого процесса // Модели речевого процесса в норме и патологии: доклады и сообщения Всесоюзн. симпозиума, Грозный, 13-15 июня 1979г. – Л., 1980. – С.13–14.
45. Сорокин В.Н. К теории речеобразования // Речевое общение в автоматизированных системах. – М., Наука, 1975. – С. 103–127.
46. Сорокин В.Н. Шумы турбулизации в речеобразовании // Автоматическое распознавание и синтез речи: тезисы докл. Всесоюзн. научной школы-семинара АРСО-12, Киев, 9-16 июня 1982г. – Киев, 1982. – С. 432–435.
47. Maeda S. The role of sinus cavities in the production of nasal vowels // Acoustic, Speech and Signal processing: proceedings of IEEE International conference ICASSP-82, Paris, France, 3-8 September 1982. – Paris, 1982. – P. 911–914.
48. Klatt D. H. Review of text-to-speech conversion for English // J. Acoust. Soc. Am. – 1987. – Vol. 82. – № 3. – P. 737–793.
49. Allen J. M., Sharon Hunnicutt M., Klatt D.H. From text to speech – the MITalk system. – Cambridge, MIT Press, 1987. – 216 p.
50. Фланаган Дж. Вычислительные машины говорят и слушают // Речевое общение человека с машиной. ТИИЭР. – Т. 64. – № 4. – С. 5–17.
51. Лобанов Б.М. Исследование и разработка методов автоматического синтеза речи по фонемному тексту: дисс. на соискание учёной степени доктора технических наук: 05.13.01. – Мн, 1983. – 323 с.
52. Stevens K.N. Control parameters for synthesis by rule // Speech synthesis: proceedings of the International conference ESCA'90, Autrans, 25 September 1990. – Autrans, 1990. – P. 27–37.
53. Rabiner R. et al. Digital hardware for speech synthesis // Congress on Acoustics: proceedings of the 7-th Int. Congress, Budapest, 10-15 September 1971. – Budapest, 1971. – P. 157–159.
54. Klatt D. Software for cascade/parallel formant synthesizer // J. Acoust. Soc. Am. – 1980. – Vol. 67. – № 3. – P. 971–995.
55. Flanagan J., Langraf L. Self-oscillating source for vocal-tract synthesizer // IEEE Transactions on Audio and Electronics. – 1968. – Vol.10. – № 1. – P. 57–64.
56. Toshio T. “Analysis-by-synthesis” method utilizing spectral features of voice source and measurement of glottal waveform parameters // J. Radio Res. Lab. – 1971. – No 97. – P. 209–220.

57. Rothenberg M. et al. A Three-parameter voice source for speech synthesis // *Speech Communication: proceedings of International Seminar, Stockholm, 3-5 March 1974.* – Stockholm, 1974. – P. 235–243.

58. Бухтилов Л.Д., Гурьянов Н.И., Лобанов Б.М., Минкевич В.В. Цифровой формантный синтезатор // *Автоматическое распознавание и синтез речи: тезисы докл. Всесоюзн. научной школы-семинара АРСО-12, Киев, 9-16 июня 1982г.* – Киев, 1982. – С. 445–448.

59. Darsinos V., Galanis D., Kokkinakis G. A method for fully automatic analysis and modeling of voice source characteristics // *Speech Communication and Technology: proceedings of the 4-th European conference EUROSPEECH'95, Madrid, Spain, 18-21 September 1995.* – Madrid, 1995. – P. 413–416.

60. Laprie M. Berger. A new paradigm for reliable automatic formant tracking // *Acoustics, Speech, and Signal Processing: proceedings of IEEE International conference ICASSP-94, Adelaide, Australia, April 19-22, 1994.* – Adelaide, 1994. – Vol 2. – P. 201–204.

61. Fant G., Hertegård S., Kruckenberg A., Liljencrants J. Covariation of subglottal pressure F_0 and glottal parameters // *Speech Communication and Technology: proceedings of the 5-th European conference EUROSPEECH'97, Rhodes, Greece, 22-25 September 1997.* – Rhodes, 1997. – P. 453–456.

62. Fant G., Hertegård S., Kruckenberg A. Covariation of subglottal pressure, F_0 and intensity // *Speech Communication and Technology: proceedings of the 9-th European conference InterSpeech'2005, Lisbon, Portugal, 4-8 September 2005.* – Lisbon, 2005. – P. 1061–1064.

63. Darsinos D.G., Kokkinakis G. Towards automatic speaker imitation by text-to-speech systems // *Speech and Computer: proceedings of International Workshop SPECOM'96, Saint-Peterburg, Russia, 28-31 October 1996.* – St-Petersburg, 1996. – P. 139–142.

64. *Progress in speech synthesis / editors: J. Van Santen, R. Sproat, J. Olive, J. Hirshberg.* – Springer Verlag, New York, 1997. – 598 p.

