
SIS II

Модули идентификации

STC-S521

Руководство пользователя
ЦВАУ.00432-01 94

АННОТАЦИЯ

Данный документ «SIS II. Модули идентификации. STC-S521. Руководство пользователя. ЦВАУ.00432-01 94» (далее – руководство) предназначен для операторов-экспертов, использующих **SIS II** для идентификации личности по голосу.

Документ содержит:

1. Общие сведения о наборе модулей.
 2. Порядок идентификации дикторов с использованием всего набора модулей.
 3. Порядок автоматической оценки характеристик сигнала.
 4. Порядок автоматического детектирования речи.
 5. Порядок автоматического разделения речи дикторов в диалоге.
 6. Порядок автоматического сравнения дикторов.
 7. Порядок идентификации дикторов путём сравнения формант.
 8. Порядок идентификации дикторов путём сравнения основного тона.
 9. Порядок идентификации дикторов путём сравнения речевых признаков.
 10. Порядок принятия общего идентификационного решения о различии или тождестве дикторов на основе решения каждого отдельного модуля.
 11. Описание сообщений оператору при ошибках в работе и рекомендаций по правильным действиям.
- В приложении приведены:
- А. Пояснения к характеристикам качества фонограмм.
 - Б. Пояснения к доверительным интервалам идентификации дикторов.
 - В. Описание элементов мелодического контура и их параметров.
 - Г. Перечень терминов и определений.
 - Д. Перечень сокращений.

Ни одна из частей этого издания не подлежит воспроизведению, передаче, хранению в поисковой системе или переводу на какой-либо язык в любой форме, любыми средствами без письменного разрешения ООО «ЦРТ».

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
Общие положения	5
Требования к персоналу.....	5
Соглашения и обозначения.....	5
Товарные знаки.....	6
1 ОБЩИЕ СВЕДЕНИЯ	7
1.1 Сведения о продукте и изготовителе.....	7
1.2 Назначение	7
1.3 Состав и возможности	7
2 МЕТОДИКА	9
2.1 Работа с модулем	9
3 ОЦЕНКА ХАРАКТЕРИСТИК СИГНАЛА	15
3.1 Расчёт характеристик сигнала.....	15
3.2 Графики спектральных характеристик.....	17
3.3 Оценка и рекомендации по полученным характеристикам сигнала.....	18
4 ДЕТЕКТИРОВАНИЕ РЕЧИ	19
4.1 Выполнение детектирования	19
4.2 Результаты детектирования	20
5 РАЗДЕЛЕНИЕ РЕЧИ ДИКТОРОВ В ДИАЛОГЕ	22
5.1 Выполнение разделения речи дикторов.....	22
5.2 Результаты разделения речи дикторов	23
5.3 Коррекция разделения речи дикторов.....	24
6 АВТОМАТИЧЕСКОЕ СРАВНЕНИЕ	25
6.1 Метод идентификации дикторов на основе сравнения статистик основного тона голоса.....	25
6.2 Спектрально-формантный метод идентификации дикторов.....	26
6.3 Метод полной изменчивости.....	27
6.4 Построение общего идентификационного решения.....	28
6.5 Выполнение сравнения	29
6.6 Результаты идентификации.....	30
7 СРАВНЕНИЕ ФОРМАНТ	33
7.1 Подготовка к работе с модулем.....	33
7.2 Формирование таблицы опорных фрагментов	40
7.3 Диаграмма рассеяния	45
7.4 Сравнение выбранных фрагментов.....	48
7.5 Возможные методы выполнения идентификации с помощью модуля сравнения формант.....	52
7.5.1 Настройка порогов принятия идентификационного решения	52
7.5.2 Фонемно-спектральный метод. Сравнение гласных фонем в сопоставимых контекстах.....	53
7.5.3 Фонемно-спектральный метод. Сравнение форматных структур в типичных фонемных областях....	55
7.5.4 Метод формантного выравнивания.....	57
8 СРАВНЕНИЕ ОСНОВНОГО ТОНА	59
8.1 Подготовка к работе с модулем.....	60

8.2 Формирование фрагментов	62
8.3 Сравнение фрагментов	64
9 СРАВНЕНИЕ РЕЧЕВЫХ ПРИЗНАКОВ	66
9.1 Подготовка к работе с модулем.....	67
9.2 Сравнение фонограмм	69
10 ПРИНЯТИЕ ОБЩЕГО ИДЕНТИФИКАЦИОННОГО РЕШЕНИЯ.....	76
11 СООБЩЕНИЯ, ВЫДАВАЕМЫЕ ОПЕРАТОРУ	79
ПРИЛОЖЕНИЯ	82
Приложение А: Пояснения к характеристикам качества фонограммы	82
Приложение Б: Пояснения к доверительным интервалам.....	85
Приложение В: Элементы мелодического контура и параметры их описания	86
Приложение Г: Перечень терминов и определений	90
Приложение Д: Перечень сокращений	92
Приложение Е. Таблица весов по типам и подтипам мелодического контура.....	93

ВВЕДЕНИЕ

Общие положения

В данном руководстве приведены необходимые сведения по работе с модулями для идентификации личности по голосу в составе **SIS II**.

Порядок использования **SIS II** изложен в документе «SIS II. Специальное программное обеспечение для идентификации говорящего на русском языке STC-S521 Руководство пользователя ЦБАУ.00432-01 94».

Настоящее руководство также не заменяет учебную, справочную литературу и руководства от производителей операционной системы и общего программного обеспечения.

Требования к персоналу

Персонал, производящий установку набора модулей идентификации, должен иметь профессиональные навыки установки общего и специального программного обеспечения.

Персонал, работающий с набором модулей идентификации в составе **SIS II**, должен иметь базовые навыки работы с приложениями в среде операционных систем **Microsoft Windows**, а также специальные знания в области экспертизы фонограмм речи.

Соглашения и обозначения

В руководстве приняты следующие типографские соглашения:

Формат	Значение
Обычный	Основной текст документа
<i>Курсив</i>	Применяется для выделения первого появления <i>термина</i> , значение которого поясняется здесь же или даётся в приложении.
Полужирный	Применяется для написания наименований программных компонентов и наименований управляющих и информационных элементов интерфейса (заголовки, кнопки и т. п.).
Полужирный курсив	Применяется для написания имён файлов и путей доступа к ним.

Словосочетание «выбрать, выделить, нажать объект (или нажать на объект)» означает: «навести указатель манипулятора типа «мышь» на объект, и нажать кнопку манипулятора».

Словосочетание «дважды щёлкнуть по или на объекте» означает «навести указатель манипулятора типа «мышь» на объект, и быстро дважды нажать кнопку манипулятора».

Выбор меню может быть показан при помощи стрелки →, например, текст **Файл** → **Выход**, должен пониматься так: выбрать меню **Файл**, затем команду **Выход** из меню **Файл**.

Ниже приведены примеры оформления материала руководства, указывающие на важность сведений.



Ссылки на другие документы в основном тексте.



Примечания; важные сведения; указания на действия, которые необходимо выполнить в обязательном порядке.



Требования, несоблюдение которых может привести к некорректной работе, повреждению или выходу из строя изделий или программного обеспечения.

Товарные знаки

Наименование **SIS** является товарным знаком общества с ограниченной ответственностью «Центр речевых технологий».

Все остальные названия компаний и названия продуктов, упомянутые в документе, являются собственностью их соответствующих владельцев.

В состав программы входят модули межплатформенной среды разработки приложений **Qt 4.7.0** (<http://qt.nokia.com>), лицензия GNU LGPL 2.1 <http://www.gnu.org/licenses/lgpl-2.1.html>.

1 ОБЩИЕ СВЕДЕНИЯ

1.1 Сведения о продукте и изготовителе

Наименование	Модули идентификации
Условное наименование	STC-S521
Обозначение	ЦВАУ.00432-01
Изготовитель	Общество с ограниченной ответственностью «Центр речевых технологий»
Адрес	196084, г. Санкт-Петербург, ул. Красуцкого, д. 4, литера А
Телефон	+7 (812) 325-88-48
Факс	+7 (812) 327-92-97

1.2 Назначение

Модули идентификации в составе **SIS II** предназначены для оценки характеристик сигнала, детектирования речи и шумов, а также идентификации личности по голосу с применением как полностью автоматических, так и автоматизированных методов.

1.3 Состав и возможности

Набор включает следующие программные модули, входящие в состав **SIS II**:

- 1) Модуль «Методика», представляющий собой совокупность всех модулей.
- 2) Модуль оценки качества фонограммы «Характеристики сигнала» позволяет автоматически вычислить основные параметры звукового сигнала фонограммы в целом и оценить ее пригодность для дальнейших исследований.
- 3) Модуль разделения речи дикторов «Разделение дикторов в диалоге» позволяет выполнять автоматическую маркировку речевых сегментов по дикторам на записях естественной речи.
- 4) Модуль поиска речевых участков «Детектирование речи» позволяет автоматически выделять в фонограмме речевые, шумовые участки и паузы.
- 5) Модуль «Автоматическое сравнение» выполняет автоматическую идентификацию личности по голосу тремя методами с использованием обобщающего решения.
- 6) Модуль «Сравнение формант» позволяет автоматизировать процесс идентификации личности по голосу на основе метода сравнения формант.
- 7) Модуль «Сравнение основного тона» позволяет автоматизировать процесс идентификации личности по голосу на основе сравнения параметров мелодических контуров голоса сравниваемых дикторов.

8) Модуль «Сравнение речевых признаков» позволяет автоматизировать процесс идентификации личности по голосу на основе экспертного сравнения лингвистических характеристик речи сравниваемых дикторов.

9) Модуль «Общее решение» позволяет принять общее решение относительно идентификации личности по голосу на основе результатов работы каждого из использованных модулей идентификации дикторов.

2 МЕТОДИКА

Модуль «Методика» позволяет систематизировать порядок работы эксперта при проведении идентификационной экспертизы. В интерфейсе модуля отображаются результаты по каждому проведенному этапу исследования.

2.1 Работа с модулем



Перед запуском модуля «Методика» должен быть задан активный проект, а также подготовлены как минимум два окна с данными речи дикторов.

Для начала работы с модулем «Методика», выберите пункт **Методика** в меню **Модули**.

В открывшемся окне **Методика** (рис. 1) представлены 4 вкладки, которые становятся доступными после выбора всех необходимых параметров или проведения необходимых действий.

После выполнения всех обязательных действий иконка вкладки меняется с на .

Во вкладке **Выбор сигналов** (рис. 1) выберите файлы для сравнения в выпадающих списках **Файл 1** и **Файл 2** из списка открытых в проекте и нажмите кнопку **Далее**.

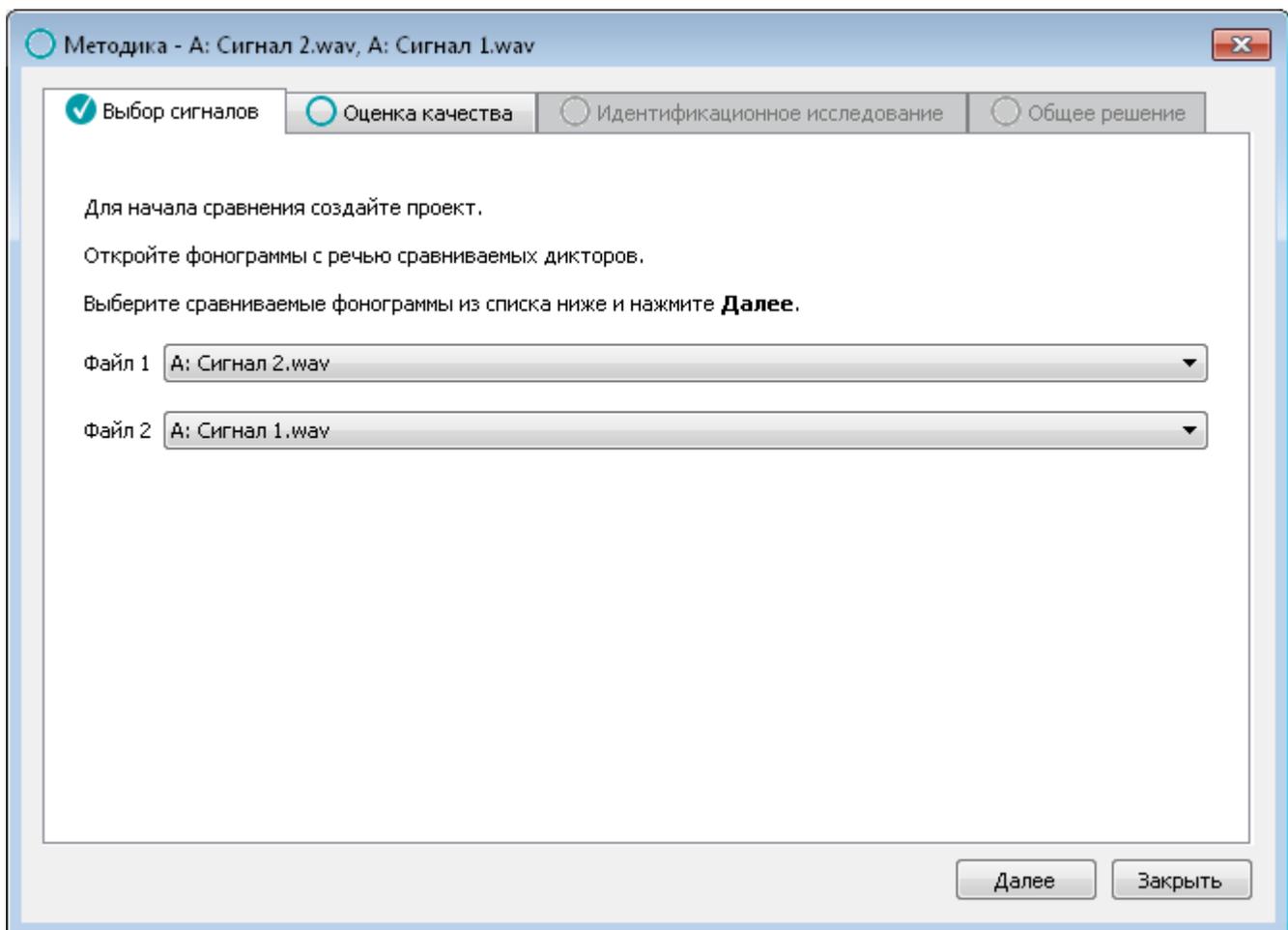


Рисунок 1 – Окно «Методика». Вкладка «Выбор сигналов»

Во вкладке **Оценка качества** (рис. 2) необходимо произвести автоматическую оценку качества каждой фонограммы.

Для этого нажмите кнопку **Начать**, расположенную под названием файла фонограммы.

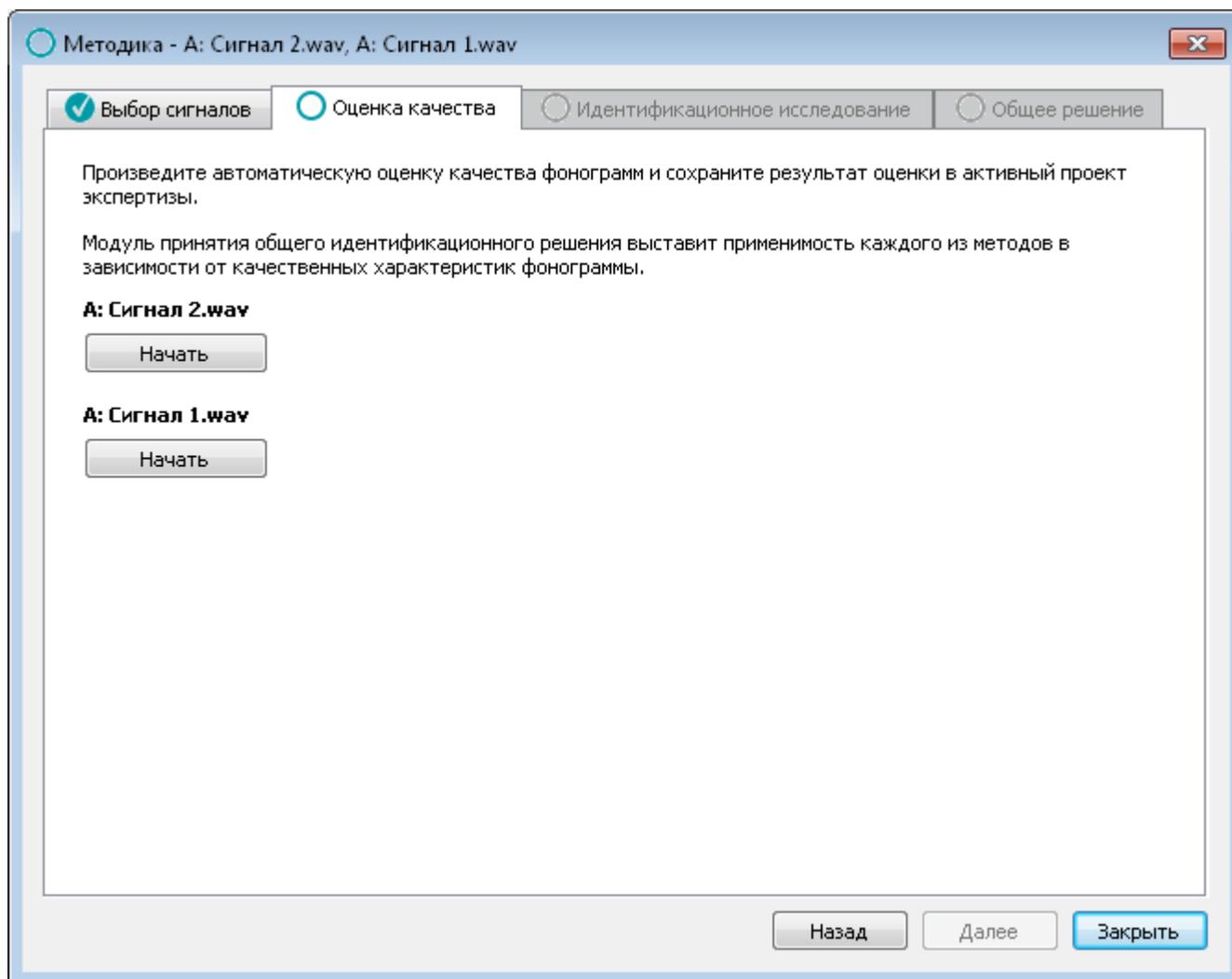


Рисунок 2 – Окно «Методика». Вкладка «Оценка качества»

В открывшемся окне **Характеристики сигнала 'Название_файла'** (рис. 3) задайте необходимые значения параметров и нажмите кнопку **Рассчитать**.

Подробное описание параметров представлено в п. 3.1 Расчёт характеристик сигнала.

Рисунок 3 – Окно «Характеристики сигнала»

После того как значения параметров будут рассчитаны, сохраните полученные данные, нажав кнопку **Сохранить в проект**.

Для закрытия окна нажмите кнопку **Заккрыть**.

Повторите процедуру расчета характеристик сигнала для второй фонограммы.

После завершения оценки параметров обеих фонограмм нажмите кнопку **Далее** (см. рис. 2) для перехода во вкладку **Идентификационное исследование**.

Во вкладке **Идентификационное исследование** (рис. 4) следует выполнить сравнение фонограмм с использованием представленных методов.

Для получения краткой справки о каждом методе нажмите кнопку  в соответствующей секции.

Для выполнения сравнения фонограмм с использованием определенного метода/модуля нажмите кнопку **Начать**, расположенную под названием искомого метода/модуля.

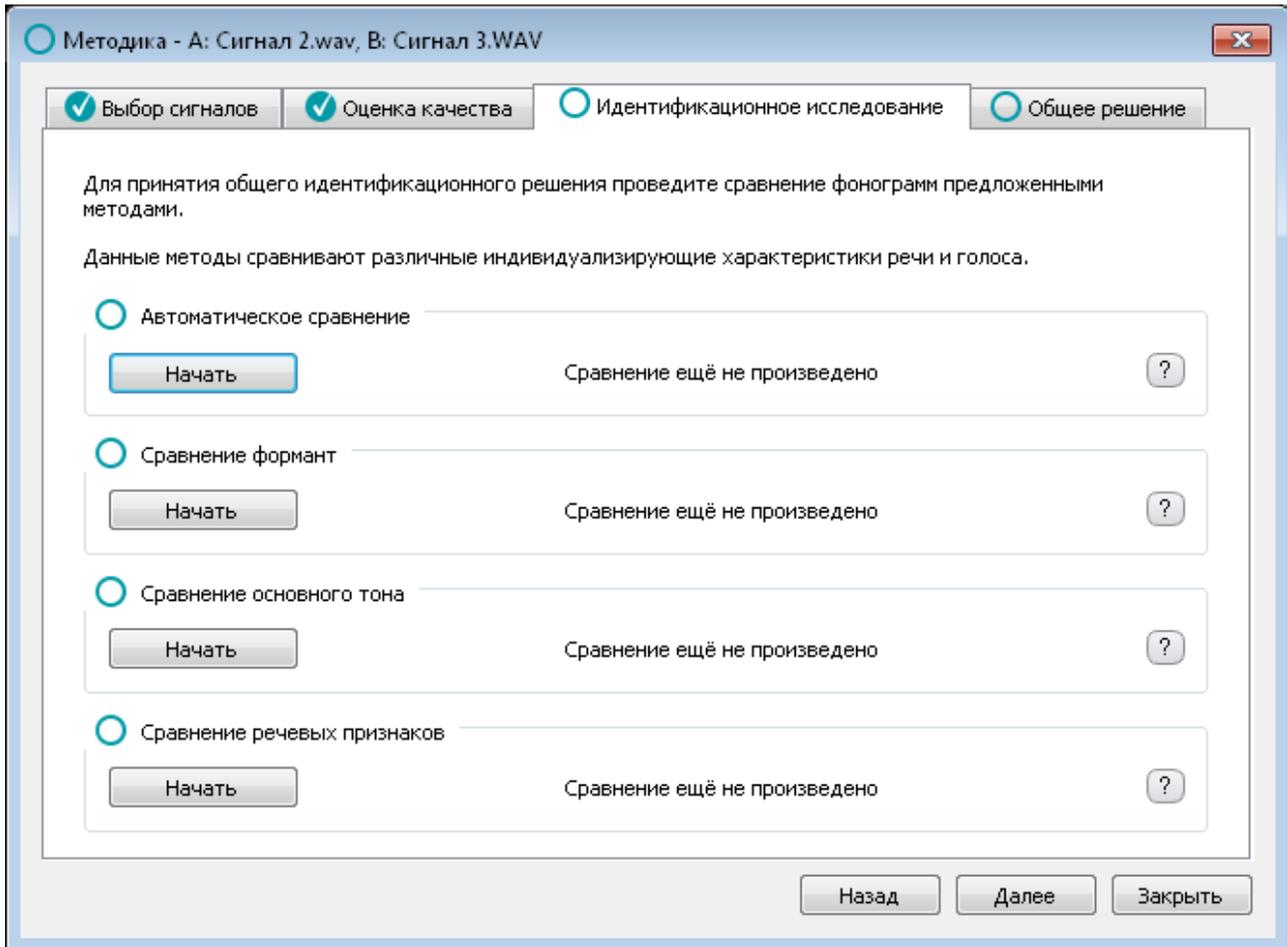


Рисунок 4 – Окно «Методика». Вкладка «Идентификационное исследование»

После нажатия кнопки **Начать** открывается окно настроек соответствующего модуля. Подробное описание работы модулей представлено далее.



Рекомендуется проводить сравнение фонограмм всеми представленными методами.

После того как получены результаты сравнения с использованием всех модулей нажмите кнопку **Далее** (рис. 5) для перехода во вкладку **Общее решение**.

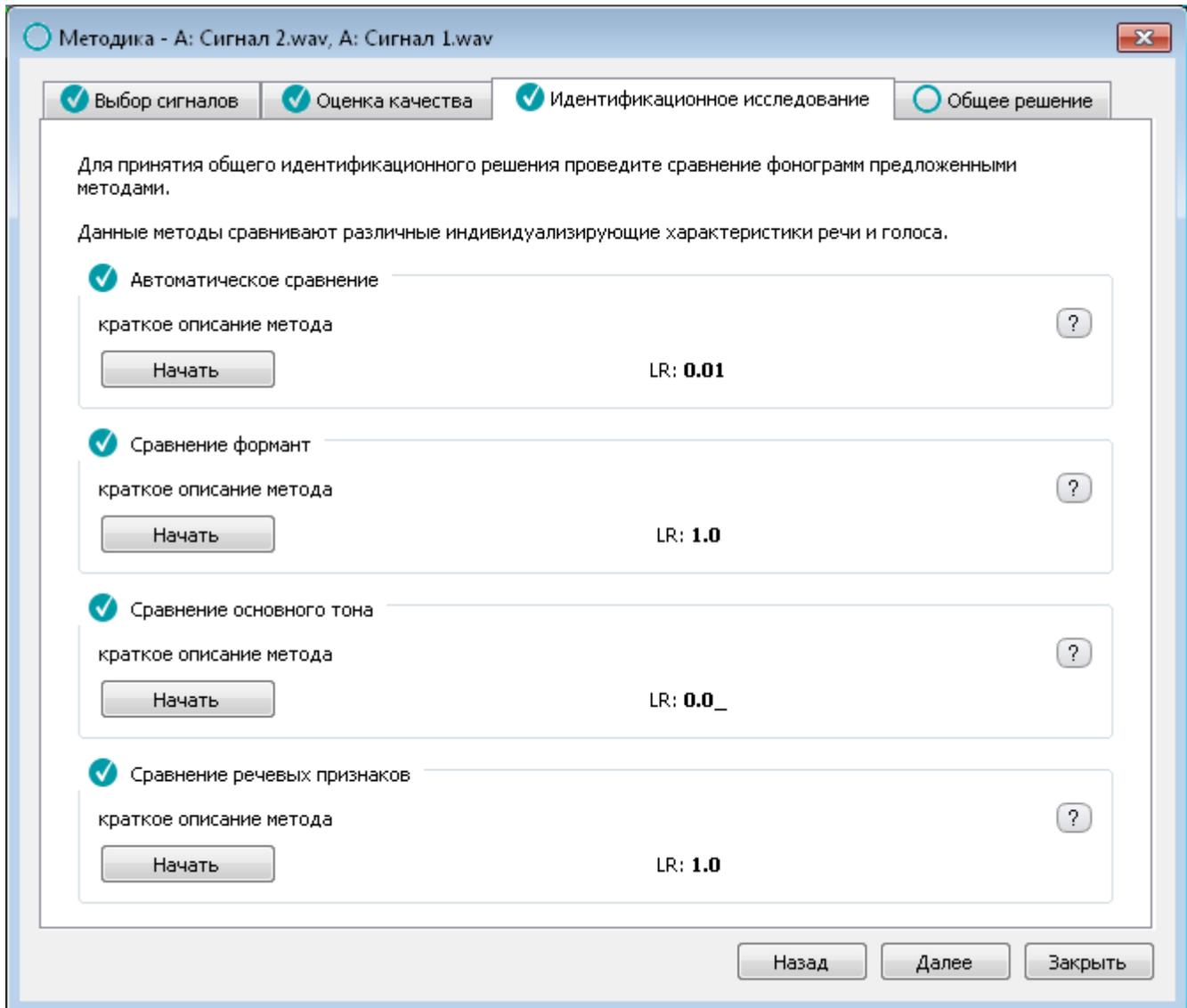


Рисунок 5 – Окно «Методика». Вкладка «Идентификационное исследование». Результаты сравнения

Во вкладке **Общее решение** (рис. 6) с использованием модуля «Общее решение» на основании результатов проведенных ранее сравнений принимается решение относительно идентификации личности по голосу. Для получения результатов нажмите кнопку **Начать**.

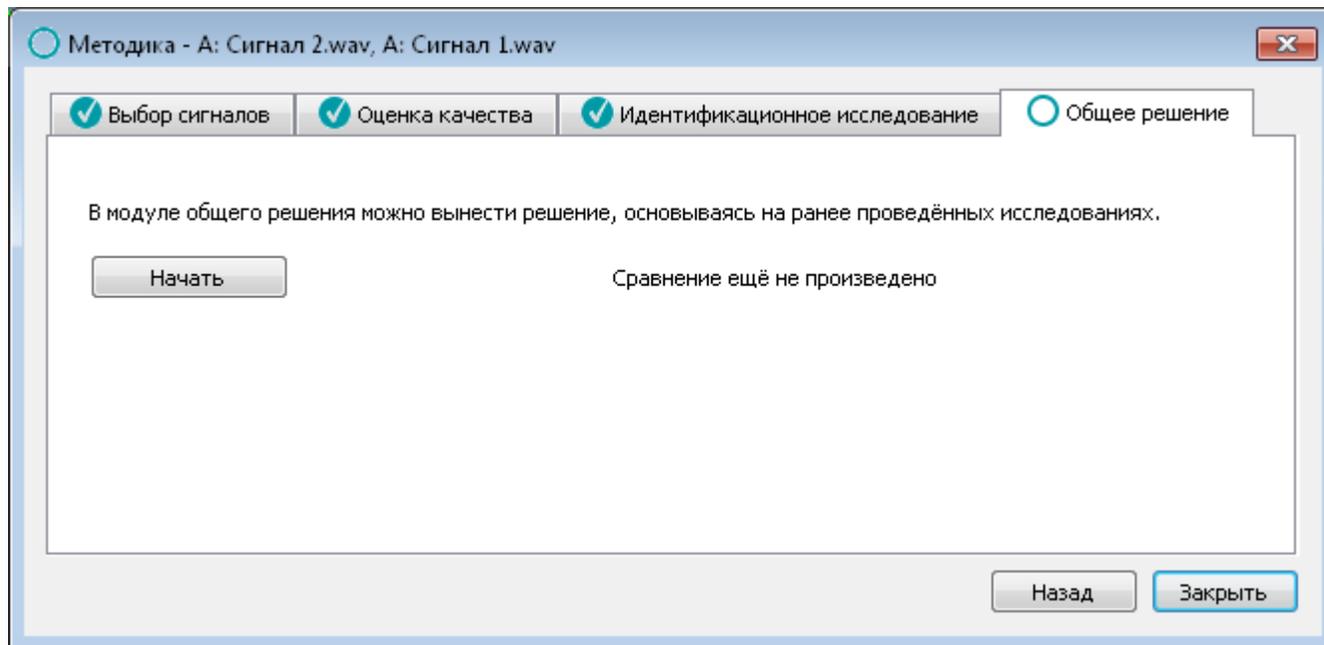


Рисунок 6 – Окно «Методика». Вкладка «Общее решение»

В открывшемся окне **Общее решение** (рис. 7) представлена таблица сравнений фонограмм с помощью различных методов. Результаты сравнения по каждому из них отражены в столбце **LR** (отношение правдоподобия).

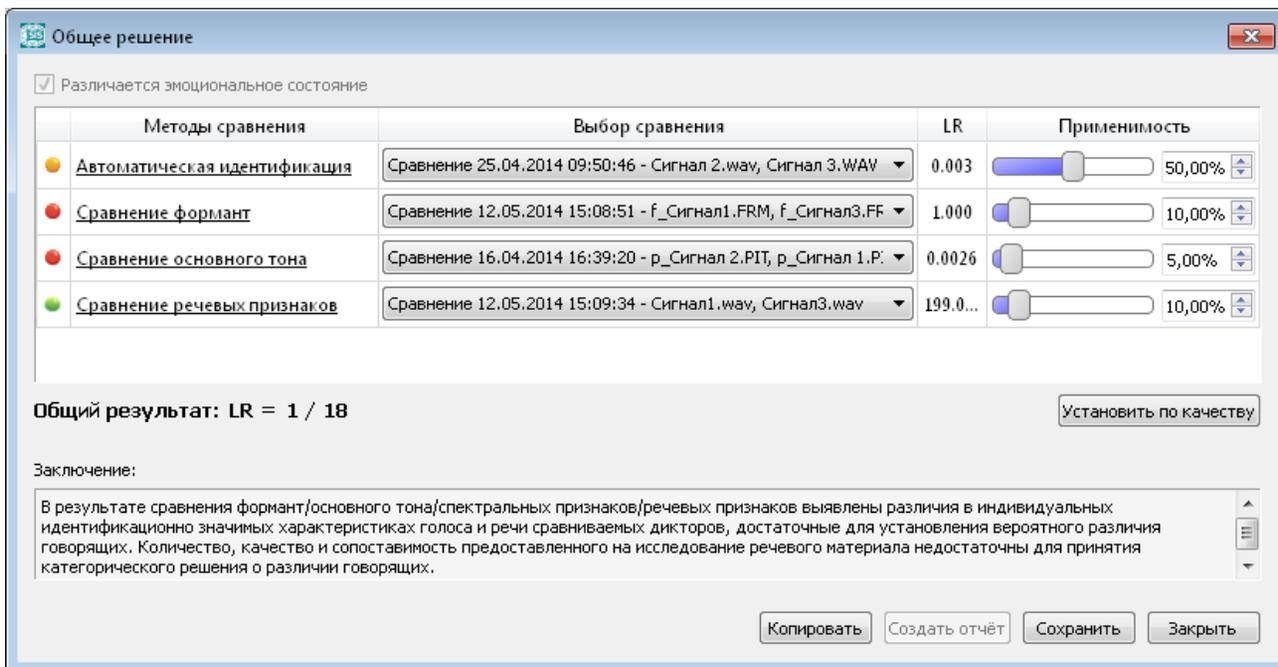


Рисунок 7 – Окно «Общее решение»

Подробное описание результатов идентификации представлено в п. 10 ПРИНЯТИЕ ОБЩЕГО ИДЕНТИФИКАЦИОННОГО РЕШЕНИЯ.

3 ОЦЕНКА ХАРАКТЕРИСТИК СИГНАЛА

Оценка характеристик сигнала позволяет определить пригодность фонограммы для экспертизы.

3.1 Расчёт характеристик сигнала

Чтобы оценить характеристики сигнала, выберите меню **Модули** → **Характеристики сигнала**.

В появившемся окне (рис. 8) указаны основные характеристики, определяющие качество оцениваемой фонограммы: амплитудно-частотная характеристика, отношение сигнал/шум, наличие или отсутствие тональных помех, среднее время реверберации и его дисперсия, наличие или отсутствие клиппирования, длительность чистой речи, канал и источник записи.

