

**Министерство образования и науки Российской Федерации**

**УНИВЕРСИТЕТ ИТМО**

**Ю.Н. Матвеев, С.В. Рыбин**

**Методические указания по выполнению  
лабораторных работ по курсу «Синтез речи»**

**РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО  
по направлению подготовки «информационные системы и  
технологии»  
в качестве учебно-методического пособия для реализации  
основных образовательных программ высшего образования  
магистратуры**

 **УНИВЕРСИТЕТ ИТМО**

**Санкт-Петербург**

**2017**

**Матвеев Ю.Н., Рыбин С.В. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ ПО КУРСУ «СИНТЕЗ РЕЧИ».**

Учебно-методическое пособие. – СПб: Университет ИТМО, 2017. – 45с.

Методические указания предназначены для магистрантов направления 09.04.02, изучающих дисциплину «Синтез речи». Лабораторные работы, представленные в методических указаниях, составляют практикум по курсу "Синтез речи" и помогут студентам подробно ознакомиться с технологией синтеза интонационной речи.

В лабораторных работах используется программный комплекс Vital Voice Tools, разработанный на базовом предприятии кафедры «Центр речевых технологий». В первом разделе пособия дано краткое описание функциональных возможностей данного комплекса. Каждая лабораторная работа начинается с теоретического введения, которое содержит основные сведения, необходимые для выполнения лабораторных работ, и составление отчетов по ним.

**Рецензент: д.ф.-м.н. Коточигов А.М.**

Рекомендовано к печати Ученым советом факультета Информационных технологий и программирования 21.10.2017 г., протокол № 10



**Университет ИТМО** – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2017

© Ю.Н. Матвеев, С.В. Рыбин, 2017

## 1. Программные комплекс Vital Voice Tools

В лабораторных работах используется программный комплекс **Vital Voice Tools** (далее VVT), разработанный на базовом предприятии «Центр речевых технологий». Приложение VVT – это графический интерфейс, предназначенный для использования и настройки процессора Vital Voice Text-To-Speech (далее TTS). Программа позволяет синтезировать произвольный текст, сформировать и редактировать пользовательский словарь, настроить произношение, задать IPA-транскрипцию для заданных фраз и слов и пр.

Приложение VVT состоит из четырех основных частей (модулей):

- Модуля синтезатора, позволяющего воспроизводить произвольный текстовый файл путем произнесения его содержимого при помощи установленного процессора Vital Voice TTS – движка синтеза.
- Модуля редактирования пользовательского словаря, позволяющего вносить изменения в словарь пользователя и добавлять в него незнакомые слова, встретившиеся в произносимом тексте.
- Модуля настройки пользовательских произношений, позволяющего пользователю настраивать произношения отдельных фраз путем модификации транскрипции и управления процедурой подбора звуковых единиц.
- Модуля редактирования словаря IPA-транскрипций, позволяющего для любого слова задать его транскрипцию, записав ее в виде последовательности символов международного фонетического алфавита IPA.

Каждый модуль представлен в виде отдельной вкладки на основном окне приложения.

### 2.1. Запуск приложения

Приложение VVT запускается без параметров. При запуске происходит загрузка модуля «Синтез» на первой вкладке основного окна приложения (рис.1.1). Загрузка остальных модулей происходит при первом переключении на вкладку соответствующего модуля.

Состояние интерфейса приложения (положение и размер окна, положение разделительных полос, ширина столбцов таблиц, выбранный голос и т. п.) сохраняется в ini-файле в папке

*ApplicationData\Speech Technology Center\VVTools*

где

*ApplicationData = C:\Documents and Settings\(user\_name)\Application Data*

Программный комплекс поддерживает управление процессом синтеза речи с помощью стандартного для движков TTS языка разметки: Speech Synthesis Markup Language (далее SSML).

SSML представляет собой основанный на XML язык разметки для приложений синтеза речи. SSML основан на языке разметки Java Synthesis Markup Language (JSML), разработанном компанией Sun Microsystems, хотя нынешние рекомендации были разработаны главным образом производителями синтезаторов речи.

Он охватывает практически все аспекты синтеза речи, хотя в некоторых областях остались неопределенные аспекты, поэтому каждый производитель принимает иной вариант языка. Кроме того, в отсутствие разметки, синтезатор, как ожидается, должен выполнить свое собственное толкование этого текста. Так SSML не является таким жестким в плане синтаксиса, как язык C, или даже HTML.

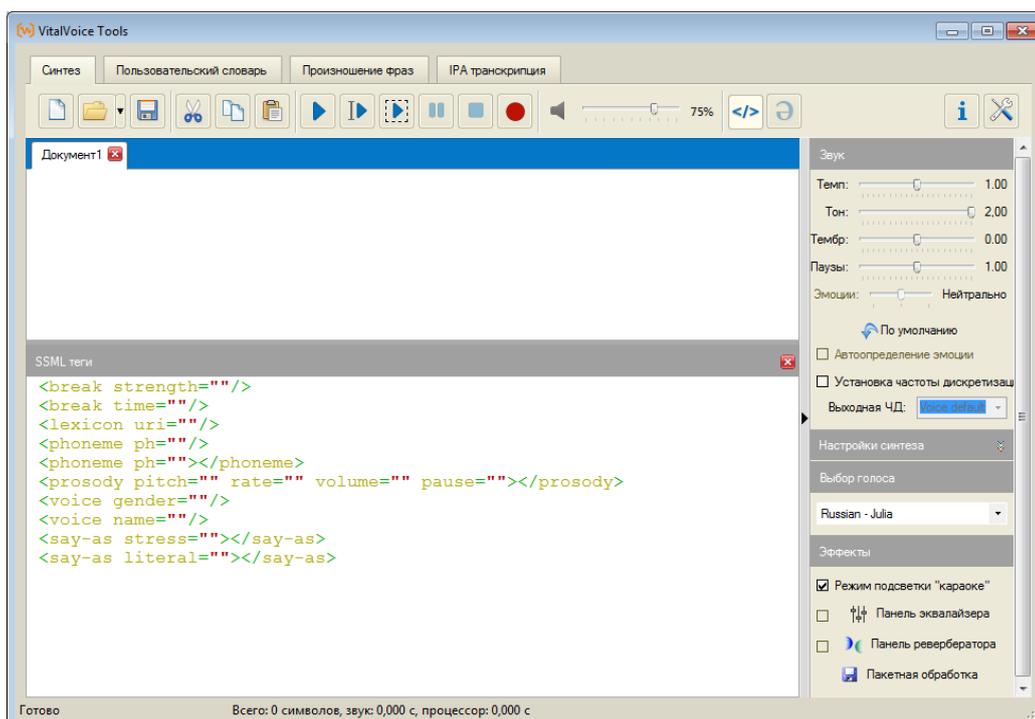


Рис. 1.1. Основное окно приложения Vital Voice Tools.

## 2.2. Модуль синтеза речи

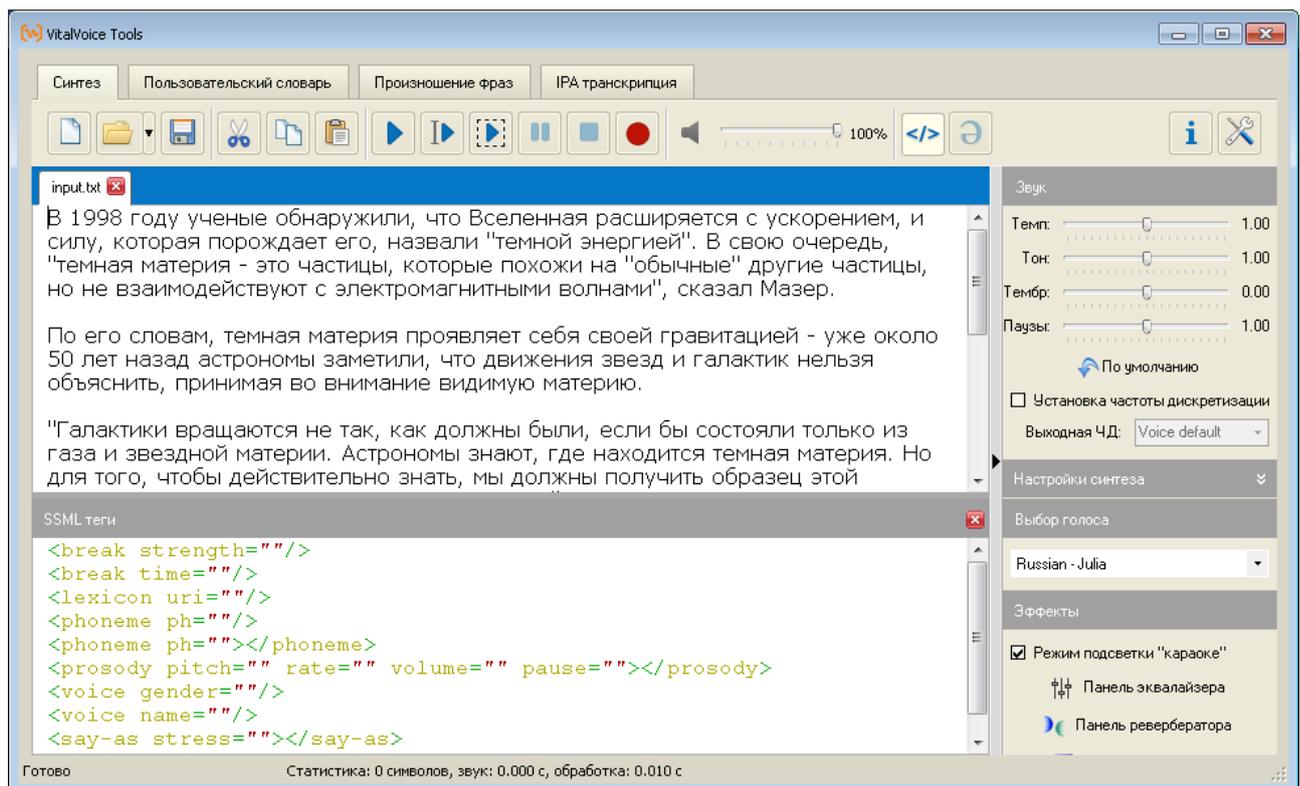
Модуль представляет собой графический интерфейс для воспроизведения текстов. Синтез позволяет воспроизводить произвольный текстовый файл путем произнесения его содержимого при помощи установленного движка синтеза.

Функциональность Синтеза:

- Открытие, отображение и редактирование текста в формате ТХТ
- Сохранение текста в формате ТХТ с возможностью выбора имени и пути сохранения файла (по умолчанию при сохранении указывается последний путь сохранения файла)
- Воспроизведение текста с возможностью предварительного выбора голоса

- Возможность установки громкости воспроизведения синтеза в диапазоне от 0% до 100%
- Изменение темпа звучания от 0.5 до 4.0
- Изменение тона звучания от 0.5 до 2.0
- Изменение тембра звучания от -1 до +1
- Изменение длительности пауз от 0.1 до 5.0
- Изменение частоты дискретизации сигнала от 8000 до 44100 Гц
- Поддержка тегов SSML
- Поддержка алфавита IPA
- Управление синтезом: воспроизведение, воспроизведение от курсора, воспроизведение выделенного фрагмента, остановка воспроизведения, пауза.
- Сохранение синтезированного звука в файл без воспроизведения с выбором имени файла, пути сохранения, формата (MP3 32-160kbps или WAV).

Графический интерфейс модуля синтеза речи представлен на *рис. 1.2*.



*Рис. 1.2. Графический интерфейс модуля синтеза речи.*

Графический интерфейс модуля предоставляет собой:

- окно отображения текста,
- панель инструментов с кнопками,
- нижнюю панель, в которой отображается окно ввода SSML тегов и виртуальная клавиатура IPA,

- боковую панель, содержащую элементы управления звучанием.

### **1.2.1. Окно отображения текста**

Окно предназначено для отображения текста, вводимого пользователем или загружаемого из файла. Непосредственно над окном отображения текста находится панель документов, каждая вкладка которой соответствует открытому или созданному пользователем документу. Окно может одновременно отображать только один текстовый документ, при этом выбор отображаемого документа производится путем выбора соответствующей вкладки на панели документов.

Текст, отображаемый в окне, может быть воспроизведен путем нажатия кнопок на панели инструментов.

### **1.1.2. Панель инструментов**

Панель инструментов содержит кнопки управления документами (создать документ, открыть документ, сохранить документ), кнопки редактирования текста (вырезать, копировать, вставить), кнопки управления воспроизведением текста (произнести весь текст, произнести начиная от курсора, произнести выделенный фрагмент, пауза, остановить), регулятор громкости воспроизведения, кнопки управления нижней панелью и кнопку вызова диалогового окна конфигурации модуля.

Кнопки управления документами и редактирования текста работают обычным образом и дополнительных комментариев не требуют. Следует отметить особенность работы кнопки сохранения документа: всякий раз при сохранении документа требуется указать в появляющемся диалоге имя файла – кнопка работает по принципу «сохранить как». Нажатие на «стрелку вниз», находящуюся справа от кнопки открытия документа, открывает список из 10 последних открытых или сохраненных документов; щелчок по названию документа из этого списка открывает выбранный документ.

При нажатии кнопки произнесения текста (воспроизведение, Play) синтезатор речи начинает произносить текст, отображаемый в окне. При нажатии кнопки «произнести от курсора» или «произнести выделенный текст» происходит воспроизведение соответствующей части текста.

Кнопка «Стоп» прерывает воспроизведение текста, при последующем нажатии кнопки «Play» текст будет воспроизводиться от начала.

Кнопка «Пауза» приостанавливает воспроизведение текста. Повторное нажатие кнопки «Пауза» возобновляет воспроизведение текста с того места, на котором оно было остановлено.

Кроме воспроизведения текста из открытого документа приложение позволяет воспроизводить текст из буфера обмена Windows. Для этого следует скопировать текст в буфер обмена и нажать горячую клавишу Ctrl-F11. Будет создан виртуальный документ «\*/clipboard», в него будет скопирован текст из буфера обмена, после чего будет воспроизведен текст из этого документа, причем

текущий процесс воспроизведения, если таковой продолжался, будет предварительно остановлен.