65. Лобанов Б.М. Синтез речи по тексту // *Четвёртая международная школа-семинар по искусственному интеллекту: сб. науч. тр., Браславские озёра, Беларусь, 29 июня – 6 июля 2000г.* – Мн.: БГУ, 2000. – С. 57–76.

66. Charpentier F., Stella M.G. Diphone synthesis using an overlap-add technique for speech waveform concatenation // *Acoustics, Speech, and Signal Processing: proceedings of the IEEE international Conference ICASSP-86, Tokyo, Japan, 13-20 April 1986.* – Tokio, 1986. – P. 2015–2018.

67. Moulines E., Charpentier F. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones // *Speech Communication*. – 1990. – Vol. 9. – P. 453–467.

68. Hamon C., Moulines E., Charpentier F. A Diphone Synthesis based on time-domain prosodic modification of speech // *Acoustics, Speech, and Signal Processing: proceedings of IEEE International conference ICASSP-89, Glasgow, UK, 23-26 May 1989*. – Glasgow, 1989. – P. 238–241.

69. Лобанов Б.М. Микроволновой синтез речи по тексту // *Анализ и синтез речи: сб. науч. трудов / научн. ред. Б.М. Лобанов*. – Мн.: Институт технической кибернетики АН БССР, 1991. – С. 57–73.

70. Laroche J., Stylianou Y., Moulines E. HNS: Speech modification based on a harmonic + noise model // *Acoustics, Speech, and Signal Processing: proceedings of IEEE International conference ICASSP-93, Minneapolis, USA, 27-30 April 1993*. – Minneapolis, 1993. – P. 550–553.

71. Dutoit T., Leich H. MBR-PSOLA: Text-to-speech synthesis based on an MBE re-synthesis of the segments database // *Speech Communication*. – 1993. – Vol. 13. – № 3-4. – P. 435–440.

72. Richard G., d'Alessandro C. Time-domain analysis-synthesis of the aperiodic component of speech signals // *Speech synthesis: proceedings of 2-nd ESCA/IEEE Workshop, New York, USA, 12 September 1994*. – New York, 1994. – P. 5–8.

73. Hunt A., Black A. Unit selection in a concatenative speech synthesis system using a large speech database // *Acoustic, Speech and Signal processing: proceedings of IEEE International conference ICASSP-96, Atlanta, USA, 7-10 May 1996*. – Atlanta, 1996. – V. 1. – P. 373–376.

74. Black A., Campbell N. Optimizing selection of units from speech databases for concatenative synthesis // *Speech Communication and Technology: proceedings of the 4-th European conference EUROSPEECH'95, Madrid, Spain, 18-21 September 1995*. – Madrid, 1995. – P. 581–584.

75. Lyudovyk T., Sazhok M. Unit Selection Speech Synthesis Using Phonetic-Prosodic Description of Speech Databases // *Speech and Computer: proceedings of the 9-th International conference SPECOM'2004, Saint-Petersburg, 20–22 Sept. 2004 / Institute of Informatics and Automation of RAS, Speech Informatics Group*. – St.-Petersburg: Anatolia, 2004. – P. 594–599.

76. Coorman G., Fackrell J., Rutten P., Van Coile B. Segment selection in the L&H Realspeak laboratory TTS system // *Spoken Language Processing: proceedings of 6-th International conference ICSLP'2000, Beijing, China, 16-20 October 2000*. – Beijing, 2000. – V. 2. – P. 395–398.

77. Silverman K. et al. TOBI: a standard for labelling english prosody // Spoken Language Processing: proceedings of 2-nd International conference ICSLP'92, Alberta, Canada, 13-16 October 1992. – Alberta, 1992. – P. 867–870.
78. Hart J., Collier R., Cohen A. A Perceptual study of intonation: an experimental phonetic approach to speech melody. – Cambridge: Cambridge University Press, 1990. – 227 p.
79. Fujisaki H. Dynamic characteristics of voice fundamental frequency in speech and singing // The Production of Speech / H. Fujisaki [et al.]; ed. by P.F. MacNeilage. – NY: Springer-Verlag, 1983. – Ch. 3. – P. 39–47.
80. Taylor P. Analysis and synthesis of intonation using the tilt model // J. Acoust. Soc. America. – 2000. – V. 107. – № 3. – P. 1697–1714.
81. Lobanov B. The phonemophon text-to-speech system // International Congress of Phonetic Sciences: proc. of the 11-th seccion ICPHS'87, Tallin, USSR, 6-10 August 1987. – Tallin, 1987. – V.1. – P. 120–124.
82. Ross K., Ostendorf M. A Dynamical system model for generating F_0 for synthesis // Speech synthesis: proceedings of 2-nd ESCA/IEEE Workshop, New York, USA, 12 September 1994. – New York, 1994. – P. 131–134.
83. Black A., Hunt A. Generating F_0 contours from ToBI labels using linear regression // Spoken Language Processing: proceedings of the 4-th International conference ICSLP'96, Philadelphia, USA, October 3-6, 1996. – Philadelphia, 1996. – P. 1385–1388.
84. Hart J. F_0 Stylization in speech: straight lines versus parabolas // J. Acoust. Soc. America. – 1990. – V. 90. – № 6. – P. 3368–3370.
85. de Pijper J.R. Modelling british english intonation. – Dordrecht: Foris Publications, 1983. – 150 p.
86. Ode C. Russian Intonation: a perceptual description. – Amsterdam, Rodopi B.V.Editions. 1989. – 304 p.
87. Nespor M., Vogel I. Prosodic phonology. – Dordrecht: Foris Publications, 1986. – 180 p.
88. Grønnum N. Superposition and subordination in intonation: a non-linear approach // International Congress of Phonetic Sciences: proceedings of the 13-th seccion ICPHS'95, Stockholm, Sweden, August 13-19, 1995. – Stockholm, 1995. – V. 2. – P. 124–131.
89. Van Santen J., Shih Ch., Mobius B. Intonation // Multilingual Text-to-Speech Synthesis / ed. by R. Sproat. – Kluwer Academic Publishers, 1998. – Chapter 6. – P. 141–190.
90. Pavlova E., Pavlov Y., Sproat R., Shih Ch., Van Santen J. Bell laboratories Russian text-to-speech system // Speech Communication and