Характеристики сигнала 'Сигнал1'

Эмоциональное состояние: Нормальное

Спектральные характеристики

АЧХ: --- Гц [Построить график]

ОСШ: 100 --- 3300 Гц --- дБ [Построить график]

Тональные помехи: В сигнале не обнаружены стационарные тональные помехи

Время реверберации

Среднее время: --- мс Дисперсия: ---

Клиппирование: В сигнале не обнаружено клиппирование

Речь

Всего: --- с Всего: --- %

Обработка

Все данные: 206.80 секунд

Выделенный фрагмент: 0.0_секунд

Источник

Микрофон Телефон

Сохранить в проект Копировать **Рассчитать** Заккрыть

Рисунок 8 – Окно «Характеристики сигнала» до выполнения команды «Рассчитать»

Для расчёта характеристик сигнала укажите область обработки (**Все данные** либо **Выделенный фрагмент**), укажите эмоциональное состояние диктора и нажмите кнопку **Рассчитать**.

Процесс вычисления характеристик сигнала требует определённого времени и отображается в окне **Выполнение заданий** (рис. 9). Процесс можно прервать, если нажать кнопку  справа от индикатора хода выполнения задания.

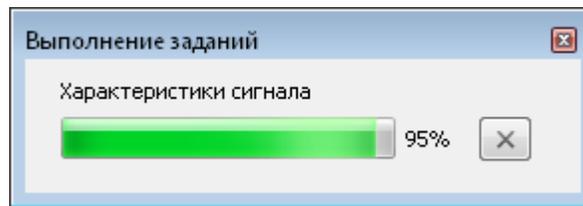


Рисунок 9 – Индикация хода выполнения задания

Окно **Характеристики сигнала** после выполнения команды **Рассчитать** представлено на рисунке 10. В зависимости от проанализированных характеристик фонограмма автоматически определяется как **пригодная**, **непригодная** или **ограниченно пригодная** для идентификации.

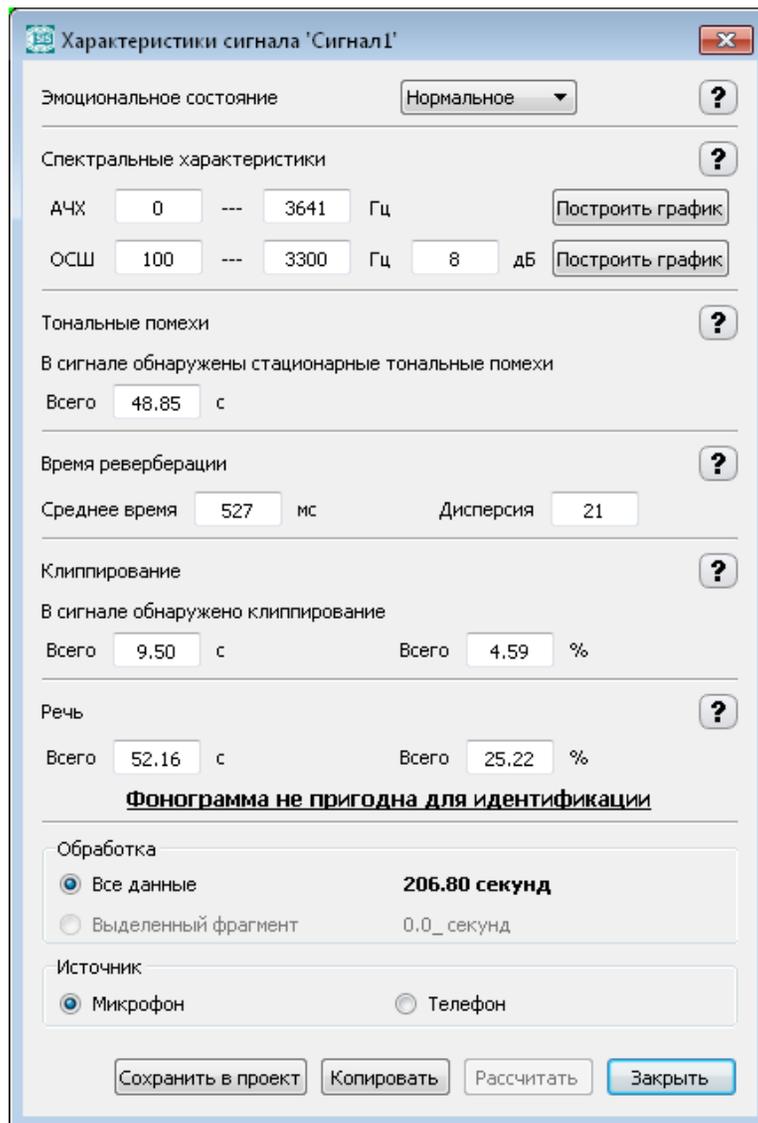


Рисунок 10 – Окно «Характеристики сигнала» после выполнения команды «Рассчитать»

3.2 Графики спектральных характеристик

Чтобы построить графики амплитудно-частотной характеристики (АЧХ) сигнала или отношения сигнал/шум (ОСШ), нажмите кнопку **Построить график** в строке АЧХ или ОСШ (см. рис. 10).

Для АЧХ строится спектр быстрого преобразования Фурье (БПФ). Пример графика представлен на рисунке 11. Пунктирными линиями выделены определённые модулем границы АЧХ.

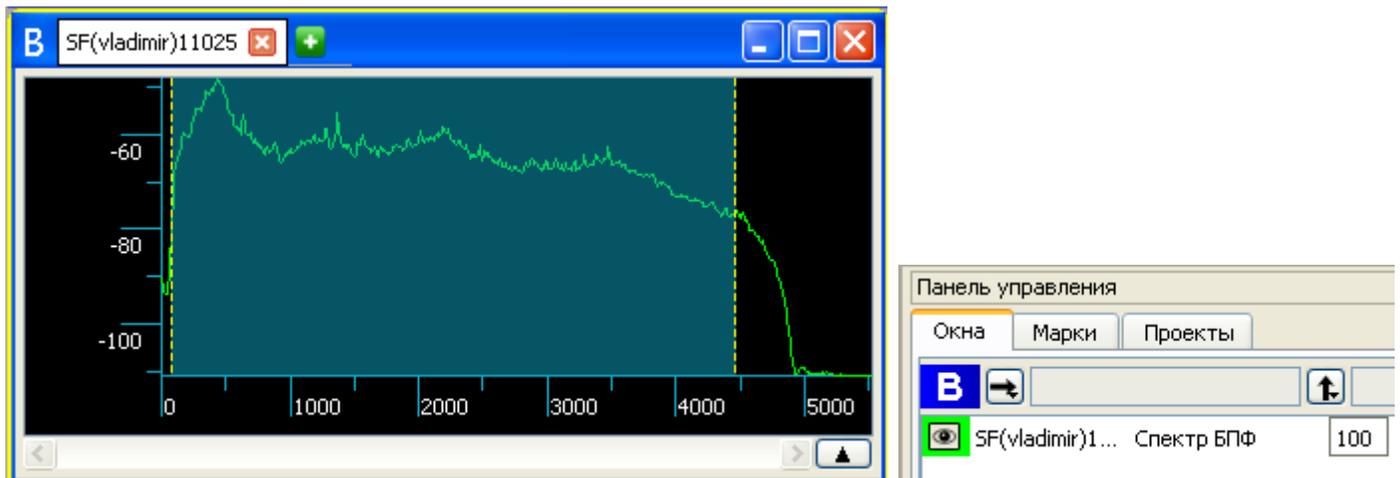


Рисунок 11 – Пример спектра БПФ для АЧХ

Для отношения сигнал/шум строится его частотное распределение (рис. 12).

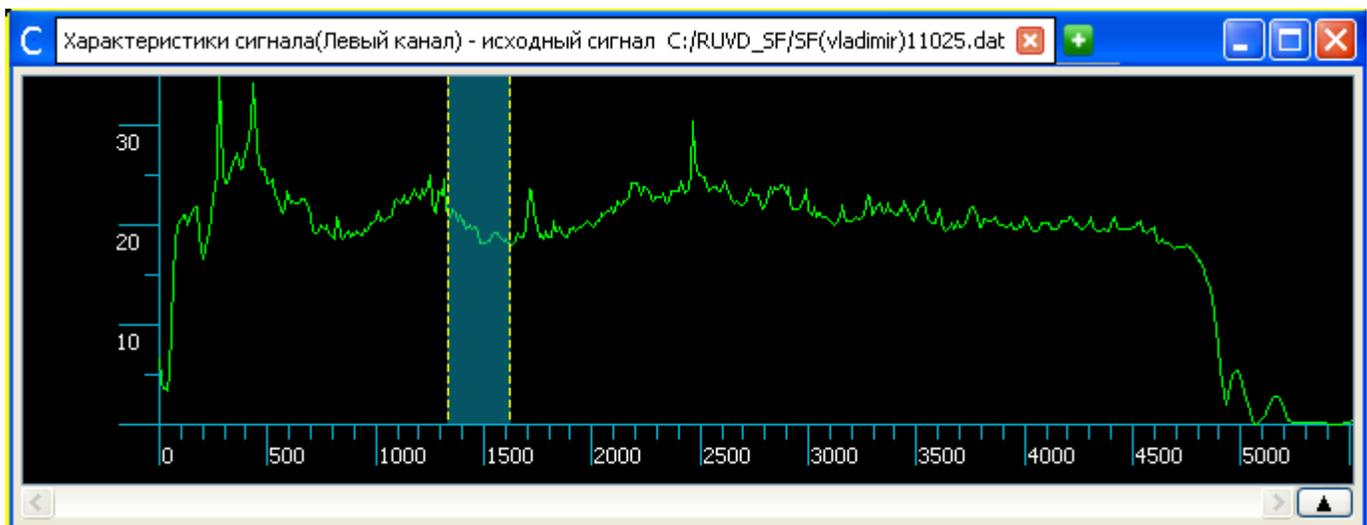


Рисунок 12 – Окно с графиком ОСШ и выделенным новым фрагментом

Пунктирными линиями выделены границы, в пределах которых вычислено интегральное значение ОСШ, приведённое в окне «Характеристики сигнала». По умолчанию интегральное значение ОСШ вычисляется в диапазоне от 100 до 3300 Гц. Если на графике ОСШ выделить другой фрагмент (рис. 12), то в окне «Характеристики сигнала» соответствующим образом изменяются значения частот, а интегральное значение ОСШ пересчитывается (рис. 13).

ОСШ 1337.19 --- 1628.28 Гц 20.06 дБ

Рисунок 13 – Выделенный частотный фрагмент и интегральное значение ОСШ для этого фрагмента

3.3 Оценка и рекомендации по полученным характеристикам сигнала

Чтобы получить рекомендации по возможности дальнейшего использования данного речевого сигнала при работе модулей идентификации дикторов для каждой группы оцениваемых характеристик, нажмите кнопку .

Пример рекомендации к характеристикам сигнала приведен на рисунке 14.

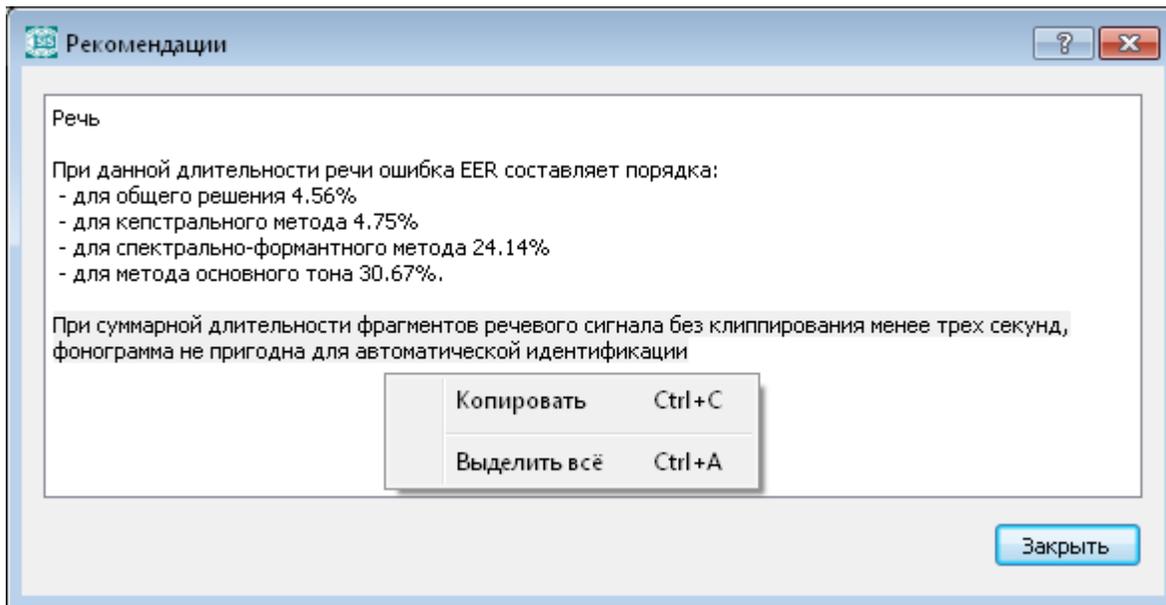
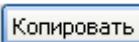


Рисунок 14 – Окно «Рекомендации» по оценке длительности речи с открытым контекстным меню

Текст, находящийся в окнах **Рекомендации** можно скопировать в буфер обмена для последующего размещения в отчётном документе. Для копирования текста служит контекстное меню, которое вызывается нажатием на правую кнопку мыши (рис. 14) или сочетание клавиш клавиатуры **Ctrl+C**.

Чтобы поместить вычисленные модулем характеристики сигнала в буфер обмена текстового редактора, нажмите кнопку  окна **Характеристики сигнала**.

В текстовом редакторе выполните операцию **Вставить**.

4 ДЕТЕКТИРОВАНИЕ РЕЧИ

Детектирование речи требуется для предварительного сегментирования фонограмм, поступающих на обработку в модуль идентификации дикторов. Дополнительно пользователь имеет возможность провести детектирование шумовых участков: служебные телефонные сигналы, щелчки, участки клиппирования.

4.1 Выполнение детектирования

Чтобы автоматически выделить речевые участки фонограммы, выберите в меню **Модули** → **Детектирование речи** и нажмите кнопку **Вычислить**.

Данный программный модуль выполняет поиск речевых участков, пригодных для автоматической идентификации дикторов, и помещает их в группу марок VAD (Voice Activity Detection, «Выделение голосовых участков»). Только эти участки сигнала далее анализируются при запуске модуля автоматической идентификации дикторов. Участки, отмеченные как шум, пауза или неречевые сигналы из рассмотрения исключаются. При необходимости выполненная модулем автоматическая сегментация сигнала может быть откорректирована вручную путем изменения положения границ участков разного типа или удаления/вставки тех или иных границ сегментов.



Подробная работа с марками интервалов представлена в документе «SIS II. Специальное программное обеспечение для идентификации говорящего на русском языке STC-S521 Руководство пользователя ЦБАУ.00432-01 94».

Данный программный модуль может осуществлять детектирование гудков, границ участков клиппирования и щелчков. Марки данных видов сигналов помещаются в общую группу марок – «Шумы».

Для детектирования гудков, участков клиппирования или щелчков, установите флажки в соответствующих позициях: **Детектировать гудки**, **Детектировать клиппирование**, **Детектировать щелчки**.

В диалоговом окне **Детектирование речи** (рис. 15) можно изменить временной интервал между зашумлёнными участками, при котором они будут объединяться в один.

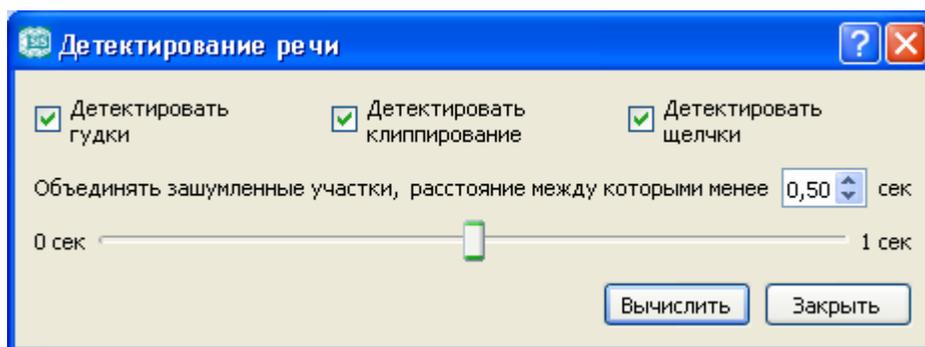


Рисунок 15 – Окно «Детектирование речи»

4.2 Результаты детектирования

После выполнения операции детектирования в окне данных появятся марки интервалов, отмечающие участки фонограммы, содержащие речь диктора и шумы. В списке марок во вкладках **VAD** и **Шумы** будут перечислены интервалы, относящихся к данным группам (рис. 16).

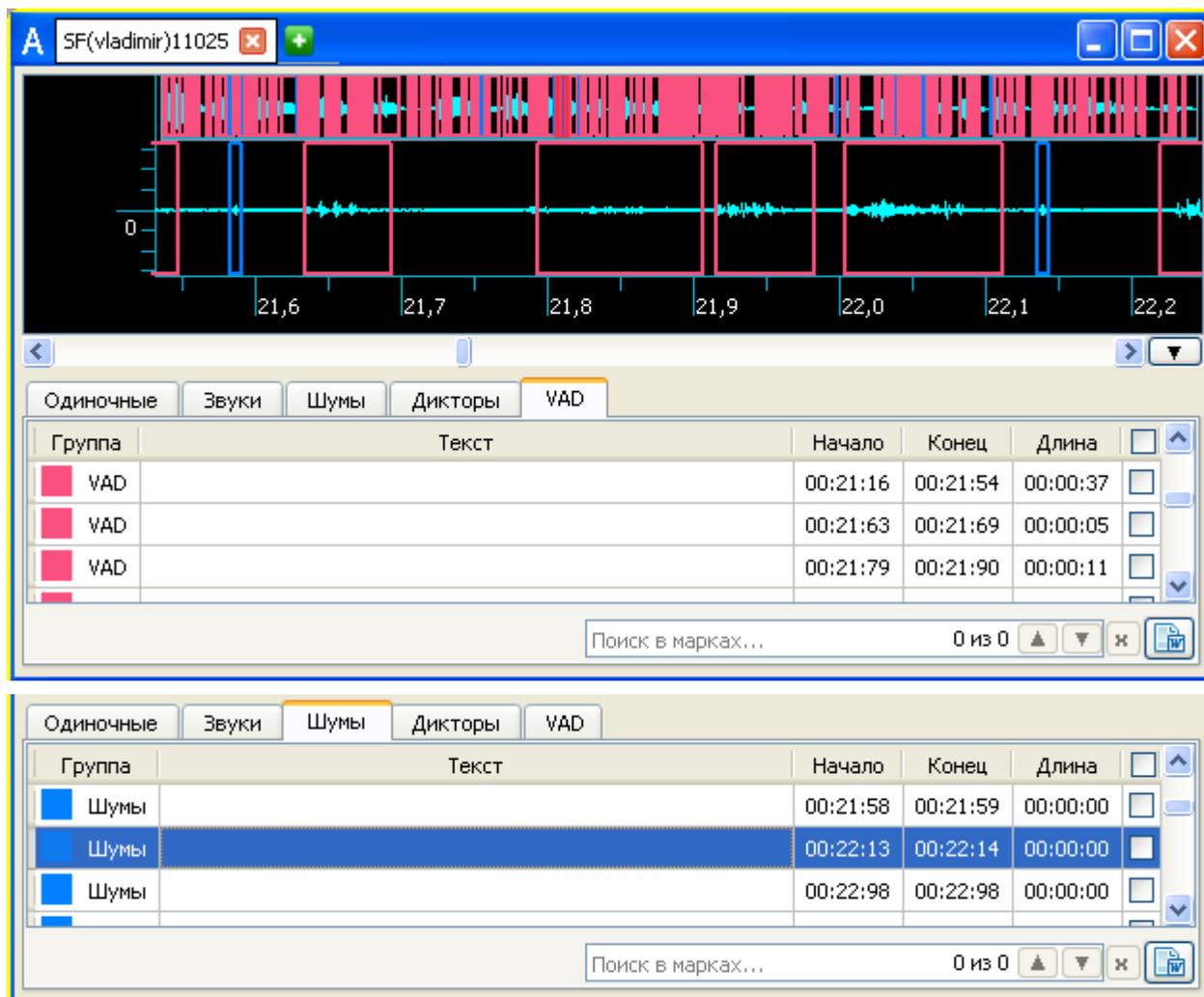


Рисунок 16 – Окно данных, а также вкладки «Шумы» и «VAD» после выполнения процесса детектирования

Общие количество марок и общую их длину можно получить, если выбрать группы **VAD** или **Шумы** на вкладке **Марки** панели управления (рис. 17).

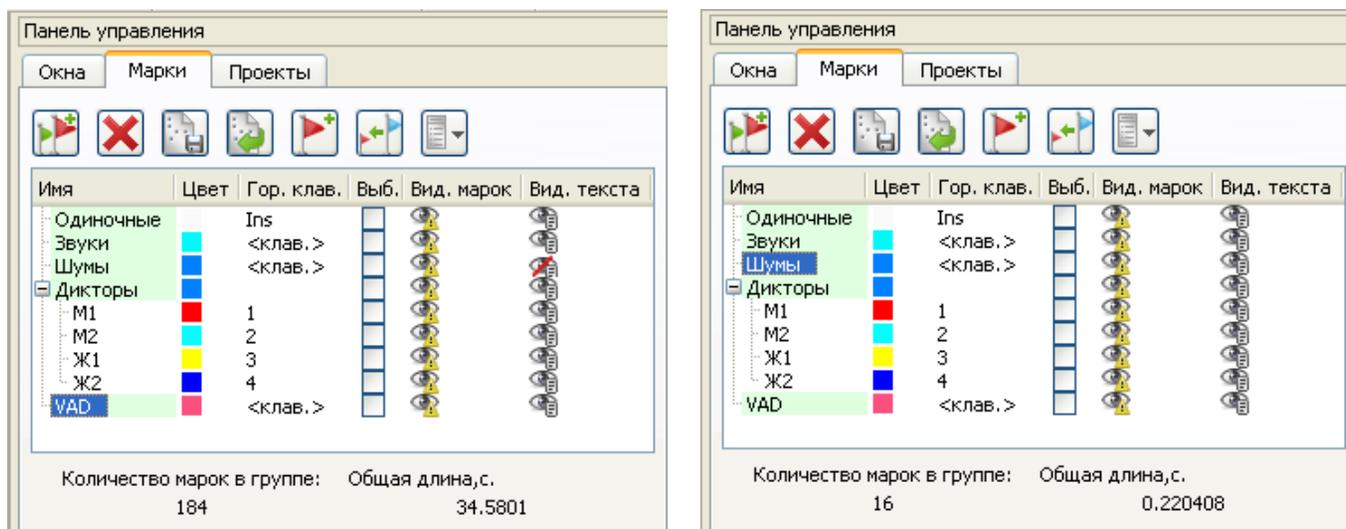


Рисунок 17 – Марки, отмечающие участки фонограммы, содержащие речь диктора и шумы

5 РАЗДЕЛЕНИЕ РЕЧИ ДИКТОРОВ В ДИАЛОГЕ

Модуль разделения дикторов в диалоге предназначен для автоматической маркировки речевых сегментов по дикторам на записях естественной речи.

Надежность работы модуля разделения дикторов в значительной степени зависит от степени качества речи входных фонограмм. Максимальное качество разделения речи дикторов (93-95% правильной маркировки по дикторам) достигается при ОСШ не менее 20 дБ и длительности речи каждого из дикторов на фонограмме не менее 16 секунд.

5.1 Выполнение разделения речи дикторов

Чтобы выполнить автоматическую маркировку речи дикторов в диалоге по репликам, выберите в меню **Модули** → **Разделение дикторов в диалоге**.

В диалоговом окне **Разделение дикторов** (рис. 18) установите максимальный промежуток времени между речевыми участками одного диктора, при котором данные участки будут считаться единым фрагментом.

Задайте параметры **Имена марок диктора 1** и **Имена марок диктора 2**, которыми будут обозначаться фрагменты речи каждого диктора.

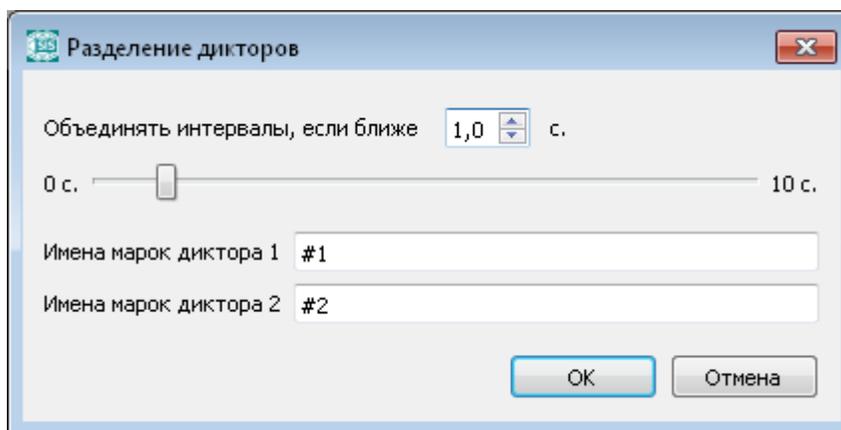


Рисунок 18 – Окно «Разделение дикторов»

Для запуска процесса разделения речи дикторов нажмите кнопку **OK**.

5.2 Результаты разделения речи дикторов

После выполнения операции разделения речи дикторов в окне данных появятся марки интервалов, отмечающие участки фонограммы, содержащие речь двух дикторов. В списке марок во вкладках с именами марок для двух дикторов будут перечислены интервалы, относящихся к данным группам (рис. 16).

Оператор может откорректировать интервалы вручную средствами **SIS II**.

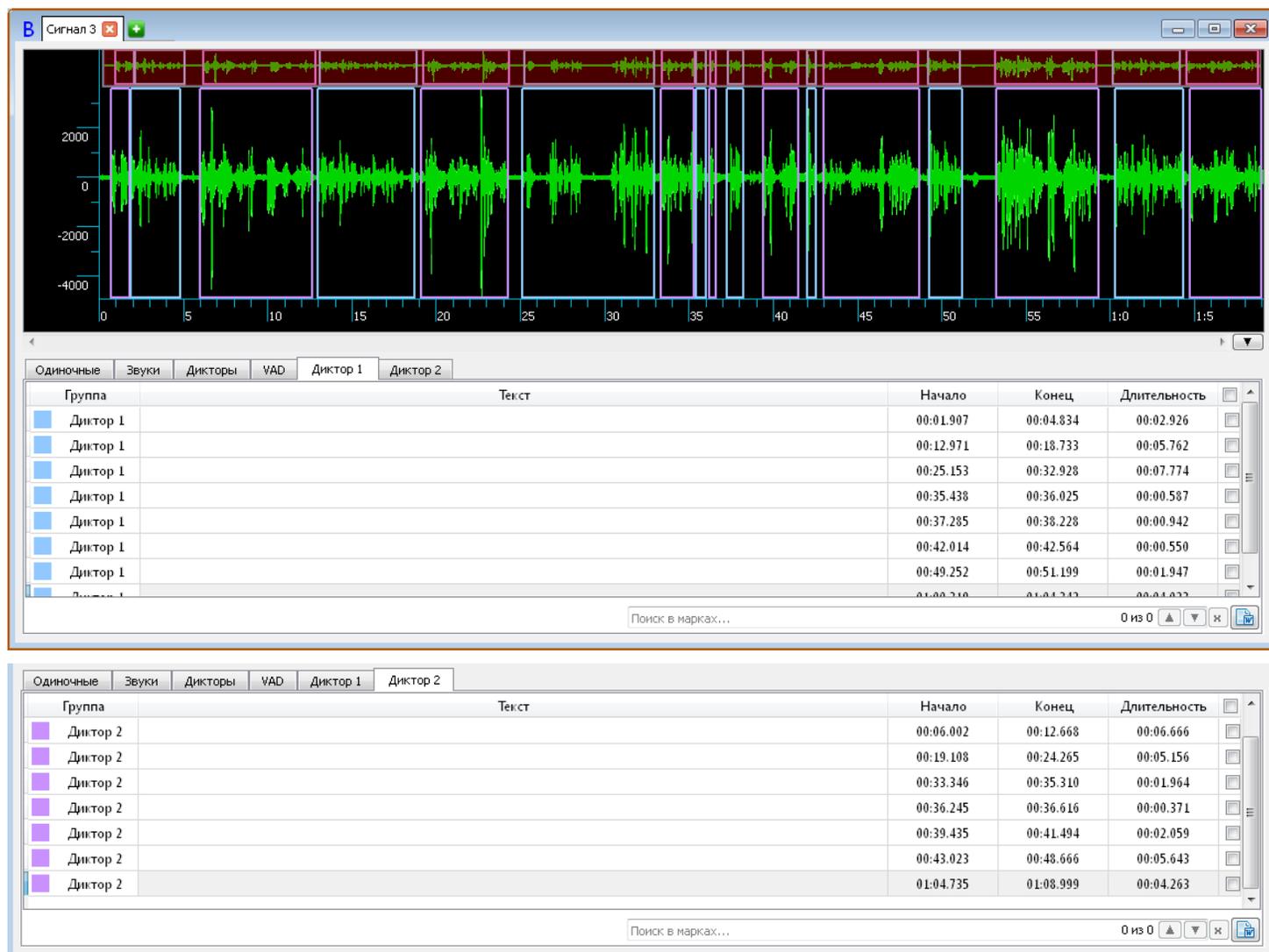


Рисунок 19 – Окно данных и вкладки списка фрагментов речи дикторов



Использование модуля на фонограммах, в которых записана речь одного или более двух дикторов не предусмотрено настройками алгоритма.

Общие количество марок и общую их длину можно получить, если выбрать группы с названием меток для каждого диктора на вкладке **Марки** панели управления (рис. 17).

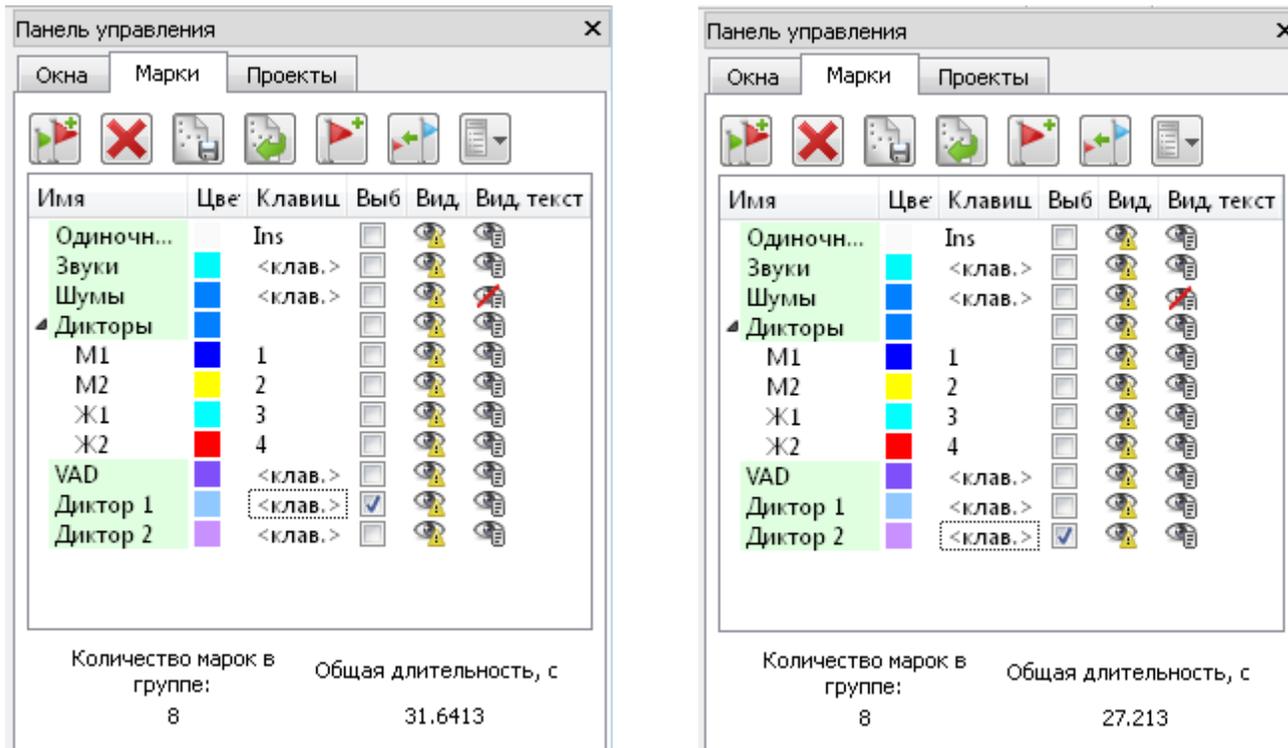


Рисунок 20 – Марки, отмечающие участки фонограммы, содержащие речь каждого диктора

5.3 Коррекция разделения речи дикторов

В том случае, если разделение речи дикторов выполнено неточно, т.е. границы фрагментов речи на осциллограмме не соответствуют их реальному положению, существует возможность проведения коррекции.

Для сдвига одной из границ интервала наведите курсор мыши на эту границу, нажмите на клавиатуре клавишу **Shift** и, удерживая нажатыми клавишу **Shift** и левую кнопку мыши, переместите границу парной марки в нужное место.

Для сдвига всего интервала наведите курсор мыши на одну из границ интервала, нажмите на клавиатуре клавишу **Ctrl** и удерживая нажатыми клавишу **Ctrl** и левую кнопку мыши, переместите марку интервала в нужное место.

Чтобы удалить марку интервала, воспользуйтесь одним из способов:

- 1) Наведите курсор мыши на одну из границ удаляемой марки и нажмите клавиши **Ctrl+Delete**.
- 2) Наведите курсор мыши на одну из границ удаляемой марки, нажмите правую кнопку мыши и в появившемся контекстном меню выберите пункт **Удалить марку**.