В процессе воспроизведения текста в основном окне выделяется слово, произносимое в данный момент (режим «караоке»). Этот режим отключается соответствующей галочкой на боковой панели.

Регулятор громкости производит установку громкости воспроизведения произносимого текста. Громкость регулируется независимо от системного микшера. Если при установке регулятора громкости на максимум громкость воспроизведения недостаточна, следует проверить настройки системного микшера.

Кнопки управления нижней панелью («Добавление SSML тегов» и «Виртуальная клавиатура IPA») представляют собой кнопки с фиксацией («залипающие»), при нажатии которых отображаются соответственно панель ввода SSML тегов и виртуальная клавиатура IPA.

Кнопка «Настройки» вызывает диалоговое окно конфигурации модуля, содержащего некоторые дополнительные настраиваемые параметры.

### 1.2.3. Нижняя панель

В нижней панели могут отображаться либо панель ввода SSML тегов, либо виртуальная клавиатура IPA, в зависимости от положения текстового курсора в окне отображения текста.

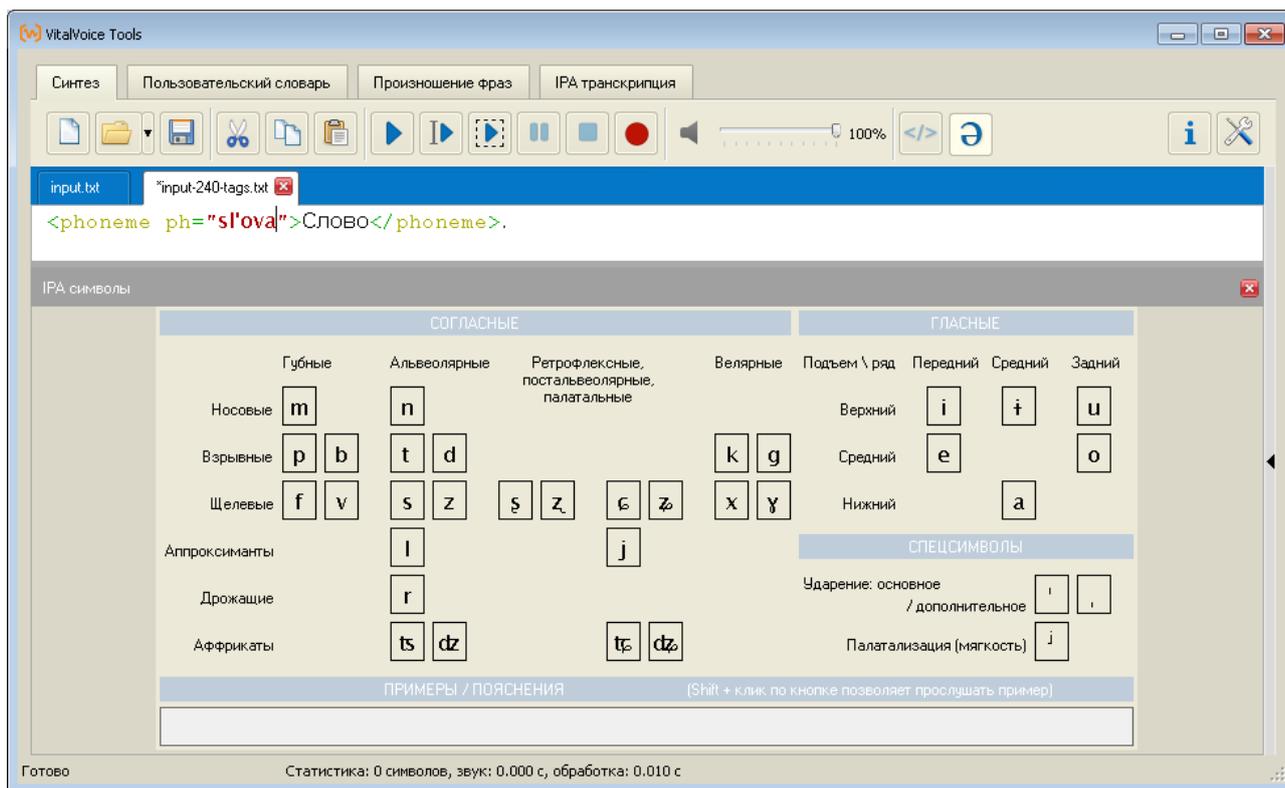
При нахождении курсора где-либо в тексте вне SSML тегов становится доступна панель ввода SSML тегов, при этом кнопка «Добавление SSML тегов» активна и управляет отображением этой панели. При нажатии этой кнопки панель показывается, при отпуске – скрывается.

По двойному щелчку на какой-либо строке в панели ввода SSML тегов соответствующий SSML тег добавляется в окно ввода текста в позицию курсора. При этом если в окне был выделен текст, то он либо замещается, либо обрамляется вводимым тегом. Замещение происходит в том случае, если тег закрытый, например `<break time="" />`, в случае открытого тега выделенный текст вставляется между открывающим и закрывающим тегами, например `<phoneme ph="">` выделенный текст `</phoneme>`.

При нахождении текстового курсора в основном окне внутри уже введенного SSML тега панель ввода SSML тегов скрывается, так как новый тег не может быть введен внутрь уже имеющегося тега.

При нахождении текстового курсора в основном окне внутри атрибута `ph` тега `phoneme` (`<phoneme ph="">`) становится доступна виртуальная клавиатура IPA (рис. 1.3).

При этом кнопка «Виртуальная клавиатура IPA» активна и управляет отображением этой клавиатуры. При нажатии этой кнопки клавиатура показывается, при отпуске – скрывается.



*Рис. 1.3. Виртуальная клавиатура IPA.*

Если нижняя панель (панель ввода SSMML тегов или виртуальная клавиатура IPA) была открыта пользователем, то при перемещении текстового курсора внутри текста управление отображением панели происходит автоматически.

#### 1.2.4. Боковая панель

Боковая панель содержит регуляторы темпа и тона произнесения текста, тембра голоса диктора, длительности пауз, элементы управления преобразованием частоты дискретизации выходного звукового сигнала, раскрываемую панель настроек синтеза, выпадающий список выбора голоса синтеза, выключатель режима «караоке», а также кнопки вызова панелей эквалайзера и ревербератора.

Регулятор темпа устанавливает быстроту произнесения текста относительно естественной для выбранного диктора (голоса). Естественная быстрота произнесения соответствует значению темпа 1.00. Доступно регулирование в пределах от 0.5 (медленно) до 4.0 (очень быстро).

Регулятор тона устанавливает частоту основного тона произносимого текста относительно естественной для выбранного диктора (голоса). Естественная частота основного тона соответствует значению тона 1.00. Доступно регулирование в пределах от 0.5 (вдвое ниже) до 2.0 (вдвое выше).

Регулятор тембра изменяет тембр голоса для выбранного диктора. Естественный тембр соответствует значению 0.0. Доступно регулирование в пределах от -1.0 («монстр») до +1.0 («Буратино»).

Регулятор длительности пауз устанавливает длительность пауз при произнесении текста относительно естественной для выбранного диктора

(голоса). Этот параметр также влияет на быстроту произнесения текста. Естественная длительность пауз соответствует значению 1.00. Доступно регулирование в пределах от 0.1 (очень короткие) до 5.0 (очень длинные).

Кнопка «По умолчанию» восстанавливает значения по умолчанию для темпа, тембра, основного тона и длительности пауз.

Левый щелчок по названию любого из регуляторов на боковой панели восстанавливает для данного регулятора значение по умолчанию.

Галочка «ресэмплинг» включает преобразование частоты дискретизации выходного сигнала к желаемому значению. Значение выбирается в выпадающем списке и может быть задано любым в диапазоне от 8 до 44.1 кГц, либо может быть установлено в значение по умолчанию для данного голоса, как если бы галочка «ресэмплинг» была выключена.

Раскрываемая панель настроек синтеза позволяет управлять параметрами расстановки пауз и обработки переводов строки. Функциональность элементов управления понятна из их названия. Следует отметить, что границы параграфов влияют на длительность пауз: пауза между параграфами более длительна, чем пауза между предложениями.

Выпадающий список выбора голоса синтеза позволяет выбрать голос (диктора), которым будет синтезироваться произносимый текст. Количество голосов в списке зависит от числа установленных в системе голосов для синтеза.

Галочка «режим подсветки «караоке»» включает выделение произносимого в данный момент слова в окне отображения текста.

Кнопки «панель эквалайзера» и «панель ревербератора» вызывают соответственно панель управления параметрами эквалайзера и панель управления параметрами реверберации.

Боковая панель может быть скрыта нажатием кнопки, расположенной непосредственно слева от этой панели.

### 1.2.5. Диалог конфигурации модуля

Окно диалога конфигурации модуля (рис. 1.4) вызывается путем нажатия кнопки «Настройка» на панели инструментов.

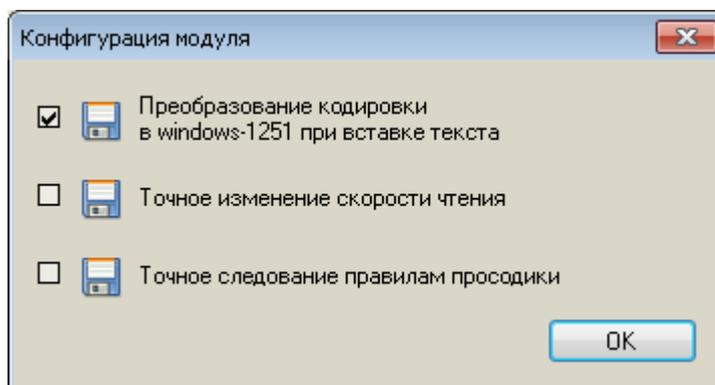


Рис. 1.4. Диалог конфигурации модуля.

Данное окно конфигурации предназначено для изменения параметров конфигурации модуля. Возможно изменение следующих параметров.

Галочка «Преобразование кодировки в windows-1251 при вставке текста». При включенной галочке перекодировка выполняется, при отключенной – нет. Перекодировку следует включать для русского языка при копировании текста из текстовых редакторов типа MS Word во избежание появления в тексте невидимых символов, таких как символ переноса, которые могут негативно влиять на синтез речи, вызывая нежелательные паузы. Отключение перекодировки имеет смысл для сохранения в копируемом тексте диакритик и прочих символов Unicode, например, для английского языка.

Галочка «Точное изменение скорости чтения». При включении галочки темп речи изменяется посредством модуля модификации звука; в этом режиме производится точная модификация, но возможно появление роботизированного звучания. При отключении галочки темп речи изменяется посредством изменения параметров интонационной модели; в этом режиме естественность звучания сохраняется.

Галочка «Точное следование правилам просодики». При включении задействуется особый режим полной модификации звука по всем параметрам с точной привязкой к интонационному контуру.

### 1.2.6. Трехполосный эквалайзер

В окне эквалайзера (рис. 1.5) пользователь может управлять параметрами эквалайзера: усилением по частотам и границами частотных полос.



Рис. 1.5. Трехполосный эквалайзер.

Эквалайзер можно совсем отключить, сняв галочку «Включить» на панели управления. Соответствующей кнопкой можно установить значения по умолчанию: усиление = 0, границы частотных полос = 300 и 3000 Гц.

### 1.2.7. Параметры реверберации

В окне управления параметрами реверберации (рис. 1.6) пользователь может управлять параметрами реверберации, либо выбирая одну из предлагаемых настроек, либо вручную регулируя задержку сигнала «эхо».

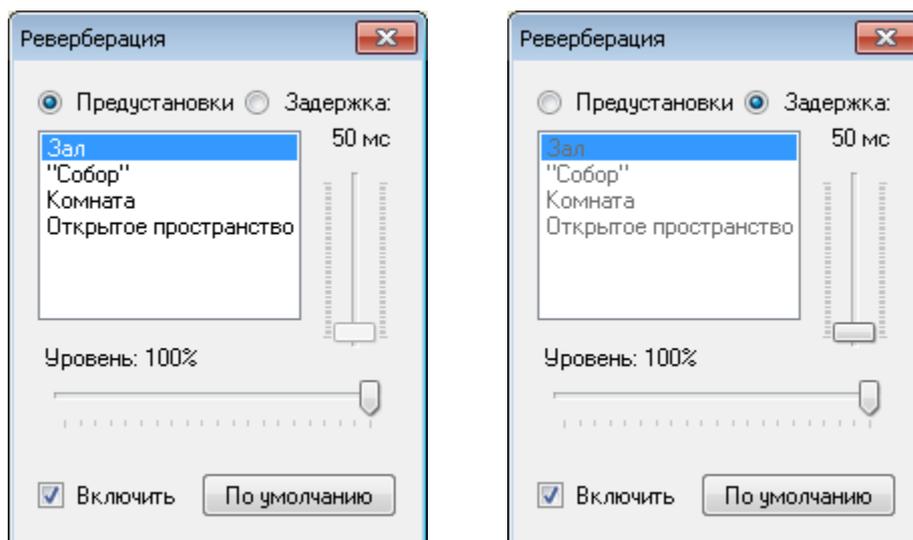


Рис. 1.6. Параметры реверберации.

В обоих случаях пользователь может управлять уровнем отраженного сигнала, используя соответствующий регулятор.

Реверберацию можно совсем отключить, сняв галочку «Включить» на панели управления. Соответствующей кнопкой можно установить значения по умолчанию: использование первой предустановки из списка с уровнем 100%.

## 2. Лабораторная работа 1.

Целью лабораторной работы является исследование влияния пользовательского словаря в системе синтеза интонационной речи.