Technology: proceedings of the 5-th European conference EUROSPEECH'97, Rhodes, Greece, 22-25 September 1997. – Rhodes, 1997. – P. 2451–2454.

91. Lobanov, B., Tsirulnik, L., Sizonov, O. AUP's Modeling of Speaker Specific Intonation Contour Peculiarities // Speech and Computer: proceedings of the 12-th International conference SPECOM'2007, Moscow, Russia, 15-18 October, 2007 / Moscow State Linguistic University. – Moscow, 2007. – V.1. – P. 312-317.

92. Лобанов Б.М., Елисеева О.Е. Речевой интерфейс интеллектуальных систем. Мн.: БГУИР, 2006. - 152 с.

93. Beutnagel M., Conkie A., Syrdal A.K. Diphone synthesis using unit selection // Speech Synthesis: proceedings of the 3-rd ESCA/COCOSDA International Workshop, Jenolan Caves, Australia, 11-14 November 1998. – Jenolan Caves, 1998. – V. 1. – P. 185–190.

94. Lambert T., Breen A. A Database design for a TTS synthesis system using lexical diphones // Speech Communication and Technology: proceedings of the 9-th European conference InterSpeech'2004, Jeju Island, Korea, October 4-8, 2004. – Jeju Island, 2004. – P.1381–1384.

95. Bunnell H., Hoskins S.R., Yarrington D. Prosodic vs. segmental contributions to naturalness in a diphone synthesizer // Speech Synthesis: proceedings of the 3-rd ESCA/COCOSDA International Workshop, Jenolan Caves, Australia, 11-14 November 1998. – Jenolan Caves, 1998. – V. 5. – P. 1723–1726.

96. Skrelin P. Allophone-based concatenative speech synthesis system for russian // Text, Speech and Dialog: proceedings of the 2-nd international workshop TSD'99, Pilsen, Czech Republic, 13-17 September 1999. – Pilsen, 1999. – P. 156–159.

97. Krivnova O.F. Automatic synthesis of russian speech // International Congress of Phonetic Sciences: proceedings of the 14-th section ICPHS'99, San Francisco, USA, 1-7 August 1999. – San Francisco, 1999. – Vol. 1. – P. 507–510.

98. Matoušek J. Building a new czech text-to-speech system using triphone-based speech units // Text, Speech and Dialog: proceedings of the 3-rd international workshop TSD'2000, Brno, Czech Republic, 13-16 September 2000. – Brno, 2000. – P. 223–228.

99. Skrelin P. Allophone- and sub-allophone-based speech synthesis system for Russian // Text, Speech and Dialog: proceedings of the 3-rd international workshop TSD'2000, Brno, Czech Republic, 13-16 September 2000. – Brno, 2000. – P. 271–276.

100. Торсуев Г.П. Константность и вариативность в фонетической системе (на материале английского языка). – М.: Наука, 1977. – 125 с.