Подробная работа с марками интервалов представлена в документе «SIS II. Специальное программное обеспечение для идентификации говорящего на русском языке STC-S521 Руководство пользователя ЦБАУ.00432-01 94».

6 АВТОМАТИЧЕСКОЕ СРАВНЕНИЕ

Модуль «Автоматическое сравнение» осуществляет автоматическую идентификацию и сравнение дикторов на основе следующих методов:

- 1) Метод статистик основного тона (ОТ).
- 2) Спектрально-формантный метод (СФ).
- 3) Метод полной изменчивости (СГР).

Идентификация диктора на основе первых трёх методов включает два основных этапа: построение соответствующих методу моделей речевого сигнала диктора и сравнение построенных моделей с принятием идентификационного решения.

По результатам использования автоматических методов идентификации принимается обобщенное решение и его результат является наиболее точным. Обобщенное решение является взвешенным усреднённым результатом идентификации по каждому из использованных частных методов, указанных выше.

6.1 Метод идентификации дикторов на основе сравнения статистик основного тона голоса

Данный метод использует шестнадцать различных статистических характеристик кривой основного тона (ОТ) голоса для всего исследуемого речевого материала: среднее значение ОТ, максимальное, минимальное значения ОТ, максимум -3% значений ОТ, минимум +1% значений ОТ, процент участков с возрастающим тоном, дисперсия логарифма ОТ, медиана, асимметрия и эксцесс гистограммы логарифма тона и гистограммы производной логарифма ОТ, средняя скорость изменения, возрастания и уменьшения тона, дисперсия производной логарифма тона.

Значение равновероятностной ошибки пропуска/отклонения целевого диктора (EER) для метода сравнения статистик основного тона голоса в зависимости от длительности сравниваемых речевых фрагментов составляет 16–20%.

Реализация данного алгоритма стала возможной благодаря созданию специалистами ООО «ЦРТ» полностью автоматического высокоточного алгоритма выделения основного тона.

Характеристики статистик ОТ голоса зависят от эмоционально-психологического и физического состояния диктора в момент произнесения, от стиля речи (чтение – спонтанная речь – публичное выступление/подготовленная речь), от окружающей акустической обстановки (Ломбард-эффект). Поэтому при сравнении дикторов важно учитывать сопоставимость речевого материала сравниваемых дикторов и, при наличии факторов, существенно влияющих на изменение ОТ, учитывать это при принятии общего идентификационного решения.

Построение модели голоса диктора производится следующим образом.

Выделение основного тона (ОТ) осуществляется с помощью спектрального анализа сигнала фонограммы. Основу метода составляет алгоритм, который базируется на анализе значений гармоник ОТ в спектре сигнала. Алгоритм направлен на преодоление проблем, связанных с подавлением части спектра сигнала в канале.

Например, в телефонном канале в полосе от 0 до 300 Гц сигнал отсутствует, однако именно в этом диапазоне чаще всего находится значение частоты ОТ. В этом случае алгоритм выделения ОТ позволяет решить поставленную задачу за счёт учёта гармоник частоты ОТ в полосе от 0,3 до 3,4 кГц.

Данный метод идентификации достаточно устойчив и к отношению сигнал/шум. Если уровень гармоник частоты основного тона хотя бы на несколько децибел превышает уровень шума, алгоритм с помощью специальных методов «подчеркивания» гармоник ОТ позволяет определить искомую частоту.

На этапе вычисления статистических характеристик ОТ вычисляется набор значений статистик ОТ по полученным кривым ОТ, который содержит такие характеристики как среднее, максимальное и минимальное значения ОТ; значения частоты, выше которой находятся 3 % значений ОТ, и частоты, ниже которой находятся 1 % значений ОТ; медиана; процент участков с возрастающим ОТ; дисперсия; асимметрия и эксцесс гистограммы логарифма ОТ и гистограммы производной логарифма ОТ; средняя скорость изменения ОТ, усредненная скорость изменения, нарастания и убывания частоты ОТ. По получаемым статистическим оценкам характеристик ОТ строится модель диктора с использованием аппроксимации данных с использованием Смеси Гауссовых Распределений (СГР).

На этапе принятия решения об идентификации производится сравнение статистических характеристик ОТ, вычисляются значения ошибок первого и второго рода (FA и FR) по каждому из полученных значений статистик ОТ, а также значения FA и FR по результирующей метрике статистик ОТ как взвешенная сумма всех относительных отклонений значений статистик ОТ для двух сравниваемых фонограмм. Производится бэббирование по длительности речевого сигнала (оценки FR/FA зависят от длительности речевого сигнала, по которому построена каждая из сравниваемых СГР-моделей).

6.2 Спектрально-формантный метод идентификации дикторов

Данный метод основан на тезисе об уникальности геометрии вокального тракта у каждого человека и отражении данного факта в различных спектральных характеристиках речи. Наиболее явно различие спектральных характеристик проявляется во взаимном расположении формант в частотной области. Данный метод основан и на выделении и сравнении положения и динамики поведения трёх и более формант (защищён российским патентом).

Применение спектрально-формантного метода обеспечивает значение величины EER до 6–7 %. Значение этого показателя для конкретного случая зависит от длительности и качества сравниваемых речевых фрагментов.

Спектрально-формантный метод довольно устойчив к широкополосному равномерному шуму типа белого или шума помещения, а также к различиям в сигналах за счет различных типов каналов звукозаписи.

Построение модели голоса диктора производится следующим образом.

Строится средний спектр речевого сигнала, используемого для построения идентификационной модели диктора. Далее на его основе для данной фонограммы строится нормализующая функция, используемая для вычисления мгновенных спектров (кадров) речевого сигнала. На каждом нормализованном кадре определяется положение всех максимумов, из которых выделяются формантные

частоты. Учитывается относительная непрерывность траектории формант и их типовые частотные области расположения каждой из формант для мужских и женских голосов.

В качестве идентификационных признаков используются положения 3-х спектральных максимумов, наиболее соответствующих формантам речевого сигнала на каждом спектральном срезе, где их удалось достаточно надёжно выделить. Эти признаки в наибольшей степени определяют индивидуальные особенности геометрии голосового тракта.

Плотность распределения идентификационных признаков моделируется с использованием СГР.

Непосредственное принятие решения об идентификации производится на основе использования SVM-классификатора. По результирующей дистанции SVM вычисляются ошибки первого/второго рода FR/FA. Производится бэггирование по длительности речевого сигнала (оценки FR/FA зависят от длительности речевого сигнала, по которому построена каждая из сравниваемых СГР-моделей).

6.3 Метод полной изменчивости

Метод, названный в данной системе СГР или TotV (Total Variability, «полная изменчивость»), является наиболее современным методом идентификации по голосу. Построение модели голоса диктора производится следующим образом.

Вычисляются речевые признаки MFCC (Mel Fourier Cepstrum coefficients, кепстральные коэффициенты по шкале мел) – характеристики, описывающие покадровое спектральное представление речевого сигнала на тональных участках речи. Плотность распределения идентификационных признаков моделируется с использованием смеси гауссовых распределений (СГР).

Параметры СГР-модели далее с помощью специально адаптированного факторного анализа представляются в виде вектора малой размерности в так называемом пространстве «полной изменчивости», в котором строятся подпространства так называемых «собственных каналов» и «собственных голосов», используемых в алгоритме JFA (Joint Factor Analysis, объединённый факторный анализ). Отличительной особенностью такого представления СГР-модели является её высокая информативность и малый размер данных.

На этапе сравнения моделей дикторов соответствующие им вектора признаков сравниваются с помощью алгоритма на основе SVM-классификатора (Support Vector Machine, машина опорных векторов), хорошо зарекомендовавшего себя с точки зрения скорости и качества идентификационного решения. По результирующей дистанции SVM вычисляются ошибки первого/второго рода FR/FA. Производится бэггирование по длительности речевого сигнала (оценки FR/FA зависят от длительности речевого сигнала, по которому построена каждая из сравниваемых СГР-моделей). Таблицы для определения вероятностей совпадения, различия дикторов по полученной дистанции между сравниваемыми сигналами получаются на этапе обучения системы путем обработки большого массива фонограмм совпадающих и разных дикторов (тысячи дикторов, различные каналы звукозаписи).

6.4 Построение общего идентификационного решения

Обобщённый метод идентификации дикторов вычисляет общее идентификационное решение по результатам, полученным одним или несколькими частными методами идентификации вне зависимости от их количества. Результатом является псевдовероятность P принадлежности двух сравниваемых фонограмм одному диктору, лежащая в диапазоне от 0 до 100 %. 0 % – обозначает минимально возможное сходство дикторов на фонограммах, 100 % – полное соответствие голосов дикторов.

Алгоритм расчёта обобщённого решения основан на использовании метода взвешенного голосования:

$$P = \sum_{i=1}^N w_i \cdot f(FR_i, FA_i)$$

где N – количество методов идентификации, по которым получено решение;

w_i – весовой коэффициент i -го метода. Значение весового коэффициента автоматически меняется в зависимости от качества сигнала;

FR_i / FA_i – ошибки первого/второго рода i -го метода. Величины FR_i / FA_i исчисляются в процентном выражении и имеют диапазон значений от 0 до 100 %.

$f(FR_i, FA_i) = [(Fri - Fai)/2 - 50]$ – функция псевдовероятности. Псевдовероятностью называется мера сходства голоса на спорной фонограмме и целевой модели, не являющаяся точной математической вероятностью, но обладающая похожими с вероятностью свойствами.

6.5 Выполнение сравнения

Для автоматической идентификации диктора, выберите пункт **Автоматическое сравнение** в меню **Модули**.

В диалоговом окне **Автоматическое сравнение** (рис. 21) выберите файлы для сравнения в выпадающих списках **Файл 1** и **Файл 2**, выберите источник сигнала (**Микрофон** или **Телефон**) и нажмите кнопку **Сравнить**.

Сравнение на основе метода ОТ возможно только в случае совпадения эмоционального состояния.

Для выбора скорректированного основного тона для сравнения установите флаг **Использовать для модели ОТ**.

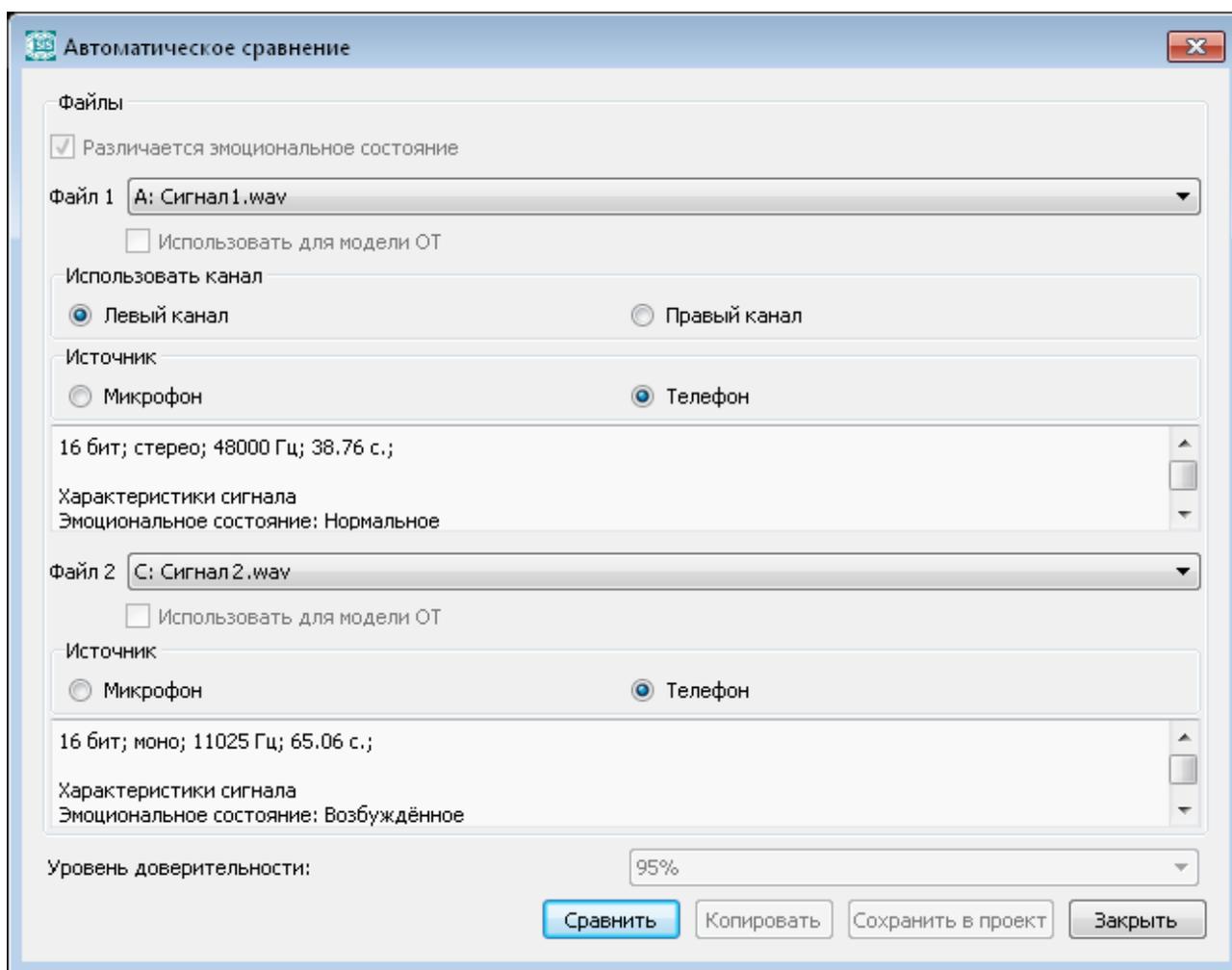


Рисунок 21 – Пример окна «Автоматическое сравнение» до начала выполнения операции сравнения

Для отмены операции идентификации нажмите кнопку **Заккрыть**.

Перед началом работы с данным модулем рекомендуется запустить модуль **Детектирования речи** для каждого сигнала, предполагаемого для дальнейшей процедуры автоматической идентификации. Данный модуль проводит автоматическую сегментацию сигналов на участки речи, пауз и шума. Обычно целесообразно проверить и при необходимости откорректировать вручную результаты работы этого модуля.

При сегментации все шумы и паузы будут исключаться из речевого сигнала аналогично тому, как это происходит в модуле детектирования речи (см. рис. 15) при установке всех флажков на его панели управления. Результат детектирования шумов за пределы модуля не выдаётся и в данном модуле не может быть проверен и откорректирован оператором. В случае слишком короткого или низкокачественного сигнала чистой речи может оказаться недостаточно для идентификации.

В этом случае целесообразно просмотреть и по возможности откорректировать результаты сегментации речи с использованием модуля «Детектирование речи».

Процесс идентификации требует определённого времени для построения моделей и отображается в окне **Выполнение заданий** и в заголовке окна **Автоматическое сравнение**.

Процесс можно прервать, если нажать кнопку  справа от индикатора хода выполнения задания в окне **Выполнение заданий**.

6.6 Результаты идентификации

После окончания идентификации результаты отобразятся в таблице окна **Автоматическое сравнение** (рис. 22).

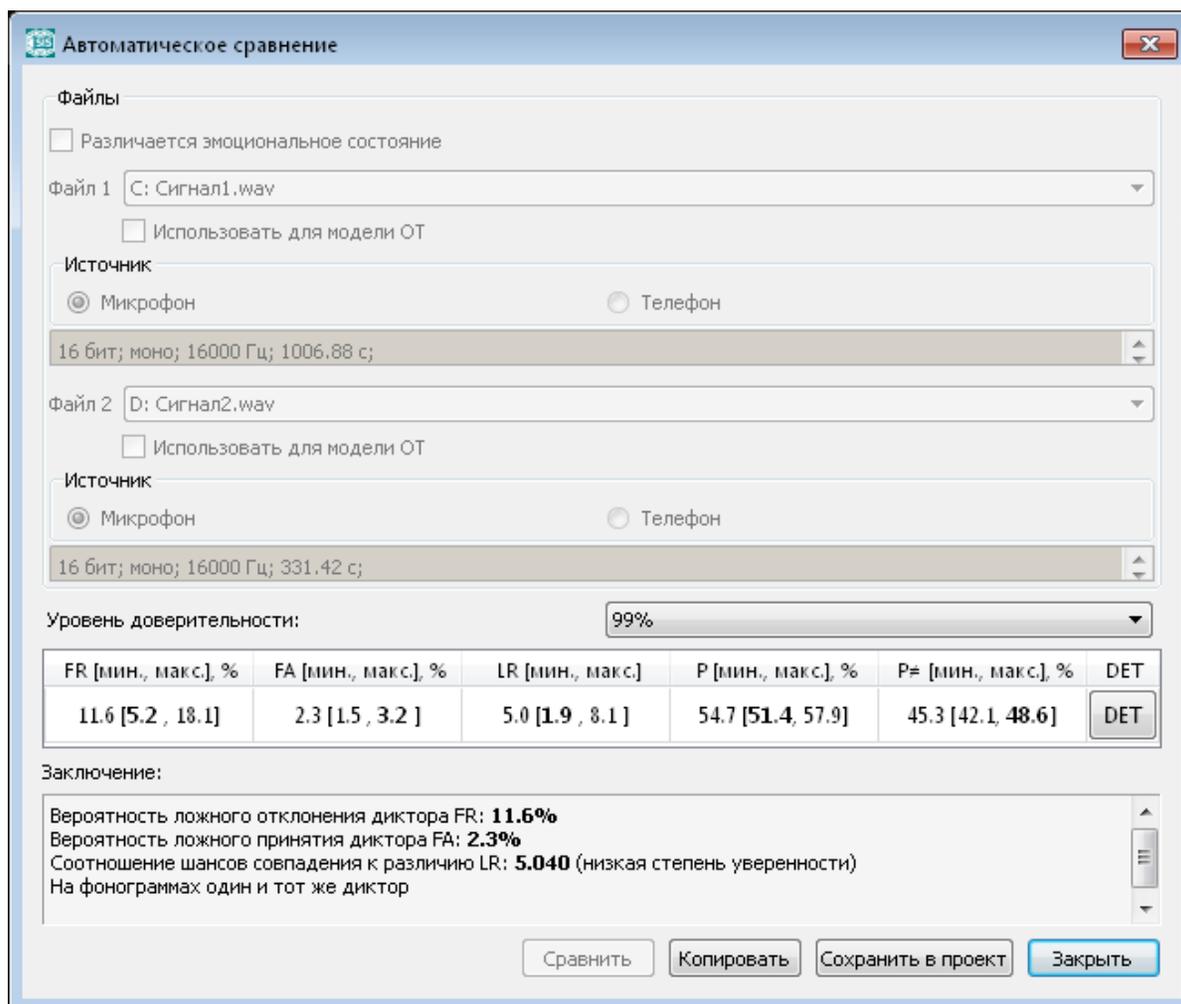


Рисунок 22 – Пример окна «Автоматическое сравнение» после выполнения операции сравнения

Для получения подробных результатов для различных методов наведите курсор на поле результатов, после чего отобразится всплывающее окно (рис. 23).

Уровень достоверности: 99%						
FR [мин., макс.], %	FA [мин., макс.], %	LR [мин., макс.]	P [мин., макс.], %	P≠ [мин., макс.], %	DET	
11.6 [5.2, 18.1]	2.3 [1.5, 3.2]	5.0 [1.9, 8.1]	54.7 [51.4, 57.9]	45.3 [42.1, 48.6]	DET	
Методы	FR [мин., макс.], %	FA [мин., макс.], %	LR [мин., макс.]	P [мин., макс.], %	P≠ [мин., макс.], %	
<input checked="" type="checkbox"/> СФ	46.6 [36.5, 56.6]	9.4 [7.7, 11.0]	5.0 [3.7, 6.3]	68.6 [63.5, 73.7]	31.4 [26.3, 36.5]	
<input checked="" type="checkbox"/> ОТ	80.1 [72.0, 88.1]	13.6 [11.7, 15.6]	5.9 [4.9, 6.8]	83.2 [79.1, 87.3]	16.8 [12.7, 20.9]	
<input checked="" type="checkbox"/> СГР	5.4 [0.9, 10.0]	6.6 [5.2, 8.0]	0.8 [0.1, 1.5]	49.4 [47.0, 51.8]	50.6 [48.2, 53.0]	

Рисунок 23 – Результаты операции сравнения для различных методов

Все результаты идентификации носят вероятностный характер, и поэтому для каждой характеристики имеется некоторый диапазон, куда при обнаруженной мере близости между сравниваемыми сигналами попадает ее значение при выбранном уровне статистической достоверности, который задается в выпадающем списке **Уровень достоверности**. В таблице результатов приводятся средние и возможные при данном уровне достоверности максимальные и минимальные значения величин. Пояснения относительно достоверных интервалов приводятся в приложении Б. Соответствующие принятому идентификационному решению значения каждой характеристики отмечены жирным шрифтом. Например, при установленном тождестве дикторов внутри интервалов возможных значений для FR и P выбираются (являются результатом идентификации) минимальные значения, а для остальных характеристик – максимальные.

В данной таблице:

FR – False Rejection («ложное отклонение»), интегральная вероятность ложного принятия совпадающих дикторов за различающихся.

FA – False Acceptance («ложное принятие»), интегральная вероятность ложного принятия различающихся дикторов за совпадающих.

LR – Likelihood Ratio («отношение правдоподобия»), отношение шансов (вероятности) совпадения дикторов к шансам (вероятности) различия дикторов.

P – псевдовероятность совпадения дикторов.

P≠ – псевдовероятность различия дикторов.

Результаты сравнения можно скопировать в буфер обмена, нажав на кнопку **Копировать**, и вставить в отчёт в текстовом редакторе. Также результаты сравнения могут быть сохранены в текущий проект. Для этого нажмите кнопку **Сохранить в проект**.



Чтобы сохранить результаты автоматического сравнения в проект, в программе должен быть, по крайней мере, один активный проект.

Для получения графиков FR-FA и DET (рис. 24) по каждому из методов нажмите кнопку **DET** в строке данного метода.

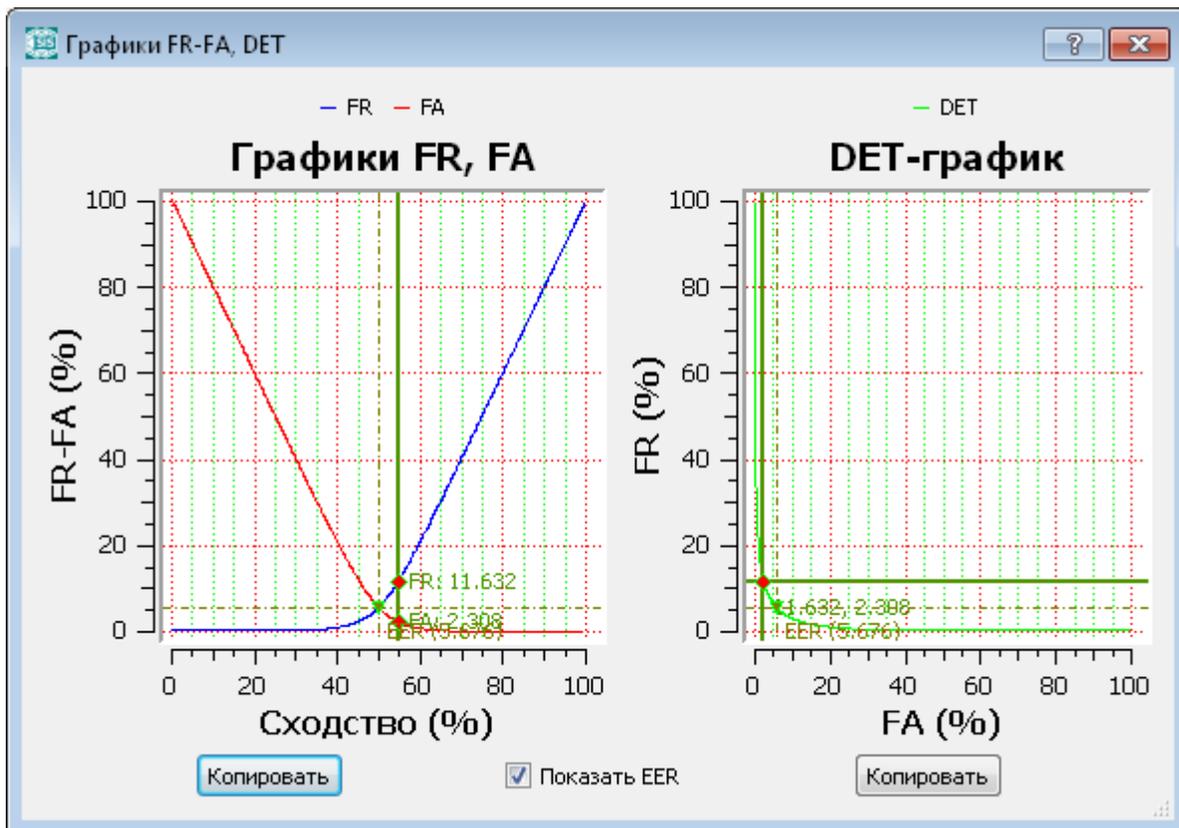


Рисунок 24 – Примеры графиков FR-FA и DET

Установка флажка **Показать EER** отображает или убирает на графиках значение EER.

Кнопки **Копировать** позволяют копировать в буфер обмена каждый из графиков, чтобы затем вставить их в отчёт в текстовом редакторе.

7 СРАВНЕНИЕ ФОРМАНТ

Модуль «Сравнение формант» позволяет автоматизировать различные методы сравнения формантных характеристик речевого сигнала. Возможно сравнение формантных частот как для совпадающих фонем (фонетико-спектральный подход, микроанализ гласных в сопоставимых трезвучиях) так и для одинаковых мгновенных артикуляций (конфигураций речевого тракта) в рамках произвольного фонетического контекста. Анализ соотношений формантных частот на сопоставимых опорных фрагментах речевого сигнала может применяться для идентификации диктора, говорящего на языке, не знакомом эксперту, а также для сравнения разноязычных фонограмм.

7.1 Подготовка к работе с модулем



Перед запуском модуля «Сравнение формант» должен быть задан активный проект, а также подготовлены как минимум два окна с данными типа «форманты» (с вычисленными формантами).

Надёжность метода сравнения формант в существенной степени зависит от того, насколько корректно построены формантные траектории. Автоматическое определение формантных траекторий осуществляется в **SIS II** путём применения модуля «Выделение формант методом БПФ» (см. раздел 12 документа «SIS II. Специальное программное обеспечение для идентификации говорящего на русском языке STC-S521 Руководство пользователя ЦБАУ.00432-01 94»). Имеется возможность поправить вычисленные форманты вручную (см. раздел 8.3 там же), а также обозначить траектории формант самостоятельно, без включения модуля выделения формант.

Стандартно процедура построения формант и их коррекции состоит из следующих шагов:

- Вычисление среднего спектра исследуемого сигнала.
- При наличии существенных неоднородностей среднего спектра, в частности, «завалов» на низких (<500Гц) и высоких (>3400Гц) частотах, – выравнивание спектра путем построения в окне спектра инверсного фильтра (рис. 20).

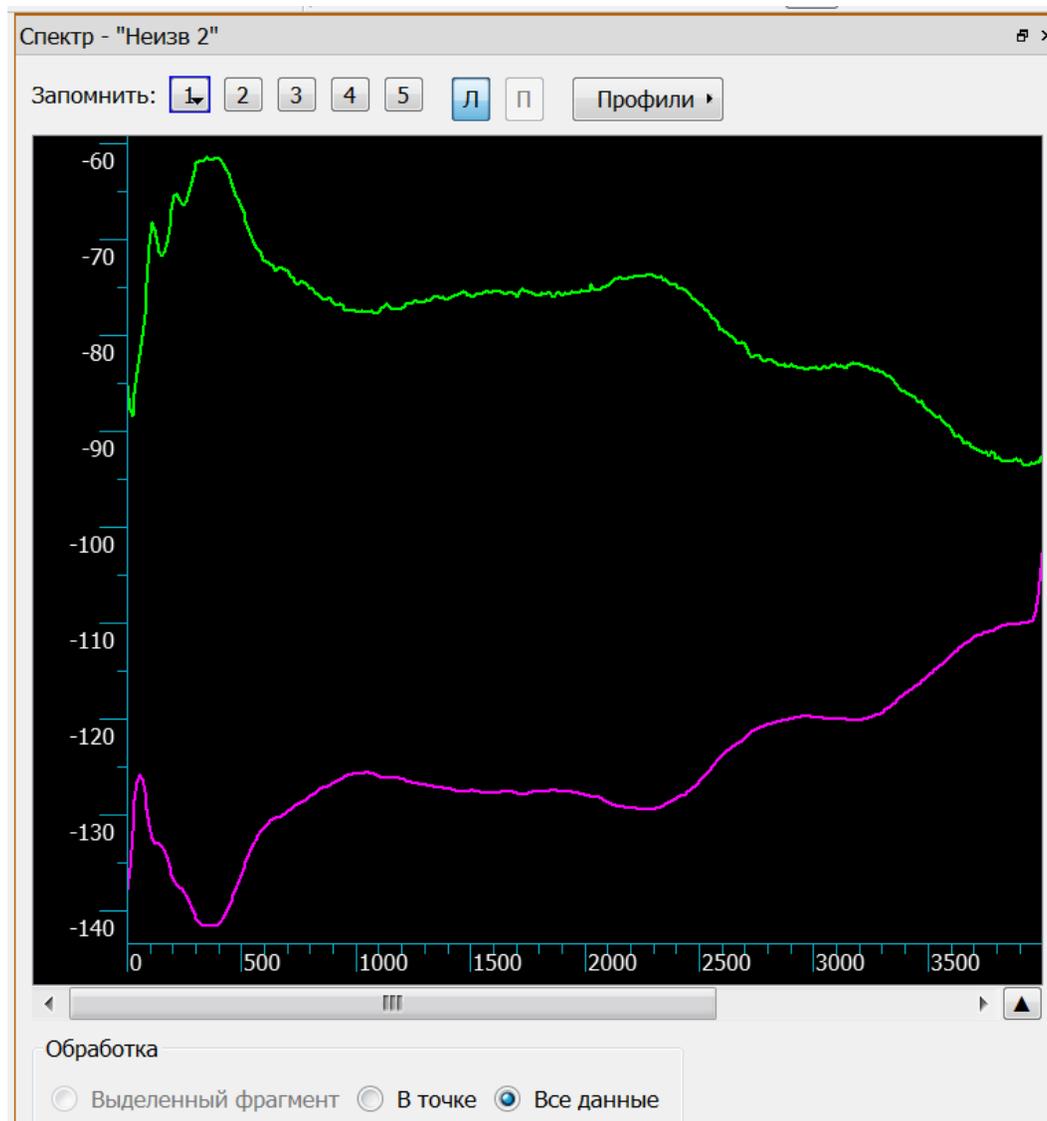


Рисунок 25 – Выравнивание спектра

Для этого необходимо щелчком правой клавиши мыши в окне спектра вызвать контекстное меню и в нем выбрать пункт **Преобразовать в фильтр → инверсный**. Далее, активизировав окно с осциллограммой сигнала, необходимо выбрать пункт меню **Обработка → Применить фильтр → Все данные → Применить**. Полученный после этого сигнал отнормировать по амплитуде и использовать для построения спектрограммы.

- Вычисление спектрограммы исследуемого сигнала. Чаще всего целесообразно использовать установки параметров вычисления спектрограмм через профиль **Форманты**, либо в режиме вычисления спектрограммы БПФ (3D БПФ), либо спектрограммы КЛП (3D КЛП). Полученное изображение оптимизировать по яркости. Спектрограмму КЛП — оптимизировать по числу КЛП (обычно от 9 до 80). Подъем высоких частот обычно следует установить в 0 позицию.
- Вычисление для текущего сигнала формант (для этого необходимо активизировать окно осциллограммы). Выбор установки реального типа источника звукозаписи **Микрофон/телефон** не всегда приводит к оптимальному выделению формантных траекторий. Рекомендуется попробовать оба типа установки и выбрать лучший вариант.

- Корректировка положения форматных траекторий вручную. Если речевой материал короткий (не более 10 секунд речи), то целесообразно проверить и откорректировать форманты на всем речевом материале, во всяком случае, на всех гласных звуках и глайдах. Для относительно длительного материала обычно имеет смысл корректировать форманты на конкретных отдельных участках, или только на наиболее четко произнесенных ударных гласных, а остальной материал – по мере необходимости.

Желательно выделение четырех низкочастотных формант, особенно для мужских голосов. Для женских голосов в телефонном канале 4-ая форманта часто не обнаруживается, однако и в этом случае целесообразно попробовать найти, пусть даже редкие, участки сигнала с ее появлением.

К ключевым факторам, влияющим на эффективность метода на основе анализа соотношений формантных частот на опорных фрагментах речевого сигнала, относятся:

1. Количество опорных фрагментов с хорошо читаемой формантной структурой для разных типов звуков (не менее трех различных типов гласных).

2. На сравниваемых фрагментах должно быть выделено не менее 3 (желательно – более 3) формант.

Чтобы применить метод сравнения формант, выберите меню **Модули** → **Сравнение формант**.

В диалоговом окне (рис. 26) укажите пару сравниваемых сигналов с уже вычисленными формантами. Если их в программе всего два, модуль выбирает их автоматически. Если сигналов больше двух, то сравниваемые сигналы выбираются оператором из выпадающих списков.

Окна, в которых находятся выбранные сигналы, будут автоматически связаны по вертикали.

Кроме того, если ранее в данном проекте уже выполнялись сравнения с выбранным сигналом, то таблицы фрагментов для этого сигнала можно загрузить из ранее сохранённых сравнений, воспользовавшись выпадающим списком **загрузить из**.

Выберите один из предустановленных профилей обработки с помощью выпадающего списка. Профиль представляет собой готовую таблицу значений границ форматных частот для разных типовых случаев: мужские и женские голоса для русского языка, таблицы для английского и испанского языков, а также параметры принятия решения о сходстве/различия дикторов.