### 2.1. Особенности работы приложения VVTools

- Для получения протоколов работы приложения VVTools нужно запустить с ключом /x
- Протоколы работы модулей TTS и состояние интерфейса сохраняется в папке:  
`...\Users\...\AppData\Roaming\Speech Technology Center\VVTools`
- Формат имени протокола:  
`result_№.Имя модуля.Result.xml`

- Сам пользовательский словарь находится по маршруту, прописанному в реестре:

[HKEY\_LOCAL\_MACHINE\SOFTWARE\Wow6432Node\Speech Technology Center\TTS\Languages]

"UserDataPath"="d:\\TTS\\Development\\trunk\\UserData\\"

## 2.2. Функции пользовательского словаря

Редактор пользовательского словаря используется для решения следующих задач:

- ✓ Добавление в пользовательский словарь слов с указанием ударения, необходимого пользователю;
- ✓ Добавление в пользовательский словарь замен для различных последовательностей слов или символов (например, задание прочтения английских слов; расшифровка сокращений и аббревиатур; чтение тех или иных специальных символов);
- ✓ Установка знака пользовательского ударения, который затем может быть использован в тексте, подаваемом на синтез.

## 2.3. Общий интерфейс пользовательского словаря.

Основное окно пользовательского словаря имеет две вкладки. На первой вкладке «Словарь» (загружаемой по умолчанию) показано текущее содержание словаря: словарные статьи и замены (рис.2.1).

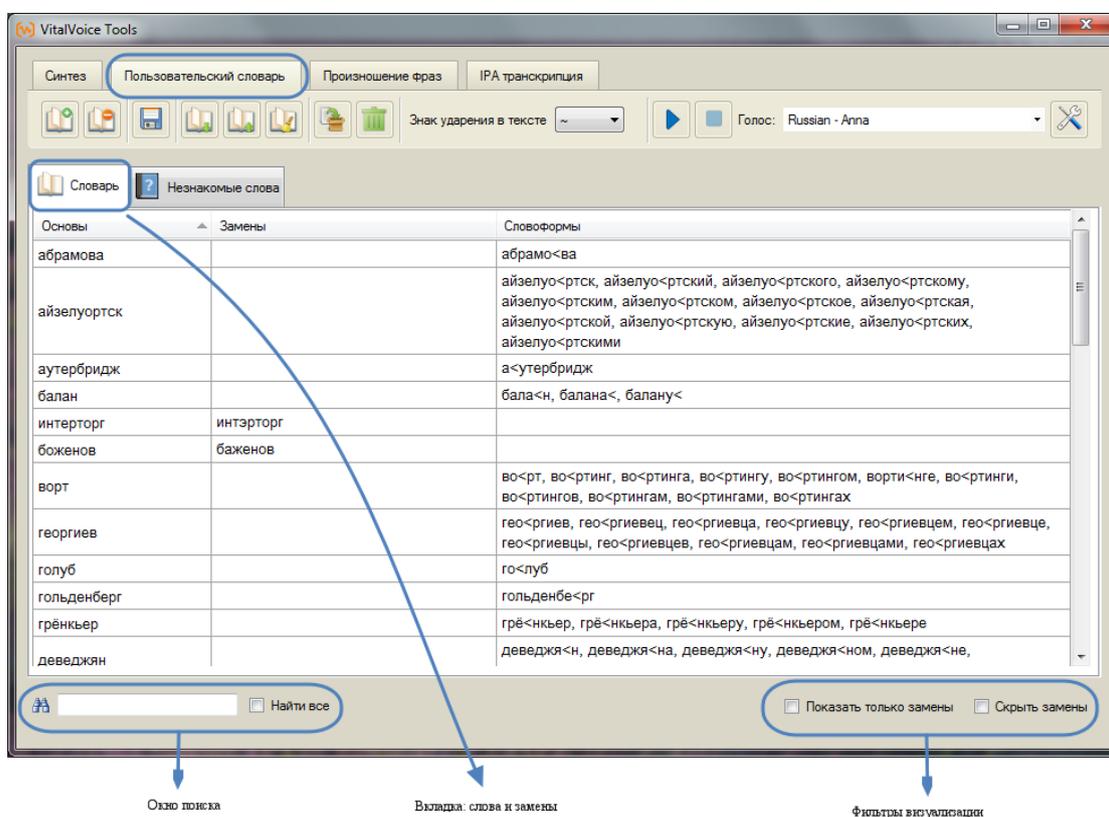


Рис.2.1. Вкладка «Словарь» основного окна пользовательского словаря.

Для слов используются первый и третий столбцы таблицы. В первом столбце показана основа, в третьем столбце - через запятую все словоформы данной словарной статьи. Знаком «<» обозначено основное ударение, которое может находиться только после гласной. Например,

|     |  |   |
|-----|--|---|
| Хоп |  | хопѐ<р, хопра<, хопру<, хопро<м, хопре< |
|-----|--|---|

Для замен используются первый и второй столбцы. В первом столбце – исходный текст, в последнем – его расшифровка. Например,

|        |         |  |
|--------|---------|--|
| invade | инвэ<йд |  |
|--------|---------|--|

В левом нижнем углу расположено окно поиска слов. При пустом окне поиск производится по всему содержимому.

Далее, расположены переключатели, регулирующие формат показа слов и замен. Оба переключателя могут быть одновременно выключены или включены поочередно, но одновременно не могут находиться в положении «включено».

Предусмотрена сортировка содержимого таблицы по любому из трех столбцов. По умолчанию, при загрузке производится сортировка по основам.

Удалить слова и замены можно выделив нужные элементы в таблице и нажав кнопку «Delete». Удаление элементов происходит по запросу (рис.2.2.):

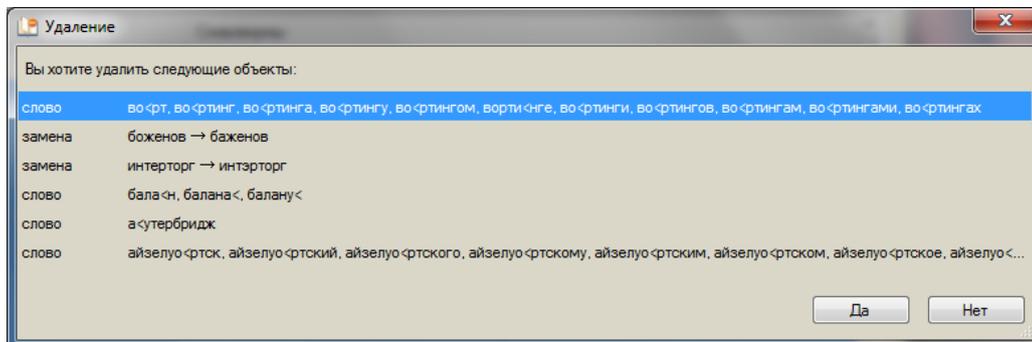


Рис.2.2.

На второй вкладке основного окна пользовательского словаря (рис.2.3) показаны русские и английские слова, которые **не были найдены** в словарях. Все слова расположены в алфавитном порядке. Также на этой вкладке можно использовать окно поиска.

## 2.4. Добавление и редактирование слов и замен

Для добавления и редактирования слов и замен нужно нажать кнопку «**Добавить слово**» в верхней левой части главного окна (рис.2.4). Загрузится окно «**Добавления и редактирования слов**». Окно состоит из четырех вкладок (каждая для своего режима добавления и редактирования).

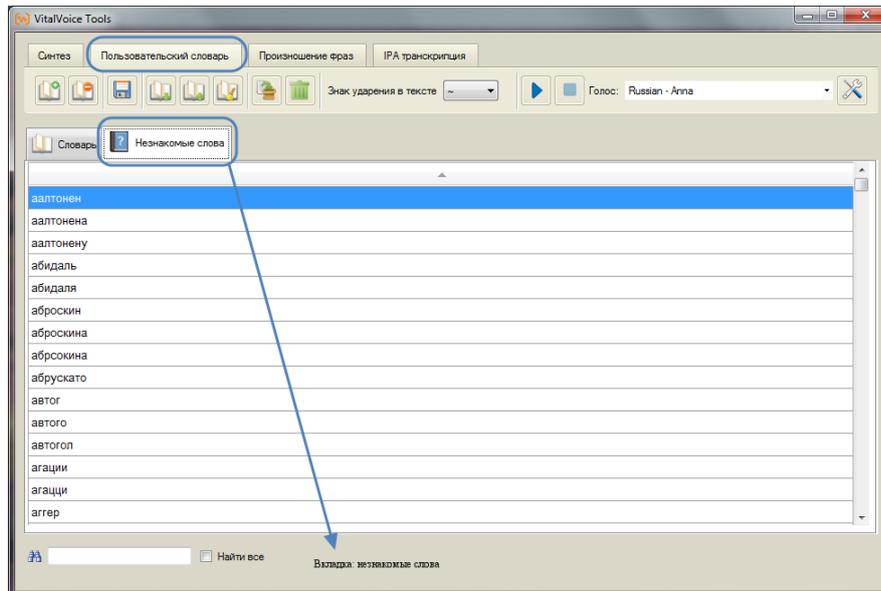


Рис.2.3. Вкладка «Незнакомые слова» окна пользовательского словаря

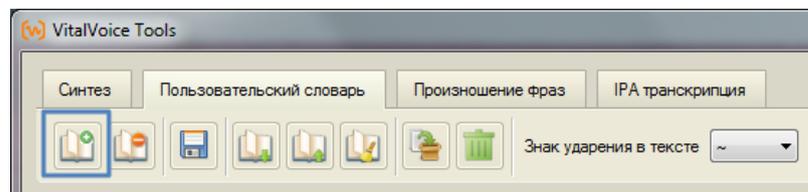


Рис.2.4.

### 2.4.1. Добавление нового слова

1. Введите новое слово в поле редактирования на первой вкладке «**Новое слово**» окна «**Добавление и редактирование слова**» (рис. 2.5). Не забудьте знак основного ударения (символ <)!

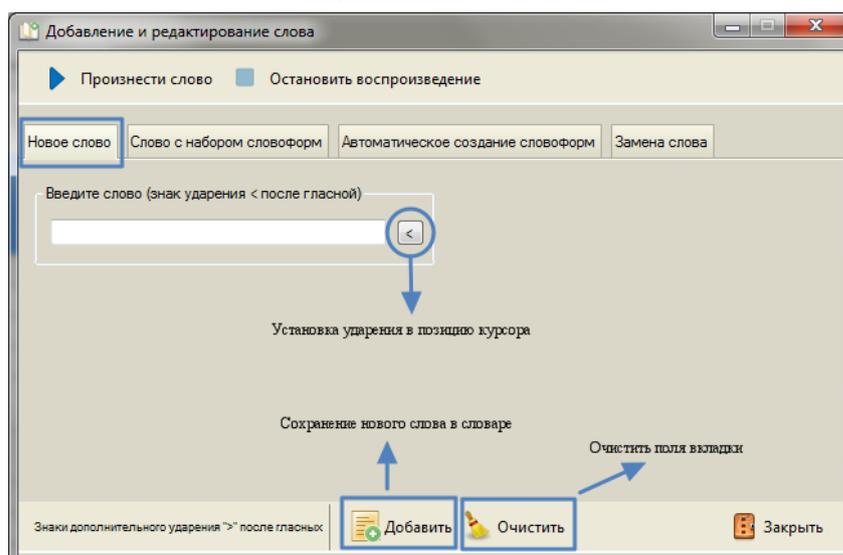


Рис.2.5.

2. Чтобы не переключать раскладку при установке знака ударения с русской на английскую, предусмотрена кнопка установки ударения в текущей позиции курсора.
3. Нажмите кнопку **«Сохранить»**, находящуюся в нижнем левом углу окна. Слово будет отображено и выделено (голубым «LightSkyBlue») цветом в списке слов пользовательского словаря.
4. Если нужно ввести в словарь несколько слов, то повторите шаги 1-3 для добавления каждого слова.
5. Для выхода из окна без сохранения слова нажмите клавишу Esc.

#### 2.4.2. Добавление слова с набором словоформ

Если вы хотите добавить словарную статью, включающую несколько форм одного и того же слова, то перейдите на закладку «Слово с набором форм» (рис. 2.6)

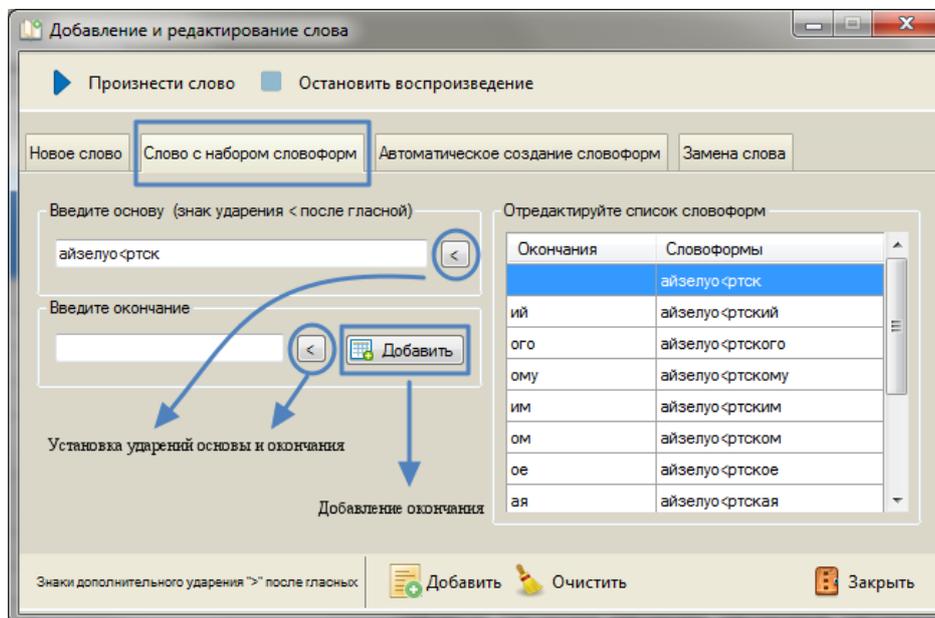


Рис.2.6.