101. Tanaka K. A. et al. Japanese Text-to-speech system based on multiform units with consideration of frequency distribution in Japanese // *Speech Communication and Technology: proceedings of the 6-th European conference EUROSPEECH'99*, Budapest, Hungary, 5-9 September 1999. – Budapest, 1999. – P. 839–842.
102. Kopecek L. Syllable segments in Czech // *Baltijskaya filologija: proceedings of International Student conference*, Saint-Petersburg, Russia, 10-12 March 1998. – St.-Petersburg, 1998. – V. 10. – P. 60–64.
103. Rajesh V. et al. On the development of text to speech system for Hindi // *International Congress of Phonetic Sciences: proceedings of the 13-th seccion ICPHS'95*, Stockholm, Sweden, 13-19 August 1995. – Stockholm, 1995. – V. 2. – P. 354–357.
104. Lewis E., Tatham M. Word and syllable concatenation in text-to-speech synthesis // *Speech Communication and Technology: proceedings of the 6-th European conference EUROSPEECH'99*, Budapest, Hungary, 5-9 September 1999. – Budapest, 1999. – P. 615–618.
105. Koyama T., Takahashi J. A V-CV Waveform based speech synthesis using global minimization of pitch conversion and concatenation distortion in V-CV unit selection // *Speech Communication and Technology: proceedings of the 6-th European conference EUROSPEECH'99*, Budapest, Hungary, 5-9 September 1999. – Budapest, 1999. – P. 2311–2314.
106. Conkie A. A robust unit selection system for speech synthesis. – *J. Acoust. Soc. America.* – 1999. – Vol. 105. - № 2. – P. 978–984.
107. Coorman G. et al. Segment selection in the L&H Realspeak laboratory TTS system // *Proc. of ICSLP*, 2000. – V. 2. – P. 395–398.
108. Taylor P., Black A. Speech synthesis by phonological structure matching // *Speech Communication and Technology: proceedings of the 6-th European conference EUROSPEECH'99*, Budapest, Hungary, 5-9 September 1999. – Budapest, 1999. – V. 2. – P. 623–626.
109. Godfrey J., Zampolli A. *Language Resources // Survey of the State of the Art in Human Language Technology.* – Cambridge University Press, 1996. – Chapter 12. – P. 357–384.
110. Zinovieva N. Phonetically sufficient allophonic database for concatenation synthesis of russian speech // *International Congress of Phonetic Sciences: proceedings of the 13-th seccion ICPHS'95*, Stockholm, Sweden, 13-19 August 1995. – Stockholm, 1995. – V. 2. – P. 358–361.
111. Fitt S., Stephen I. Synthesis of regional english using a keyword lexicon // *Speech Communication and Technology: proceedings of the 6-th European conference EUROSPEECH'99*, Budapest, Hungary, 5-9 September 1999. – Budapest, 1999. – V. 2. – P. 823–826.

112. Isard S., Miller D. Diphone synthesis techniques // *Speech Input/Output: techniques and applications: proceedings of IEE International Conference*, London, UK, May 2-8, 1986. – London, 1986. – P. 77–82.

113. Lenzo K. A., Black A. Diphone collection and synthesis // *Spoken Language Processing: proceedings of 6-th International conference ICSLP'2000*, Beijing, China, 16-20 October 2000. – Beijing, 2000. – V. 3. – P. 306–309.

114. Fotinea S.-E., Tambouratzis G., Carayannis G. Constructing a segment database for greek time domain speech synthesis // *Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001*, Aalborg, Denmark, 3-7 September 2001. – Aalborg, 2001. – V. 3. – P. 2075–2078.

115. Lyudovyk T., Sazhok M. Speech databases used for concatenative speech synthesis // *Обробленне сигналів і зображень та розпізнавання образів: праці 7-ї міжнародної конференції УкрОбраз-2004*, Київ, Україна, 11-15 октября 2004. – Київ, 2004. – С. 111–114.

116. Закревский А. Д., Поттосин Ю.В., Черемисинова Л.Д. Основы логического проектирования. Книга 1. Комбинаторные алгоритмы дискретной математики. – Мн., ОИПИ НАН Беларуси. – С 76–82.

117. Cormen T., Leiserson Ch., Rivest R. *Introduction to Algorithms*. – Cambridge, 1990.

118. Hue X. *Genetic algorithms for optimization*. – Edinburgh, Edinburgh Parallel Computing Centre Press, 1997. – 276 p.

119. Matoušek J., Krůta J. Design of speech corpus for text-to-speech synthesis // *Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001*, Aalborg, Denmark, 3-7 September 2001. – Aalborg, 2001. – V. 3. – P. 2047–2050.

120. Кривнова О. Ф., Захаров Л.М., Строкин Г.С. Подбор текстового материала и статистический инструментарий для создания речевых корпусов // *XI сессия Российского акустического общества: сб. трудов*. – М., Геос., 2001. – Т. 3. Акустика речи. Медицинская и биологическая акустика. – С. 65–68.

121. François H., Boëffard O. Design of an optimal continuous speech database for text-to-speech synthesis considered as a set covering problem // *Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001*, Aalborg, Denmark, 3-7 September 2001. – Aalborg, 2001. – P. 829–832.

122. Van Santen J., Buchsbaum A.L. Methods for optimal text selection // *Speech Communication and Technology: proceedings of the 5-th European conference EUROSPEECH'97*, Rhodes, Greece, 22-25 September 1997. – Rhodes, 1997. – P. 553–556.

123. Van Santen J. Combinatorial issues in text-to-speech synthesis // *Speech Communication and Technology: proceedings of the 5-th European conference EUROSPEECH'97, Rhodes, Greece, 22-25 September 1997.* – Rhodes, 1997. – V. 5. – P. 2511–2514.