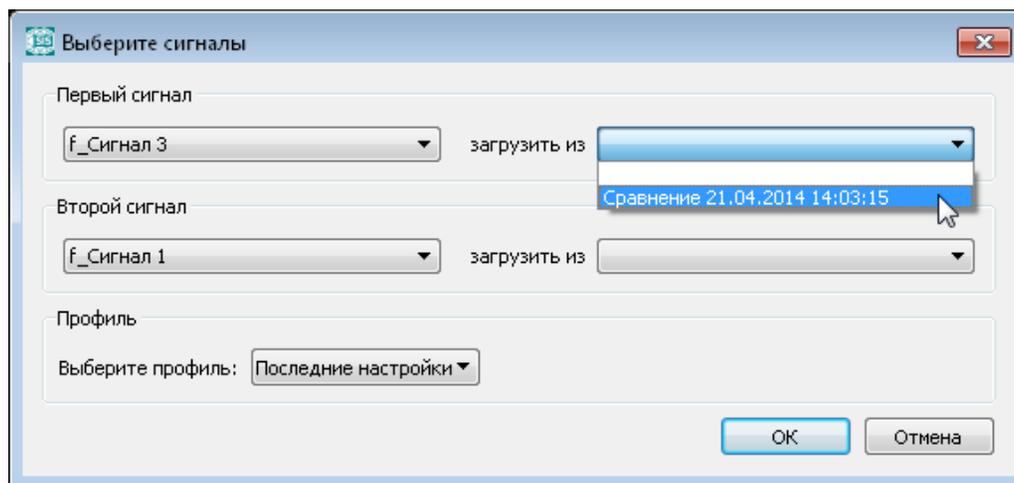


Рисунок 26 – Выбор сравниваемых сигналов

После выбора сравниваемых сигналов нажмите кнопку **ОК**.

Откроется панель работы с модулем. Панель прикрепляется к нижней части главного окна **SIS II**, при этом открытые внутри него окна с данными, панели управления или окно спектра пропорционально меняют свой размер (рис. 27).

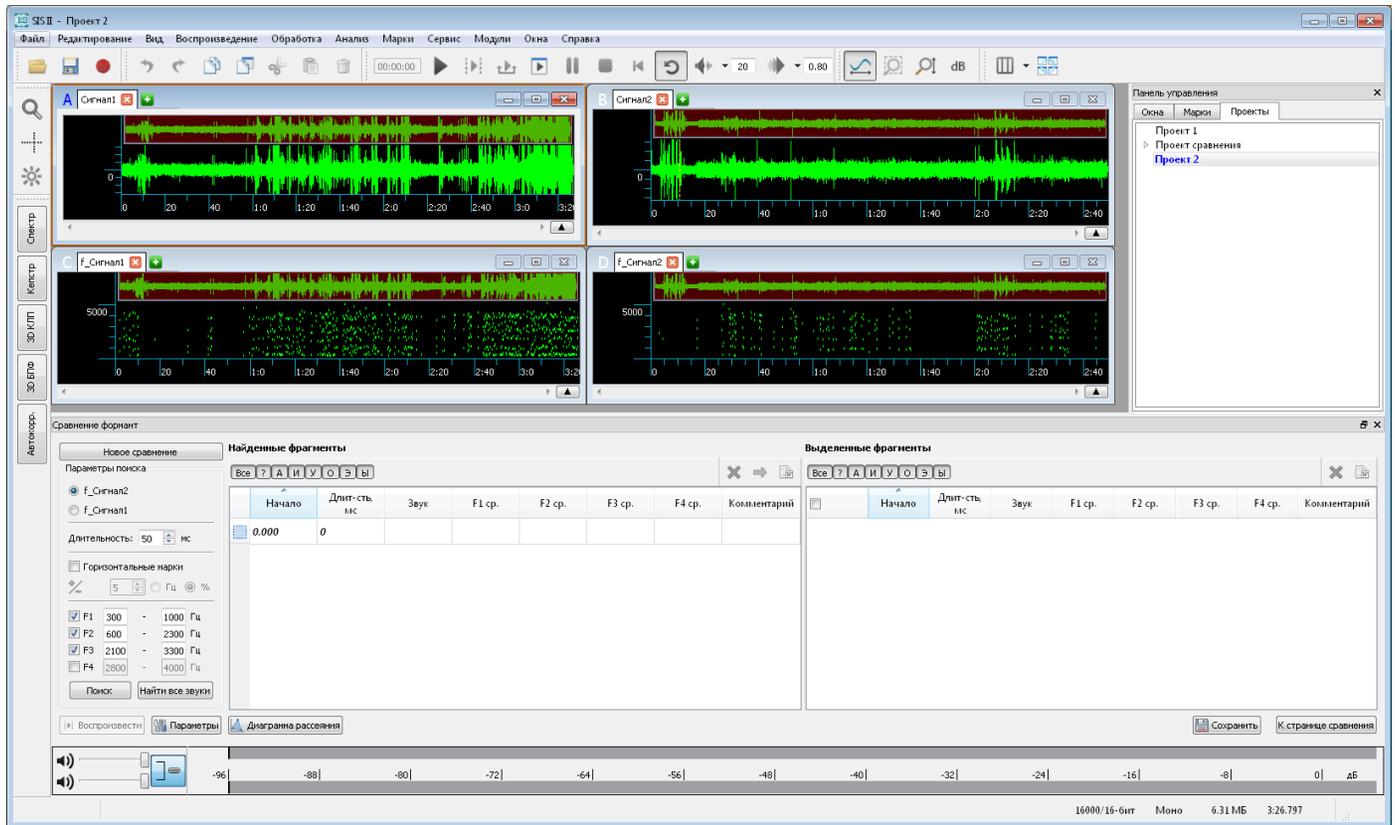


Рисунок 27 – Панель модуля формантного выравнивания внутри главного окна звукового редактора

Чтобы сделать панель модуля независимой от главного окна **SIS II** (рис. 28), следует нажать на значок  в правом верхнем углу панели или дважды щёлкнуть левой кнопкой мыши на заголовке панели. Панель можно перемещать в любое место на экране монитора, удерживая нажатой на заголовке левую кнопку мыши.

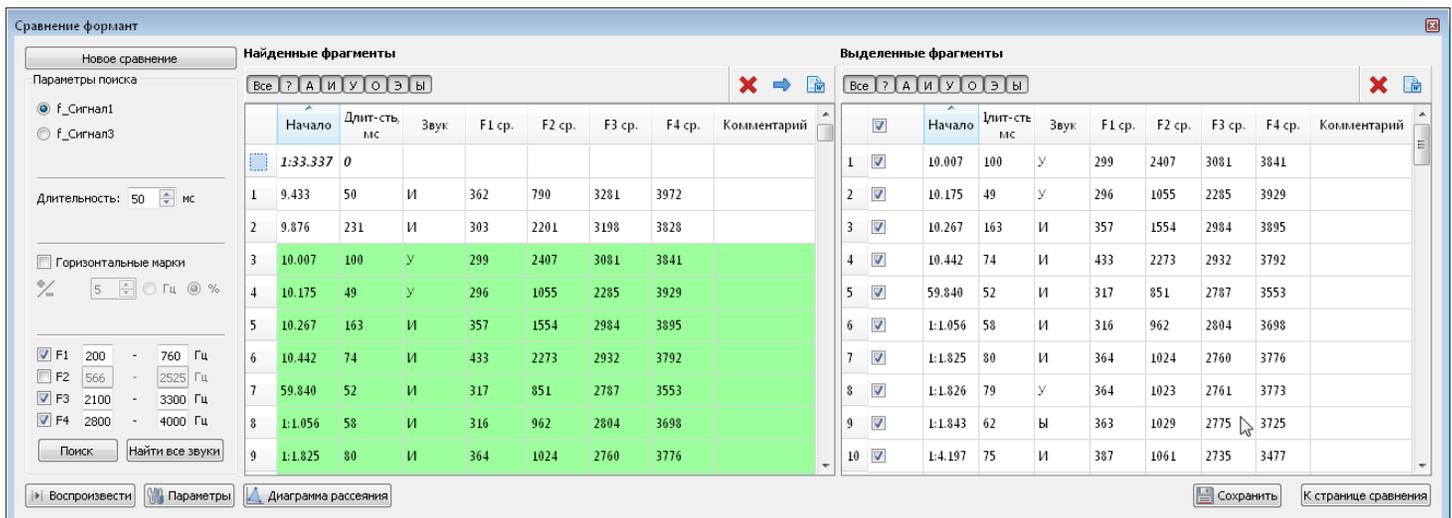


Рисунок 28 – Панель модуля в качестве отдельного независимого окна

При последующих запусках модуля предыдущее положение панели сохраняется. Чтобы прикрепить панель обратно к нижней части главного окна **SIS II**, следует снова дважды щёлкнуть на заголовке панели.

Чтобы выбрать другие сигналы (форманты) для сравнения, нажмите кнопку

Новое сравнение

и в диалоговом окне **Выберите сигналы** (см. рис. 26) осуществите выбор.

Для работы с данным модулем целесообразно использовать два монитора или монитор с большим экраном.

Во время работы с модулем может потребоваться задать отдельные параметры. Чтобы открыть окно параметров сравнения формант (рис. 29), нажмите кнопку **Параметры**.

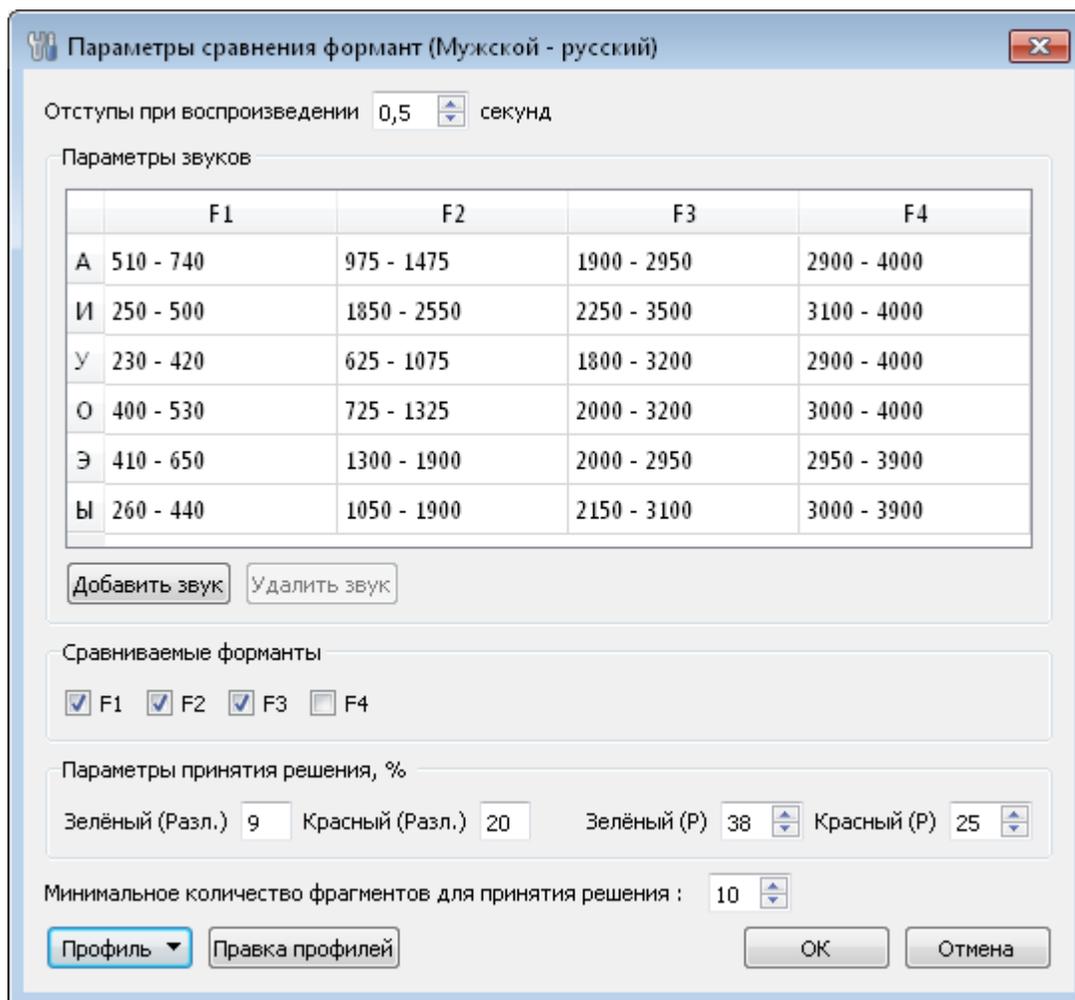


Рисунок 29 – Параметры сравнения формант

Окно позволяет задавать следующие параметры работы модуля:

1. Продолжительность проигрывания фонограммы в секундах до начала фрагмента и после него при прослушивании выбранного фрагмента. По умолчанию выставлено 0,5 секунды.

Воспроизведение выделенного в левой или правой половине таблицы фрагмента со сдвинутыми на заданную величину границами производится циклически при нажатии кнопки **Воспроизвести**. При повторном нажатии кнопки воспроизведение прекращается.

2. Любое сравнение формант в данном модуле производится только для занесенных в таблицу фрагментов, которым приписан один и тот же тип фонемы. Таким образом, до начала работы с модулем сравнения формант следует задать параметры типичных (обобщённых) формантных векторов для выбранных типов фонем. Диапазоны значений для образцов звуков приведены в таблице **Параметры звуков**. Например, при выборе профиля **Мужской > русский** для формант задаются диапазоны, показанные на рисунке 29. Необязательно задавать значения всех четырёх формант.

Оператор может убрать из таблицы отдельные звуки, если выделит соответствующую ячейку и нажмёт кнопку **Удалить звук**. Он также может добавить новые звуки. Для этого следует выполнить следующие действия:

а) Нажать на кнопку **Добавить звук**.

б) В поле **Название звука** диалогового окна **Новый звук** (рис. 30) указать название звука и нажать кнопку **ОК**.

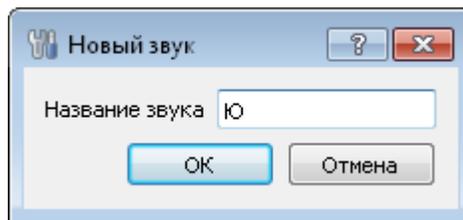


Рисунок 30 – Диалоговое окно нового звука

в) Дважды щёлкнуть левой кнопкой мыши в ячейке соответствующей форманты добавленного звука и в появившихся полях ввести границы диапазона (рис. 31).



Рисунок 31 – Ввод значений диапазона форманты

г) Для сохранения введённых границ, щёлкнуть в любой соседней ячейке таблицы **Параметры звуков**.

Для исправления текущего значения той или иной формантной границы следует сделать двойной щелчок левой кнопки мыши, поместив указатель на выбранную в таблице ячейку. В появившемся окошке следует откорректировать желаемое значение частоты форманты и нажать **Enter**.

Можно задать минимальное количество фрагментов, участвующих в сравнении. Для этого укажите необходимое значение параметра **Минимальное количество фрагментов для принятия решения**.

Если количество участвующих в сравнении фрагментов для определенного звука меньше, чем указанное число, то решение по этому звуку не принимается и строка со звуком в таблице результатов сравнения окрашивается серым.

Можно загрузить имеющиеся в программе готовые таблицы значений границ формантных частот для разных типовых случаев: мужские и женские голоса для русского языка, таблицы для английского и испанского языков. Для этого нужно в панели **Параметры** справа внизу нажать на надпись **Профиль** и выбрать желаемый профиль из выпадающего списка.

Чтобы создать и сохранить собственный профиль сначала необходимо задать в таблице **Параметры звуков** необходимые значения границ формант и наименования звуков. Границы диапазонов можно задать для одной или более формант. Затем следует нажать на надпись **Правка профилей** и затем **Сохранить текущий как**. Далее необходимо задать имя нового профиля. Записанный ранее профиль можно вызвать через кнопку **Профиль → пользовательские**.

Профили записываются в служебном каталоге AppData для данного пользователя, в папке **Roaming/Speech Technology Center/SISII** в виде файла **FormantComparePresets.xml**.



При обновлении модуля «Сравнение формант» возможно появление следующего предупреждающего сообщения: «Файл настроек модуля заменен на более новую версию от производителя». Сообщение появляется в том случае, если до обновления уже имелись какие-либо пользовательские настройки профиля, сохраненные в файле FormantComparePresets.xml. В этом случае файл настроек полностью обновляется в соответствии с новой версией. Оригинальный файл с пользовательскими настройками сохраняется под новым именем (к имени файла добавляется v001). В случае необходимости пользователь может отредактировать файл в соответствии с новой версией, чтобы сохранить настройки.

3. В поле **Сравниваемые форманты** выставлением флажков определяется, какие форманты будут использоваться при сравнении.

4. В поле **Параметры принятия решения** задаются критерии (границы в %) принятия решения при сравнении таблиц формант для двух дикторов (подробнее о критериях принятия решения см. п. 7.4). Они относятся к таблице результатов.

7.2 Формирование таблицы опорных фрагментов

Выбор опорных фрагментов производится последовательно для обоих сигналов. Текущий сигнал выбирается либо на панели модуля слева вверху (рис. 32) либо в соответствующем окне данных главного окна программы. В последнем случае выбор также дублируется на панели модуля.



Рисунок 32 – Выбор текущего сигнала

Для поиска задается минимальная длительность искомых фрагментов в миллисекундах. Типовая длительность обычно находится в пределах 5-80 мсек.

Далее поиск фрагментов с желаемой формантной структурой выполняется одним из пяти способов:

1. **По заданным границам диапазона формант** (рис. 33). В этом режиме выставлением флажков определяются используемые для поиска форманты, для каждой из них задается интервал значений, в котором должен осуществляться поиск. Интервалы по умолчанию: F1 = 300–1000; F2 = 600–2300; F3 = 2100–3300, F4 = 2800–4000 Гц.

<input checked="" type="checkbox"/>	F1	300	-	1000	Гц
<input checked="" type="checkbox"/>	F2	600	-	2300	Гц
<input checked="" type="checkbox"/>	F3	2100	-	3300	Гц
<input checked="" type="checkbox"/>	F4	2800	-	4000	Гц

Рисунок 33 – Выбор параметров для поиска по заданным границам значений формант

Для начала поиска звуков в этом режиме необходимо нажать кнопку **Поиск** под таблицей введенных диапазонов частот формант для поиска. Полученные результаты помещаются в левую таблицу **Найденные фрагменты** страницы поиска. Каждый фрагмент в таблице помечается знаком фонемы, к которой он относится в соответствии с описаниями форматных частот фонем, заданными в панели **Параметры**.

2. **По диаграмме рассеяния** (рис. 34). В этом режиме необходимо в окне **Диаграмма сравнения** навести указатель мыши на марку, и, удерживая нажатыми клавишу **Ctrl** и левую кнопку мыши, изменить положение марки.

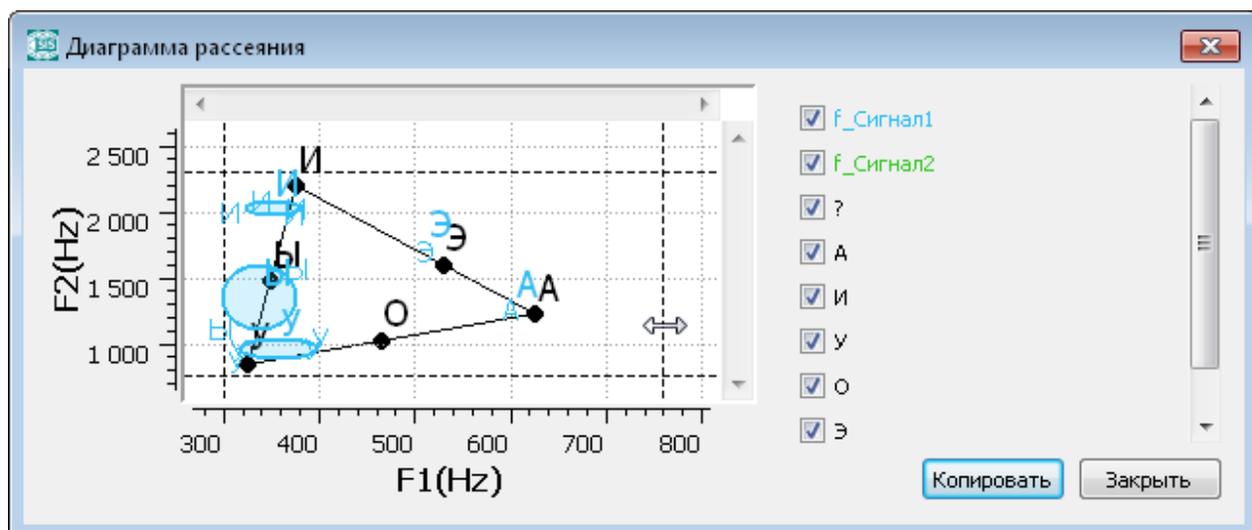


Рисунок 34 – Выбор параметров для поиска по диаграмме рассеяния

3. **По горизонтальным маркам** (рис. 35): производится поиск всех фрагментов, которые находятся на малом расстоянии по частоте от установленных в окне с формантами горизонтальных марок. Допустимое отклонение от частот, на которых стоят марки, может быть задано в герцах либо в процентах от частоты (по умолчанию это значение равно 5%). Для поиска программа выбирает самые низкочастотные горизонтальные марки в окне с формантами по числу отмеченных флажками формант.

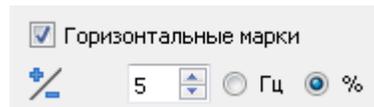


Рисунок 35 – Выбор параметров для поиска по горизонтальным маркам

Для начала поиска звуков в этом режиме необходимо нажать кнопку **Поиск** под таблицей введенных диапазонов частот формант для поиска. Полученные результаты помещаются в левую таблицу **Найденные фрагменты** страницы поиска. Каждый фрагмент в таблице помечается знаком фонемы, к которой он относится в соответствии с описаниями форматных частот фонем, заданными в панели **Параметры**.

4. **По всем звукам**. После нажатия кнопки **Найти все звуки** автоматически находятся все реализации звуков, форманты которых попадают в границы диапазонов формант для каждого звука, заданного в таблице параметров на панели, вызываемой через кнопку **Параметры**. Эта таблица меняется в соответствии с загружаемыми профилями. По умолчанию поиск осуществляется по границам диапазонов первых двух формант. При необходимости поиска фрагментов по другим формантам, установите соответствующие флаги.

5. Можно выбрать конкретный фрагмент в сигнале с помощью стандартных вертикальных временных марок. Сразу после выделения фрагмента в окне сигнала его параметры появляются в верхней строке таблицы **Найденные фрагменты**. Можно выделить эту строку таблицы, щелкнув по ней левой кнопкой мыши (строка подкрасится синим цветом). Далее, нажав вверху таблицы кнопку , можно перенести этот фрагмент направо, в таблицу выбранных для сравнения фрагментов.

Длительность фрагмента задаётся оператором одна для всех вариантов поиска (по умолчанию она равна 50 мс).

Параметры поиска, установленные оператором, сохраняются между запусками модуля.

Результаты поиска отображаются в таблице **Найденные фрагменты** (рис. 36) в следующем формате: начальная позиция в секундах от начала фонограммы и длительность фрагмента в миллисекундах, предполагаемый тип звука, средние для выбранного фрагмента значения формант F1–F4, комментариев. В последней строке таблицы, помеченной значком **X**, приводятся средние значения формант для всех отображаемых в таблице фрагментов.

Найденные фрагменты

Все ? А И У О Э Ы

	Начало	Длит-сть, мс	Звук	F1 ср.	F2 ср.	F3 ср.	F4 ср.	Комментарий
	1.804	0						
1	7.967	84	И	322	869	2870	3844	
2	7.984	65	У	315	867	2883	3852	
3	9.945	82	И	363	1027	2759	3774	
4	9.946	81	У	362	1027	2759	3771	
5	9.963	64	Ы	362	1033	2773	3724	
6	16.304	69	И	379	2020	2544	3830	
7	16.375	81	И	343	2087	3059	3721	
8	18.226	118	И	389	1554	2607	3634	
9	18.854	58	И	444	1392	2726	3777	

Рисунок 36 – Таблица найденных фрагментов

Чтобы добавить несколько фрагментов из левой таблицы в правую, их следует выделить указателем мыши и нажатием левой кнопки, удерживая клавишу **Shift** на клавиатуре. Чтобы выделить все фрагменты в таблице следует щелкнуть левой кнопкой мыши по первой ячейке в верхней строке таблицы. При этом все строки таблицы должны выделиться синим цветом. Далее, чтобы скопировать выделенные фрагменты из левой таблицы **Найденные фрагменты** в таблицу **Выделенные фрагменты**, нужно нажать справа над заголовком на значок  **Выбрать**. При этом скопированные фрагменты отмечаются также в окне слева зеленым цветом.

Аналогично производится удаление выбранных фрагментов из таблицы: они отмечаются, а затем нажимается значок  вверху таблицы.

Имеется возможность изменить содержание таблиц фрагментов при помощи флажков, выставляемых в контекстном меню (рис. 37). Контекстное меню появляется после щелчка правой кнопкой мыши по верхней строке таблицы.

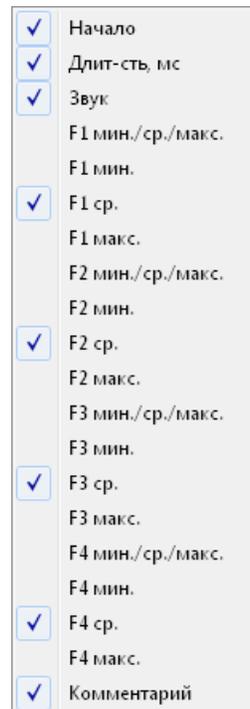


Рисунок 37 – Контекстное меню таблицы фрагментов

Содержание таблицы выбранных для сравнения выделенных фрагментов (ВФ) аналогично таблице найденных фрагментов: начальная позиция фрагмента на сигнале, его длительность, тип звука, средние по фрагменту значения формант F1–F4, комментарий (рис. 38).

Выделенные фрагменты									
Все ? А И У О Э Ы									
	<input checked="" type="checkbox"/>	Начало	Длит-сть мс	Звук	F1 ср.	F2 ср.	F3 ср.	F4 ср.	Комментарий
1	<input checked="" type="checkbox"/>	12.326	71	И	303	1062	2757	3490	
2	<input checked="" type="checkbox"/>	15.759	67	И	375	1116	2833	3921	
3	<input checked="" type="checkbox"/>	16.144	57	И	403	1038	2612	3338	
4	<input checked="" type="checkbox"/>	16.203	51	И	378	1978	2653	3835	
5	<input checked="" type="checkbox"/>	21.036	87	И	310	1823	2629	3790	
6	<input checked="" type="checkbox"/>	21.185	61	У	353	710	2054	3061	
7	<input checked="" type="checkbox"/>	21.352	67	У	319	723	1891	3088	
8	<input checked="" type="checkbox"/>	21.438	63	И	323	785	2822	3237	
9	<input checked="" type="checkbox"/>	21.924	59	И	399	1911	2923	3639	
10	<input checked="" type="checkbox"/>	22.202	49	Э	581	1663	2744	3409	

Рисунок 38 – Таблица фрагментов, выбранных в качестве опорных

В последней строке таблицы, помеченной значком **X**, приводятся средние значения формант для всех фрагментов в соответствующем столбце таблицы.

Чтобы сохранить в проекте сформированную таблицу ВФ, нажмите кнопку .

Фрагменты, добавленные в таблицу опорных фрагментов, в таблице **Найденные фрагменты** выделяются зелёным цветом. При выборе фрагмента в одной из таблиц, он также будет выделяться в другой таблице.

Оператор имеет возможность изменить тип звука, выбрав его в контекстном меню, а также добавить комментарий к фрагменту. Реализована возможность сортировки по любому из столбцов в обеих таблицах.

По умолчанию в таблицах отображаются фрагменты по всем звукам. Оператор может отключить показ и использование фрагментов с определёнными типами звуков нажатием на соответствующую кнопку с наименованием звука вверху таблицы (рис. 39). Для включения отображения фрагментов «отключенного» звука следует повторно нажать на соответствующую ему кнопку или на кнопку **Все**. В последнем случае в таблице снова будут отображаться фрагменты по всем звукам.



Рисунок 39 – Кнопки включения и выключения отображения фрагментов

Выделенный в таблице фрагмент можно удалить, нажав на значок **X Удалить выбранные**, или скопировать в буфер обмена, нажав на значок  **Скопировать в буфер выбранные**.



По умолчанию в отображении на диаграмме рассеяния, а также при сравнении, участвуют все фрагменты, добавленные в таблицу выделенных фрагментов. Чтобы исключить использование фрагмента, следует снять флажок в начале его строки.

При двойном щелчке левой кнопки по строке фрагмента в окне размещения формант для данного диктора будет показан участок спектрограммы с данным фрагментом, а он сам будет отмечен временными марками и постоянной парной маркой, а также подсвечен. На данном участке возможно стандартными средствами изменить положение временных и постоянных марок на границах фрагмента, откорректировать вручную график форматных траекторий. Парные марки, выделяющие выбранные в модуле СФ фрагменты сигнала, видны и могут использоваться только при работе модуля СФ. Возможно и целесообразно сохранить данную разметку, нажав кнопку сохранить внизу панели модуля.

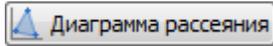
7.3 Диаграмма рассеяния

Анализ характеристик распределения форматных векторов и их сравнение для разных дикторов может быть выполнено на основе использования графического представления этих распределений в виде диаграммы рассеяния, или скаттерограммы. Для визуальной сравнительной оценки близости форматных распределений дикторов на плоскости с осями, соответствующими частотам двух формант, отображаются положения каждого формантного вектора, а также средние значения и СКО для каждого типа звука. Такое изображение позволяет наглядно представить и оценить формантные параметры, на основе которых происходит принятие идентификационных решений, увидеть «выпадающие» из общего тренда данные, оценить степень индивидуальности и близость формантной структуры речи сравниваемых дикторов в целом.

В качестве осей геометрического представления форматных параметров выбраны значения частот первой и второй формант (или по выбору любых других из четырех имеющихся). Такое представление полезно, поскольку:

- положения формант являются наиболее прозрачными и хорошо изученными акустическими корреляторами протекающих во время речепроизводства артикуляционных процессов;
- типовые значения формант для гласных фонем разных языков известны и образуют на плоскости первой-второй формант так называемый «треугольник гласных» (треугольник Фанта);
- отклонения индивидуальных положений формант гласных от типовых является устойчивым идентифицирующим признаком каждого диктора;
- значения формант достаточно устойчивы к изменениям амплитудно-частотных характеристик канала записи, изменению состояния диктора и стиля речи;
- значения формант успешно используются в отечественной и мировой фоноскопической экспертизе при проведении сравнительного анализа акустических признаков фонем уже более 20 лет. Их сравнение является обязательной составной частью таких методик, как «Диалект», «Этнос» и др.

Для визуального представления использован особый вид графика, на котором отмечены все индивидуальные значения, принимаемые переменными – диаграмма рассеяния.

Чтобы открыть окно диаграммы рассеяния (рис. 40), нажмите кнопку  в нижней части панели модуля СФ.

В координатах частот первой и второй форманты на диаграмме рассеяния изображён треугольник гласных, построенный по средним значениям формант образцов наиболее удаленных друг от друга звуков: А, У, И.

По умолчанию на треугольник накладываются отображения форматных векторов всех выбранных фрагментов обоих анализируемых сигналов (для каждого сигнала своим цветом). Опции «Показывать область СКО» и «Показывать положение звуков» также по умолчанию отмечены. Для каждого звука рисуется закрашенный его цветом эллипс с центром в точке координат со средним значением формант и размерами, пропорциональными СКО по данной оси координат.

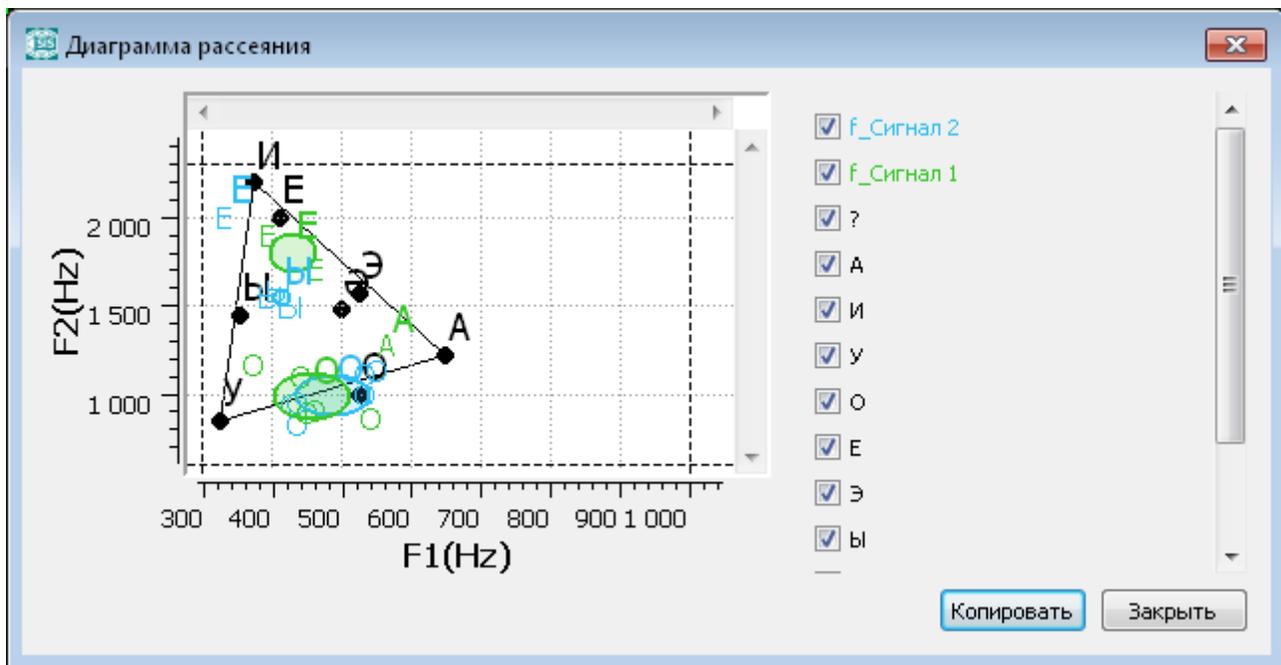


Рисунок 40 – Окно диаграммы рассеяния

Убирая те или иные флаги, оператор имеет возможность удалять с изображения диаграммы рассеяния отдельные звуки и области СКО, а также все звуки и области СКО, принадлежащие тому или иному диктору.