1. Введите основу слова (ту часть слова, которая остается неизменной во всех формах) в поле «Основа». Например, если вы хотите добавить формы «Ива́нов, -а, -ы» введите ива<нов.
2. Введите окончание словоформы в поле «Окончание». Например, для формы «Ива́нова» введите а.
3. Нажмите кнопку «Добавить». Словоформа появится в списке форм слова.
4. Для добавления формы с нулевым окончанием, нажмите кнопку «Добавить» при пустом поле «Окончание».

5. Не забудьте знак ударения (символ <)! Ударение может стоять либо на основе, либо на окончании. Ударение в окончании можно изменить, щелкнув левой кнопкой мышки по соответствующей строке окончания (рис.2.7).

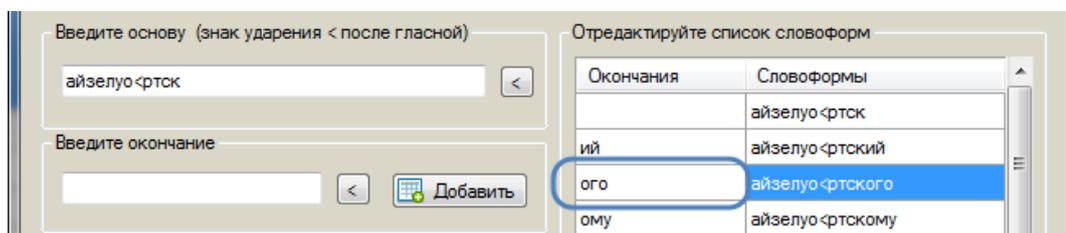


Рис.2.7.

Повторите шаги 2-5 для каждой из форм, которую вы хотите добавить в словарную статью.

Вы можете удалить из словарной статьи лишние формы, нажав клавишу «Delete». Будет выдан соответствующий запрос на удаление. Для добавления словарной статьи в словарь нажмите кнопку «Сохранить». Для выхода из окна без сохранения слова нажмите клавишу Esc.

### 2.4.3. Автоматическое создание словоформ

Вкладка «Автоматическое создание словоформ» позволяет пользователю сохранить полную парадигму добавляемого в словарь слова, выбрав набор окончаний по образцу одной из парадигм, хранящихся в памяти словаря (рис.2.8).

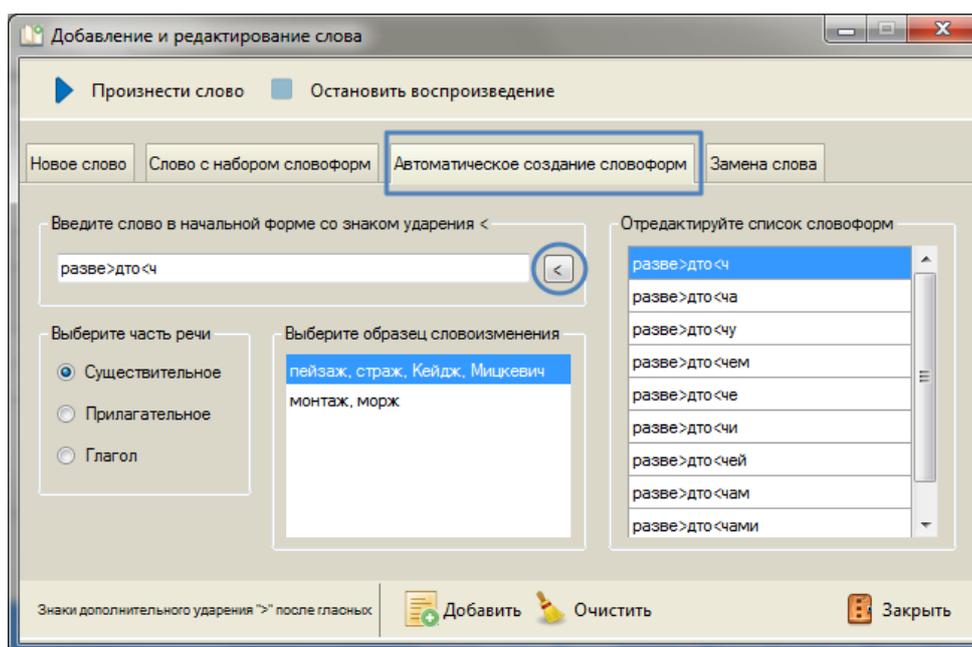


Рис.2.8.

Для добавления слова нужно выполнить следующие шаги:

1. Вести слово в начальной форме. Для существительного это именительный падеж единственного числа (например, "стол"), для прилагательного - именительный падеж единственного числа мужского рода (например, "красивый"), для глагола - неопределенная форма (инфинитив) (например, "делать"). Обязательно поставьте знак ударения. Для обозначения ударного гласного используется символ < (знак «меньше»), который ставится после ударной гласной. Вы также можете задать дополнительное ударение (одно или несколько) при помощи символа > (знак «больше»).
2. Выберите часть речи вводимого слова (существительное, прилагательное или глагол). Если вы добавляете причастие, выберите прилагательное. Парадигма глагола не включает причастия, поэтому при необходимости добавить к глаголу причастия их нужно добавлять в словарь отдельно. Имена собственные, такие как фамилии, личные имена, географические названия, являются существительными.
3. Выберите образец словоизменения из тех, что представлены в списке. При однократном щелчке мышью по выбранному образцу, в таблице словоформ в правой части окна будет показано словоизменение вводимого слова по выбранной модели.
4. Проверьте правильность выбранной парадигмы. По умолчанию ударение во всех формах слова ставится на том же слоге, что и в начальной форме. Щелкнув два раза по форме в списке, вы можете изменить место ударения в конкретной форме (рис. 2.9).

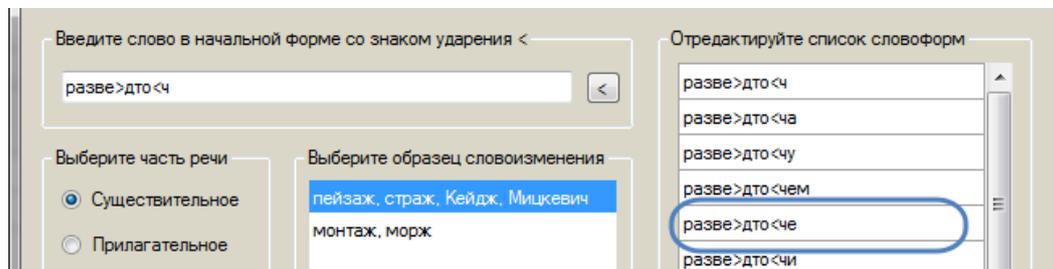


Рис.2.9.

5. Вы также можете удалить из парадигмы лишние формы, нажав клавишу «Delete». Будет выдан соответствующий запрос на удаление.
6. Нажмите кнопку "Сохранить" для завершения редактирования словарной статьи. Слово будет добавлено в пользовательский словарь.

#### 2.4.4. Замена слова

Вкладка «Замена слова» имеет следующий вид (рис. 2.10).

С помощью этой вкладки можно:

- Задать расшифровку сокращений или аббревиатур;
- задать чтение слов, написанных латиницей;

- задать чтение специальных символов

Последовательность действий следующая:

1. Введите текст в поле для исходного текста. Вы можете использовать также знаки препинания и другие символы. Например, если нужно расшифровать «пн. - вс.» как «с понедельника по воскресенье», то введите **пн. - вс.**
2. Введите желаемую расшифровку в поле замены. Например, **с понедельника по воскресенье.**
3. Вы можете задать желаемые ударения в расшифровке, используя символы <.
4. Нажмите кнопку «**Сохранить**».

Если нужно ввести в словарь несколько замен, то повторите шаги 1-3 для добавления каждого из них.

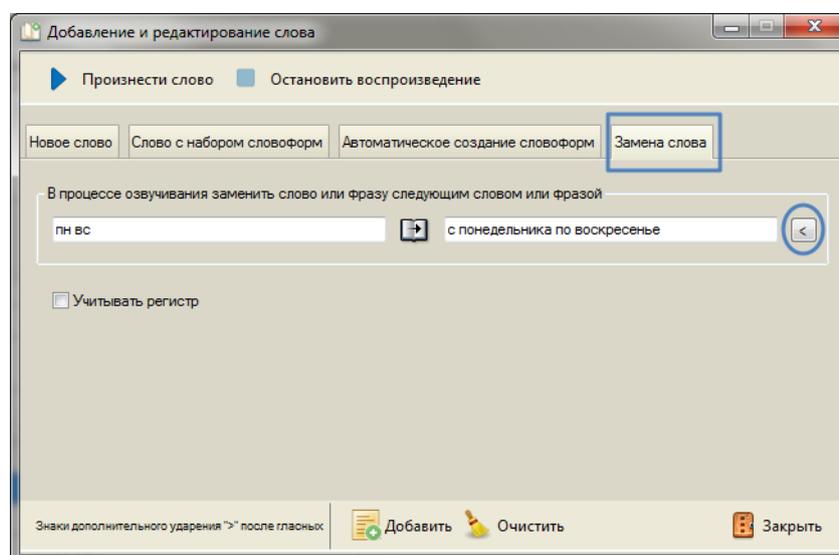


Рис.2.10.

## 2.5. Редактирование слов и замен

Для редактирования содержащихся в словаре слов или замен дважды щелкните мышью на соответствующей строке в списке главного окна.

Для элемента **слово** окно редактирования откроется на второй вкладке «**Слово с набором словоформ**». На этой вкладке уже будет загружено слово со всеми словоформами. Кроме того, это слово будет загружено и готово к редактированию в первой (как одиночное слова) и в третьей (автоматическое создание словоформ) вкладках.

Для элемента **замена** окно редактирования откроется на последней вкладке «**Замена слова**». В текстовые поля будет загружены исходный текст и текст замены.

## 2.6. Ударение в слове

Ударение в слове задается символом < (знак «меньше»), который ставится после ударной гласной. Если в слове не задано ударение, слово в словаре сохранено не будет, будет выдано предупреждение «В слове нет знака ударения».

Вы также можете задать в слове дополнительные ударения при помощи символа > (знак «больше»).

В слове может быть только одно основное ударение, но дополнительных ударений может быть несколько. При наличии дополнительных ударений в слове также обязательно должно быть основное ударение. При попытке сохранить слово, в котором есть дополнительное ударение (одно или несколько), но нет основного ударения, будет выдано предупреждение «В слове нет основного ударения».

Ударение может быть поставлено только после гласной буквы. При попытке поставить знак ударения после согласной знак печататься не будет.

## 2.7. Знак ударения в тексте

Вы можете выбрать знак ударения из списка в поле «Знак ударения в тексте» (рис. 2.11).

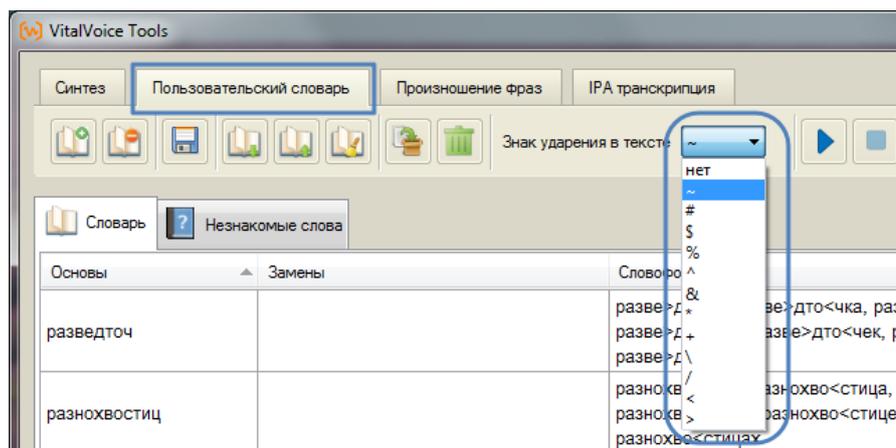


Рис.2.11.

Если один из предоставленных знаков выбран в качестве знака пользовательского ударения, то этот знак может быть использован для изменения ударения в слове при подаче текста на синтез речи.

Например, если выбран знак ~ то при чтении текста:

"по до~роге едет тракто~р." синтезатор произнесет: "по дОроге едет трактОр."

Если в поле «Знак ударения» выбрано значение «Нет», то не существует знака, который мог бы быть использован для изменения ударения в слове при подаче текста на синтез речи.

## 2.8. Инструменты

Панель инструментов находится в верхней части главного окна пользовательского словаря (рис. 2.12).

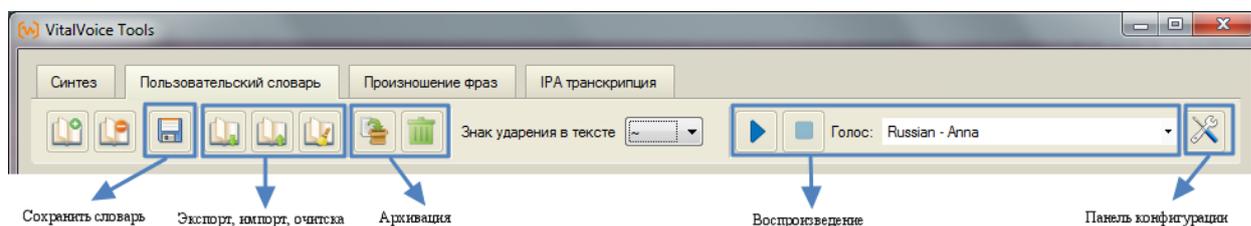


Рис.2.12.

- **Экспорт словаря** позволяет сохранить текущий словарь слов и замен как текстовый файл по заданному пользователем маршруту.
  - При **импорте словаря** словарь по заданному пользователем маршруту будет загружен в главное окно.
  - При **очистке словаря** список слов в главном окне будет очищен.
- Кнопка **«архивация»** позволяет сохранить текущее состояние словаря слов и замен по стандартному для движка синтеза маршруту.

## 2.9. Чтение слова с учетом данных пользовательского словаря

- ✓ При наличии слова в пользовательском словаре, ударение в данном слове всегда будет проставлено так, как оно задано в пользовательском словаре.
- ✓ При наличии в написании слова в пользовательском словаре буквы Ё, данное слово будет прочитано с Ё, даже если в тексте оно будет написано через Е.