124. Boëffard O., Emerard F. Application-dependent prosodic models for tts and automatic design of learning database corpus using genetic algorithm // *Speech Communication and Technology: proceedings of the 5-th European conference EUROSPEECH'97, Rhodes, Greece, 22-25 September 1997.* – Rhodes, 1997. – V. 5. – P. 2507–2510.

125. Black A., Taylor P. Automatically clustering similar units for unit selection in speech synthesis // *Speech Communication and Technology: proceedings of the 5-th European conference EUROSPEECH'97, Rhodes, Greece, 22-25 September 1997.* – Rhodes, 1997. – V. 2. – P. 601–604.

126. Kim S. H., Lee Y.J., Hirose K. Pruning of redundant synthesis instances based on weighted vector quantization // *Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001, Aalborg, Denmark, 3-7 September 2001.* – Aalborg, 2001. – V. 3. – P. 2231–2234.

127. Zhao Y. et al. Custom-tailoring TTS voice font-keeping naturalness when reducing database size // *Speech Communication and Technology: proceedings of the 8-th European conference EUROSPEECH'2003, Geneva, Switzerland, 1-4 September 2003.* – Geneva, 2003. – P. 2957–2960.

128. Kumar, R. Automatic pruning of unit selection speech databases for synthesis without loss of naturalness / R. Kumar, S. Kishore // *Speech Communication and Technology: proceedings of the 9-th European conference InterSpeech'2004, Jeju Island, Korea, October 4-8, 2004.* – Jeju Island, 2004. – P. 1377–1380.

129. Tambouratzis G., Fotinea S.-E., Carayannis G. On the systematic construction of high-quality segment databases for Greek TTS systems // *Noise Control: proceedings of the 4-th European Conference Euronoise'2001, Patras, Greece, 8-12 January 2001.* – Patras, 2001. – P. 608–614.

130. Богданов Д. С., Кривнова О.Ф., Подрабинович А.Я., Фарсобина В.В. База речевых фрагментов русского языка «ISABASE». – *Интеллектуальные технологии ввода и вывода информации.* – М., 1998. – С. 20–23.

131. Вольская Н., Коваль А., Коваль С., Опарин И., Погарева Е., Скрелин П., Смирнова Н., Таланов А. Синтезатор русской речи по тексту нового поколения // *Компьютерная лингвистика и интеллектуальные технологии: труды междунар. конф. Диалог'2005, Звенигород, 1–7 июня 2005 г./ Институт проблем информатики РАН; отв. ред.: И.М. Кобозева [и др.].* – М.: Наука, 2005. – С. 84–85.

132. Людовик Т.В. Синтез речи с моделированием особенностей произношения на основе анализа индивидуальных речевых баз данных большого объема [Электронный ресурс]. – 2004. – Режим доступа: <http://www.dialog-21.ru/Archive/2004/Lyudovyk.htm>. – Дата доступа: 29.01.07.

133. Nefti S., Voëffard O. Acoustical and topological experiments for an HMM-based speech segmentation system // *Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001*, Aalborg, Denmark, 3-7 September 2001. – Aalborg, 2001. – V. 3. – P. 1711–1714.

134. Kominek J., Black A.W. A Family-of-models approach to HMM-based segmentation for unit selection speech synthesis // *Speech Communication and Technology: proceedings of the 9-th European conference InterSpeech'2004*, Jeju Island, Korea, 4-8 October 2004. – Jeju Island, 2004. – V. 3. – P. 1146–1149.

135. Mermelstein P. Automatic segmentation of speech into syllabic units // *J. Acoust. Soc. America*. – 1975. – V. 58. – № 4. – P. 880–883.

136. Lewis E., Tatham M. Automatic segmentation of recorded speech into syllables for speech synthesis // *Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001*, Aalborg, Denmark, 3-7 September 2001. – Aalborg, 2001. – V. 3. – P. 1703–1706.

137. Itakura F. Line spectrum representation of linear predictive coefficients of speech signals // *J. Acoust. Soc. America*. – 1975. – V. 57. – № 2. – P. 537–540.

138. Chou F.C. et al. Automatic segmental and prosodic labeling of mandarin speech database // *Spoken Language Processing: proceedings of 5-th International conference ICSLP'98*, Sydney, Australia, 30 November – 4 December 1998. – Sydney, 1998. – V. 4. – P. 1263–1266.

139. Toledano D.T., Hernández Gómez L.A. Local refinement of phonetic boundaries: a general framework and its application using different transition model // *Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001*, Aalborg, Denmark, September 3-7, 2001. – Aalborg, 2001. – V. 3. – P. 1695–1698.

140. Cox S. et al. Techniques for accurate automatic annotation of speech waveform // *Spoken Language Processing: proceedings of 5-th International conference ICSLP'98*, Sydney, Australia, November 30 – December 4, 1998. – Sydney, 1998. – V. 5. – P. 1947–1950.

141. Angelini B., Barolo C., Falavigna D., Omologo M., Sandri S. Automatic diphone extraction for an Italian text-to-speech synthesis system // *Speech Communication and Technology: proceedings of the 5-th European*

conference EUROSPEECH'97, Rhodes, Greece, September 22-25, 1997. – Rhodes, 1997. – V. 2. – P. 581–584.