Размеры окна диаграммы рассеяния можно менять при помощи мыши так же, как любого окна операционной системы.

Имеется возможность, удерживая нажатой левую кнопку мыши, выделить для детального рассмотрения определённый фрагмент диаграммы рассеяния. При отпускании кнопки в окне отобразится только выделенный фрагмент (рис. 41).

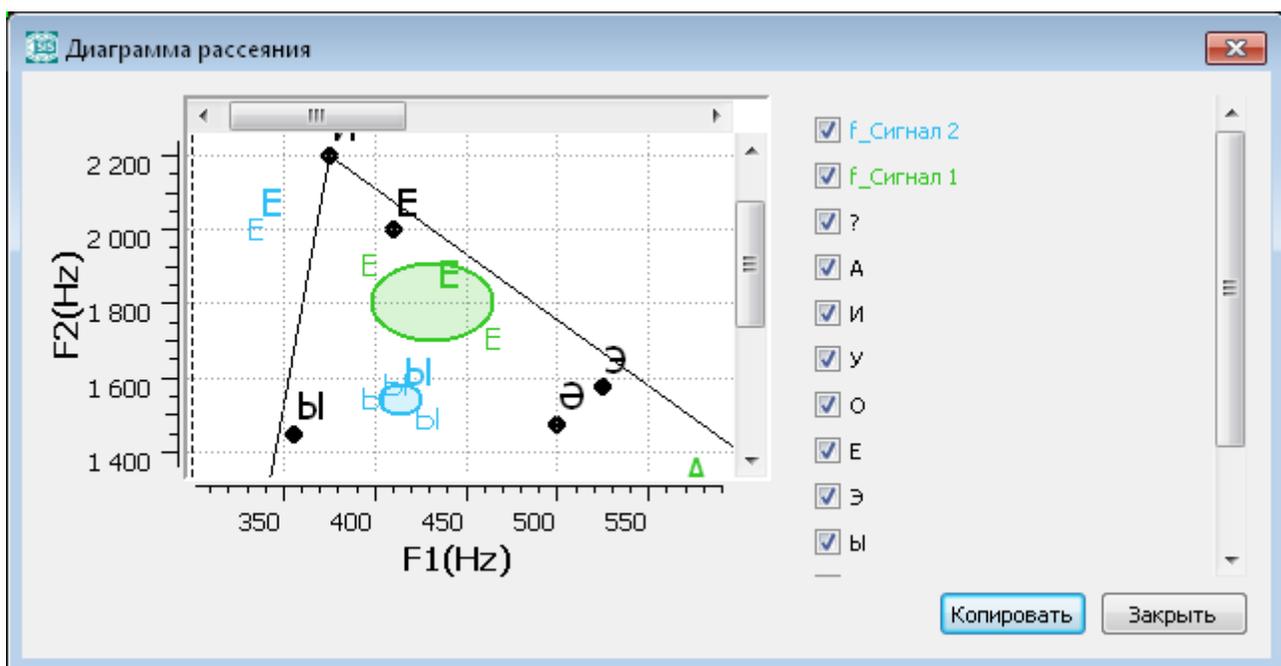


Рисунок 41 – Отображение выделенной части окна диаграммы рассеяния

В данном случае перемещение по диаграмме рассеяния становится возможным при помощи полос горизонтальной и вертикальной прокрутки. Для возвращения к полному изображению следует щелкнуть правой кнопкой мыши при положении курсора в окне диаграммы. В появившемся контекстном меню диаграммы рассеяния выбрать пункт **Показать целиком**.

Чтобы скопировать изображение диаграммы рассеяния в буфер обмена, воспользуйтесь пунктом контекстного меню **Копировать в буфер** или нажмите кнопку **Копировать**. Чтобы закрыть окно скатерограммы, нажмите кнопку **Закреть**.

Имеется возможность поменять оси отображения диаграммы рассеяния на любую пару осей с 1 по 4 формант. Для этого следует щелчком правой кнопки мыши вызвать контекстное меню, а в нем выбрать пункт **Изменить оси**, а затем выбрать нужную пару формант для построения новой диаграммы (рис. 42).

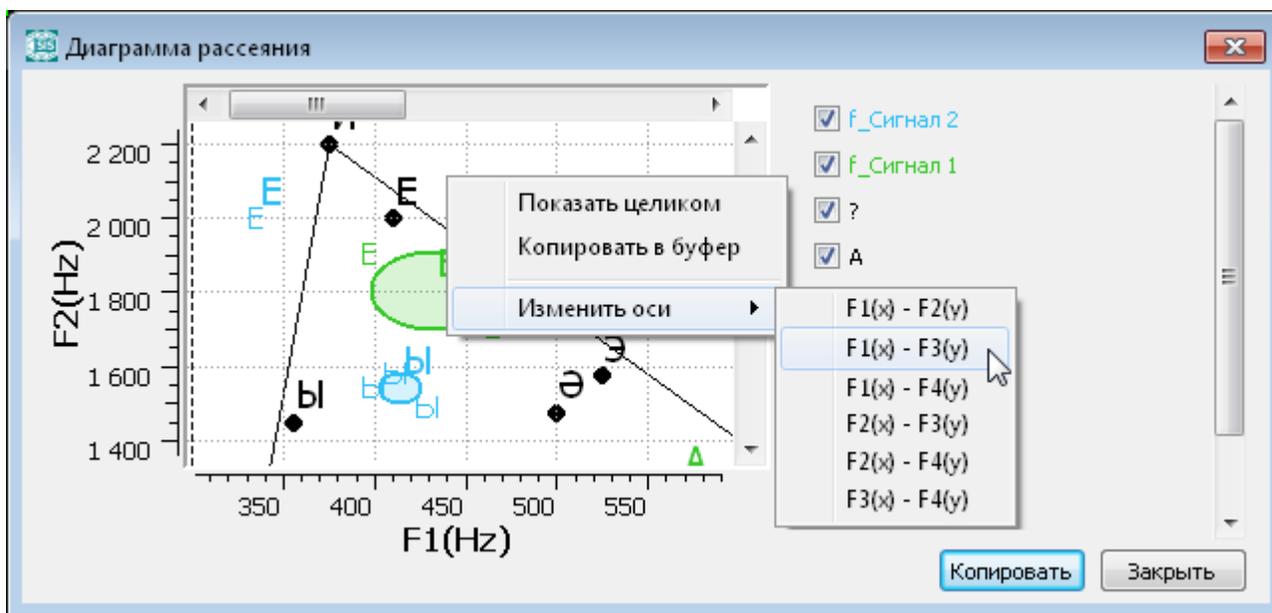


Рисунок 42 – Выбор новой пары осей для построения другой диаграммы рассеяния.

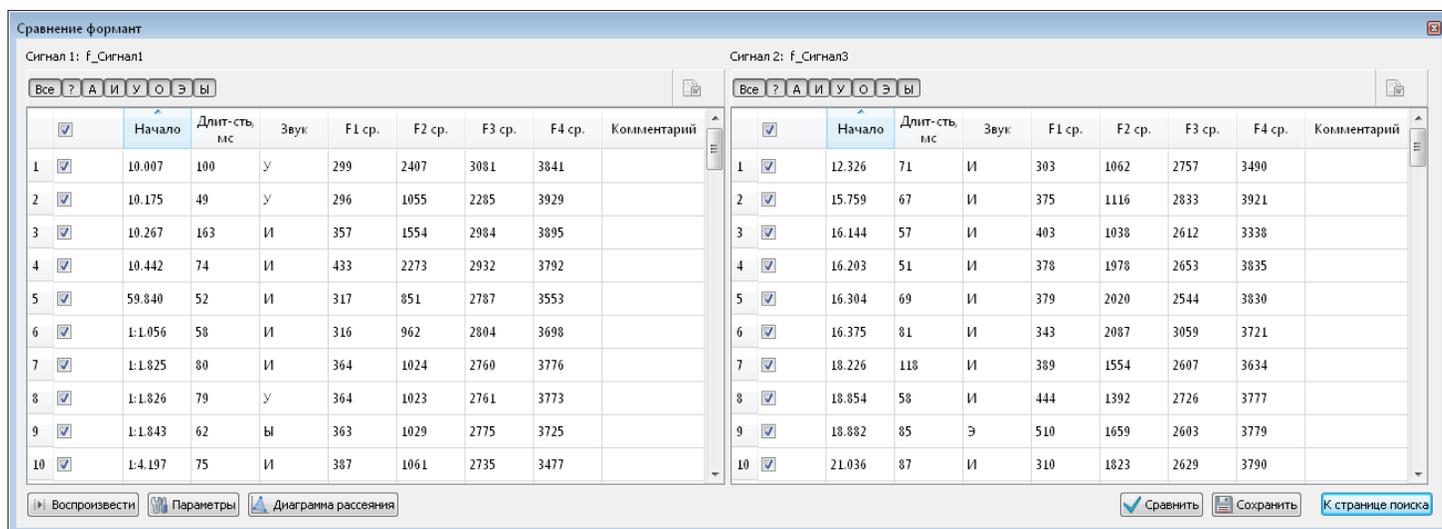
Новое изображение тех же фрагментов будет построено на формантной плоскости для выбранной пары осей.

Просмотр исследуемого речевого материала с помощью диаграммы рассеяния позволяет визуально определить степень совпадения/различия формантных структур в речи дикторов, уточнить положение оптимальных границ типов звуков в таблице параметров сравнения, выбрать для численного сравнения формант наиболее представительные фрагменты. График диаграммы рассеяния, помещенный в отчет, может служить наглядной и убедительной иллюстрацией выводов эксперта.

Именно использование диаграммы рассеяния позволяет легко установить и проиллюстрировать существенное различие в реализации дикторами того или иного типа звука.

7.4 Сравнение выбранных фрагментов

Чтобы перейти к сравнению выбранных опорных фрагментов, нажмите кнопку [К странице сравнения](#). В рамках панели модуля будут открыты таблицы опорных фрагментов сигнала 1 и сигнала 2 (рис. 43).



	Начало	Длительность, мс	Звук	F1 ср.	F2 ср.	F3 ср.	F4 ср.	Комментарий
1	10.007	100	У	299	2407	3081	3841	
2	10.175	49	У	296	1055	2285	3929	
3	10.267	163	И	357	1554	2984	3895	
4	10.442	74	И	433	2273	2932	3792	
5	59.840	52	И	317	851	2787	3553	
6	1:1.056	58	И	316	962	2804	3698	
7	1:1.825	80	И	364	1024	2760	3776	
8	1:1.826	79	У	364	1023	2761	3773	
9	1:1.843	62	Ы	363	1029	2775	3725	
10	1:4.197	75	И	387	1061	2735	3477	

	Начало	Длительность, мс	Звук	F1 ср.	F2 ср.	F3 ср.	F4 ср.	Комментарий
1	12.326	71	И	303	1062	2757	3490	
2	15.759	67	И	375	1116	2833	3921	
3	16.144	57	И	403	1038	2612	3338	
4	16.203	51	И	378	1978	2653	3835	
5	16.304	69	И	379	2020	2544	3830	
6	16.375	81	И	343	2087	3059	3721	
7	18.226	118	И	389	1554	2607	3634	
8	18.854	58	И	444	1392	2726	3777	
9	18.882	85	Э	510	1659	2603	3779	
10	21.036	87	И	310	1823	2629	3790	

Рисунок 43 – Панель модуля в режиме сравнения

Чтобы вернуться к поиску опорных фрагментов, нажмите кнопку [К странице поиска](#).

Чтобы выполнить сравнение опорных фрагментов, нажмите кнопку [Сравнить](#). В сравнении участвуют только те опорные фрагменты, которые отмечены флагом. По умолчанию отмечены все опорные фрагменты сигналов (фонограмм) 1 и 2.

В сравнении участвуют фрагменты, звуки которых имеются на обеих фонограммах. Для каждого звука вычисляются вариативность формантных векторов для первой фонограммы, вариативность формантных векторов для второй фонограммы, дистанция – вариативность расстояния между обоими дикторами, ее среднеквадратическое отклонение (СКО), вероятность совпадения распределений формант по критерию Стьюдента.

Для подсчета вариативности вычисляется нормированное Эвклидово расстояние между средними значениями частот формант для каждой пары выбранных фрагментов сравниваемых фонограмм. Вариативность формантных частот данного типа звука данного диктора вычисляется на фрагментах одной фонограммы. Дистанция между обоими дикторами для данного типа звука вычисляется для всех возможных пар фрагментов первой и второй фонограммы. Показатель **P** вычисляется как вероятность совпадения формантных распределений по критерию Стьюдента.

Решение о сходстве/различии дикторов принимается на основе совокупности всех показателей в таблице для каждого типа звука отдельно. Правила принятия решений сформулированы на основе анализа работы модуля на базе телефонных переговоров формата GSM (более 1000 дикторов, по 3-12 сессий).

Для принятия решения в настройках параметров задаются пороги: зелёный и красный.

В таблице результатов параметр **Вариативность** для первого и второго дикторов считается как нормированное среднеквадратичное отклонение (СКО) в % для всех формантных векторов для каждого типа фонем по отдельности. Нормировка осуществляется путем деления каждой формантной частоты на

среднее значение этой форманты для данного типа фонемы. В скобках указывается количество опорных фрагментов, участвующих в сравнении.

Параметр **Дистанция** считается как среднее нормированное Эвклидово расстояние в % для всех возможных пар сравнения форматных векторов каждого из дикторов для данного типа фонем. Для нормировки каждое расстояние между двумя формантными значениями частот делится на среднюю частоту двух сравниваемых формант.

Параметр **СКО** – среднеквадратичное отклонение для Дистанции.

Величина **Р** рассчитывается с помощью критерия Стьюдента как вероятность совпадения для данных дикторов функций плотности распределения их форматных векторов для данного типа фонемы. Все характеристики вычисляются только для одной-четырех формант, отмеченных в **Параметры > Сравниваемые форманты**.

Вводятся пороги принятия идентификационного решения для двух управляющих характеристик: **Различия** и вероятности (**Р**). Величины вводимых порогов автоматически изменяются при изменении числа учитываемых формант. Пересчет ведется следующим образом: если пороги заданы для 4-х выбранных формант, то для 3-х они будут по **Р** в 1,25 раза больше, а для **Различия** – только 0,85 от заданных.

Коэффициенты пересчета:

Число формант	Для Р	Для Различия
4	1	1
3	1,25	0,85
2	1,65	0,8
1	1,9	1

Если в таблице результатов сравнения дикторов **Различие** менее чем зеленый порог (Разл.), то решение для данного типа звука считается зеленым.

Если в таблице результатов сравнения дикторов **Различие** превышает красный порог (Разл.) принятия решения, то решение для данного типа звука считается красным.

Если **Дистанция** меньше 3%, то решение будет зеленым независимо от Р.

Если в таблице результатов сравнения вероятность **Р** в % превышает зеленый порог (Р), то решение для данного типа звука считается зеленым по вероятностному критерию.

Если в таблице результатов сравнения дикторов вероятность **Р** в % не превышает красный порог (Р), то решение для данного типа звука считается красным по вероятностному критерию.

Если для данного типа звука результат сравнения формант удовлетворяет «зеленым» условиям обоим критериев, то строка результата подсвечивается зеленым цветом, что отражает высокую степень сходства форматных характеристик речи сравниваемых дикторов для данного типа звука.

Если для данного типа звука результат сравнения формант удовлетворяет «красному» условию хотя бы по одному критерию, то строка результата подсвечивается красным цветом, что отражает различие форматных характеристик речи сравниваемых дикторов для данного типа звука.

В остальных случаях строка результата подсвечивается желтым цветом, что отражает некоторое сходство форматных характеристик речи сравниваемых дикторов для данного типа звука, недостаточное для принятия определенного идентификационного решения. Такая степень сходства может встретиться как у совпадающих, так и у различающихся дикторов.

По умолчанию пороги для дикторов-мужчин для 4 формант и выбора всех типов звука по всему файлу примерно таковы:

Зеленый (Разл.)=9%, Красный (Разл.) = 26%, Зеленый (P)=24%, Красный (P) = 17%

В каждом конкретном исследовании целесообразно выставить индивидуальные пороги принятия решения при каждом конкретном способе отбора опорных фрагментов для сравнения. Для этого необходимо сравнить каждого диктора самого с собой, желательно либо на двух разных файлах с его речью, либо на двух частях одного длинного файла. Пороги по **P** можно выставить, сравнивая форманты для двух копий одного и того же файла. При этом так выставляется зеленый порог, чтобы в результате сравнения для 5 типов гласных было не более 1 желтого решения. Красный порог по **P** в таком случае ставится примерно на 30% ниже зеленого (**P**).

Результаты сравнения отображаются в правой части панели в виде таблицы (рис. 44).

	Вариативн.1 (N), %	Вариативн.2 (N), %	Дистанция/ Различие, %	СКО, %	P, %
О	20.87 (78)	22.43 (40)	22.10 / -1.46	9.89	26.63
А	18.90 (17)	16.57 (33)	18.77 / -0.69	7.88	23.44
У	18.25 (56)	18.89 (29)	19.48 / 3.11	8.73	26.73
И	13.79 (35)	14.67 (29)	14.38 / -1.98	7.25	32.18
Ы	11.99 (15)	8.24 (15)	12.70 / 5.89	7.53	39.96
Э	1.39 (2)	0.00 (1)	12.59 / 803.78	1.36	13.67

В результате сравнения формант выявлены различия в индивидуальных идентификационно значимых характеристиках голоса и речи сравниваемых дикторов, достаточные для установления вероятного различия говорящих.

Рисунок 44 – Результаты сравнения

Каждому звуку, участвовавшему в сравнении, на основании рассчитанных значений показателей близости формантных характеристик и установленных порогов присваивается один из цветов: «красный», «жёлтый» или «зелёный». Кроме того, звуку может быть присвоен «серый» цвет в том случае, если опорных фрагментов менее 10, в этом случае получить идентификационное решение невозможно. Для каждого конкретного случая идентификации необходимо проводить настройку индивидуальных порогов принятия решения по **P** (см. Следующий раздел).

Идентификационное решение по паре дикторов при сравнении многих типов гласных принимается на основе следующих правил:

1. Определяется вес каждого звука:
 - для каждого красного звука присваивается результат **+2**;
 - для каждого желтого звука присваивается результат **+1**;
 - для каждого зеленого звука присваивается результат **-1**;
 - для каждого серого звука присваивается результат **0**.
2. Подсчитывается сумма по всем звукам и принимается одно из следующих решений:
 - при сумме меньшей или равной **-6**, делается вывод о *категоричном совпадении дикторов*.
 - при сумме равной **-5** или **-4** делается вывод о *вероятном совпадении дикторов*.
 - при сумме лежащей в интервале **от -3 до +3** невозможно получить идентификационное решение.
 - при сумме равной **4..6** делается вывод о *вероятном различии дикторов*.
 - при сумме большей или равной **7**, делается вывод о *категоричном различии дикторов*.

В строках таблицы у каждого из звуков отображается значение их вариативности (в процентах) для каждого из дикторов, дистанция между ними, различие, а также среднее квадратическое отклонение (СКО) и вероятность.

Ниже приводится частное идентификационное решение по результатам сравнения формантных характеристик речевых сигналов.

Выделенные строки таблицы результатов можно скопировать в буфер обмена, нажав на значок .

Чтобы выделить текст частного идентификационного решения, щёлкните на нём правой кнопкой мыши и выберите пункт контекстного меню **Выделить все** или воспользуйтесь сочетанием клавиш **Ctrl+A** на клавиатуре.

Чтобы скопировать текст частного идентификационного решения в буфер обмена, щёлкните на нём правой кнопкой мыши и выберите пункт контекстного меню **Копировать** или воспользуйтесь сочетанием клавиш **Ctrl+C** на клавиатуре.

Отобранные для сравнения данной пары сигналов фрагменты можно сохранить в проекте, если нажать кнопку  **Сохранить**. Чтобы запустить модуль и открыть таблицы фрагментов из ранее сохранённых сравнений, следует левой кнопкой мыши дважды щёлкнуть на пункте **Сравнение...** в активном проекте (рис. 45).

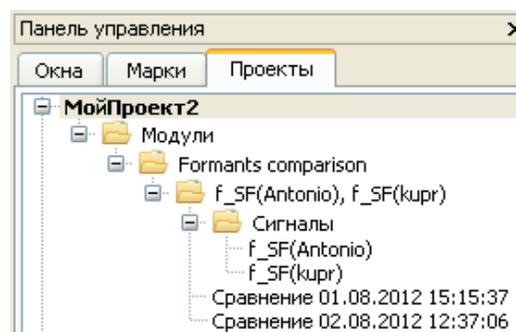


Рисунок 45 – Сохранённые в проекте сравнения формант

Чтобы завершить работу с модулем и закрыть панель, нажмите на значок **×** или  в правом верхнем углу панели.

7.5 Возможные методы выполнения идентификации с помощью модуля сравнения формант

Возможны различные варианты использования модуля для целей идентификации дикторов.

В любом случае, до начала работы с модулем СФ для обеих выбранных фонограмм следует построить спектрограммы в тех же окнах наложить на них форманты для всего речевого сигнала. Обычно целесообразно строить и визуально оптимизировать спектрограммы БПФ или КЛП в режимах «Форманты» с указанием пола диктора. Далее целесообразно провести просмотр и в случае необходимости коррекцию траекторий формант вручную либо для всего речевого материала, если он короткий, либо только для участков с используемыми далее звуками. Возможна коррекция формант и только уже на этапе выбора фрагментов сигнала для занесения в таблицы модуля СФ и только на именно этих фрагментах.

7.5.1 Настройка порогов принятия идентификационного решения

Независимо от используемого метода сравнения форматных структур фрагментов фонограмм для более надежного принятия решения рекомендуется проводить индивидуальную настройку порогов принятия решения. Для этого необходимо сравнить выделенные фрагменты для одного и того же диктора между собой. Для этого можно скопировать выделенные форманты каждого из дикторов в новое окно — они автоматически приобретают при этом новое имя. В этом модуль СФ, можно загрузить форманты одного и того же диктора, но под разными именами. Далее следует в каждом файле выделить те фрагменты, которые потом будут использованы при исследовании. Для них следует выполнить сравнение, определить те цифры по **P**, которые являются минимальными границами в тех режимах сравнения, которые предполагается далее использовать. Обычно целесообразно менять в панели параметры число сравниваемых формант и для каждой их комбинации записать пороговое значение **P**.

Следует определить индивидуальные пороги для сравнения по четырем формантам, по F1 и F2, по F3 и F4, по F4, по F3. Пороги автоматически пересчитываются при изменении числа сравниваемых формант, но этот пересчет приблизителен и может оказаться не оптимальным. Поэтому целесообразно, далее для каждого конкретного числа сравниваемых формант устанавливать пороги, найденные при сравнении формант диктора с самим собой.

Далее при сравнении дикторов между собой Зеленый порог по **P** следует установить максимально большим, но таким, чтобы при сравнении каждого из пары сравниваемых дикторов с самим собой все результаты сравнения были зелеными. Красный порог по **P** выставляется как)примерно 0,7 от величин γ зеленого порога по **P** для каждого сочетания сравниваемых формант индивидуально.

Порог по Дистанции **D** можно установить индивидуально только если сравнивать формантную структуру действительно разных файлов в речью одного и того же диктора. Если речи в фонограмме достаточно много (30 секунд и более), то возможно сравнение формантной структуры двух частей одного и того же файла.

7.5.2 Фонемно-спектральный метод. Сравнение гласных фонем в сопоставимых контекстах

В таблицу «Выделенные фрагменты» заносятся центральные участки формально одинаковых гласных звуков речи, выбираемых в тождественном или сопоставимом фонетическом контексте. Занесение проводится по одному фрагменту, выбираемому на спектрограмме с помощью временных марок. При работе в модуле СФ этот фрагмент автоматически помещается в верхней строке левой части таблицы на странице поиска. Если фрагмент окончательно выбран, форманты на нем при необходимости откорректированы, то его можно занести в правую часть таблицы, выделив данную строку щелчком указателя мыши и нажав стрелку → вверху над таблицей.

Текст фонограммы прослушивается, в нем выбираются фонемы, фонетическое качество которых (то есть и звучание) должно быть одинаковым в выбранных для сравнения фрагментах исследуемых фонограмм.

Наиболее просто этого можно достичь, если выбирать гласные, стоящие в обеих фонограммах под словесным (а желательно и фразовым) ударением в окружении одинаковых фонем. То есть, следует выбирать центральные участки гласных в совпадающих триадах: Согласный -Гласный- Согласный (СГС)

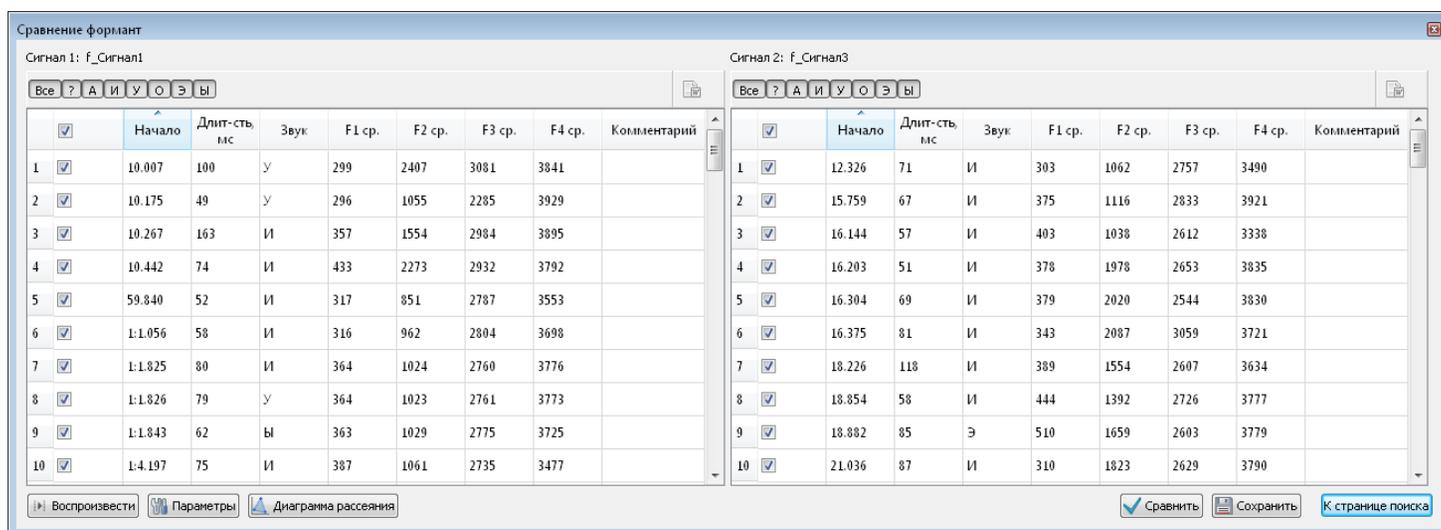
Вместо согласного с гласным в триадах может соседствовать и пауза или гласный, Важно, чтобы такой контекст был в обеих сравниваемых триадах.

Если в речевом материале фонограмм не удастся найти одинаковые триады, то возможно сравнение гласных в фонетически сопоставимых триадах, то есть, триадах с фонетически близкими контекстами справа и слева от гласного. Близкими по влиянию на соседний гласный можно считать согласные, совпадающие по мягкости, огубленности, веляризации, месту образования, назальности, желательно, по звонкости. Пауза и не смягченные аллофоны гласных А, Э, Ы образуют отдельную контекстную группу, также, как и огубленные У, О и Е, И.

Таким образом, триады, например, фАх и пАк можно считать для данного подхода сопоставимыми. Контексты любых мягких согласных, кроме назальных, можно считать сопоставимыми. Расположенный справа от гласного звук влияет на него меньше, чем расположенный слева. Наиболее подвержен контекстному влиянию гласный А, а наименее – гласный И.

Кроме влияния соседних звуков при выборе сопоставимых фонем следует обращать внимание на сопоставимость темпа и громкости речи, эмоционального и физического состояния диктора, влияние стиля и жанра речи и обстоятельств акта коммуникации (акустическая обстановка, поза, предметы во рту и др.).

Для принятия достоверного решения необходимо сравнивать не менее трех типов гласных, максимально разнесенных по признакам ряда и подъема, То есть на плоскости F1*F2, гласные должны располагаться в разных, максимально разнесенных частях. Рекомендуется выбирать гласные из всех групп на тексте, обозначаемых следующими буквами: (А, Э), (И, Е), (У, О), (Я, Ю, Ё) (Ы). Обязательным является наличие гласных из первых трех групп. Чем больше групп и типов гласных представлено в выбранных фрагментах, тем более надежным является принятое идентификационное решение. Обычно для достоверного принятия решения достаточно 15-20 реализаций гласных в сопоставимых триадах. Пример таблиц фрагментов, занесенных в таблицы, приведен на рисунке 4б:



	Начало	Длительность, мс	Звук	F1 ср.	F2 ср.	F3 ср.	F4 ср.	Комментарий
1	10.007	100	У	299	2407	3081	3841	
2	10.175	49	У	296	1055	2285	3929	
3	10.267	163	И	357	1554	2984	3895	
4	10.442	74	И	433	2273	2932	3792	
5	59.840	52	И	317	851	2787	3553	
6	1:1.056	58	И	316	962	2804	3698	
7	1:1.825	80	И	364	1024	2760	3776	
8	1:1.826	79	У	364	1023	2761	3773	
9	1:1.843	62	Ы	363	1029	2775	3725	
10	1:4.197	75	И	387	1061	2735	3477	

	Начало	Длительность, мс	Звук	F1 ср.	F2 ср.	F3 ср.	F4 ср.	Комментарий
1	12.326	71	И	303	1062	2757	3490	
2	15.759	67	И	375	1116	2833	3921	
3	16.144	57	И	403	1038	2612	3338	
4	16.203	51	И	378	1978	2653	3835	
5	16.304	69	И	379	2020	2544	3830	
6	16.375	81	И	343	2087	3059	3721	
7	18.226	118	И	389	1554	2607	3634	
8	18.854	58	И	444	1392	2726	3777	
9	18.882	85	Э	510	1659	2603	3779	
10	21.036	87	И	310	1823	2629	3790	

Рисунок 46 – Пример фрагментов центральных участков ударных звуков речи, занесенных в таблицы модуля СФ. Дикторы читают один и тот же текст.

Если автоматическое определение типа звука не соответствует нужному, то можно значок типа звука поменять, нажав правой кнопкой мыши на ячейку со значком и выбрав из выпадающего меню требуемый знак. Для сравнения всех типов фонем более точно иногда целесообразно ввести новый список границ фонем, например, расширив его до типовых значений формант гласных в окружении мягких согласных.

Далее можно сравнить форматные структуры как для всех звуков сразу, так и для отдельных подтипов звуков:

Сравнение можно проводить как по каждому типу/подтипу гласных в целом, так и по каждому типу триад по отдельности. В данном виде исследования сравниваются произносительные навыки дикторов, вклад в которые имеют и анатомическое сходство/различие их произносительных органов.

Предполагается, что в одном и том же контексте диктор произносит ударные гласные одинаково. Важно отмечать выделяемый участок гласного одним и тем же способом. Например, вблизи максимума амплитуды гласного, на максимально стационарном участке форматных траекторий. Следует избегать использования участков сильно назализованных гласных с искаженной форматной структурой. На фрагментах с быстрым изменением форматных частот следует быть максимально осторожным в выборе сопоставимых фрагментов. В этой ситуации целесообразно одновременно выводить на экран спектрограммы и форманты обоих сравниваемых фрагментов двух фонограмм\ и согласовано, исходя из единых соображений, выбирать положение границ фрагментов.

Сравнение выбранных фрагментов целесообразно проводить по всем имеющимся формантам. При близости дикторов по таким характеристикам возможно провести уточняющее сравнение и только по области 1-2 формант (сравнение фонетических целей), а в случае сходства дикторов, далее только по 3-4 или только 3 или 4 формантам – сравнение фонемонеразличительной, наиболее индивидуальной части форматной структуры гласных.

Метод может быть распространен и на оценку форматной структуры безударных гласных и отдельных согласных. Безударные целесообразно оценивать также только в сопоставимых контекстах, а также отдельно от ударных, по типам: первый предударный, второй и более предударный, заударный, заударный конечный.

Отметим, что положительный результат идентификации при данном методе исследования обеспечивает существенно более высокую надежность, чем отрицательный. Причинами различия в произнесении ударных звуков в том же контексте могут быть различия в темпе, громкости, стиле и жанре речи и т. п.

7.5.3 Фонемно-спектральный метод. Сравнение форматных структур в типичных фонемных областях

В таблицу выделяемых фрагментов заносятся все фрагменты с четко определяемой формантной структурой 3-4-х формант и относящиеся к гласным или плавным согласным. Для выделения таких фрагментов необходимо вызвать модуль СФ, указать два файла выделенных формант, наложенных на спектрограммы. Далее для каждого файла устанавливается длительность фрагментов поиска (начать с 40 мсек, если участков слишком мало, то можно последовательно изменять эту длительность в пределах 50-5 мсек. Должен быть загружен соответствующий профиль гласных звуков для данного языка.

Далее можно искать все необходимые фрагменты, нажав кнопку **Найти все звуки**. В этом случае среди форматных траекторий выбранной фонограммы автоматически будут найдены и помещены в таблицу найденных фрагментов все участки сигнала с формантными траекториями, попадающими на протяжении заданной длительности в границы формант звука, заданные в таблице **Параметры звука** в разделе **Параметры модуля СФ**.

Далее, полученный список фрагментов в таблице нужно откорректировать. Обычно целесообразно удалить фрагменты с пропущенной 3-й, а для мужчин и 4-й формантой. Далее, каждый фрагмент обычно целесообразно просмотреть на спектрограмме с наложенными формантными траекториями и либо откорректировать положение формант и границ фрагмента вручную, либо исключить его из рассмотрения.