Обратите внимание, что слово, добавленное в пользовательский словарь, будет читаться с желаемым ударением только в тех формах, в которых оно сохранено в словаре. Например, если в пользовательский словарь добавлено слово «ива<нов» (с ударением на «а»), и это единственная сохраненная в пользовательском словаре форма, то форма «иванова» будет прочитана по умолчанию (с ударением на «о»). Чтобы все формы слова читались с желаемым ударением, нужно все их добавить в словарь.

## 2.10. Обработка пользовательских замен

Обработка пользовательских замен и расшифровок при синтезе речи производится с учетом заданных знаков препинания. Например, если задана расшифровка «пн. - вс.» как «с понедельника по воскресенье», то написание «пн-вс» не будет расшифровано заданным образом. Этот вариант следует ввести отдельно.

Знаки препинания, стоящие в поданном на синтез тексте перед или после заданного для расшифровки элемента, будут обработаны в обычном режиме.

Если заданный для расшифровки элемент будет разделен в тексте знаком конца абзаца (переводом каретки), то расшифровки не произойдет. Так, в рассмотренном примере следующий текст не будет расшифрован заданным образом:

пн.-

пт.

Обработка пользовательских замен производится без учета регистра. Так, если в словарь введена замена слова «anne» на «a<nна», то в тексте слово «Anne» будет также прочитано.

## 2.11. Обработка знака пользовательского ударения в тексте

Знак пользовательского ударения, заданный через словарь пользователя, может быть затем использован в тексте, подаваемом на синтез, для задания ударения в отдельных словах.

Знак пользовательского ударения должен стоять в тексте после гласной. Например, если задан знак пользовательского ударения ~, то при подаче на синтез текста «Здравствуйте, господин Ива~нов» слово «Иванов» будет прочитано с ударением на «а».

Если знак пользовательского ударения не определен, то при подаче текста на синтез нельзя задать ударение с помощью знака после гласной. В этом случае для изменения ударения в синтезируемом тексте можно использовать соответствующий тег.

## 2.12. Описание экспериментов

1. Ввести в пользовательский словарь одиночное новое слово.
2. Добавить в пользовательский словарь слово с ручным набором словоформ.
3. Добавить в пользовательский словарь слово с автоматической генерации словоформы по заданному шаблону.
4. Добавить в пользовательский словарь замены для нескольких различных последовательностей слов и символов.
5. Установить знак пользовательского ударения для **всего** текста.
6. Провести синтез текстов с учетом введенных слов, замен и пользовательского ударения.

## 2.13. Оформление результатов

- ✓ Необходимо приложить текущий пользовательский словарь **Dict.UserWords.dic**, в котором зафиксированы все внесенные изменения по пунктам 1-6 предыдущего раздела.
- ✓ Нужно приложить протокол работы движка синтеза речи для последнего модуля лингвистического процессора – это модуль Accent (**result\_2.6.Accent.Result.xml**)

При каждом запуске процесса синтеза речи в приложении VVTools протоколы работы **создаются заново**.

### 3. Лабораторная работа 2.

Целью работы является управление лингвистическим процессором системы синтеза речи с помощью тегов пользовательской разметки **SSML**.

Основным инструментом для управления лингвистикой при работе движка TTS является тег “say-as”.

#### 3.1. Общая структура тега say-as

Тег say-as является открытым тегом. Синтаксис тега следующий:

|                                    |       |           |
|------------------------------------|-------|-----------|
| <say-as информационная часть тега> | текст | </say-as> |
|------------------------------------|-------|-----------|

Информационная часть тега представляет собой список вида:

|                                    |
|------------------------------------|
| Атрибут тега = “Значение атрибута” |
|------------------------------------|

На одном слове может быть не более одного тега. Вложенность тегов say-as запрещена. Тег работает с целым словом (или с группой слов: разделы “**telephone**” и “**characters**”). При установке тега на часть слова или вообще вне слова – соответствующий тег будет проигнорирован.

#### 3.2. Перечень атрибутов тега.

В теге say-as возможны следующие атрибуты:

- a. literal
- b. stress
- c. interpret-as
- d. format
- e. detail

Наличие одного из первых трех перечисленных атрибутов: **literal**, **stress** или **interpret-as** является обязательным. Обязательные атрибуты literal и stress могут присутствовать в теге как вместе, так и по отдельности, обязательный атрибут interpret-as вместе с другими обязательными атрибутами не устанавливается. В случае такой некорректной установки в теге будет обработан первый встретившейся обязательный атрибут (или, возможно, пара для случая literal и stress).

#### 3.3. Атрибут literal

Обязательный атрибут literal задаёт способ чтения выделенного слова в буквенном виде. Формат следующий:

|   |                |                              |
|---|----------------|------------------------------|
| <code>&lt;say-as literal="Новое слово"&gt;</code> | Исходное слово | <code>&lt;/say-as&gt;</code> |
|---|----------------|------------------------------|

В этом случае, в начале обработки исходное слово будет заменено на новое, указанное в качестве значения атрибута. И далее обрабатываться будет уже новое слово. Например,

|  |        |                              |
|--|--------|------------------------------|
| <code>&lt;say-as literal="Здравствуйте"&gt;</code> | Привет | <code>&lt;/say-as&gt;</code> |
|--|--------|------------------------------|

Слово «Привет» в данном месте текста будет заменено на слово «Здравствуйте».

### 3.4. Атрибут `stress`

Обязательный атрибут `stress` задаёт номер гласной<sup>1</sup>, на которой на слово ставится основное ударение. Нумерация считается по гласным слова, начиная с 1. Формат следующий:

|  |                |                              |
|--|----------------|------------------------------|
| <code>&lt;say-as stress="номер гласной"&gt;</code> | Исходное слово | <code>&lt;/say-as&gt;</code> |
|--|----------------|------------------------------|

В случае несоответствия значения атрибута количеству гласных в слове атрибут будет проигнорирован. Например,

|  |        |                              |
|--|--------|------------------------------|
| <code>&lt;say-as stress="2"&gt;</code> | Компас | <code>&lt;/say-as&gt;</code> |
|--|--------|------------------------------|

Вместо словарного слова «ко<мпас», будет обрабатываться слово «компа<с»<sup>2</sup>. Значения атрибута больше 2 в данном примере будут игнорироваться.

При совместной обработке атрибутов `stress` и `literal` сначала исходное слово будет заменено на новое, согласно значению атрибута `literal`, а затем к этому слову будет применено значение атрибута `stress`. Например,

|  |          |                              |
|--|----------|------------------------------|
| <code>&lt;say-as literal="шофер" stress="1"&gt;</code> | водитель | <code>&lt;/say-as&gt;</code> |
|--|----------|------------------------------|

Исходно слово «водитель» будет заменено на слово «шо<фер».

### 3.5. Атрибут `interpret-as`

Формат тега с данным атрибутом следующий:

<sup>1</sup> Для английского TTS атрибут `stress` задает номер слога, на который ставится основное ударение.

<sup>2</sup> Знак < в слове означает, что на предшествующую знаку гласную устанавливается основное ударение.

|   |       |           |
|---|-------|-----------|
| <say-as interpret-as ="значение атрибута" [format] [detail] > | Слово | </say-as> |
|---|-------|-----------|

Возможные значения атрибута `interpret-as`:

- a. date
- b. time
- c. telephone
- d. characters
- e. cardinal
- f. ordinal

В случае наличия в теге нескольких атрибутов `interpret-as` будет обработан первый, остальные проигнорированы. Вместе с обязательным атрибутом `interpret-as` могут присутствовать необязательные атрибуты `format` и `detail` (вместе или по отдельности).

### 3.5.1. Значение “date”

Устанавливает значение даты в грегорианском стиле<sup>3</sup>. Текст внутри тега задается в виде числовых полей с разделителями. Разделителем может быть точка, дефис, двоеточие или слеш. При этом необходимо наличие атрибута **format**, значением которого является одна из следующих строк:

- a. "mdy" – месяц, день, год
- b. "dmy" – день, месяц, год
- c. "ymd" – год, месяц, день
- d. "md" – месяц, день
- e. "dm" – день, месяц
- f. "ym" – год, месяц
- g. "my" – месяц, год
- h. "m" – месяц
- i. "d" – день
- j. "y" – год

Для значения дня возможны одна или две цифры (первая из цифр может быть нулем). Для значения месяца возможны одна или две цифры (первая из цифр может быть нулем). Для значения года возможны от одной до четырех цифр (запись может начинаться с нулей). Пробелы внутри слова-даты запрещены. Отсутствие разделителей внутри слова-даты разрешено только для «одинокных» форматов: “m”, “y” или “d”.

В случае некорректного значения слова или атрибута `format`:

- a. несоответствия значения атрибута `format` содержимому слова;
- b. неподдерживаемого значения атрибута `format`;
- c. некорректного задания значений дня, месяца или года;
- d. наличия разных разделителей в слове-дате;

тег будет проигнорирован.

<sup>3</sup> Для английского TTS дата расшифровывается в следующем формате: месяц → день (порядковое числительное) → год (количественное числительное)

Для одиночного формата дня (“d”) слово будет заменено на порядковое числительное единственного числа, среднего рода, в именительном падеже. Например, для тега

|   |    |           |
|---|----|-----------|
| <say-as interpret-as="date" format="d"> | 25 | </say-as> |
|---|----|-----------|

получим строку «двадцать пятое».

Для одиночного формата месяца слово будет заменено на название соответствующего месяца в именительном падеже. Например, для тега

|   |   |           |
|---|---|-----------|
| <say-as interpret-as="date" format="m"> | 2 | </say-as> |
|---|---|-----------|

получаем слово «февраль».

Для одиночного формата года (“y”) слово будет заменено на порядковое числительное единственного числа, мужского рода, в именительном падеже. Например, для тега

|   |      |           |
|---|------|-----------|
| <say-as interpret-as="date" format="y"> | 1995 | </say-as> |
|---|------|-----------|

получим «тысяча девятьсот девяносто пятый».

Для форматов “dm”, “md” расшифровка всегда происходит в порядке *день, месяц*. При расшифровке для дня устанавливается родительный падеж, для месяца – родительный падеж. Например, для тега

|  |     |           |
|--|-----|-----------|
| <say-as interpret-as="date" format="md"> | 3/6 | </say-as> |
|--|-----|-----------|

получим выражение «шестого марта».

Для форматов “my”, “ym” расшифровка всегда происходит в порядке *месяц, год*. При этом при отсутствии самого слова «год» после даты оно будет добавлено. Если в позиции для года выделено две цифры (XX), он будет раскрыт до полной даты. Если число XX меньше 20, то будет 20XX, иначе 19XX. При расшифровке для месяца устанавливается именительный падеж, для года – родительный падеж. Например, для тега

|  |      |           |
|--|------|-----------|
| <say-as interpret-as="date" format="dy"> | 3/02 | </say-as> |
|--|------|-----------|

получим выражение «март две тысячи второго года»

Для полных форматов “mdy”, “dmy” или “ymd” расшифровка всегда происходит в порядке *день, месяц, год*. При этом при отсутствии самого слова «год» после даты оно будет добавлено. Если в позиции для года выделено две цифры (XX), он будет раскрыт до полной даты. Если число XX меньше 20, то

будет 20XX, иначе 19XX. При расшифровке для дня устанавливается родительный падеж. Например, для тега

|   |        |           |
|---|--------|-----------|
| <say-as interpret-as="date" format="mdy"> | 3/6/02 | </say-as> |
|---|--------|-----------|

получим выражение «шестого марта две тысячи второго года».

### 3.5.2. Значение “time”

Устанавливает значение времени<sup>4</sup>. Текст внутри тега задается в виде числовых полей с разделителями или без них, в последовательности: часы, минуты, секунды. Значение для часов является обязательным. Секунды указываются только при наличии минут. Разделителем может быть точка, дефис, двоеточие или слеш. Наличие двух различных разделителей в слове-времени считается ошибкой.

Для значения часов возможны одна или две цифры в диапазоне от 0 до 23 (первая из цифр может быть нулем). Для значения минут и секунд возможны одна или две цифры в диапазоне от 0 до 59 (первая из цифр может быть нулем). Пробелы внутри слова-времени запрещены.

При отсутствии разделителей слово-время должно иметь одну из следующих длин:

- 2 (интерпретируется как значение только часов),
- 4 (интерпретируется как значение часов и минут),
- 6 (интерпретируется как значение часов, минут и секунд)

Другая длина слова-времени при отсутствии разделителей считается ошибкой. Например, для тега

после 

|                              |      |           |
|------------------------------|------|-----------|
| <say-as interpret-as="time"> | 2230 | </say-as> |
|------------------------------|------|-----------|

получим строку «после двадцати двух».

При этом полного (часы, минуты, секунды) и краткого (часы, минуты) форматов времени добавляются соответственно служебные слова «час...», «минут...» и (для полного формата) «секунд...». Например, для тега

сейчас 

|                              |         |           |
|------------------------------|---------|-----------|
| <say-as interpret-as="time"> | 9:21:30 | </say-as> |
|------------------------------|---------|-----------|

получим строку «сейчас девять часов двадцать одна минута тридцать секунд», а для тега

время отправления поезда 

|                              |      |           |
|------------------------------|------|-----------|
| <say-as interpret-as="time"> | 1:15 | </say-as> |
|------------------------------|------|-----------|

<sup>4</sup> Для английского TTS полное время расшифровывается в следующем формате: число - число - and - число - second(s). Секунды без окончания для числа 1, остальные случаи - с окончанием s.

получим строку «время отправления поезда один час пятнадцать минут».

**Замечание.** *Незначащие нули* в часах, минутах и секундах не читаются, например, тег

|                          |                              |       |           |
|--------------------------|------------------------------|-------|-----------|
| время отправления поезда | <say-as interpret-as="time"> | 01:07 | </say-as> |
|--------------------------|------------------------------|-------|-----------|

будет читаться, как «время отправления поезда один час семь минут».