142. Brugnara F., Falavigna D., Omologo M. Automatic segmentation and labeling of speech based on hidden markov models // *Speech Communication*, 1993. – V. 12. - № 4. – P. 357–370.

143. Talkin D., Wightman C.W. The Aligner: Text-to-speech alignment using markov models and a pronunciation dictionary // *Speech synthesis: proceedings of 2-nd ESCA/IEEE Workshop*, New York, USA, 12 September 1994. – New York, 1994. – P. 82–92.

144. Toledano D.T. et al. Trying to mimic human segmentation of speech using HMM and fuzzy logic post-correction rules // *Speech synthesis: proceedings of 3-rd ESCA/IEEE Workshop*, Sydney, Australia, 26-29 November 1998. – Sydney, 1998. – P. 207–212.

145. Toledano D. T. Neural network boundary refining for automatic speech segmentation // *Acoustic, Speech and Signal processing: proceedings of IEEE International conference ICASSP-2000*, Istanbul, Turkey, 5-9 June, 2000. – Istanbul, 2000. – P. 3438–3441.

146. Лобанов Б.М., Слуцкер Г.С., Тизик А.П. Автоматическое распознавание звукоочетаний в текущем речевом сигнале: сб. науч. трудов НИИР. – Вып. 4. – М., 1969. – С. 67–75.

147. Слуцкер, Г.С. Нелинейный метод анализа речевых сигналов: сб. науч. трудов НИИР. – Вып. 4. – М., 1969. – С. 76–82.

148. Malfrère F., Dutoit T. High quality speech synthesis for phonetic speech segmentation // *Speech Communication and Technology: proceedings of the 5-th European conference EUROSPEECH'97*, Rhodes, Greece, September 22-25, 1997. – Rhodes, 1997. – P. 2631–2634.

149. Слуцкер, Г.С., Хабаров, В.И. Автоматизация предъявления ЭВМ реализаций звуков речи-2: сб. науч. трудов НИИР. – Вып. 1. – М., 1970. – С. 89–94.

150. Лобанов, Б.М., Цирульник, Л.И. Внутрисловные и межсловные правила обработки фонемного текста для разговорного стиля речи // *Функциональные стили звучащей речи: труды междунар. конф.*, Москва, 20–22 сент. 2005 г. / *Филологический факультет МГУ им. М.В. Ломоносова*. – М., 2005. – С. 80–82.

151. Лобанов, Б.М., Цирульник, Л.И. Моделирование внутрисловных и межсловных фонетико-акустических явлений полного и разговорного стилей речи в системе синтеза речи по тексту «Мультифон» // *Анализ разговорной русской речи: труды первого междисциплинарного семинара АР³*, Санкт-Петербург, 29 августа 2007 г. – СПб.: ГУАП, 2007. – С. 57-71.

152. Лобанов, Б.М., Цирульник, Л.И. Правила разметки речевого корпуса на фонетические сегменты и стратегия выбора элементов компиляции при синтезе речи // Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции Диалог'2007, Бекасово, 30 мая – 3 июня 2007 г. – М.: Издательский центр РГГУ, 2007. – С. 393-399.

153. Lobanov, B., Tsirulnik, L., Karnevskaya, E. Speech Corpus Phonetic Segmentation for TTS Synthesis // Speech and Computer: proceedings of the 12-th International conference SPECOM'2007, Moscow, Russia, 15-18 October, 2007 / Moscow State Linguistic University. – Moscow, 2007. – V.2. – P. 615-618.

154. Лобанов, Б.М., Пьорковска, Б.Э., Рафалко, Я.Ш., Цирульник, Л.И. Реализация межъязыковых различий интонации завершенности и незавершенности в синтезаторе русской и польской речи по тексту // Компьютерная лингвистика и интеллектуальные технологии: труды междунар. конф. Диалог'2005, Звенигород, 1–7 июня 2005 г./ Институт проблем информатики РАН; отв. ред.: И.М. Кобозева [и др.]. – М.: Наука, 2005. – С. 356–362.

155. Лобанов, Б.М., Цирульник, Л.И. Персональные особенности синтагматического членения речи телеведущего Ю.Сенкевича // Компьютерная лингвистика и интеллектуальные технологии: труды междунар. конф. Диалог'2004, «Верхневолжский», 2–7 июня 2004 г./ Институт проблем информатики РАН; отв. ред.: И.М. Кобозева [и др.]. – М.: Наука, 2004. – С. 419–423.

156. Lobanov, B., Tsirulnik, L. Statistical study of speaker's peculiarities of utterances into phrases segmentation // Speech Prosody: proceedings of the 3-rd International conference, Dresden, Germany, May 2–5, 2006. – Dresden, 2006. – V. 2. – P. 557–560.