Обычно целесообразно исключать из общего рассмотрения участки назальных и назализованных звуков (их реализация часто нестабильна и у одного и того же диктора), а также слишком тихие, зашумленные, неясные по формантной реализации звуки.

Далее следует перенести проверенные фрагменты в правую часть панели, в таблицу Выбранные фрагменты.

До начала сравнения дикторов следует сравнить каждого диктора с самим собой и уточнить значения порогов для принятия решения по каждому виду сравнения (то есть при задании для сравнения разных сочетаний формант от 1 до 4).

После занесения всех фрагментов в таблицы и уточнения порогов следует перейти на страницу сравнения, задать параметры сравнения и получить результат идентификации.

Следует сравнивать и по 4 формантам, и по первым двум, и по 3-4 вместе. В большинстве ситуаций наличие хотя бы одного красного решения при достаточной представительности речевого материала (более 10 реализаций на каждый тип звука) характерно для случая различия дикторов.

Ориентируясь на диаграмму рассеяния, можно выбрать область, наиболее характерную для различия дикторов и провести анализ и сравнение фрагментов только для нее. Если речевой материал сопоставим по длительности и стилю речи, то сравнение форматных распределений в каждой конкретной частотной области является индивидуализирующим признаком. Например, два диктора, форманты речи которых

изображены на рисунке 47 ниже, имеют совершенно различное проявление звуков с формантами в этой области.

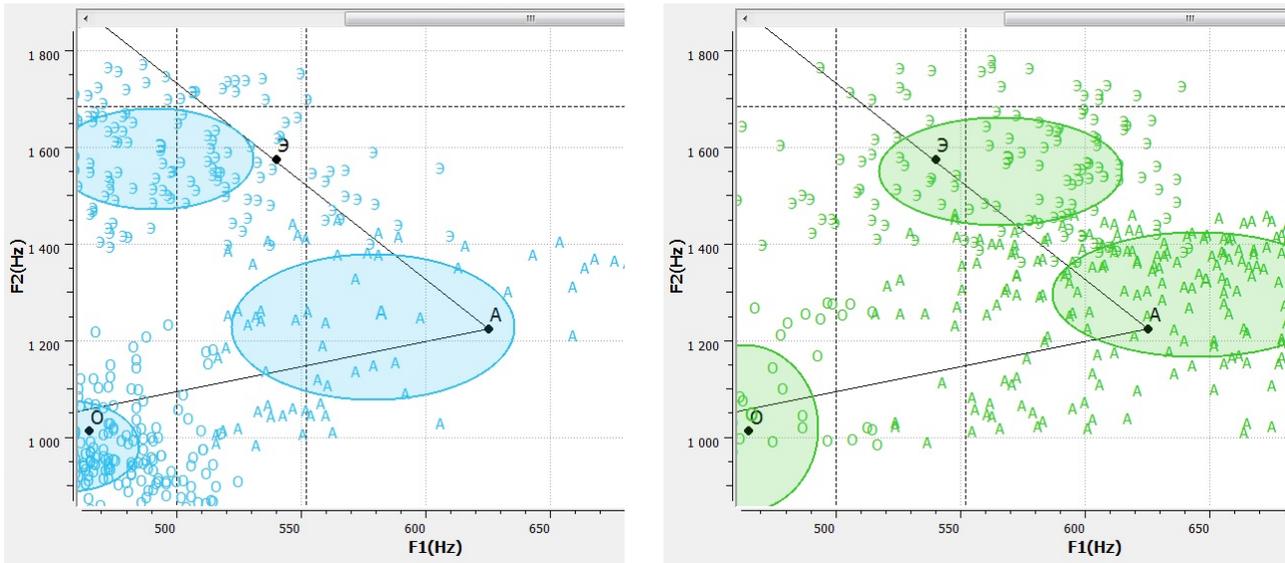


Рисунок 47 – Форманты гласных для двух дикторов, выделенные в одной и той же частотной области: F1=450-650Гц, F2=110-2000Гц

При сравнении их формантной структуры только в этой области и только для 1 и 2 формант, можно увидеть численное подтверждение этого факта (рис. 48):

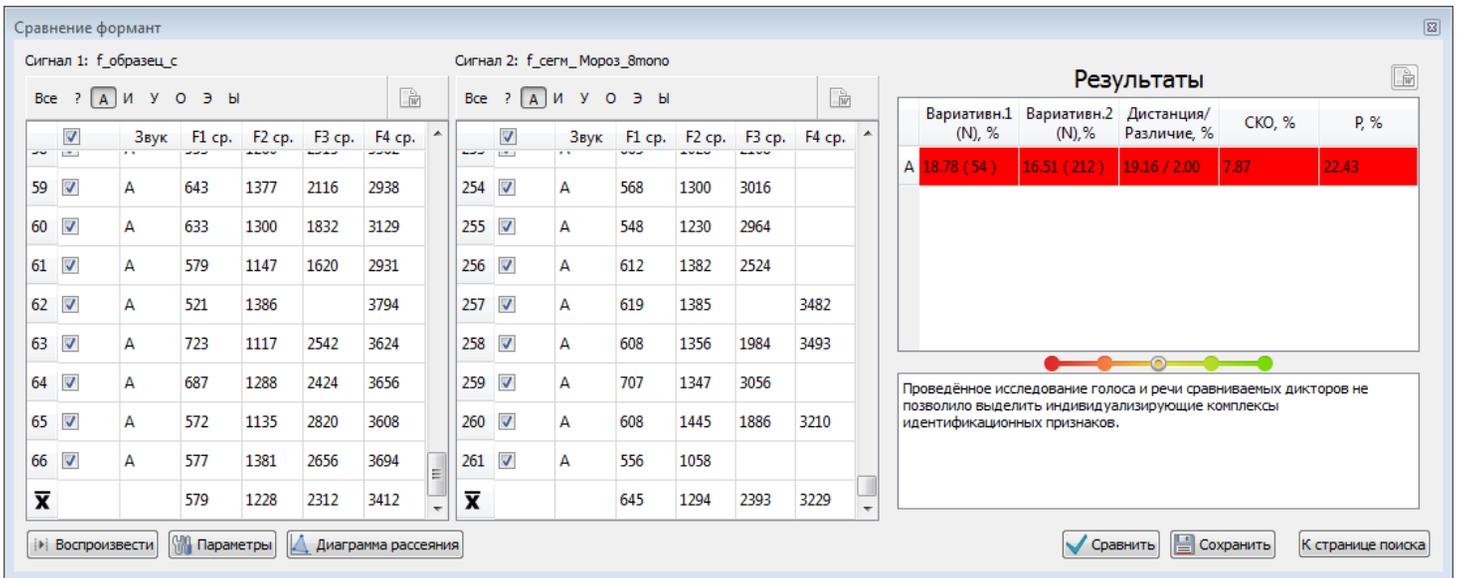


Рисунок 48 – Результаты сравнения формант гласных для двух дикторов

7.5.4 Метод формантного выравнивания

При работе по этому методу выбор опорных фрагментов происходит специфическим образом. В первой фонограмме следует выбрать фрагмент с четко прослеживаемой, соответствующей теоретическим представлениям, формантной структурой. Затем следует на спектрограмме с наложенными формантами отметить горизонтальными курсорами положение формант в выбранном фрагменте и провести поиск аналогичных по частоте формант фрагментов в данной фонограмме. Для этого следует отметить флажком пункт **Горизонтальные марки** в панели модуля СФ. Далее отметить ниже те форманты, по которым проводится поиск, и нажать кнопку **Поиск**. В ячейках таблички рядом с номерами формант F1-F4 появляются числовые границы диапазонов формантных частот, в которых проводится поиск. В случае необходимости можно их откорректировать прямо в ячейках.

Поиск следует вести по 4-м формантам, если это в данной фонограмме не осуществимо — то по трем или двум. Порог отличия целесообразно в начале поиска ставить минимальным (1-2%), постепенно увеличивая его, если фрагменты с малым порогом отсутствуют.

Каждый найденный фрагмент со сходной формантной структурой следует просмотреть на спектрограмме, и, в случае необходимости откорректировать форманты вручную. Следует удалять из списка найденных сомнительные фрагменты с назализованной или неясной по другим причинам формантной структурой. Если не обнаруживаются фрагменты, близкие к первоначально выбранному, то в процессе просмотра фонограммы следует выбрать наиболее типичный, относительно часто встречающийся тип формантной структуры, максимально близкий к первоначально установленному, переместить горизонтальные курсоры на положение его формант и повторить поиск.

Если такие фрагменты найдены, то при отличиях в формантах не более $\pm 3\%$ их можно считать тождественными.

Если не удастся найти близкие фрагменты в данной фонограмме, то целесообразно попробовать выбрать для анализа звук с несколько отличной формантной структурой и повторить поиск. В ситуации коротких фонограмм допустимо проводить исследование далее и только с одной реализацией типа звука, хотя в этом случае надежность идентификации может понизиться.

Далее следует по установленным горизонтальным курсорам найти аналогичные фрагменты во второй фонограмме. Если это удастся, и в двух фонограммах обнаружено по три и более фрагментов, совпадающих по 4 (3) формантам в диапазоне отличий не более 3%, то, с точки зрения данного метода, сравниваемые дикторы имеют близкую, сходную геометрию акустического речеобразующего тракта для данного типа звука.

Если сразу найти близкие по частотам формант фрагменты не удастся, то следует изменить процедуру поиска. Целесообразно сначала попробовать найти фрагменты с совпадением трех формант (любых трех из четырех, для дикторов-женщин – двух из трех) с тем же порогом точности совпадения (т.н. формантно выровненные фрагменты).

Обнаруженное различие для невыровненной форманты является критерием сходства/различия геометрий вокального тракта сравниваемых дикторов для данного типа звука.

Для каждого найденного фрагмента следует проверить и откорректировать форманты, а затем найти отклонение в % по форманте, не вошедшей в число выровненных формант.

Например, это можно сделать последовательным изменением порога точности совпадения формант при поиске и установлением границы расхождения, прямым численным сравнением средних значений формант или в режиме **Сравнение** при установке сравнения только не выровненной форманте.

Обязательно следует проверить, нет ли подобной артикуляции с новым положением четвертой форманты и в первой фонограмме. Для этого целесообразно установить горизонтальный курсор на новое значение форманты во второй фонограмме и провести поиск фрагментов по 4 формантам в первой фонограмме. Возможно наличие у диктора так называемых компенсаторных артикуляций, когда один и тот же звук формируется несколькими (обычно не более двух) способами задания акустических свойств речевых органов. Кроме того, случаи заметного различия формантных векторов при совпадении 2-3 формант возможны за счет изменения сигнала в канале связи либо искажением АЧХ, либо алгоритмами компрессии сигнала.

Не обязательно выравниваемые форманты должны быть первыми тремя, допустимо использовать все возможные комбинации из 4 по 3.

После установления частного идентификационного вывода по одному типу звука следует перейти к другим типам звука, определяемым по положению первых двух формант.

Обязательным является использование позиций для 4 и более различных частотных областей. Должны использоваться диапазоны, характерные для гласных А, У или О, И или Е, а также иные диапазоны частот на плоскости $F1 \times F2$, отстоящие друг от друга приблизительно более, чем на 200 Гц по $F1$ или 400 Гц по $F2$. Всего следует использовать для сравнения не менее 15-20 фрагментов. Чем ниже качество речевого материала, тем больше фрагментов следует использовать для сравнения.

Если для двух и более типов звуков имеются существенные расхождения в положении невыровненных формант, то это основание считать данных дикторов различающимися по идентификационно значимым геометрическим свойствам речеобразующего аппарата и/или речевым навыкам.

Для дикторов с малой длиной речевого тракта возможно использование метода при выравнивании двух формант и сравнении третьей. В этом случае следует увеличивать число сравниваемых сегментов и принимать во внимание ряд специфических факторов: например, что при близости частот $F2$ и $F3$ их взаимное положение уже не независимо; высокий основной тон иногда не позволяет точно установить правильное положение первой форманты и др.

Для дикторов с принципиально различным типом произношения большинства звуков иногда возможно провести выравнивание частот только для двух формант. В этом случае сравнение следует проводить по остальным формантам в векторе.

Следует также обратить внимание, что у родственников возможно чрезвычайно точное совпадение формантных векторов для большинства звуков речи. В этом случае надо более внимательно просмотреть все возможные типы звуков: иногда заметное различие может быть только в одном из них.

8 СРАВНЕНИЕ ОСНОВНОГО ТОНА

Модуль «Сравнение основного тона» позволяет автоматизировать процесс идентификационного исследования речевых сигналов на основе структурного анализа мелодических контуров речевого сигнала сравниваемых фонограмм. Модуль предназначен для сравнения речевых образцов на основе параметров реализации однотипных элементов структуры мелодического контура. Для анализа предусмотрено 18 типов фрагментов контура и 15 параметров их описания. Для разных типов используются разные наборы параметров.

Параметры выделяемых экспертом на графике основного тона однотипных участков контура, сопоставимых с точки зрения реализационной специфики, коммуникативной и эмоционально-стилистической нагрузки, вычисляются автоматически, данные заносятся в таблицу.

Сравнение осуществляется по каждому из используемых типов на основе пороговых значений, рассчитанных для параметров каждого типа элемента контура. Далее результаты сравнения отдельных типов суммируются для получения общего решения относительно тождества либо различия голосов. При вычислении общего результата учитываются весовые коэффициенты, приписанные каждому параметру и типу фрагмента мелодического контура.

Если показатель оценки сходства существенно превышает показатель различий, то высока вероятность тождества голосов. При высокой степени различий между фонограммами формулируется решение о различии голосов. При незначительных различиях в оценках (менее 10 %) выдаётся сообщение о том, что решение принять невозможно.

Типы фрагментов контура, параметры их описания, а также порядок принятия идентификационного решения на их основе более детально описаны в приложении В.

8.1 Подготовка к работе с модулем



Перед запуском модуля «Сравнение основного тона» должен быть задан активный проект, а также подготовлены как минимум два окна с данными типа «основной тон» (с определённым основным тоном).

Предварительная подготовка фонограмм для сравнения данным методом заключается в получении корректных графиков основного тона (ОТ) голоса. Получение данных графиков осуществляется в **SIS II** автоматически путём применения модуля «Выделение основного тона».

Для достижения заявленной надёжности идентификации по каждому из типов фрагментов мелодического контура должны быть соблюдены определённые условия, которые излагаются ниже:

1. Метод предназначен для сравнительного анализа фонограмм, содержащих речь нейтрального стиля, без ярко выраженной эмоционально-модальной окраски.

2. Заявленная надёжность метода достигается при сравнении фрагментов утвердительных высказываний одного стиля (спонтанная речь сравнивается со спонтанной речью, чтение – с чтением).

Результат сравнения типа «спонтанная речь – чтение» зависит от индивидуальных особенностей чтения конкретного диктора.

3. Ограничения по длительности имеются только для длительных речевых фрагментов – в стандартном случае не менее 60 секунд; в отдельных случаях возможно использование фонограмм длительностью 30–60 секунд.

Для остальных типов фрагментов контура важна не длительность сама по себе, а представленность в фонограмме целевых структур для анализа – ядерных тонов, предшквал, шкал, пауз гезитации и т. д. В зависимости от темпа речи, от акцентно-ритмической структуры и сегментного состава достаточная длительность фонограммы может варьироваться от 15–20 секунд до нескольких минут.

Рекомендуемое число фрагментов каждого типа для анализа составляет 7–10. При низкой степени вариативности параметров возможно использование 5–7 фрагментов, однако дальнейшее сокращение числа измерений ведёт к существенному снижению надёжности метода.

4. Поскольку основой получения данных для анализа является кривая ОТ, с технической точки зрения метод имеет те же ограничения, что и автоматический метод сравнения статистики ОТ: желательное соотношение сигнал/шум не менее 10–15 дБ, отсутствие реверберации на обеих фонограммах.

Чтобы применить метод анализа мелодического контура, выберите меню **Модули → Сравнение основного тона**.

В диалоговом окне (рис. 49) укажите пару сравниваемых сигналов (с определённым основным тоном). Если их в программе всего два, модуль выбирает их автоматически. Если сигналов больше двух, то сравниваемые сигналы выбираются оператором из выпадающих списков.

Кроме того, если ранее в данном проекте уже выполнялись сравнения с выбранным сигналом, то таблицы фрагментов для этого сигнала можно загрузить из ранее сохранённых сравнений, воспользовавшись выпадающим списком **загрузить из**.

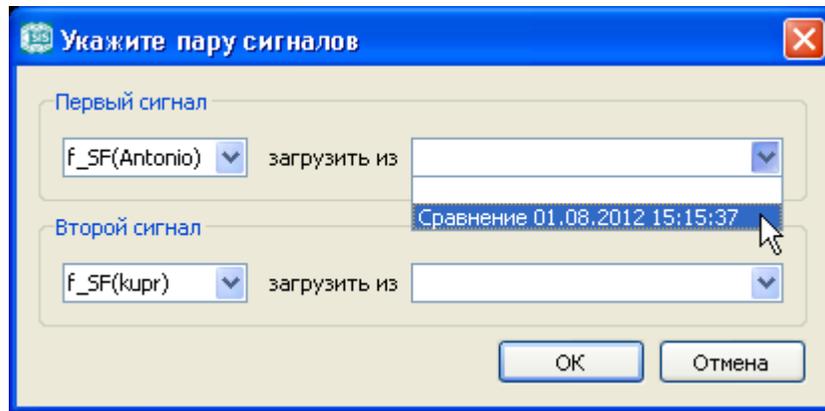


Рисунок 49 – Выбор сравниваемых сигналов

После выбора сравниваемых сигналов нажмите кнопку **ОК**. Откроется панель работы с модулем. Панель прикрепляется к нижней части главного окна **SIS II**, при этом открытые внутри него окна с данными, панели управления или окно спектра пропорционально меняют свой размер (рис. 50).

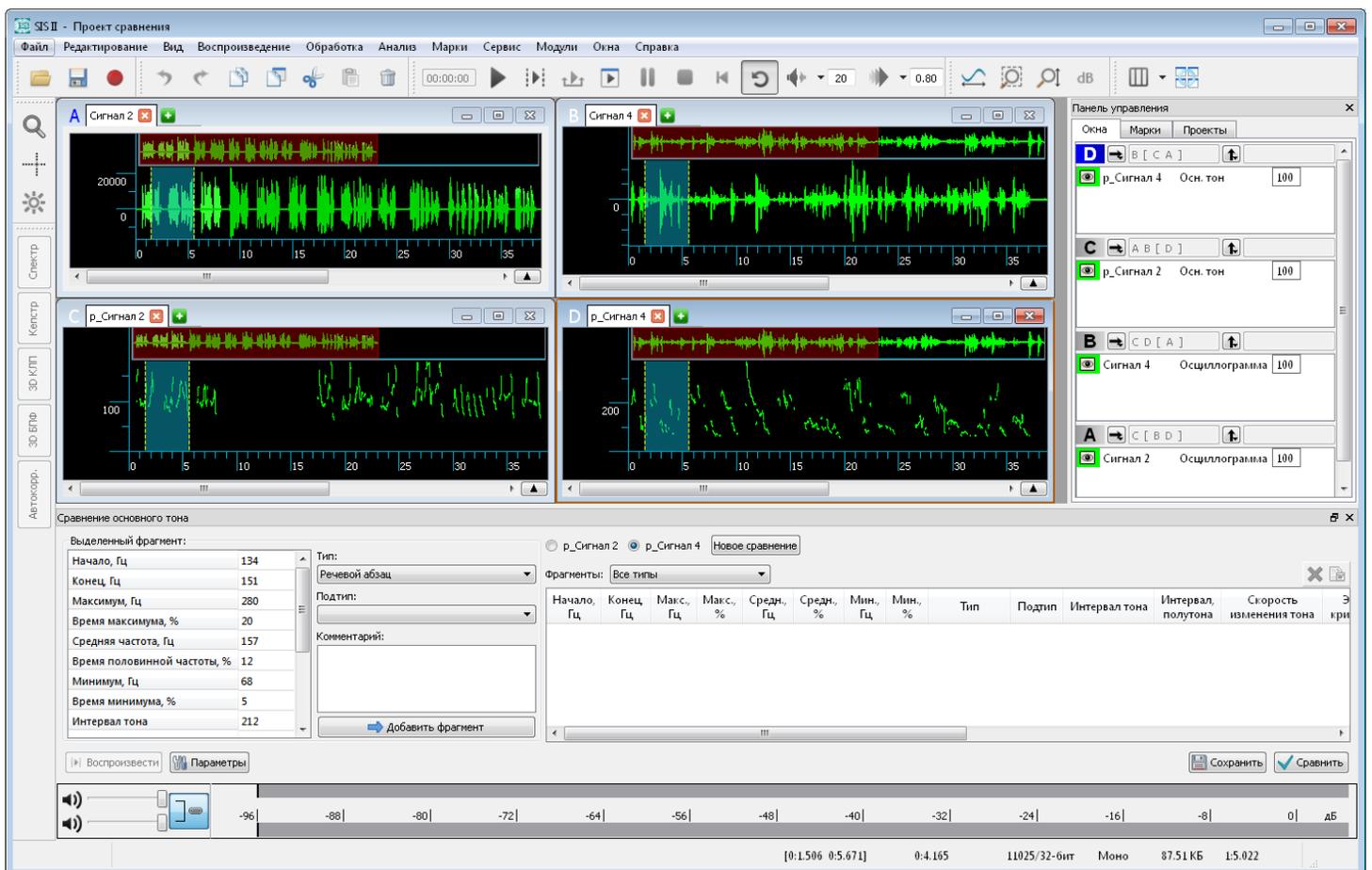


Рисунок 50 – Панель модуля внутри главного окна звукового редактора

Чтобы сделать панель модуля независимой от главного окна **SIS II** (рис. 51), следует нажать на значок  в правом верхнем углу панели или дважды щёлкнуть левой кнопкой мыши на заголовке панели. Панель можно перемещать в любое место на экране монитора, удерживая нажатой на заголовке левую кнопку мыши.

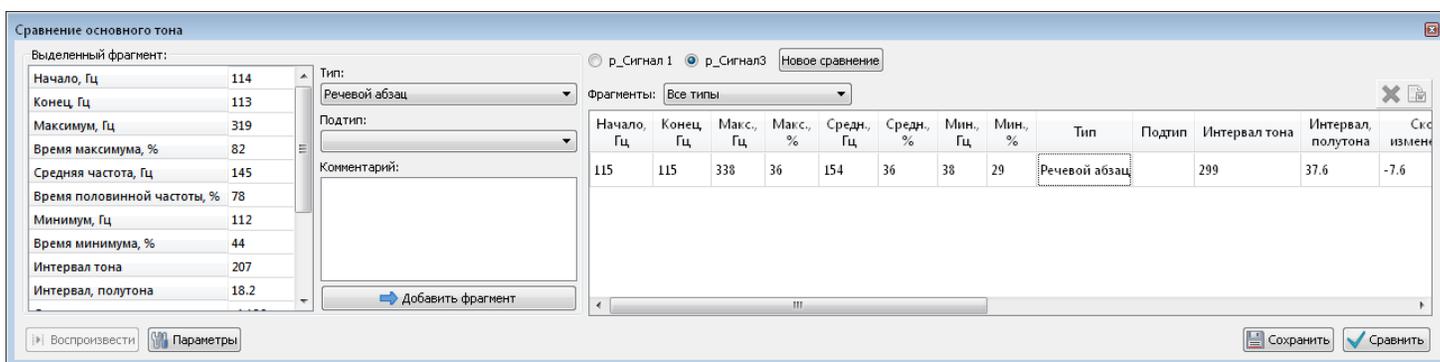


Рисунок 51 – Панель модуля в качестве отдельного независимого окна

При последующих запусках модуля предыдущее положение панели сохраняется. Чтобы прикрепить панель обратно к нижней части главного окна **SIS II**, следует снова дважды щёлкнуть на заголовке панели.

Чтобы выбрать другие сигналы (форманты) для сравнения, нажмите кнопку **Новое сравнение** и в диалоговом окне **Укажите пару сигналов** (см. рис. 49) осуществите выбор.

Перед началом работы с модулем может потребоваться задать отдельные параметры. Чтобы открыть окно параметров (рис. 52), нажмите кнопку **Параметры**.

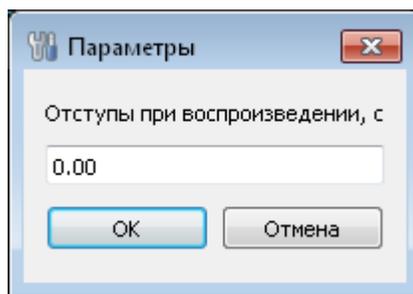


Рисунок 52 – Отступы при воспроизведении

Окно позволяет задавать продолжительность проигрывания фонограммы в секундах до начала опорного фрагмента и после него при прослушивании выбранного опорного фрагмента. По умолчанию выставлено 0,5 секунды.

Воспроизведение выбранного фрагмента производится циклически при нажатии кнопки **Воспроизвести**. При повторном нажатии кнопки воспроизведение прекращается.

Настройки текущего мелодического контура выставляются автоматически.

8.2 Формирование фрагментов

Текущий сигнал выбирается либо на панели модуля (рис. 53), либо в соответствующем окне данных главного окна программы. В последнем случае выбор также дублируется на панели модуля.



Рисунок 53 – Выбор текущего сигнала

Для выделенного в окне данных фрагмента основного тона рассчитывается и отображается в таблице **Выбранный фрагмент** набор параметров, приведённый на рисунке 54 и описанный в приложении В.

Далее оператор из выпадающих списков должен задать выбранному фрагменту **Тип** и **Подтип**, а также может дать комментарий (рис. 54).

После этого, чтобы добавить фрагмент в общую таблицу фрагментов, следует нажать кнопку



Выделенный фрагмент:

Начало, Гц	114
Конец, Гц	113
Максимум, Гц	319
Время максимума, %	82
Средняя частота, Гц	145
Время половинной частоты, %	78
Минимум, Гц	112
Время минимума, %	44
Интервал тона	207
Интервал, полутона	18.2
Скорость изменения тона	-4.186

Тип: Ядерный слог

Подтип: Нисходящий

Комментарий:



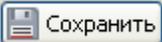
Рисунок 54 – Статистическая информация, тип, подтип и комментарий к фрагменту

В таблице отображается статистическая информация для фрагмента с указанием типа и подтипа, а также комментария оператора. Оператор может отобразить фрагменты только определённого типа: для этого над таблицей есть выпадающий список с возможностью выбрать тип фрагмента.

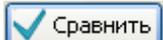
Оптимальное число однотипных опорных фрагментов для получения статистики составляет 10–15.

На основе данного набора статистических параметров эксперт может оценить степень представительности и однородности выборок, предназначенных для обобщения и сравнения, и при необходимости скорректировать наборы элементов.

Выделенный в таблице фрагмент можно удалить, нажав на значок , или скопировать в буфер обмена, нажав на значок . Предусмотрена возможность изменить тип и/или подтип, а также комментарий у фрагмента, добавленного в таблицу.

Чтобы сохранить в проекте сформированную таблицу фрагментов, нажмите кнопку .

8.3 Сравнение фрагментов

После завершения выбора фрагментов для двух фонограмм нажмите кнопку .

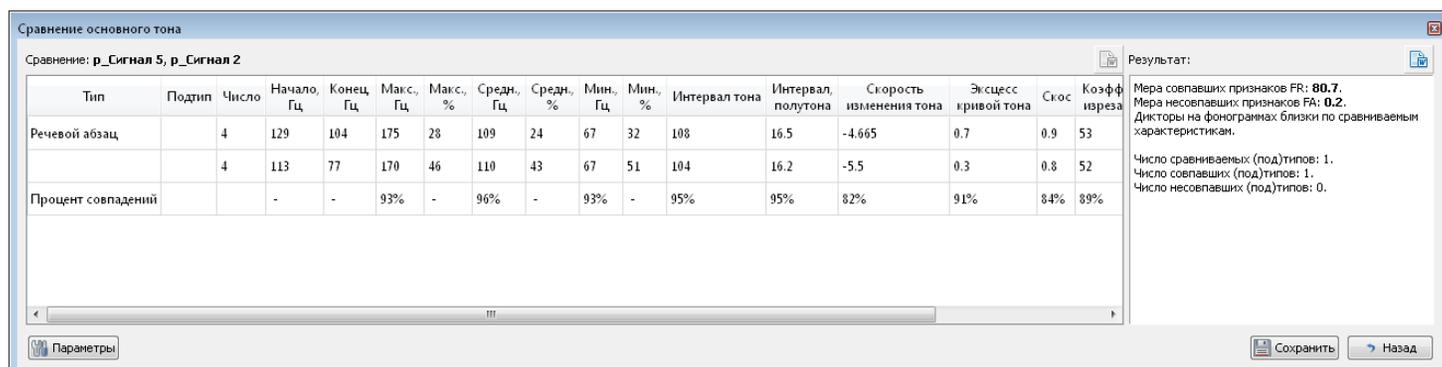
Модуль автоматически выполняет сравнение выбранных оператором фрагментов и выдаёт результат, представленный на рисунке 55.

В таблице результатов сравнения отображается информация о количестве сравниваемых фрагментов для каждого типа/подтипа. Отображается разброс значений для каждого из статистических параметров сравниваемых фонограмм и процент совпадения.

Сравнение фонограмм производится по следующему алгоритму:

- Усреднение нескольких фрагментов единой комбинации «тип/подтип/источник».
- Циклический поиск фрагментов одного типа/подтипа в двух источниках и проверка достаточности числа фрагментов (по умолчанию не менее 5) нужного типа/подтипа для каждого из сигналов.
- Циклическая проверка использования сочетания «тип/подтип и параметр» для пороговых значений; расчёт процента сходства/различия двух дикторов по одному признаку; считывание весовых коэффициентов в зависимости от факта сходства или различия, а также проверка использования сочетания «тип/подтип и параметр» для весовых коэффициентов.
- Расчёт значений «За» и «Против» как результирующих характеристик по всем параметрам одного варианта типа/подтипа.
- Статистическая проверка данных.
- Расчёт единого для источника значения «За» и «Против».
- Расчёт степени сходства/различия дикторов в зависимости от достаточного числа длительных фрагментов.
- Определение значений FR/FA по степени сходства/различия и качественный вывод о сходстве/различии дикторов.
- Проверка данных на высокую типичность и окончательное формирование всех выходных данных.

В окне **Результат:** отображаются FR и FA, общее количество сравниваемых типов (подтипов), число совпавших и не совпавших типов (подтипов), формулировка заключения.



Сравнение: р_Сигнал 5, р_Сигнал 2

Тип	Подтип	Число	Начало, Гц	Конец, Гц	Макс., Гц	Макс., %	Средн., Гц	Средн., %	Мин., Гц	Мин., %	Интервал тона	Интервал, полутона	Скорость изменения тона	Экссесс кривой тона	Снос	Коефф. изреза
Речевой абзац		4	129	104	175	28	109	24	67	32	108	16.5	-4.665	0.7	0.9	53
		4	113	77	170	46	110	43	67	51	104	16.2	-5.5	0.3	0.8	52
Процент совпадений			-	-	93%	-	96%	-	93%	-	95%	95%	82%	91%	84%	89%

Результат:

Мера совпавших признаков FR: **80.7**.
 Мера несовпавших признаков FA: **0.2**.
 Дикторы на фонограммах близки по сравниваемым характеристикам.

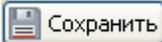
Число сравниваемых (под)типов: 1.
 Число совпавших (под)типов: 1.
 Число несовпавших (под)типов: 0.

Параметры Сохранить Назад

Рисунок 55 – Результат сравнения основного тона

Выделенную в таблице результатов строку, а также содержание окна **Результат**: можно скопировать в буфер обмена, нажав на значок .

Чтобы вернуться к выбору фрагментов, нажмите кнопку .

Чтобы сохранить в проекте сформированную таблицу фрагментов, нажмите кнопку .

Чтобы запустить модуль и открыть таблицы фрагментов из ранее сохранённых сравнений, следует левой кнопкой мыши дважды щёлкнуть на пункте **Сравнение...** в активном проекте (рис. 56).

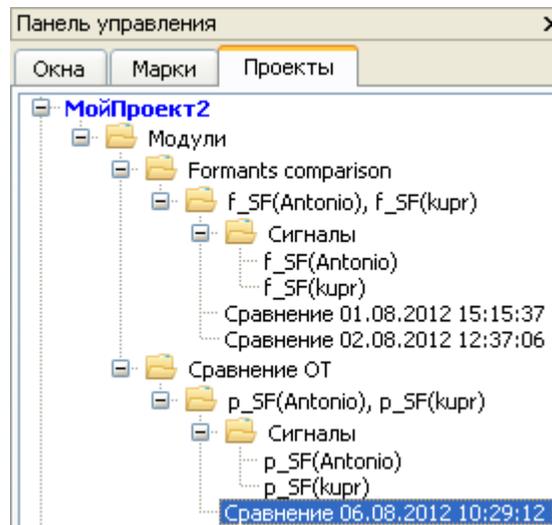


Рисунок 56 – Сохранённые в проекте сравнения основного тона

Чтобы завершить работу с модулем и закрыть панель, нажмите на значок  или  в правом верхнем углу панели.

9 СРАВНЕНИЕ РЕЧЕВЫХ ПРИЗНАКОВ

Метод речевого анализа лишь частично автоматизирован. Вся подготовительную работу по отбору материала, оценке его приемлемости для анализа, выделению признаков и оценке их выраженности в сравниваемых фонограммах выполняет эксперт. Поэтому следует помнить, что надёжность конечного результата речевой идентификации в значительной мере определяется квалификацией эксперта и правильностью оценки им речевых признаков.

Фонограммы, предназначенные для сравнения методом речевого анализа, должны быть приемлемого качества, т.е. обеспечивать адекватное восприятие на слух существенных для речевого анализа характеристик гласных и согласных.

Затруднению речевого анализа (вплоть до невозможности его применения) способствует высокий уровень фонового шума на фонограмме, перегрузка сигнала, высокая реверберированность, различия в канале записи (телефон и микрофон) и/или звукозаписывающей аппаратуре. Кроме того, снизить надёжность идентификационного сравнения могут различия в речевом стиле (чтение/спонтанная речь), в манере произнесения (аккуратная или небрежная, энергичная или вялая), в темпе речи. Наличие или отсутствие визуального контакта между собеседниками, принимающими участие в разговорах на сравниваемых фонограммах (разговор по телефону или при личной встрече собеседников) также может играть роль.