### 3.5.3. Значение “telephone”

Задаёт чтение данного слова или группы слов как номеров или номера телефонов<sup>5</sup>. Слово номер телефона может содержать знак «+» и круглые скобки. Номер читается как количественное числительное в именительном падеже. При этом происходит разбиение номера на двух- и трехзначные числа. Нечисловые слова, попавшие в область действия тега, при этом обрабатываются обычным способом. Например, для тега

|                                   |                            |           |
|-----------------------------------|----------------------------|-----------|
| <say-as interpret-as="telephone"> | +(8)2284, у секретаря 2597 | </say-as> |
|-----------------------------------|----------------------------|-----------|

получим «плюс восемь двадцать два восемьдесят четыре у секретаря двадцать пять девяносто семь».

### 3.5.4. Значение “characters”

Задаёт чтение данного слова или группы слов по буквам. При этом буквы читаются как алфавитные, числительные по цифрам, как количественные в именительном падеже, специальные символы и знаки препинания заменяются соответствующими словами. Прописные и строчные буквы при этом не различаются. Например, для тега

|                                    |                |           |
|------------------------------------|----------------|-----------|
| <say-as interpret-as="characters"> | Б2а24-В!Zx?#7X | </say-as> |
|------------------------------------|----------------|-----------|

получим строку «бэ два а два четыре дефис би восклицательный знак зэт икс вопросительный знак решетка семь икс».

### 3.5.5. Значение “cardinal”

Задаёт чтение данного слова как количественного числительного. Атрибут **format** может содержать разделитель десятичной дроби. Атрибут **detail** может содержать разделитель тысяч. Разделителями могут служить символы точки, запятой или точки с запятой. При наличии атрибута разделитель данного типа в

<sup>5</sup> Для английского TTS читаем по цифрам.

слове может быть только один. Кроме разделителей, в слове допускаются только цифры.

Если есть разделитель тысяч, то часть для тысяч должна быть не длиннее 12 знаков. Если при этом есть еще и разделитель дроби, то целая часть дроби должна быть не более трех знаков.

Если разделителя тысяч нет, то целая и дробная часть слова (при наличии разделителя дроби) не должны быть длиннее 15 знаков.

При наличии разделителя дроби число расшифровывается и согласуется, как десятичная дробь (с добавлением служебных слов), например, для тега

|   |            |                              |
|---|------------|------------------------------|
| <code>&lt;say-as interpret-as="cardinal" format="." detail=";"&gt;</code> | 335;75.945 | <code>&lt;/say-as&gt;</code> |
|---|------------|------------------------------|

Получим строку: «триста тридцать пять тысяч семьдесят пять целых и девятьсот сорок пять тысячных».

В отсутствие разделителя дроби число интерпретируется как количественное числительное. Падеж числительного выбирается в зависимости от грамматического контекста (форм слов перед числительным и после него). При отсутствии контекста по умолчанию выбирается именительный падеж. Например, для тега

|          |   |     |                              |     |
|----------|---|-----|------------------------------|-----|
| не более | <code>&lt;say-as interpret-as="cardinal"&gt;</code> | 325 | <code>&lt;/say-as&gt;</code> | км. |
|----------|---|-----|------------------------------|-----|

получим строку: «не более трехсот двадцати пяти километров».

### 3.5.6. Значение “ordinal”

Задаёт чтение данного слова как порядкового числительного. Атрибуты `format` и `detail` не используются. Падеж, число и род числительного выбирается в зависимости от грамматического контекста (форм слов перед числительным и после него). При отсутствии контекста по умолчанию выбирается именительный падеж, мужской род, единственное число. Например, для тега

|  |    |                              |      |
|--|----|------------------------------|------|
| <code>&lt;say-as interpret-as="ordinal"&gt;</code> | 21 | <code>&lt;/say-as&gt;</code> | дом. |
|--|----|------------------------------|------|

получим строку: «двадцать первый дом».

## 3.6. Описание экспериментов

1. Отработка атрибута **literal** для русского и английского языков через интерфейс VVTools.
2. Отработка атрибута **stress** для русского и английского языков через интерфейс VVTools.

3. Отработка атрибута **interpret-as** для русского и английского языков через интерфейс VVTools.
4. Отработка атрибута **date** для русского и английского языков через интерфейс VVTools.
5. Отработка атрибута **time** для русского и английского языков через интерфейс VVTools.
6. Отработка атрибута **telephone** для русского и английского языков через интерфейс VVTools.
7. Отработка атрибута **characters** для русского и английского языков через интерфейс VVTools.
8. Отработка атрибута **cardinal** для русского и английского языков через интерфейс VVTools.
9. Отработка атрибута **ordinal** для русского и английского языков через интерфейс VVTools.

### 3.7. Оформление результатов

Нужно приложить протоколы работы движка синтеза речи (по каждому пункту предыдущего раздела отдельно) для последнего модуля лингвистического процессора – это модуль Accent (**result\_2.6.Accent.Result.xml**).

При каждом запуске процесса синтеза речи в приложении VVTools протоколы работы **создаются заново**.

## 4. Лабораторная работа 3.

Цель работы:

1. Управление работой просодического процессора системы синтеза речи с помощью тегов пользовательской разметки **SSML**.
2. Разбиение текста на синтагмические единицы.
3. Исследование интонационных контуров синтагм в системе синтеза русской речи (согласно классификации Е.А. Брызгуновой [1, с.96-122])

Основными инструментами для управления просодикой лингвистикой при работе движка TTS являются теги **“break”**, **“emphasis”**, **“s”** и **“procody”**.

### 4.1. Структура тега break

Тег break является **закрытым** тегом. Синтаксис тега следующий:

```
<break атрибут тега="значение"/>
```

Вложенность тегов break запрещена. В теге break возможны следующие атрибуты:

- a. time

## b. strength

## 4.1.1. Атрибут time

Атрибут time задаёт

- Целое число – значение длительности в миллисекундах.
- Целое число с суффиксом ms – значение длительности в миллисекундах; целое число с суффиксом s – значение длительности в секундах.

В результате будет выставлена пауза заданной длины в месте установки тега. Например,

***Я отойду на минутку. <break time="5000"/> Вы еще здесь?***

Между предложениями будет поставлена пауза 5с.

***Я отойду на минутку. <break time="4000ms"/> Вы еще здесь?***

Между предложениями будет поставлена пауза 4с.

## 4.1.2. Атрибут strength

Атрибут stress задаёт паузу заданной длины. Значения атрибута могут быть следующие:

- "none",
- "x-weak",
- "weak",
- "medium" (по умолчанию),
- "strong",
- "x-strong"

В результате будет выставлена пауза заданной длины в месте установки тега. Например,

***Я отойду <break strength="weak"/> на минутку.***

Между словами будет поставлена небольшая пауза.

## 4.2. Структура тега emphasis

Тег emphasis является **закрытым** тегом. Синтаксис тега следующий:

```
<emphasis level="значение">
```

Вложенность тегов emphasis запрещена. Атрибут level задает выделение слова логическим ударением. Значения атрибута могут быть следующие:

- "strong",
- "moderate" (по умолчанию),
- "none",
- "reduced".

Пример:

***Это <emphasis level="strong">огромный</emphasis> банковский счёт.***

На слово «огромный» будет поставлено логическое ударение.

### 4.3. Структура тега prosody

Тег prosody является **открытым** тегом. Синтаксис тега следующий:

|                              |       |            |
|------------------------------|-------|------------|
| <prosody атрибут="значение"> | текст | </prosody> |
|------------------------------|-------|------------|

В данной лабораторной работе мы рассмотрим только один атрибут тега имеющий просодическое значение: **pause**. В отличие от тега break тег prosody с атрибутом pause не устанавливает, а изменяет значение длительности паузы, устанавливаемой стандартными методами движка TTS. Значение атрибута – это коэффициент изменения длительности паузы в процентах. Например,

*<prosody pause="50"> Доброе утро. Как дела?</prosody>*

Пауза между предложениями будет уменьшена в два раза.

*<prosody pause="200"> Доброе утро. Как дела?</prosody>*

Пауза между предложениями будет увеличена в два раза.

### 4.4. Структура тега s

Тег s является **открытым** тегом. Синтаксис тега следующий:

|     |       |      |
|-----|-------|------|
| <s> | текст | </s> |
|-----|-------|------|

У тега нет атрибутов, он служит для разделения текста на предложения в месте установки. Например,

*<s>Ёжики кололись</s> и плакали.*

Часть "Ёжики кололись" выделится в отдельное предложение.

### 4.5. Определение синтагм

Синтагмы (полученные отрезки текста между паузами) в дальнейшем используются как область реализации интонационных контуров.

Правильная сегментация текста по синтагмам очень важна для достижения естественности звучания синтезируемого текста. Неправильная расстановка пауз, не соответствующая синтаксической и смысловой структуре предложения, всегда приводит к ухудшению восприятия текста, а зачастую - и к искажению его смысла. К тому же, восприятие длинных участков текста без пауз сильно затруднено. Поэтому правильной расстановке пауз и членению текста на синтагмы уделяется большое внимание.

Часть интонационных границ расставляется в зависимости от знаков препинания. Однако такой подход недостаточен, поскольку в тексте, как правило, все равно остаются большие последовательности слов, не разделенных пунктуационными знаками, которые требуется разбить на синтагмы. Расстановка интонационных границ в таких последовательностях осуществляется на основе грамматических связей между словами. Кроме того, учитывается длина отрезка между паузами.

В данной лабораторной работе предлагается сравнить проанализировать результаты расстановки синтагм только по знакам препинания (режим включается соответствующим флагом в панели «**Настройки синтеза**» в главном окне приложения VVTools и определением синтагм по заранее заданным правилам для движка TTS.

Границы синтагмы можно определить по протоколу работы модуля Segment. После каждой синтагмы ставится тег интонации:

`<intonation type="номер контура"/>`

Кроме того, по тегу

`<stress power="N" type="M"/>`

можно определить слово и гласную в нем, куда ставится фразовое ударение. Для этого слова N=2, M=1.

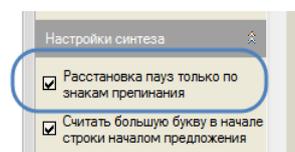


Рис.4.1.

#### 4.6. Исследование интонационных контуров

При выборе интонационной модели используется информация о знаках препинания, а также некоторые другие характеристики синтагмы, такие как наличие вопросительных слов. При определении интонационных типов за основу была взята классификация интонационных конструкций, принадлежащая Е.А. Брызгуновой [1], в модифицированном и расширенном виде. В программе VVTools для русского языка выделяется 13 интонационных типов.

- Повествовательное предложение (синтагма). Выставляется, если синтагма заканчивается точкой или если следует предложения без пунктуационного знака, а также на синтагму перед "но" в сложносочинённых предложениях.
- Повествовательное предложение (синтагма заканчивается точкой или точкой с запятой), в котором есть слово с особым выделением (напр., «Именно *Иван* пришел на работу».)
- Вопрос с вопросительными словами (именно с вопросительными местоимениями, а не союзами). Сюда также относятся вопросы с *ли/ль*, *Не мог(ла) бы ты...* и вторая часть вопросительной синтагмы с "или".
- Восклицательное предложение (синтагма).
- Восклицательное предложение с вопросительными словами (например, «Какая погода!»)
- Незавершенность — синтагма, заканчивающаяся многоточием.
- Простой вопрос, а также первая часть вопросительной синтагмы с "или" (кроме вопросов, начинающихся с союза "а" и при этом не содержащих глагольной формы).

- Вопрос со значением противопоставления - присваивается вопросительным синтагмам, если в них есть союз "а", но нет глагольной формы.
- Синтагма, заканчивающаяся на непарную запятую-тире, либо синтагма перед второй парной запятой-тире.
- Синтагма, заканчивающаяся на двоеточие, закрывающие скобки, второе парное тире.
- Синтагма, заканчивающаяся на непарном тире либо первую парную запятую-тире. Также эта интонационная модель проставляется на синтагму перед запятой и "а", при том что слева от запятой есть частица «не» (конструкция со значением противопоставления, напр. "Это газета, а не журнал").
- Синтагма, заканчивающаяся на запятую, открывающие скобки, первое парное тире.
- Синтагма, заканчивающаяся на паузу без пунктуационных знаков.

Присвоение синтагмам интонационных моделей происходит следующим образом: определяется последнее слово синтагмы, затем определяется, следует ли после него какой-либо знак препинания, и если да, то какой; далее в зависимости от найденного знака исследуется лексический состав синтагмы и другие знаки в пределах предложения.

#### 4.7. Описание экспериментов

1. Отработка атрибута **time** тега **break** через интерфейс VVTools.
2. Отработка атрибута **strength** тега **break** через интерфейс VVTools.
3. Отработка тега **emphasis** через интерфейс VVTools.
4. Отработка тега **procodey** через интерфейс VVTools.
5. Разбивка текста тегом **s** через интерфейс VVTools.
6. Определение синтагм по правилам и по знакам пунктуации через интерфейс VVTools.
7. Определение различных интонационных контуров через интерфейс VVTools.

#### 4.8. Оформление результатов

Нужно приложить протоколы работы движка синтеза речи (по каждому пункту предыдущего раздела отдельно) для последнего модуля просодического процессора – это модуль Segment (**result\_3.3.Segment.Result.xml**). При каждом запуске процесса синтеза речи в приложении VVTools протоколы работы создаются заново.

### 5. Лабораторная работа 4.

Цель работы:

1. Управление работой просодического процессора системы синтеза речи с помощью пользовательской разметки **SSML**

2. Исследование международного фонетического алфавита транскрипций IPA [2] с помощью приложения VVTools.

### 5.1. Фонетические SSML теги

Основными инструментами для управления фонетической обработкой движка TTS являются теги “**phoneme**” и “**lexicon**”.

Элемент `<phoneme>` используется для задания (с помощью транскрипционных символов) особого произношения выделенного текста. Синтаксис тега следующий

```
<phoneme alphabet="ipa" ph="транскрипция">слово</phoneme>
```

Вложенность тегов `phoneme` запрещена.