157. Лобанов, Б.М., Цирульник, Л.И. Алгоритм персонализированного синтагматического членения текста для TTS-систем клонирования речи // Теория и практика речевой коммуникации: тезисы докладов междунар. конф. к 250-летию основания МГУ им. М.В. Ломоносова и 40-летию образования лаборатории фонетики и речевой коммуникации филологического факультета, Москва, 7–9 сент. 2004 г./ Филологический факультет МГУ им. М.В. Ломоносова. – М.: Макс-пресс, 2004. – С. 79–83.

158. Lobanov, B.M., Tsirulnik, L.I., Fiodorov, A.M. Study of idiosyncrasy of syntagmatic segmentation for personal speaking manner cloning by TTS // Speech analysis, synthesis and recognition: Application of Phonetics: proceedings of the International conference. Krakow, July 9-13, 2005. – Krakow, 2005. – P. 69–72.

159. Lobanov, B., Tsirolnik, L., Zhadinets, D., Karnevskaya, E. Language- and speaker specific implementation of intonation contours in multilingual TTS synthesis // *Speech Prosody: proceedings of the 3-rd International conference, Dresden, Germany, May 2–5, 2006.* – Dresden, 2006. – V. 2. – P. 553-556.

160. Златоустова, Л.В. и др. Алгоритмы преобразования русских орфографических текстов в фонемную запись. – М.: МГУ, 1970. - 120 с.

161. Лобанов, Б.М., Панченко, Б.В. Преобразователь графема-фонема для синтеза речи по орфографическому тексту // *Автоматическое распознавание и синтез речи: тезисы докл. Всесоюзн. научной школы-семинара АРСО-8, Львов, 10-17 августа 1974г.* – Львов, 1974. – Ч. 4. – С. 15–18.

162. Зиновьева, Н.В., Кривнова, О.Ф. Прикладные системы с использованием фонетических знаний // *Проблемы фонетики: сб. статей.* – М.: Прометей, 1993. – С. 288–300.

163. Захаров, Л.М. Акустическая вариативность звуковых единиц в русской речи // *Язык и речь: проблемы и решения: сб. науч. трудов к юбилею профессора Л.В. Златоустовой / под ред. Г.Е. Кедровой и В.В. Потапова.* – М.: Макс-Пресс, 2004. – С. 240–269.

164. Русская разговорная речь / Академия Наук СССР. Институт русского языка; под ред. Е.А. Земской. – М.: Наука, 1973. – 485 с.

165. Voersma, P. Praat: doing phonetics by computer / P. Voersma, D. Weenink // [Electronic resource]. – Mode of access: <http://www.praat.org/>.

166. Щерба Л.В. Избранные работы по языкознанию и фонетике. Л., 1958.

167. Бондарко, Л.В. Слоговая структура речи и дифференциальные признаки фонем (экспериментально-фонетическое исследование на материале русского языка): автореф. дис. ...докт. филол. наук: 660; Ленинградский Государственный университет. – Л., 1969. – 32 с.

168. Цирульник, Л.И. Алгоритмы синтеза и технология компьютерного клонирования персональных характеристик речи: дисс. на соискание учёной степени кандидата технических наук: 05.13.01. – Мн, 2007. – С. 172-182.

169. Lobanov, B., Tsirolnik, L., Zhadinets, D., Piorkovska, B., Rafalko, J., Szpilevsky, E. Language-specific application of intonation contours in russian and polish multilingual TTS Synthesis // *Speech and Computer: proceedings of the 10-th International conference SPECOM'2005, Patras, Greece, 20–22 Sept. 2005 / Saint-Petersburg Institute of Informatics and Automation of RAS, Speech Informatics Group.* – St.-Petersburg: Anatolia, 2005. – P. 565–570.

170. Lobanov, B., Kiselov, V., Tsirolnik, L., Shpilewski, E., Piurkowska, B., Rafalko J. Polish TTS in multi-voice slavonic languages speech synthesis system // Speech and Computer: proceedings of the 9-th International conference SPECOM'2004, Saint-Petersburg, 20–22 Sept. 2004 / Institute of Informatics and Automation of RAS, Speech Informatics Group. – St.-Petersburg: Anatolia, 2004. – P. 565–570.

171. Брызгунова, Е.А. Практическая фонетика и интонация русского языка – М., 1963. – 307 с.

172. Кодзасов, С.В. Размещение тональных акцентов в русском слове // Язык и речь: проблемы и решения: сб. науч. трудов к юбилею профессора Л.В. Златоустовой / под ред. Г.Е. Кедровой и В.В. Потапова. – М.: Макс-Пресс, 2004. – С. 62–76.

173. Передача речи по трактам радиотелефонной связи. Требования к разборчивости речи и методы артикуляционных измерений: ГОСТ 16600-72. – Введ. 27.09.1972. – Москва: Государственный комитет стандартов Совета Министров СССР, 1973. – 90 с.

174. Зализняк А.А. Грамматический словарь русского языка. Словоизменение. – М.: Русский язык. – 1987. – 880 с.