Одно из основных требований правильной идентификации – одинаковый канал записи спорной и образцовой фонограммы и схожее качество записи. Для сравнения слуховым методом микрофонных и телефонных фонограмм рекомендуется предварительно обработать микрофонные фонограммы эквалайзером с полосой пропускания, аналогичной полосе фонограммы телефонного канала.

Кроме требований к качеству записи, важно учитывать и другие факторы, влияющие на степень надёжности идентификационных выводов, в частности, на наличие и степень выраженности отдельных групп слуховых признаков. Важно учитывать влияние разных ситуативных факторов, таких как жанр, время и место разговора, наличие требований и ограничений на поведение дикторов. Одним из главных факторов является эмоциональное состояние диктора.

Данные факторы могут повлиять на следующие группы признаков:

- громкость;
- интонация (в т. ч. темп речи);
- речевые сбои;
- манера речи;
- частотность и долгота хезитационных пауз.

В случае качественной и стилистической сопоставимости фонограмм ведущим фактором, определяющим надёжность сравнения, является их фонетико-лингвистическая представительность. Под *представительностью* понимается наличие в обеих сравниваемых фонограммах контекстов реализации речевых признаков в количестве, достаточном для получения надёжных данных для сравнения. При этом важную роль играет длительность исследуемых фонограмм, поскольку короткие фонограммы (менее 20 секунд), как правило, не обеспечивают должной фонетической представительности.

При длине фонограммы более одной минуты проведение идентификации слуховым методом возможно по всем группам признаков.

При фонограммах меньшей длины (от 10 секунд) идентификация также возможна, но её результат может быть менее точным из-за того, что для принятия экспертом более обоснованного решения о наличии или отсутствии большинства признаков в речи диктора требуется более длительный речевой отрезок. Например, на более длительном речевом отрезке может быть варьирование темпа, высоты и громкости речи диктора, которое не будет представлено на коротком отрезке.

Кроме того, выделение большинства признаков, характеризующих наличие либо отсутствие каких-либо особенностей речи диктора (хезитационные паузы, нарушения речи), может также требовать фонограмм большей длины.

Для получения решения могут остаться незаполненными не более 3 признаков основного списка. При большем количестве незаполненных полей получить решение невозможно.

9.1 Подготовка к работе с модулем



Перед запуском модуля «Сравнение речевых признаков» должен быть задан активный проект, а также подготовлены как минимум два окна со звуковыми данными.

Чтобы применить метод анализа слуховых характеристик, выберите меню **Модули → Сравнение речевых признаков**.

В диалоговом окне (рис. 57) выберите сравниваемые сигналы. Если их в программе всего два, модуль выбирает их автоматически. Если сигналов больше двух, то сравниваемые сигналы выбираются оператором из выпадающих списков.

Кроме того, если ранее в данном проекте уже выполнялись сравнения с выбранным сигналом, то таблицы фрагментов для этого сигнала можно загрузить из ранее сохранённых сравнений, воспользовавшись выпадающим списком **загрузить из**.

Можно также указать один из двух возможных источников записи фонограммы (микрофон либо телефон) и, если у диктора наблюдается явный иностранный акцент, выбрать акцент.



Выбор акцента осуществляется для всех фонограмм, поскольку сравнение фонограмм, где дикторы говорят с разными акцентами либо с акцентом и без акцента, не имеет смысла.

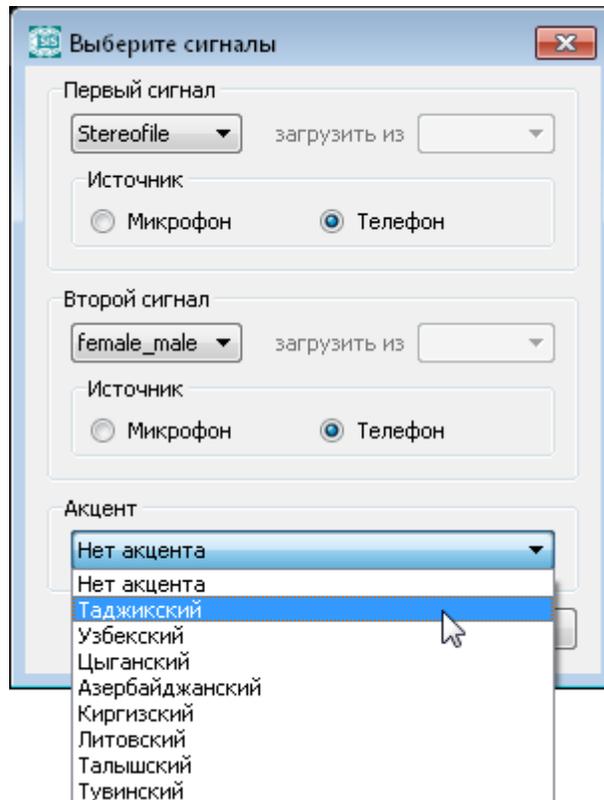


Рисунок 57 – Выбор сравниваемых сигналов

После выбора сравниваемых сигналов, источника и, если необходимо, нужного акцента нажмите кнопку **OK**. Откроется панель работы с модулем. Панель прикрепляется к нижней части главного окна **SIS II**, при этом открытые внутри него окна с данными, панели управления или окно спектра пропорционально меняют свой размер (рис. 58).

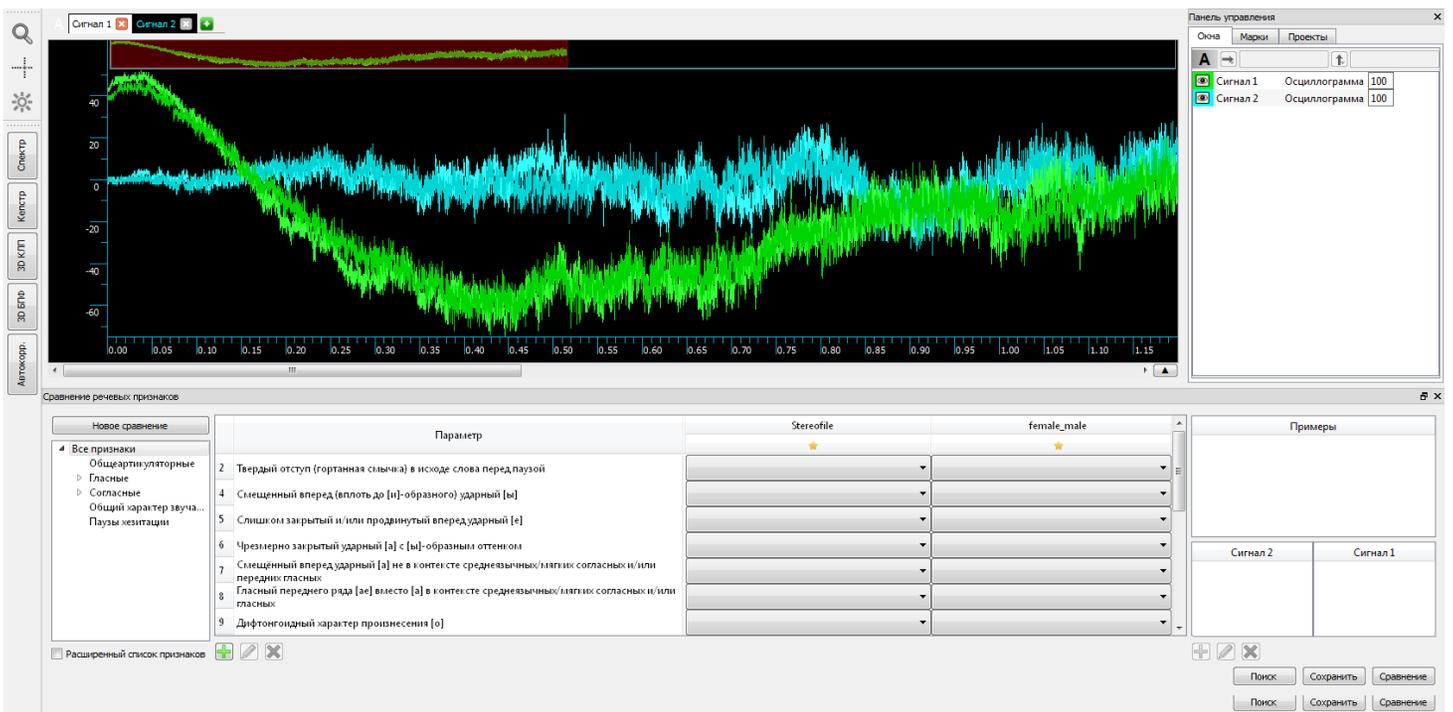


Рисунок 58 – Панель модуля внутри главного окна программы

Чтобы сделать панель модуля независимой от главного окна **SIS II** (рис. 59), следует нажать на значок  в правом верхнем углу панели или дважды щёлкнуть левой кнопкой мыши на заголовке панели. Панель можно перемещать в любое место на экране монитора, удерживая нажатой на заголовке левую кнопку мыши.

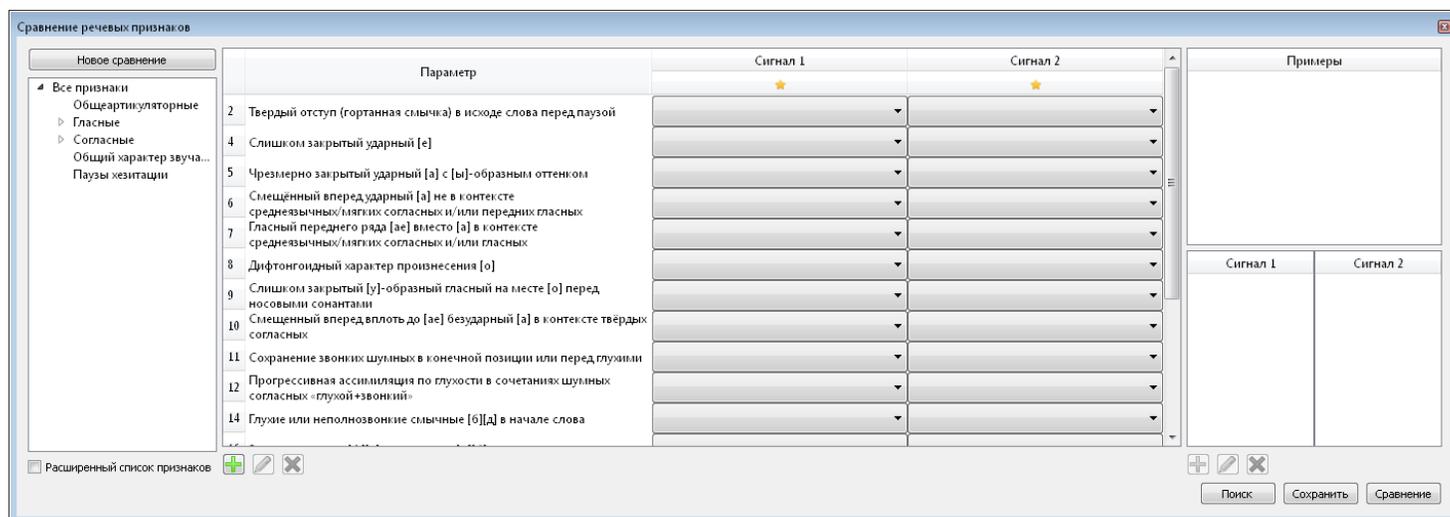


Рисунок 59 – Панель модуля в качестве отдельного независимого окна

При последующих запусках модуля предыдущее положение панели сохраняется. Чтобы прикрепить панель обратно к нижней части главного окна **SIS II**, следует снова дважды щёлкнуть на заголовке панели.

Чтобы выбрать другие сигналы для сравнения, нажмите кнопку **Новое сравнение** и в диалоговом окне **Выберите сигналы** (см. рис. 57) осуществите выбор.

9.2 Сравнение фонограмм

Для сравнения пары открытых в проекте фонограмм эксперт должен внимательно прослушать их, выделяя фрагменты, на которых реализованы идентификационно значимые лингвистические и аудитивные признаки. В модуле предусмотрен большой реестр лингвистических и аудитивных признаков, объединённых в несколько групп для более быстрого их выбора из реестра. Группы признаков организованы в виде раскрывающегося древовидного списка в левой части панели модуля.

Все лингвистические и аудитивные признаки в реестре разделены на основные и расширенные. По умолчанию в таблице в центральной части панели модуля представлены только основные, более частотные признаки. Для доступа к расширенным признакам отметьте галочкой поле **Расширенный список признаков** в левом нижнем углу панели. Чтобы выбрать интересующую группу признаков, щёлкните левой кнопкой мыши по соответствующему пункту древовидного списка. Щёлкнув левой кнопкой мыши по значку  напротив раскрывающегося пункта, можно будет увидеть подпункты, которые он содержит. Свернуть часть дерева признаков можно щелчком левой кнопки мыши по значку  напротив соответствующих пунктов (рис. 60).

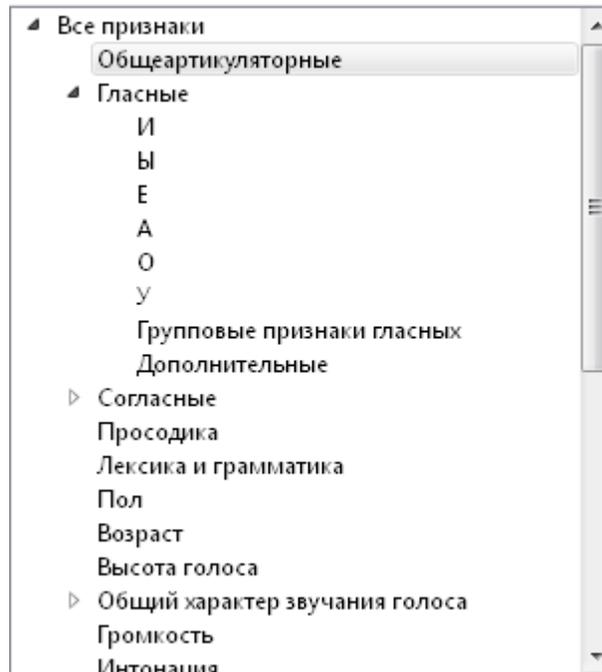


Рисунок 60 – Список групп лингвистических и аудитивных признаков

Таблица в центральной части панели модуля содержит строки с названиями признаков и столбцы, где выбирается их присутствие или отсутствие на обеих сравниваемых фонограммах.

Над заголовками столбцов присутствия/отсутствия признаков, отмеченными знаком , приведены названия файлов сравниваемых фонограмм.

При наведении курсора на название признака из расширенного списка для некоторых региональных признаков может появляться всплывающая подсказка с его описанием, дающая информацию о том, для речи жителей какого региона может быть характерен данный признак.

Несмотря на то, что в программе уже имеется обширный набор речевых признаков для анализа фонограмм, также предусмотрена и возможность добавления новых признаков пользователем, например, для детализации существующих. Для этого нужно нажать на кнопку , расположенную под списком параметров. Откроется диалоговое окно **Добавление признака** (рис. 61), в котором необходимо указать его название, описание (появляющееся в виде всплывающей подсказки), добавить соответствующий звуковой файл с примером, иллюстрирующим данный признак, и ввести текст фразы, произносимой на звуковом фрагменте примера. После этого нажмите кнопку **ОК**, и новый признак будет сохранён в общем реестре и доступен в списке в группе пользовательских признаков.

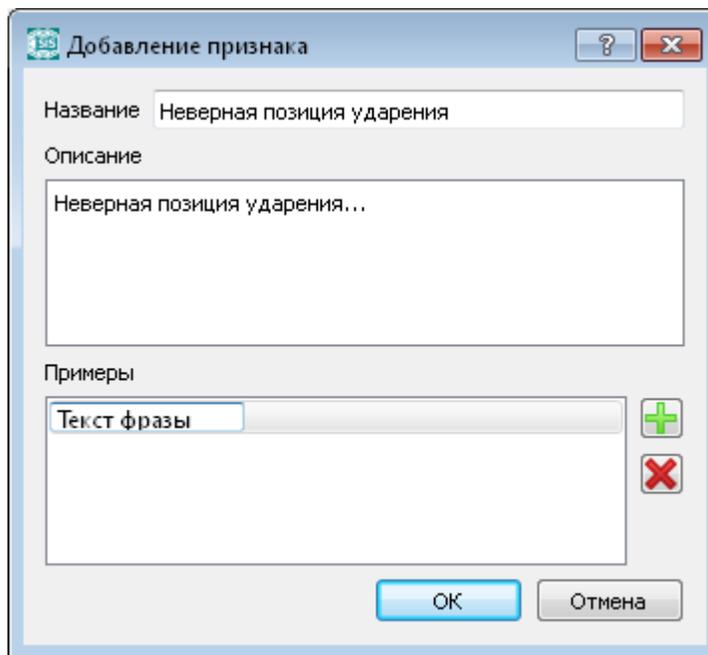


Рисунок 61 – Диалоговое окно добавления признака

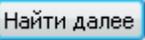
Пользователь может отредактировать признаки. Для этого нужно выделить его в таблице нажатием левой кнопки мыши и нажать на кнопку , расположенную под списком параметров. Откроется диалоговое окно **Редактирование признака**, аналогичное окну добавления признака (рис. 61), в котором можно поменять соответствующие данные признака, убрать или добавить новые файлы с примерами его реализации. Для изменения названия или описания признака наведите курсор на соответствующие поля и сделайте щелчок левой кнопкой мыши. Для корректировки текста примера сделайте двойной щелчок мыши на его фразе. Чтобы удалить звуковой пример, нажмите на кнопку . После внесения изменений нажмите кнопку **ОК**, чтобы сохранить их, или кнопку **Отмена**, чтобы отменить редактирование признака.

Чтобы удалить добавленный признак, нужно выделить его в таблице нажатием левой кнопки мыши и нажать на кнопку  внизу центральной части панели модуля.



Пользователь может удалять только добавленные им признаки.

Для быстрого поиска определенного признака нажмите кнопку .

В открывшемся окне (рис. 62) введите название (или его часть) искомого признака, установите направление поиска в списке признаков и нажмите кнопку .

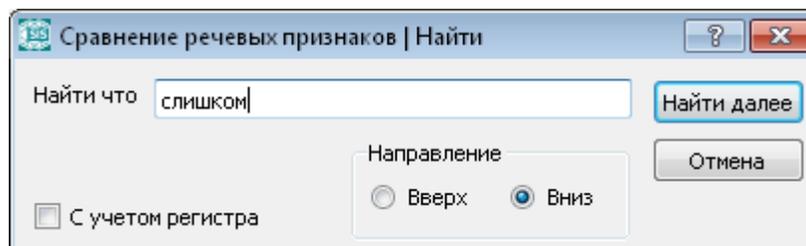


Рисунок 62 – Окно поиска признака

Признак, название которого содержит текст, заданный в качестве ключа поиска, выделяется в списке. Для поиска всех признаков, содержащих текст, введенный в качестве ключа поиска, необходимо последовательно нажимать кнопку **Найти далее**.

Если поиск среди списка не дал результатов, будет отображено информационное окно (рис. 63).

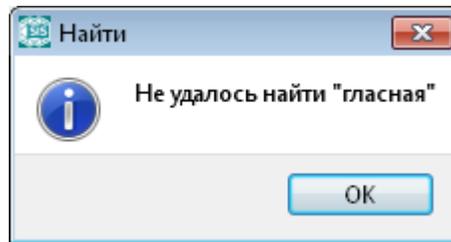


Рисунок 63 – Отсутствие результатов поиска признака

Выделив признак, можно прослушать заданные для него звуковые примеры. Для этого в верхней правой части панели модуля в поле **Примеры** необходимо дважды щелкнуть мышью на строке одного из примеров (рис. 64). Прозвучит звуковой фрагмент, иллюстрирующий данный признак.

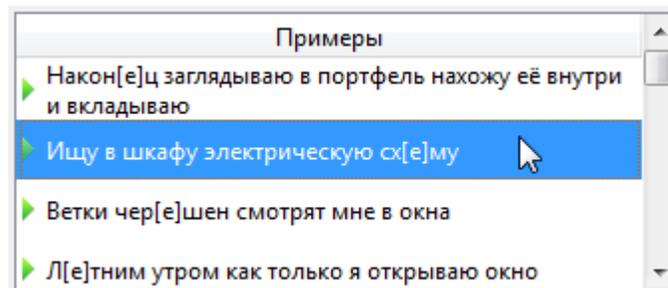


Рисунок 64 – Примеры звуковых фрагментов

Для проведения сравнения речевых признаков необходимо отметить выраженность признака (см. рис. 65).

Доступность параметров для выбора в рамках каждого конкретного признака определяется природой последнего. Так, например, для признака **Логические акценты** можно указать только его наличие, отсутствие (**Да**, **Нет**), для признака **Пол** – один из двух взаимоисключающих параметров (**Мужской/Женский**), а для признака **Темп** можно также указать один из четырех параметров его проявления (**Медленный темп**, **Средний темп**, **Быстрый темп** и **Переменный темп**).

В случае отсутствия контекстов, необходимых для оценки реализации признака, устанавливается значение **Нет реализации**.

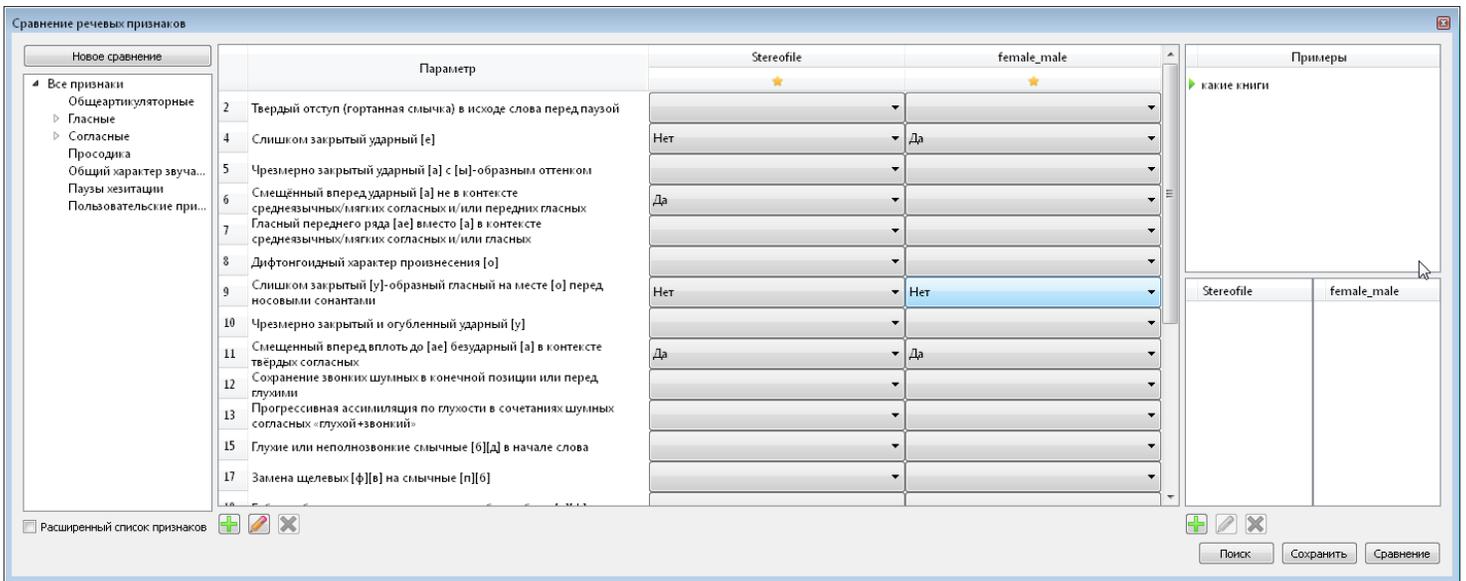


Рисунок 65 – Отметка присутствия признаков

Кроме того, реализацию каждого признака на исследуемых фонограммах можно проиллюстрировать. Для этого, выделив фрагменты сигналов на фонограммах и соответствующий им признак в таблице, в нижней правой части панели модуля в таблице реализаций признаков (рис. 66) нажмите на кнопку  под таблицей и задайте название реализации данного признака на соответствующей фонограмме (как правило, в качестве названия служит фраза, произносимая в данном фрагменте). В окне данных сигнала появится вертикальная марка с таким же названием (рис. 67).

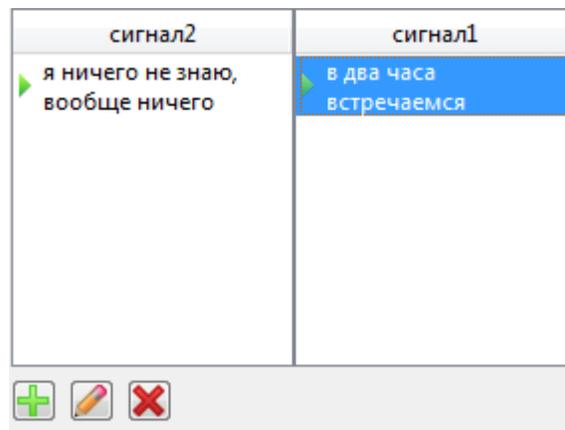


Рисунок 66 – Таблица реализации признаков



Рисунок 67 – Марка реализации признака в окне данных

Если сделать двойной щелчок мыши по строке реализации в таблице, соответствующая вертикальная марка займёт всю видимую область окна данных сигнала.

Чтобы отредактировать реализацию, нажмите на кнопку ; чтобы удалить реализацию, нажмите на кнопку .

Для сохранения выполненных процедур в проекте нажмите на кнопку .

Отметив все выявленные признаки, нажмите на кнопку , для перехода на другую страницу модуля. Эта страница содержит таблицу совпадения признаков (рис. 68) и поле результата сравнения (рис. 69).

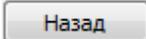


Если для сравнения использовались признаки расширенного списка, то необходимо сначала закрыть расширенный список и только после этого проводить сравнение.

В таблице указано совпадение/несовпадение признаков.

	Параметр	Совпадение
1	Твердый приступ гласного в начале слова или морфемы	Совпадение
2	Твердый отступ (гортанная смычка) в исходе слова перед паузой	Совпадение
3	Значительная назализация гласных [а] и [о] перед [н]	Совпадение
4	Слишком закрытый ударный [е]	Совпадение
5	Чрезмерно закрытый ударный [а] с [ы]-образным оттенком	Несовпадение
6	Смещённый вперед ударный [а] не в контексте среднеязычных/мягких согласных и/или передних гласных	Совпадение

Рисунок 68 – Таблица совпадения признаков

При необходимости изменить выставленные критерии или выполнить другие действия на предыдущей странице, нажмите кнопку .

Закончив определение присутствия/отсутствия и выбора значений признаков, эксперт получает статистику совпадения и несовпадения признаков и результирующую оценку идентификации диктора в поле **Результат** (рис. 69).

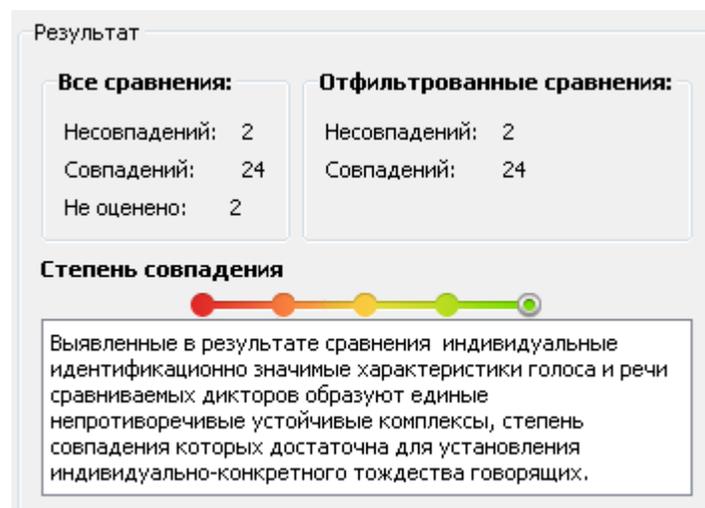


Рисунок 69 – Поле результата сравнения

Если не отмечено более 3 признаков (установлено значение **Нет реализации**), принятие определенного идентификационного невозможно. В этом случае в поле результатов сравнения красным цветом в строке **Нет реализации:** выделяется количество неотмеченных признаков (рис. 70).

Результат	
Все сравнения:	Отфильтрованные сравнения:
Несовпадений: 2	Несовпадений: 2
Совпадений: 8	Совпадений: 8
Нет реализации: 18	

Рисунок 70 – Выделение количество неотмеченных признаков

Красная зона на индикаторе **Степень совпадения** означает различие, зелёная – совпадение.

Идентификационное решение принимается на основе следующих условий:

- При несовпадении 16 и более признаков, делается вывод о *категоричном различии дикторов*.
- Если число несовпадений выбранных признаков находится в пределах от 10 до 15, делается вывод о *вероятном различии дикторов*.
- Если число несовпадений выбранных признаков находится в пределах от 5 до 9, делается вывод о *вероятном совпадении дикторов*.
- При несовпадении менее 5 признаков, делается вывод о *категоричном совпадении дикторов*.

Для сохранения результата сравнения фонограмм в проекте нажмите кнопку **Сохранить**.

Чтобы запустить модуль и открыть данные из ранее сохранённых сравнений, следует сделать двойной щелчок мыши на пункте **Сравнение...** в активном проекте (рис. 71).

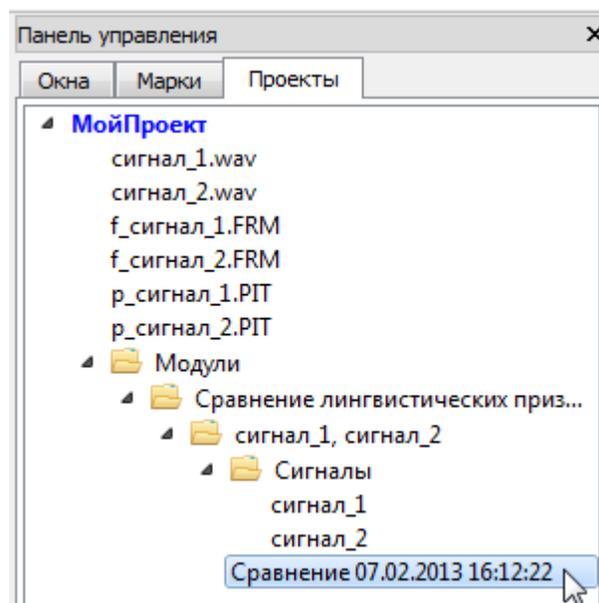


Рисунок 71 – Сохранённые в проекте сравнения лингвистических признаков

Чтобы завершить работу с модулем и закрыть панель, нажмите на значок **×** или **☒** в правом верхнем углу панели.

10 ПРИНЯТИЕ ОБЩЕГО ИДЕНТИФИКАЦИОННОГО РЕШЕНИЯ

Основываясь на результатах работы каждого из описанных выше модулей, программа позволяет принять общее идентификационное решение. При этом эксперт имеет возможность настраивать применимость каждого из отдельных методов исследования, изменяя таким образом, его вес в общем решении. Применимость также может быть выставлена автоматически по результатам расчёта качественных и количественных характеристик сравниваемых фонограмм.

Модуль просматривает сохранённые результаты сравнений по всем модулям в активном проекте. Если для исследуемой пары фонограмм имеются результаты нескольких сравнений по какому-либо из модулей, пользователь может выбрать из списка сравнение, результаты которого должны учитываться для принятия общего решения.



Перед запуском модуля «Общее решение» должен быть задан активный проект, в рамках которого уже были проведены сравнения различными методами идентификации дикторов и сохранены их результаты для исследуемой пары фонограмм. Также в программе должны быть открыты все сигналы, которые использовались в выбранных сравнениях.

Чтобы открыть модуль «Общее решение», выберите меню **Модули** → **Общее решение**.

Откроется диалоговое окно модуля (рис. 72). Окно модуля «Общее решение» содержит таблицу сравнений фонограмм с помощью различных методов. Результаты сравнения по каждому из них отражены в столбце **LR** (отношение правдоподобия). Качественные характеристики фонограмм для сравнений по каждому из методов отображаются в левом столбце в виде индикатора зелёного, жёлтого или красного цвета, соответствующего хорошему, среднему или плохому качеству данных.

Эксперт может изменить выставленное по умолчанию значение применимости для различных методов. Например, если речевой анализ проводил не очень опытный эксперт, можно понизить вес его оценки. Для изменения применимости передвиньте ползунок в нужное положение или задайте точное значение с помощью поля ввода.

Методы, в основе которых лежит анализ основного тона (**ОТ** и **Сравнение основного тона**), сильно зависят от эмоционального состояния диктора. Например, если человек возбуждён, то увеличивается не только среднее значение основного тона, но также могут сильно меняться и все остальные статистические характеристики ОТ. Поэтому если выставить или снять флажок **Различается эмоциональное состояние**, значения применимости названных методов будет пересчитываться в меньшую или большую сторону соответственно в зависимости от качества данных.

Если после ручной коррекции величин применимости нажать кнопку **Установить по качеству**, то её значения будут пересчитаны в зависимости от качества фонограммы (будут выставлены по умолчанию).