### 5.2. Атрибуты тега `phoneme`

В теге `phoneme` возможны следующие атрибуты:

- `alphabet` – имя алфавита. Атрибут необязательный, по умолчанию значение “**ipa**” – обычно используется именно этот международный фонетический алфавит.
- `ph` – строка с транскрипционными знаками. Обязательный атрибут, значение – транскрипционные значки и/или коды таких значков.

### 5.3. Примеры работы тега `phoneme`

Транскрипция слова томат:

```
<phoneme ph="t&#x259;mei&#x325;&#x27E;ou&#x325;">томат</phoneme>
```

Транскрипция слова "томат" будет заменена на "темеиру".

```
<phoneme ph="&#x712;zdras&#x690;t&#x690;i;">здравствуйте</phoneme>
```

```
<phoneme
```

```
ph="&#x642;&#x616;&#x712;s&#x690;&#x230;t">шестьдесят</phoneme>
```

### 5.4. Подключение произвольного IPA-словаря

Подключение произвольного IPA-словаря пользователя производится путем подачи `ssml`-тега **lexicon** на вход движка синтеза. При этом в атрибуте `uri` этого тега указывается путь к файлу IPA-словаря:

```
<lexicon uri="path"/>
```

Тег является закрытым и действует до появления следующего тега **lexicon**. Указание пустого атрибута `uri` или отсутствие атрибута `uri` в теге выгружает текущий словарь. Файл словаря должен находиться в корне движка.

После подключения IPA-словаря для каждого слова, подаваемого на вход движка, вместо произношения, создаваемого движком по стандартным правилам

транскрипции русского языка, будет подставлено произношение из IPA-словаря, соответствующее этому слову, при наличии такового в словаре.

## 5.5. Определение транскрипции через интерфейс VVTools

Модуль редактирования словаря IPA-транскрипций приложения VVTools предназначен для создания и редактирования транскрипций для отдельных слов. Этот словарь используется движком синтеза для модификации произношения отдельных слов. Модуль позволяет для любого слова задать его транскрипцию, записав ее в виде последовательности символов международного фонетического алфавита IPA.

IPA-транскрипции слов хранятся в пользовательском словаре IPA-транскрипций (далее – IPA-словарь). Транскрипция, созданная для конкретного слова, не зависит от голоса диктора (она не содержит информации об элементах звуковой базы), однако при синтезе она может использоваться как для всех существующих голосов, так и для одного конкретного голоса, согласно тому, как определит пользователь для данного слова.

IPA-словарь, создаваемый пользователем при помощи модуля редактирования словаря IPA-транскрипций VVTools, хранится в файле **Transform.Lexicon.txt**, который находится в папке с пользовательскими данными, путь к которой определяется ключом реестра

**HKLM\\SOFTWARE\\Speech Technology  
Center\\TTS\\Languages\\<...>\\UserDataPath**

где <...> = Russian для русского языка. При отсутствии данного ключа реестра IPA-словарь будет сохраняться в текущей директории.

## 5.6. Интерфейс модуля IPA транскрипции

Внешний вид интерфейса модуля изображен на рис.5.1.

Панель инструментов расположена в верхней части окна приложения, снизу от элемента управления для выбора модулей. Кнопки на панели инструментов позволяют производить следующие операции:

- **Генерация транскрипции** ([ ]). Создает для слова транскрипцию по умолчанию при помощи движка синтеза.
- **Произнести слово** (Play). Произносит слово с применением созданной транскрипции.
- **Произношение по умолчанию** (Play1). Произносит слово с применением произношения по умолчанию (транскрипция по стандартным правилам русского языка).
- **Стоп** (Stop). Прервать произнесение слова.
- **Сохранить** (Save). Сохраняет файл IPA-словаря.
- **Голос**. Выпадающий список выбора голоса позволяет выбрать голос для озвучивания слов и транскрипций. Выбранный пункт определяет голос, которым будет озвучена транскрипция в интерфейсе модуля, но как список

слов, так и транскрипции, созданные для каждого слова, не зависят от выбранного голоса.

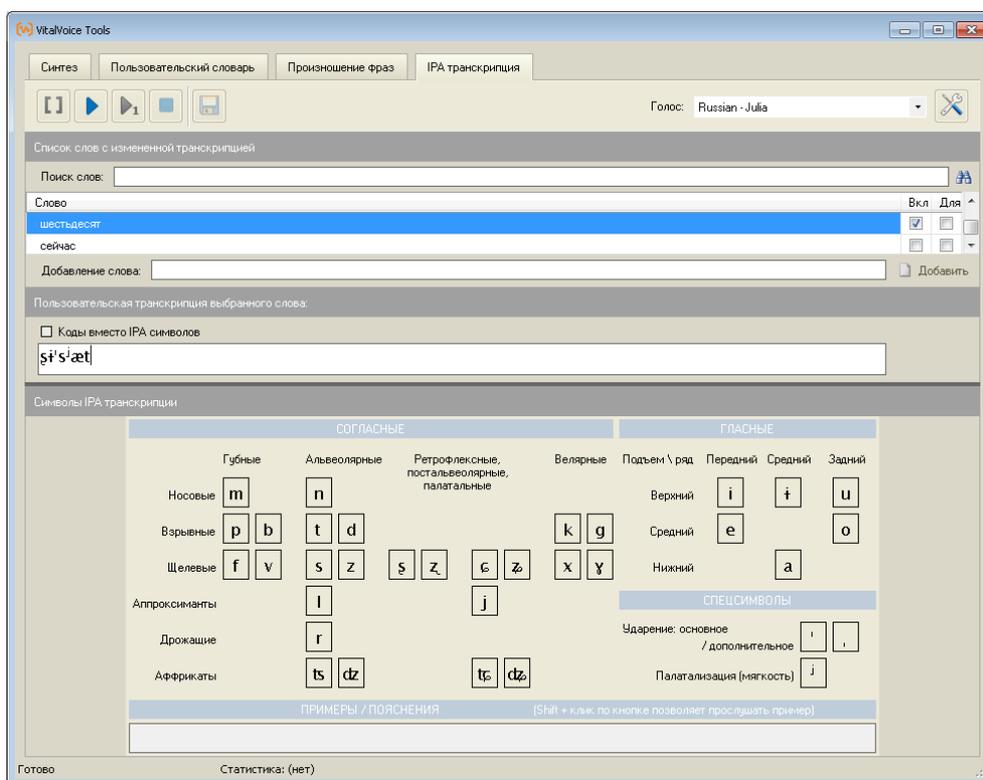


Рис. 5.1.

Табличное представление «**Список слов с измененной транскрипцией**» в верхней части окна под панелью инструментов содержит список слов, содержащихся в IPA-словаре. Добавление слова производится в отдельной строке.

Ячейки столбца «**Слово**» содержат текстовое представление слова. Редактирование слова производится прямо в ячейке. Ячейки столбца «**Вкл**» позволяют включать / отключать использование слова в движке синтеза. Ячейки столбца «**Для**» позволяют включать / отключать привязку слова к конкретному голосу в движке синтеза. Слово привязывается к тому голосу, который выбран в панели инструментов. Слова, не привязанные к голосу, используются для всех голосов. Слова, привязанные к другому голосу (не к тому, который выбран в панели инструментов), не отображаются в списке.

Строка для добавления нового слова, находящаяся под списком слов, позволяет добавлять новое слово в IPA-словарь. Кнопка справа от этой строки непосредственно осуществляет добавление слова в список, при этом для введенного слова создается транскрипция по умолчанию. Для вновь создаваемого слова привязка к голосу отсутствует.

Строка для ввода текста «**Фильтр слов**», находящаяся над списком слов, позволяет фильтровать слова по содержанию.

При вводе в строку каких-либо символов в списке слов отображаются только те слова, которые содержат в виде подстроки, введенные пользователем символы. Удобно при большом количестве слов в словаре.

Строка для ввода пользовательской транскрипции слова позволяет задавать транскрипцию для выбранного слова из списка. Транскрипция задается в виде последовательности IPA символов, которые могут быть введены при помощи виртуальной клавиатуры. Кнопка «Сохранить» непосредственно осуществляет обновление транскрипции для выбранного слова. В использовании этой кнопки нет необходимости, поскольку каждое внесенное изменение в строке ввода пользовательской транскрипции сохраняется автоматически.

В нижней части окна приложения находится виртуальная IPA-клавиатура, позволяющая вводить символы IPA-транскрипции в строку для ввода пользовательской транскрипции. Щелчок по кнопке виртуальной клавиатуры добавляет соответствующий символ в позицию текстового курсора указанной строки.

Над строкой для ввода пользовательской транскрипции слова расположена галочка «Коды вместо IPA символов». Включение этой галочки позволяет отображать транскрипцию в виде кодов, так что она может быть вставлена в виде тегов непосредственно в текст. Транскрипция может быть представлена как в чистом виде (как атрибут тега `phoneme`), так и вместе с этим тегом (рис.5.2).

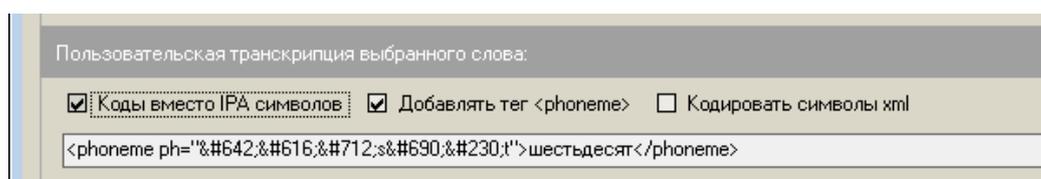


Рис. 5.2. Представление транскрипции в виде кодов.

Имеется также возможность дополнительно кодировать специальные xml символы в виде кодов: `< = &lt;`, `& = &amp;` и т.п. (рис.5.3).

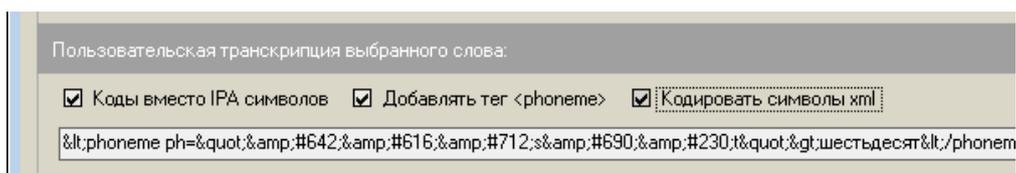


Рис. 5.3. Кодирование специальных символов xml.

Выбор голоса для озвучивания слов и их транскрипций производится путем выбора требуемого элемента выпадающего списка в выпадающем списке выбора голоса.

Выбор голоса влияет на озвучивание слов и транскрипций в интерфейсе модуля и на состав элементов, отображаемых в списке. Слова и транскрипции из IPA-словаря, не привязанные к какому-либо голосу, всегда присутствуют в списке; их использование в движке синтеза не зависит от выбранного голоса. Слова, привязанные к конкретному голосу, отображаются в списке только при выборе данного голоса, и в движке используются только совместно с ним.

Редактирование слова производится в ячейке таблицы слов в столбце «Слово» путем нажатия клавиши F2, либо левого щелчка по текущей ячейке, либо двойного щелчка по ячейке. При изменении слова в ячейке изменение его транскрипции не производится. Добавление нового слова производится путем

записи этого слова в строку добавления слова и нажатия кнопки «Добавить» справа от этой строки. Удаление слова производится путем выбора пункта «Удалить слово» контекстного меню, вызываемого правым щелчком по ячейке со словом. В модуле реализовано логическое удаление слова путем пометки его как неиспользуемого, без физического удаления из IPA-словаря. Для восстановления удаленного слова следует включить отображение удаленных слов, поставив соответствующую галочку «Показывать удаленные слова» (Check Delete) в окне настроек и затем заново включить использование удаленного слова. Физическое удаление слова из IPA-словаря возможно путем удаления неиспользуемого слова в режиме отображения удаленных слов. Физическое удаление слова не может быть отменено средствами, предоставляемыми модулем.

Галочка в ячейке столбца «Вкл» позволяет включать / выключать использование соответствующего слова в движке, аналогично логическому удалению слова. Галочка в ячейке столбца «Для» позволяет включать / отключать привязку соответствующего слова к выбранному голосу.

Изначально при добавлении нового слова для него генерируется транскрипция по умолчанию. Сгенерировать транскрипцию по умолчанию для слова можно и вручную, используя кнопку «Генерация транскрипции» (рис.5.4).

Транскрипцию слова можно модифицировать, редактируя ее в строке для ввода пользовательской транскрипции.

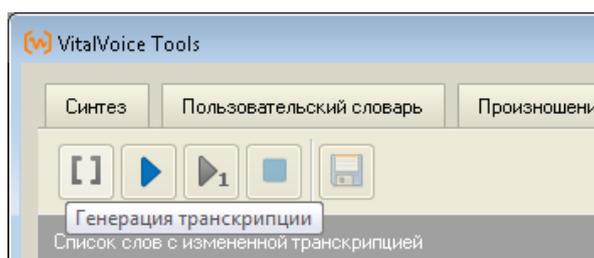


Рис. 5.4. Генерация транскрипции.

Произнесение слова с использованием созданной транскрипции производится путем нажатия кнопки «Произнести слово» (Play) панели инструментов (рис.5.5а).

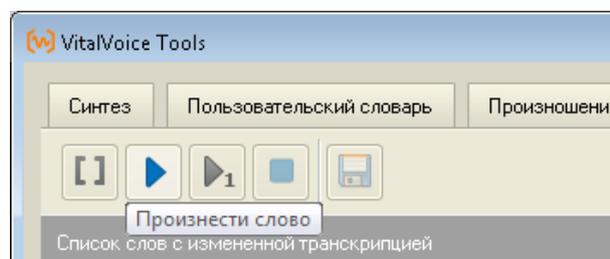


Рис. 5.5а. Произнести слово.