175. RealSpeak. Expressive, natural, multi-lingual text-to-speech [Electronic resource]. – Mode of access: <http://www.nuance.com/realspeak/demo/>.

176. Передача речи по трактам связи. методы оценки качества, разборчивости и узнаваемости: СТБ ГОСТ Р 50840-2000. – Введ. 01.04.2001. – Минск: Белстандарт, 2000. – 372 с.

177. Kiparsky, P. Phonological change. – Linguistic theory foundations. – Cambridge etc., 1988. – P.363–415.

178. Beutnagel, M. At al The AT&T next-gen TTS system // ASA, EAA, and DAGA: proceedings of joint meeting, Berlin, Germany, 15-19 March 1999. – Berlin, 1999. – P. 18–24.

179. Лобанов, Б.М., Давыдов, А.Г., Киселёв, В.В., Цирульник, Л.И. Система сегментации речевого сигнала методом анализа через синтез // Известия Белорусской инженерной академии. – 2004. – №1/1. – С.112–115.

180. Лобанов, Б.М., Цирульник, Л.И., Ли, И.В. Обработка речевых корпусов для распознавания и синтеза русской речи в системах управления // Обработка информации и управление в чрезвычайных и экстремальных ситуациях: доклады пятой междунар. конф., Минск, 24 – 26 октября 2006г. / Объединённый институт проблем информатики НАН Беларуси; научн. ред. А.В. Тузиков. – Минск, 2006. – С. 73–77.

181. Ронжин, А.Л., Карпов, А.А., Лобанов, Б.М., Цирульник, Л.И., Йокиш, О. Фонетико-морфологическая разметка речевых корпусов для

распознавания и синтеза русской речи // Информационно-управляющие системы. – 2006. – Вып. 25. – Т. 6. – С. 24–34.

182. Турбович, И.Т., Гитис, В.Г., Маслов, В.К. Опознавание образов. – М.: Наука, 1971. – С. 140–141.

183. Аванесов, Р.И. Русское литературное произношение / Р.И. Аванесов. – 5-е изд. – М.: Просвещение, 1972. – 250 с.

184. Thorpe, L. Performance of current perceptual objective speech quality measures / L. Thorpe, W. Yang // Speech coding: proceeding of IEEE Workshop, Porvoo, Finland, 20-23 June 1999. – Porvoo, 1999. – P. 144–146.

185. Chen, J.-D. Objective distance measures for assessing concatenative speech synthesis / J.-D. Chen, N. Campbell // Speech Communication and Technology: proceedings of the 6-th European conference EUROSPEECH'99, Budapest, Hungary, September 5-9 1999. – Budapest, 1999. – V. 2. – P. 611–614.

186. Chu, M. An objective measure for estimating MOS of synthesized speech / M. Chu, H. Peng // Speech Communication and Technology: proceedings of the 7-th European conference EUROSPEECH'2001, Aalborg, Denmark, September 3-7 2001. – Aalborg, 2001. – P. 2087–2090.

187. Wouters, J. A Perceptual evaluation of Distance Measures For Concatenative Speech Synthesis / J. Wouters, M. Magon // Spoken language processing: proceedings of 5-th international conference ICSLP'98, Sydney, Australia, 30 November – 4 December 1998. – Sydney, 1998. – P. 2747–2750.

188. Telephone transmission quality subjective opinion tests. A method for subjective performance assessment of the quality of speech voice output devices: ITU-T Recommendation P.85. – Enf. 21.06.1994. – Geneva: ITU-T Study Group 12, 2004. – 13 p.

189. Кендал, М. Дж. Статистические выводы и связи / М. Дж. Кендал, А. Стюарт. – М.: Наука, 1973. – 892 с.

Сведения об авторах

Лобанов Борис Мефодьевич

доктор технических наук, заведующий лабораторией распознавания и синтеза речи Объединённого института проблем информатики Национальной академии наук Беларуси, автор более 230 научных работ по проблемам компьютерного синтеза и распознаванию речи, в том числе 2-х монографий и 17-ти изобретений.

Область научных интересов – методы автоматического анализа, синтеза и распознавания речевых сигналов, человеко-машинные системы речевого общения, речевые компьютерные технологии.

Цирульник Лилия Исааковна

кандидат технических наук, старший научный сотрудник лаборатории распознавания и синтеза речи Объединённого института проблем информатики Национальной академии наук Беларуси, автор более 30 научных работ по проблемам компьютерного синтеза и клонирования речи.

Область научных интересов – методы автоматического анализа и синтеза речевых сигналов, человеко-машинные системы речевого общения, речевые компьютерные технологии.

Аннотация

Предлагаемая читателю монография является обобщением научных результатов, полученных авторами за последние годы. Она посвящена рассмотрению теоретических и экспериментальных основ, а также описанию конкретных практических результатов в решении задач компьютерного синтеза и клонирования персональных характеристик речи. Монография предназначена широкому кругу специалистов, интересующихся современными проблемами информатики, искусственного интеллекта и компьютерной лингвистики, а также для аспирантов и студентов соответствующих специальностей.