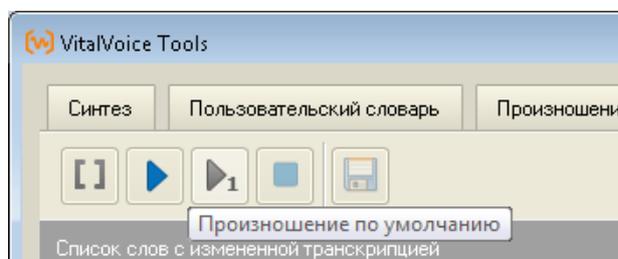


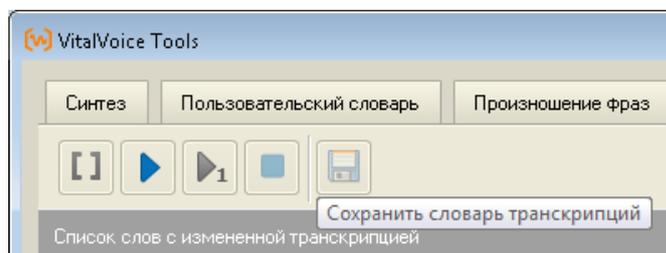
Рис. 5.5b. Произношение по умолчанию.

Произнесение слова с использованием транскрипции по умолчанию (по стандартным правилам транскрипции русского языка, без использования

созданной пользователем транскрипции) производится путем нажатия кнопки «Произношение по умолчанию» (Play1) панели инструментов (рис. 5.5b).

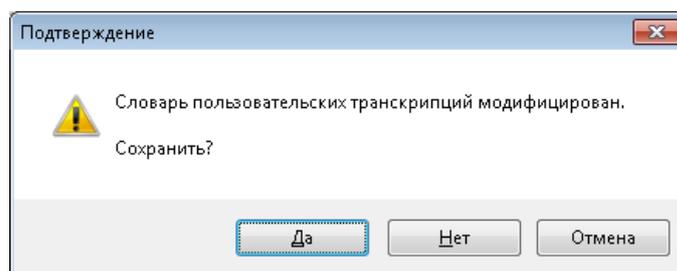
Произношение по умолчанию может быть использовано для сравнения с произношением созданной пользователем транскрипции.

Сохранение словаря IPA-транскрипций в файл производится путем нажатия кнопки «Сохранить» панели инструментов (рис. 5.6).



*Рис. 5.6. Сохранение IPA-словаря.*

Если в настройках модуля включена галочка «Автоматическое сохранение изменений», то сохранение словаря IPA-транскрипций в файл производится автоматически при любом изменении слова или его транскрипции. Если пользователь пытается выйти из приложения, а словарь транскрипций не сохранен, модуль выдает соответствующее предупреждение (рис. 5.7).



*Рис. 5.7. Предупреждение при выходе.*

## 5.7. Ограничения использования алфавита IPA в движке TTS

Алфавит IPA предоставляет весьма широкие возможности для записи транскрипций слов с множеством оттенков произношения. С помощью этого алфавита можно записать много различных звуков (все известные фонемы всех языков мира); еще большее разнообразие вариаций произношения может быть записано при помощи диакритик (специальных знаков, надстрочных и подстрочных символов): аспирация, назализация, хриплость, шепот и т.п. – всего более 30 признаков.

При этом в русском языке реализуются только некоторые из всего множества фонем, которые могут быть записаны символами IPA [2]:

Остальные символы IPA не имеют соответствия в фонетике русского языка.

Однако, в целях максимально широкого охвата всех имеющихся символов IPA, чтобы избежать отсутствия звука в том месте, где он, согласно транскрипции, должен быть, в движке синтеза было реализовано преобразование

почти всех символов IPA в звуки. Это преобразование основано на экспертном подборе русской фонемы, наиболее близкой к соответствующему IPA-символу по звучанию. Например, фонема [h] используется для озвучивания 11 различных символов IPA. Поэтому, несмотря на то, что каждый из этих 11 символов формально соответствуют разным звукам, все они будут одинаково озвучены движком синтеза в силу банального отсутствия соответствующих звуков в русской фонетике, и, следовательно, в звуковой базе данных, записанной русскоговорящим диктором.

В силу тех же причин движком синтеза не поддерживаются большинство диакритик IPA, означающих различные вариации произношения звуков. Поддерживается только основное ударение и дополнительное ударение для гласных звуков, а также палатализация (мягкость) для согласных звуков.

Символы алфавита IPA при обработке их движком синтеза преобразуются в звуки русской фонетики максимально корректно. Символы, не являющиеся символами IPA, а также некоторые особо экзотические символы IPA игнорируются при обработке движком и не дают звука при произношении.

Для удобства введены некоторые расширения. Все буквы латинского алфавита, даже те, которые не являются символами IPA, максимально корректно преобразуются в русские фонемы. При этом комбинации символов, вроде sh, zh, ch, sch, и т.п. не поддерживаются, латинские буквы обрабатываются по одной.

Основное ударение может быть обозначено «\*» перед слогом или перед гласной, дополнительное ударение – «+» перед слогом или перед гласной, палатализация – «'» после согласной.

## 5.8. Описание экспериментов

1. Отработка тега **phoneme** через интерфейс VVTools.
2. Создание пользовательского словаря транскрипций через интерфейс VVTools.

## 5.9. Оформление результатов

- ✓ Приложить протоколы работы движка синтеза речи для модуля фонетического процессора (**result\_4.1.Transform.Result.xml**).
- ✓ Приложить пользовательский словарь транскрипций Transform.Lexicon.txt.
- ✓ При каждом запуске процесса синтеза речи в приложении VVTools протоколы работы **создаются заново**.

## ЛИТЕРАТУРА

1. Брызгунова Е.А. Интонация // Русская грамматика. Том 1. «Наука». Москва, 1982.
2. [http://en.wikipedia.org/wiki/Wikipedia:IPA\\_for\\_Russian](http://en.wikipedia.org/wiki/Wikipedia:IPA_for_Russian)
3. Taylor P. "Text-to-Speech Synthesis", Cambridge University Press, pp. 426-433, United Kingdom, 2008.

## Оглавление

|  |    |
|--|----|
| 1. Программные комплекс Vital Voice Tools .....                  | 3  |
| 2.1. Запуск приложения .....                                     | 3  |
| 2.2. Модуль синтеза речи .....                                   | 4  |
| 1.2.1. Окно отображения текста .....                             | 6  |
| 1.1.2. Панель инструментов .....                                 | 6  |
| 1.2.3. Нижняя панель .....                                       | 7  |
| 1.2.4. Боковая панель .....                                      | 8  |
| 1.2.5. Диалог конфигурации модуля .....                          | 9  |
| 1.2.6. Трехполосный эквалайзер.....                              | 10 |
| 1.2.7. Параметры реверберации.....                               | 11 |
| 2. Лабораторная работа 1. ....                                   | 11 |
| 2.1. Особенности работы приложения VVTools .....                 | 11 |
| 2.2. Функции пользовательского словаря.....                      | 12 |
| 2.3. Общий интерфейс пользовательского словаря.....              | 12 |
| 2.4. Добавление и редактирование слов и замен .....              | 13 |
| 2.4.1. Добавление нового слова.....                              | 14 |
| 2.4.2. Добавление слова с набором словоформ .....                | 15 |
| 2.4.3. Автоматическое создание словоформ .....                   | 16 |
| 2.4.4. Замена слова.....   | 17 |
| 2.5. Редактирование слов и замен .....                           | 18 |
| 2.6. Ударение в слове.....                                       | 19 |
| 2.7. Знак ударения в тексте .....                                | 19 |
| 2.8. Инструменты .....   | 20 |
| 2.9. Чтение слова с учетом данных пользовательского словаря..... | 20 |
| 2.10. Обработка пользовательских замен.....                      | 20 |
| 2.11. Обработка знака пользовательского ударения в тексте .....  | 21 |
| 2.12. Описание экспериментов .....                               | 21 |
| 2.13. Оформление результатов .....                               | 21 |
| 3. Лабораторная работа 2. ....                                   | 22 |
| 3.1. Общая структура тега say-as .....                           | 22 |
| 3.2. Перечень атрибутов тега. ....                               | 22 |
| 3.3. Атрибут literal.....  | 22 |
| 3.4. Атрибут stress .....  | 23 |

|   |    |
|---|----|
| 3.5. Атрибут interpret-as.....                                | 23 |
| 3.5.1. Значение “date” .....                                  | 24 |
| 3.5.2. Значение “time” .....                                  | 26 |
| 3.5.3. Значение “telephone” .....                             | 27 |
| 3.5.4. Значение “characters” .....                            | 27 |
| 3.5.5. Значение “cardinal” .....                              | 27 |
| 3.5.6. Значение “ordinal” .....                               | 28 |
| 3.6. Описание экспериментов .....                             | 28 |
| 3.7. Оформление результатов .....                             | 29 |
| 4. Лабораторная работа 3. ....                                | 29 |
| 4.1. Структура тега break.....                                | 29 |
| 4.1.1. Атрибут time.....                                      | 30 |
| 4.1.2. Атрибут strength.....                                  | 30 |
| 4.2. Структура тега emphasis.....                             | 30 |
| 4.3. Структура тега prosody.....                              | 31 |
| 4.4. Структура тега s .....                                   | 31 |
| 4.5. Определение синтагм .....                                | 31 |
| 4.6. Исследование интонационных контуров .....                | 32 |
| 4.7. Описание экспериментов .....                             | 33 |
| 4.8. Оформление результатов .....                             | 33 |
| 5. Лабораторная работа 4. ....                                | 33 |
| 5.1. Фонетические SSML теги .....                             | 34 |
| 5.2. Атрибуты тега phoneme.....                               | 34 |
| 5.3. Примеры работы тега phoneme.....                         | 34 |
| 5.4. Подключение произвольного IPA-словаря .....              | 34 |
| 5.5. Определение транскрипции через интерфейс VVTools.....    | 35 |
| 5.6. Интерфейс модуля IPA транскрипции.....                   | 35 |
| 5.7. Ограничения использования алфавита IPA в движке TTS..... | 39 |
| 5.8. Описание экспериментов .....                             | 40 |
| 5.9. Оформление результатов .....                             | 40 |
| ЛИТЕРАТУРА .....  | 40 |

**Миссия университета** – генерация передовых знаний, внедрение инновационных разработок и подготовка элитных кадров, способных действовать в условиях быстро меняющегося мира и обеспечивать опережающее развитие науки, технологий и других областей для содействия решению актуальных задач.

---

## КАФЕДРА РЕЧЕВЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

### О кафедре

Кафедра речевых информационных систем (РИС) создана в 2011 году на факультете Информационных технологий и программирования (ФИТиП).

Организатором создания кафедры выступает «Центр речевых технологий» ([www.speechpro.ru](http://www.speechpro.ru)). Заведующий кафедрой – доктор технических наук Матвеев Юрий Николаевич.

Кафедра РИС обеспечивает подготовку докторантов, аспирантов и магистров. Для тех, кто имеет высшее образование, но хотел бы связать свое будущее с речевыми технологиями, имеются курсы дополнительного профессионального образования.

### Обучение на кафедре

Кафедра «Речевые информационные системы» (базовая кафедра «Центра речевых технологий») Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (ИТМО) в рамках направления 09.04.02 «Информационные системы и технологии» открывает прием в магистратуру по новой образовательной программе «Речевые информационные системы».

Срок обучения – 2 года. Обучение завершается защитой магистерской диссертации.

Целевая установка магистратуры – подготовка специалистов, способных участвовать в исследовательской и проектной работе в области речевых информационных технологий со специализацией в направлениях распознавания и синтеза речи, распознавания личностей по голосу, мультимодальной биометрии, в области проектирования и разработки информационных систем и программного обеспечения.

Область профессиональной деятельности выпускников кафедры РИС включает:

- исследование, разработка, внедрение речевых информационных технологий и систем;
- методы и алгоритмы цифровой обработки речевых сигналов;
- автоматизированные системы обработки речевых сигналов;

- программное обеспечение автоматизированных речевых информационных систем;

- системы автоматизированного проектирования программных и аппаратных средств для речевых информационных систем и информационной поддержки таких средств.

Объектами профессиональной деятельности выпускников кафедры РИС являются:

- информационные процессы, технологии, системы и сети, предназначенные для обработки, распознавания, синтеза речевых сигналов;

- инструментальное (математическое, информационное, техническое, лингвистическое, программное, эргономическое, организационное и правовое) обеспечение речевых информационных систем;

- способы и методы проектирования, отладки, производства и эксплуатации информационных технологий и систем в областях обработки, распознавания, синтеза речевых сигналов, телекоммуникации, связи, инфокоммуникации, медицины.

Широкий профиль подготовки, знание универсальных методов исследования и проектирования информационных систем, практические навыки работы с современным программным обеспечением – все это позволяет выпускникам кафедры найти работу в научных институтах и университетах, в фирмах, на производственных предприятиях, а также в коммерческих структурах. Учебный план предусматривает, в частности, следующие курсы:

Информационные технологии:

- Системный анализ и моделирование информационных процессов и систем;

- Управление программными проектами;

- Организация проектирования речевых информационных систем;

- Управление качеством разработки программного обеспечения.

Речевые технологии:

- Цифровая обработка сигналов;

- Цифровая обработка речевых сигналов;

- Основы речевых технологий

- Машинное обучение;

- Введение в интеллектуальный анализ данных;

- Распознавание речи;

- Синтез речи

- Распознавание диктора (говорящего по голосу);

- Мультимодальные биометрические системы.

К преподаванию привлекаются ведущие специалисты «Центра речевых технологий», преподаватели Университета ИТМО, а также специалисты, работающие в известных научных организациях (СПИИРАН), а также производственных и коммерческих организациях.

Матвеев Юрий Николаевич,  
Рыбин Сергей Витальевич

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ  
ЛАБОРАТОРНЫХ РАБОТ ПО КУРСУ «СИНТЕЗ РЕЧИ»**

Учебно-методическое пособие

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ № 4082

Тираж 50 экз.

Отпечатано на ризографе