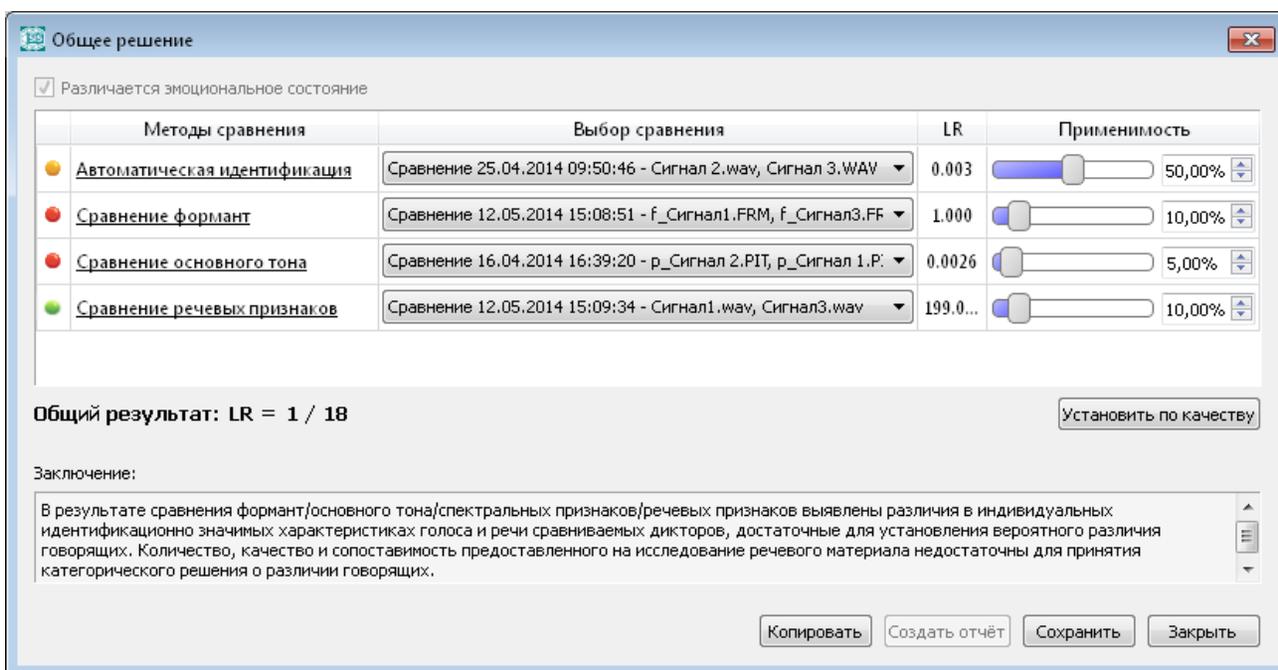


Рисунок 72 – Диалоговое окно модуля «Общее решение»

Под основной таблицей в окне модуля приводится общий результат – отношение подобия (LR) по всем использованным сравнениям – и даётся общее решение в виде экспертного заключения.

Чтобы скопировать полученные результаты в виде данных для включения в текстовый отчёт, нажмите кнопку **Копировать**.

Для сохранения результата сравнения фонограмм в проекте нажмите кнопку **Сохранить**.

Для генерирования отчета об исследовании в формате PDF нажмите кнопку **Создать отчёт**.



Кнопка создания отчета не будет доступна при выборе сравнений разных сигналов.

Чтобы запустить модуль и открыть данные из ранее сохранённых сравнений, следует сделать двойной щелчок мыши на пункте **Сравнение...** в активном проекте (рис. 73).

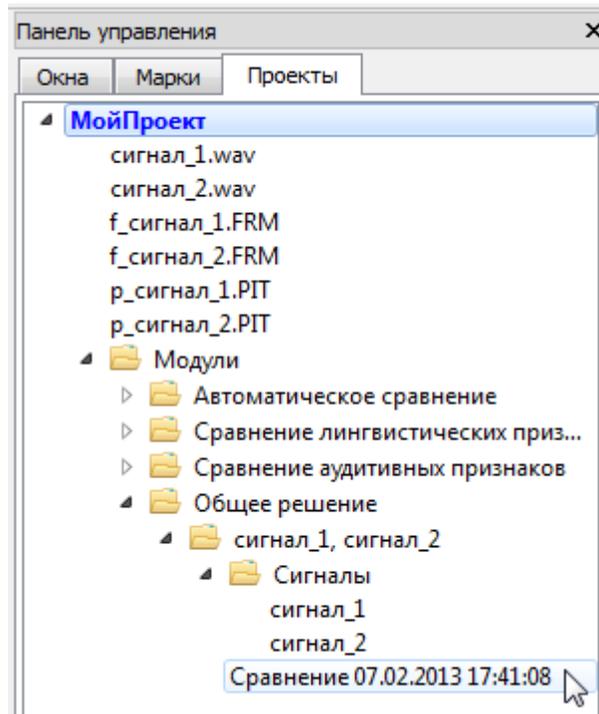
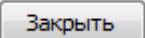


Рисунок 73 – Сохранённые в проекте общие идентификационные решения

Чтобы завершить работу с модулем, нажмите на значок  в правом верхнем углу диалогового окна или на кнопку .

11 СООБЩЕНИЯ, ВЫДАВАЕМЫЕ ОПЕРАТОРУ

Если при выборе вычисления характеристик сигнала активным является окно, не содержащее осциллограммы этого сигнала, появится сообщение, представленное на рисунке 74.

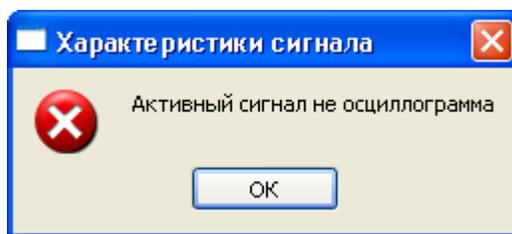


Рисунок 74 – Сообщение об ошибке выбора данных для расчёта характеристик

Нажмите кнопку **ОК** и сделайте активным окно с нужным типом сигнала.

Если для детектирования речи выбрано окно, не содержащее осциллограммы сигнала, появится сообщение, представленное на рисунке 75.

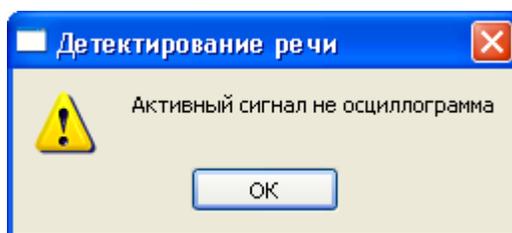


Рисунок 75 – Предупреждение об отсутствии нужного типа сигнала

Нажмите кнопку **ОК** и сделайте активным окно с нужным типом сигнала.

Если речь в выбранных для идентификации файлах не отсегментирована, появится окно предупреждения (рис. 76).

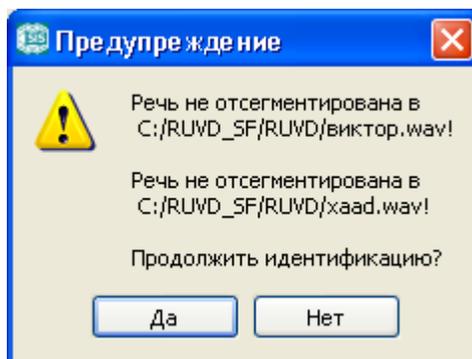


Рисунок 76 – Предупреждение о наличии несегментированной речи

Для повышения надёжности идентификации нажмите кнопку **Нет**, отсегментируйте речь при помощи модуля «Детектирование речи» и повторите идентификацию.

Если перед запуском модулей «Сравнение формант» и «Сравнение основного тона» не задан активный проект, появится окно предупреждения (рис. 77).

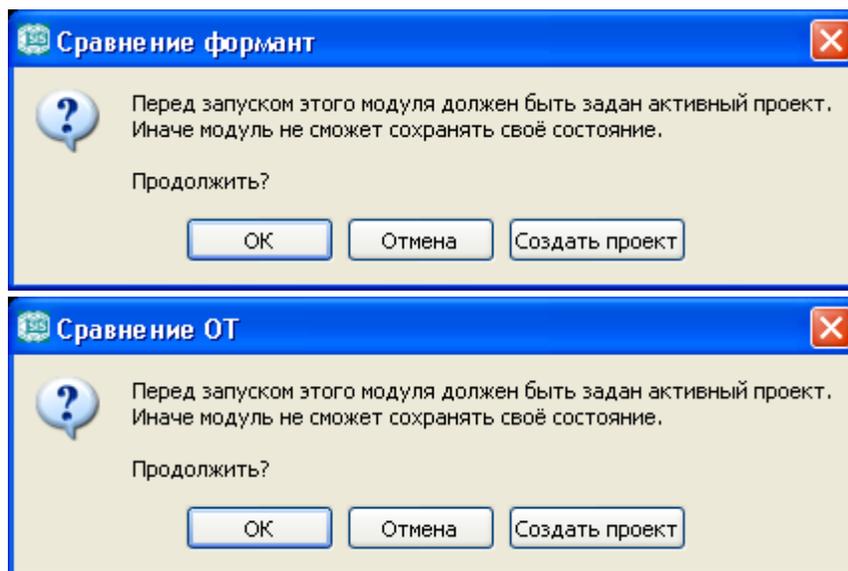


Рисунок 77 – Предупреждение об отсутствии активного проекта

Нажмите кнопку **ОК** для продолжения, если сохранения состояния модуля не требуется.

Нажмите кнопку **Отмена** и сделайте активным один из проектов, уже имеющих на вкладке **Проекты** панели управления.

Нажмите кнопку **Создать проект** и создайте проект, если сохранение результатов работы модуля должно выполняться в новом проекте.

При попытке создать новый проект с уже использованным именем появится окно предупреждения (рис. 78).

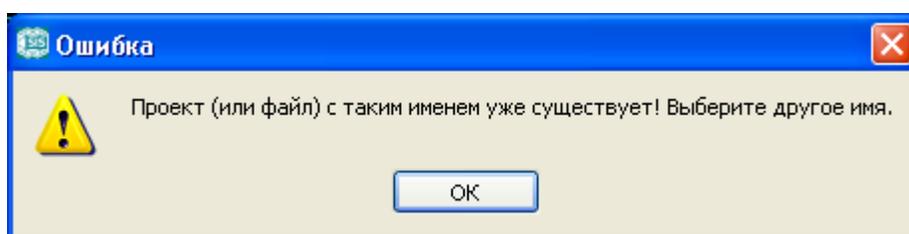


Рисунок 78 – Предупреждение о невозможности присвоить проекту указанное имя

Нажмите кнопку **ОК** и задайте другое имя проекту.

Если перед запуском модуля «Сравнение формант» или «Сравнение основного тона» в активном проекте отсутствуют вычисленные форманты или основной тон или они вычислены только для одной фонограммы, появится окно предупреждения (рис. 79).

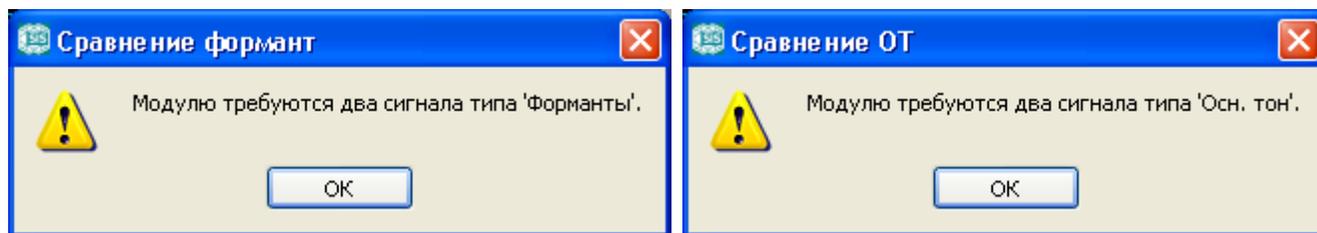


Рисунок 79 – Предупреждение об отсутствии сигналов требуемого типа

Нажмите кнопку **ОК** и выполните вычисление требуемых данных при помощи соответствующих модулей.

Если в диалоговом окне **Укажите пару сигналов** при выборе сигналов для сравнения формант или основного тона выбран один и тот же сигнал, появится окно предупреждения (рис. 80).

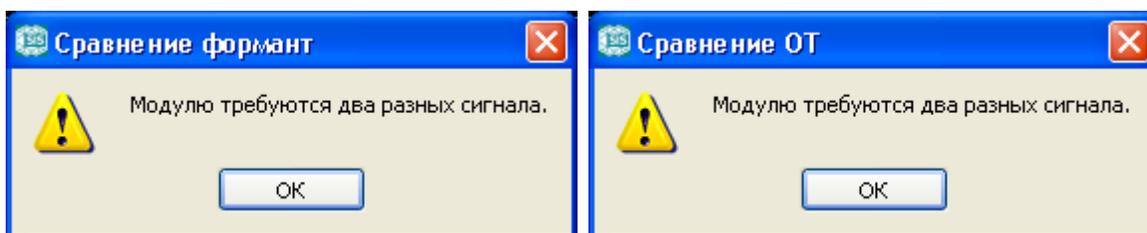


Рисунок 80 – Предупреждение о необходимости выбора для сравнения разных сигналов

Нажмите кнопку **ОК** и выполните выбор разных сигналов.

Если при выполнении сравнения ОТ выбрано недостаточно фрагментов, появится окно предупреждения (рис. 81).

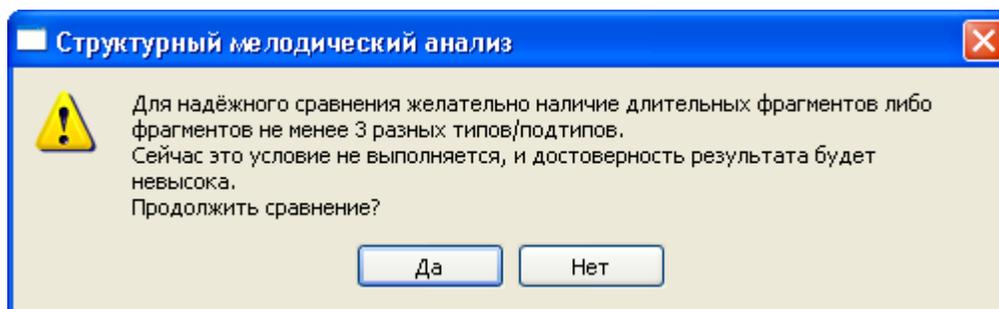


Рисунок 81 – Предупреждение о необходимости выбора большего числа фрагментов

Нажмите кнопку **Нет** и выполните более тщательный и полный выбор фрагментов.

ПРИЛОЖЕНИЯ

Приложение А: Пояснения к характеристикам качества фонограммы

А.1 Неравномерность амплитудно-частотных характеристик

Амплитудно-частотная характеристика (АЧХ) – функция, показывающая зависимость модуля некоторой комплекснозначной функции от частоты. Чаще всего означает модуль комплексного коэффициента передачи линейного четырёхполюсника. Также может рассматриваться АЧХ других комплекснозначных функций частоты, например, спектральной плотности мощности сигнала.

Неравномерность АЧХ – разность между максимальным и минимальным вносимым затуханием в рабочей полосе пропускания формирующего фильтра, это степень отклонения от прямой, параллельной оси частот.

Идеальной для воспроизведения звука была бы акустика, имеющая АЧХ в виде прямой линии. К сожалению, АЧХ реальных акустических систем представляют собой кривые, имеющие множество пиков и провалов. Появление этой неравномерности при воспроизведении звуков различной частоты вызвано не идеальностью, как компонентов, так и акустической системы в целом, наличием различного рода паразитных резонансов, вибрации стенок корпуса и т. п. Чем равномернее АЧХ, тем более естественным будет воспроизведение.

Степень неравномерности АЧХ характеризуется отношением максимального значения звукового давления к минимальному или, по другой методике, отношением максимального (минимального) значения к среднему в заданном диапазоне частот, выраженного в децибелах.

А.2 Отношение сигнал/шум

Одним из основных показателей, влияющих на качество идентификации диктора, результатов акустической обработки и т. д., является отношение уровня исходного речевого сигнала к уровню мешающего шума, так называемое *отношение сигнал/шум (ОСШ)*. В данном случае под уровнем сигнала понимается его мощность. В модуле **Характеристики сигнала** вычисляется матрица отношения сигнал/шум: $ОСШ(k, m)$ (здесь k и m , соответственно, индексы частоты и кадра), по которой в качестве выходных параметров можно получить:

- 1) Интегральное ОСШ по всему файлу в частотных полосах.
- 2) Интегральное ОСШ по всему файлу.

Алгоритм оценки ОСШ реализован следующим образом.

С помощью детектора речевой активности (*Voice Activity Detector – VAD*) помечаются те участки входного сигнала, на которых присутствует речь, и отсутствуют помехи (щелчки, всплески тона и т. п.). Речевые участки нарезаются на кадры, на которых вычисляется спектр мощности (СПМ) сигнала.

По спектру мощности вычисляются текущие оценки спектров мощностей входного сигнала $|X(k, m)|^2$ и шума $|N(k, m)|^2$ (здесь k и m , соответственно, индексы частоты и кадра). При этом из анализа исключаются участки сигнала, помеченные как «неречевые».

Далее, отсекаются малоинформативные и потенциально опасные с точки зрения возможных шумов полосы (менее 100 Гц и выше 3300 Гц), и в полученной полосе частот вычисляется отношение:

$$INR(m) = 1/K \sum_{k=1, K} [|X(k, m)|^2 / |\hat{N}(k, m)|^2],$$

представляющее усреднённую величину, характеризующую, насколько на данном временном кадре СПМ входного сигнала больше СПМ шума, т. е. некоторую общую оценку величины зашумленности на кадре.

Величина $INR(m)$ сравнивается с порогом, и на дальнейшую обработку посылаются только кадры с $INR(m)$, превышающим заданный по умолчанию порог. Далее вычисляются:

1) Интегральное ОСШ по всему файлу:

В полосе частот 100–3300 Гц вычисляется ОСШ на временных кадрах:

$$ОСШ(m) = \langle |X(k, m)|^2 \rangle / \langle |\hat{N}(k, m)|^2 \rangle - 1,$$

где $\langle \rangle$ – операция усреднения по всем частотным полосам.

Затем ОСШ на временных кадрах рекурсивно накапливается для получения оценки интегрального ОСШ по всему файлу:

$$E[ОСШ] = 1/M \sum_{m=1, M} [ОСШ(m)],$$

где M – общее число кадров в файле.

2) Интегральное ОСШ в частотных полосах по всему файлу:

По всем частотным полосам, начиная от нулевой частоты до частоты Найквиста, вычисляются ОСШ в частотных полосах на временных кадрах:

$$ОСШ(k, m) = \max\{ 0, |X(k, m)|^2 / |\hat{N}(k, m)|^2 - 1 \}.$$

Затем рекурсивно вычисляется интегральное ОСШ в частотных полосах по всему файлу:

$$E[ОСШ(k)] = 1/M \sum_{m=1, M} [ОСШ(k, m)],$$

где M – общее число кадров в файле.

Перевод каждого из видов ОСШ в децибелы осуществляется по известной формуле:

$$ОСШ_{дБ} = 10 \lg(ОСШ).$$

А.3 Стационарные тональные помехи

Помеха – нежелательное физическое явление или воздействие электрических, магнитных или электромагнитных полей, электрических токов или напряжений внешнего или внутреннего источника, которое нарушает нормальную работу технических средств или вызывает ухудшение технических характеристик и параметров этих средств.

Тональные помехи (ТП) – это периодические сигналы с частотой основного периода от нескольких десятков герц и выше.

К помехам относятся, прежде всего, сигналы, спектр которых пересекается со спектром РС, и амплитуда спектральных компонент соизмерима с амплитудами спектральных компонент РС (в противном случае они не относятся к помехам). Примеры (классы тональных помех):

- периодические импульсные процессы, порождающие в спектре гребёнку гармоник основного тона;
- тональные сетевые наводки (50–60 Гц) и их гармоники;

- акустические шумы компьютера (пики спектра в НЧ области);
- тональные импульсы в телефонной линии и т. п.;
- акустические помехи искусственного происхождения (сирены, музыка и пр.).

А.4 Время реверберации

Реверберация – это процесс постепенного уменьшения интенсивности звука при его многократных отражениях. Иногда под реверберацией понимается эмуляция данного эффекта с помощью ревербераторов. Условно принятое время реверберации – время, за которое уровень отраженного звука уменьшается на 60 дБ.

Для вычисления времени реверберации используют формулу, которая принадлежит Сэбину, первому исследователю архитектурной акустики:

$$T = \frac{0.164V}{A},$$

где V – это объём помещения, A – общий фонд звукопоглощения,

$$A = a_1S_1 + a_2S_2 + \dots$$

a_i – коэффициент звукопоглощения (зависит от материала, его дисперсных или фрикционных характеристик), S_i – площадь каждой поверхности.

А.5 Клиппирование

Клиппирование или *клиппинг* (англ. *clipping* или *clip* – обрезание, отсечение) – одна из форм искажения формы волны, которое происходит при перегрузке усилителя и при превышении выходным напряжением усилителя предела напряжения питания. Субъективно выражается в появлении «шипения» и «треска».

В цифровой обработке клиппинг происходит, когда сигнал заходит за определённый ограниченный диапазон. Например, в 16-битной системе значение 32 767 является наиболее положительным (которое может быть представлено), и если во время обработки сигнала амплитуда будет увеличена в два раза, например 32 000 должно стать 64 000, то вместо этого амплитуда усекается до максимальной – 32 767. В результате этого верхушки волн становятся плоскими и происходят грубые искажения сигнала.

А.6 Длительность речи

Речь – это «конкретное говорение, протекающее во времени и облеченное в звуковую (включая внутреннее проговаривание) или письменную форму. Под речью принято понимать как сам процесс говорения, так и результат этого процесса, т. е. и речевую деятельность, и речевые произведения, фиксируемые памятью или письмом».

В данном контексте под речью и её длительностью понимаются участки фонограммы, значимые с точки зрения содержащейся в них сугубо речевой информации и её процентного соотношения с менее значимой неречевой информацией (например, шумом).

Данная характеристика речевого сигнала является принципиальной с точки зрения идентификационной значимости речи, поскольку напрямую связана с количеством реализуемых в полезном сигнале как акустических, так и лингвистических признаков.

Приложение Б: Пояснения к доверительным интервалам

Все получаемые в модуле идентификации результаты сравнения речевых сигналов в двух файлах основаны на выделении в них идентификационно значимых признаков и вычислении меры близости полученных наборов признаков между собой. Для каждого значения этой меры близости во время так называемого периода обучения модуля идентификации были получены вероятности совпадения и различия дикторов, речь которых содержалась в сравниваемых файлах. Эти вероятности были получены разработчиками на большой обучающей выборке фонограмм: десятки тысяч дикторов, различные каналы звукозаписи, много сессий звукозаписи, разнообразный тип речевого материала.

В виду огромного количества трудно контролируемых факторов, потенциально способных повлиять на характеристики речевого сигнала, получаемые вероятности совпадения-различия дикторов точно соответствуют данным обучающей выборки только «в среднем». То есть, только при проведении множественных сравнений аналогичных файлов. Само значение получаемой вероятности совпадения-различия дикторов является случайной величиной, могущей несколько изменяться от случая к случаю. Применение статистических данных к единичному случаю сравнения файл-файл требует учёта возможного разброса получаемых значений меры близости двух файлов и соответствующей ей вероятности совпадения/различия дикторов в зависимости от различных деталей ситуации произнесения речи.

Для таких величин в математической статистике предложено использовать понятие доверительного интервала, разработанное американским статистиком Ю. Нейманом (Neyman J.), на основе идей английского статистика Р. Фишера (Fisher R.).

В контексте применения этого понятия в модуле идентификации дикторов *доверительный интервал* (англ. *credible interval* или *confidence interval*) – интервал значений рассматриваемой величины, в который она попадает при данной мере близости речевых сигналов двух сравниваемых файлов с заданной надёжностью (уровнем доверия, доверительностью, уровнем доверительности), изменяемой в процентах от 0 до 100.

Интуитивная интерпретация доверительного интервала такова, если он достаточно велик (например, 95 % или 99 %), то получаемая в результате работы модуля величина почти наверняка попадёт внутрь рассчитанного доверительного интервала значений данной величины. Значения величины внутри доверительного интервала хорошо соответствуют экспериментальным данным обучающей выборки и не противоречат им для соответственно 95 % или 99 % дикторов, фонограммы с речью которых присутствовали в выборке.

Таким образом, если, например, LR при равном 95 % имеет границы доверительного интервала от 10 до 20 при среднем значении 15, то это следует понимать так: «Если при обучении модуля при сравнении двух речевых сигналов была получена такая же мера близости между ними, как и при текущем сравнении, то для 95 % дикторов в обучающей выборке LR находилась в пределах от 10 до 20, в среднем 15». Если мы хотим использовать это значение в своих экспертных выводах далее, то на основе данного результата можно сделать вывод: сравниваемые дикторы совпадают с отношением правдоподобия LR не менее 10, что верно для не менее 95 % дикторов, могущих быть авторами речи в сравниваемых файлах.

Приложение В: Элементы мелодического контура и параметры их описания

Используемый набор элементов (типов и подтипов) мелодического контура и параметров их описания представлен ниже в таблице.

№	Тип	Подтип	Используемые параметры	Список параметров
1	1. Ядерный слог	Восходящий	1..11, 15	1. Начальная частота, Гц 2. Конечная частота, Гц 3. Максимум, Гц 4. Время максимума (%) 5. Минимум, Гц 6. Время минимума (%) 7. Интервал-Гц 8. Интервал-ПТ 9. Время половинной частоты (%) 10. Средняя частота, Гц 11. Скорость изменения тона (полутона в секунду) 12. Скос 13. Экссесс 14. Коэффициент изрезанности 15. Длительность, мс
2		Нисходящий	1..5, 7..11, 15	
3		Восходяще-нисходящий	1..13, 15	
4		Нисходяще-восходящий	1..15	
5	2. Ядерно-заядерный участок	Восходящий	1..15	
6		Нисходящий	1..5, 7..15	
7		Восходяще-нисходящий	2..5, 8, 10..15	
8		Нисходяще-восходящий	1..5, 7, 8, 10..15	
9	3. Шкала	Нисходящая	1..8, 14	
10		Восходящая	1..8, 10, 14	
11	4. Предшкала (и безударный участок контура)	Низкая	3, 5, 10	
12		Высокая	3, 10	
13	5. Синтагма	Нисходящее завершение	2..6, 8, 14	
14		Восходящее завершение	2..6, 8, 14	
15	6. Речевой абзац		3, 5, 7, 8, 10..14	
16	7. Пауза хезитации		3, 5, 10	
17	8. Предъядерный слог	Восходящий	1..10, 12..15	
18	+ядро	Нисходящий	1..4, 6..8, 10, 12..15	

В столбце «Используемые параметры» цифрами обозначены те параметры, которые участвуют в идентификационном сравнении элементов данного типа (подтипа). Текстовые соответствия содержатся в следующем столбце «Список параметров».

Общее число структурных элементов мелодического контура составляет 18, а число параметров, используемых для их описания, – 15.

Помимо стандартных составляющих мелодического контура в качестве самостоятельных единиц анализа используются заполненные паузы хезитации, а также тип «Речевой абзац» – последовательность из нескольких синтагм или высказываний, на которой осуществляются долгосрочные усреднения значений параметров (рекомендуемая длительность фрагмента – не менее 10 секунд).

Для выделенного фрагмента контура рассчитывается следующий набор параметров:

1. *Начальная частота* – значение первого отсчёта (Гц) в начальной точке выделенного фрагмента контура.

2. *Конечная частота* – значение последнего отсчёта (Гц) в конечной точке выделенного фрагмента контура.

3. *Максимальная частота* – максимальное значение частоты ОТ (Гц) в пределах выделенного фрагмента контура.

4. *Время максимума* – координата максимального значения в процентах от общей длительности выделенного фрагмента.

5. *Минимальная частота* – минимальное значение частоты ОТ (Гц) в пределах выделенного фрагмента контура.

6. *Время минимума* – координата минимального значения в процентах от общей длительности выделенного фрагмента.

7. *Интервал* – разница между максимальным и минимальным значением частоты ОТ (Гц).

8. *Интервал* – разница между максимальным и минимальным значением частоты ОТ (ПТ).

9. *Время половинной частоты* – координата значения половинной частоты (от интервала между максимумом и минимумом) в процентах от общей длительности выделенного фрагмента.

10. *Средняя частота* – среднее значение частоты ОТ (Гц) в пределах выделенного фрагмента контура.

11. *Скорость изменения тона* – средняя скорость убывания или возрастания тона на выделенном участке контура в полутонах на миллисекунду.

12. *Скос* (асимметрия) характеризует степень несимметричности распределения значений ЧОТ относительно среднего значения: положительная асимметрия указывает на отклонение распределения в сторону положительных значений; отрицательная асимметрия указывает на отклонение распределения в сторону отрицательных значений.

При анализе восходящих и нисходящих ядерных тонов (для типов «ядерный слог» и «ядерно-заядерный участок») данный параметр характеризует степень выпуклости-вогнутости мелодической кривой. Для более длительных участков речи значение скоса может указывать на наличие-отсутствие резких перепадов значений.

Для вычисления скоса используется формула:

$$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{S} \right)^3$$

13. *Эксцесс* характеризует относительную остроконечность или сглаженность распределения значений ОТ по сравнению с нормальным распределением: положительный эксцесс обозначает относительно остроконечное распределение; отрицательный эксцесс обозначает относительно сглаженное распределение.

При анализе восходящих и нисходящих ядерных тонов (для типов «ядерный слог» и «ядерно-заядерный участок») данный параметр характеризует степень равномерности изменения ЧОТ. Для более длительных участков речи значение эксцесса может указывать на наличие-отсутствие резких перепадов значений.

Для вычисления эксцесса используется формула:

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{S} \right)^4$$

14. *Коэффициент изрезанности* контура используется для оценки степени изрезанности протяжённых участков мелодического контура (синтагм, шкал, длительных речевых участков). Вычисляется как отношение суммы интервалов тона (в полутонах) по модулю к длительности речевого участка.

15. *Длительность* – длительность выделенного фрагмента в миллисекундах.

Для оценки скорости изменения тона, а также для дополнительной оценки интервалов в качестве единиц используются полутона (ПТ).

Шкала полутонов позволяет более объективно оценивать близость интервалов тона (а также крутизны-пологости тональных участков) при различиях в речевом регистре. Кроме того, использование полутонов (в отличие от герц) облегчает приспособление пороговых значений к различиям в высоте голоса (мужской-женский).

Для расчёта интервала в полутонах между двумя значениями в герцах используется следующая формула:

$$\frac{12}{\lg 2} \lg \frac{F_{0\max}}{100} - \frac{12}{\lg 2} \lg \frac{F_{0\min}}{100}$$

где $F_{0\max}$ и $F_{0\min}$ – соответственно максимальное и минимальное значение ЧОТ в герцах. Данная формула взята из справочника звукового редактора PRAAT (<http://www.fon.hum.uva.nl/praat/>).

Набор параметров для каждого используемого (под)типа устанавливался экспериментальным путём, в зависимости от целесообразности его использования при сравнении с тем или иным пороговым значением.

Оптимальное число однотипных опорных фрагментов для получения статистики составляет 10–15.

На основе данного набора статистических параметров эксперт может оценить степень представительности и однородности выборок, предназначенных для обобщения и сравнения, и при необходимости скорректировать наборы элементов.

Дальнейшее сравнение осуществляется автоматически в экспертном режиме на основе полученных средних значений по каждому параметру каждого (под)типа элементов мелодического контура.

Для каждой пары сравниваемых значений находится расстояние (разность значений), которое затем сравнивается с соответствующим порогом. Порог является условным пограничным значением, отделяющим сходные реализации от различных. Степень схождения/различия вычисляется и отображается в процентах.

При совпадении расстояния с порогом степень схождения оценивается в 50 %. При превышении порога значения лежат в области менее 50 % и убывают по мере удаления от порогового значения. Соответственно, если расстояние меньше порога, то значения превышают 50 %. Нулевое расстояние (или полное совпадение) значений параметра даёт 100 % схождения.

Каждый параметр имеет свой весовой коэффициент, определяющий его долю в общей оценке степени схождения/различия по каждому из задействованных (под)типов фрагментов контура. Присвоение весовых коэффициентов мелодическим параметрам осуществлялось с учётом степени их реализационной стабильности с одной стороны, и различающей способности с другой.

Для каждого (под)типа рассчитываются степень схождения (на основе значений более 50 %) и степень различий (на основе значений менее 50 %).

Далее по этим частным значениям (под)типов аналогичным образом, с учётом весовых коэффициентов, рассчитывается общая оценка схождения/различия речевых образцов, на основе которой формулируется общий вероятностный вывод о тождестве/различии голосов и выводятся вероятности ложного отклонения и ложного принятия диктора, рассчитанные для данного соотношения схождения/различия сравниваемых характеристик.

В качестве базового типа при принятии решения используется «речевой абзац». Это связано, прежде всего, с тем, что для корректного выделения участков мелодического контура данного типа от эксперта требуется лишь умение оценить правильность рисунка выделенного ОТ и при необходимости откорректировать его, в то время как эффективность использования данных по другим типам фрагмента контура напрямую зависит от действий эксперта, его квалификации и опыта. Кроме того, задача по выделению «локальных» фрагментов контура требует от эксперта гораздо больше усилий и времени.

Приложение Г: Перечень терминов и определений

В настоящем документе применяют следующие термины с соответствующими определениями.

Акустико-фонетические (фонетические) признаки устной речи – признаки, отражающие акустические свойства речеобразующего тракта, артикуляционные навыки человека. Эта группа признаков воспринимается на слух и выявляется с помощью технических средств, служит основой инструментального анализа фонограмм устной речи; признаки могут быть оценены количественно.

Активная вкладка – вкладка активного (выделенного) окна данных, которая служит в настоящий момент источником данных. Всегда отображается поверх других вкладок.

Амплитуда – величина максимального отклонения при колебательном процессе от некоторого среднего положения (положения равновесия), характеризует величину колебательного движения (размер перемещения). При исследовании гармонических звуковых колебаний под амплитудой понимается звуковое давление в речевом сигнале, выражаемое амплитудой тока, напряжения или другой электрической величиной.

Вкладка данных – независимые данные, помещённые в ходе работы с программой совместно с другими данными в одном окне данных.

Данные – графическое изображение в окне данных информации, полученной при записи звука, чтении файлов, работе с программой **SIS II**. Данными являются осциллограммы, спектрограммы и другие графические изображения, полученные в ходе работы с программой.

Детектор голосовой активности VAD (англ. *Voice Activity Detection*) – программное средство для отделения активной речи от фонового шума или тишины.

Диапазон – величина, характеризующая максимальные пределы изменения признаков звучащей речи; разность между максимальным и минимальным значениями признака.

Диктор – личность, человек, чья речь присутствует в фонограмме.

Звук – поток энергии, вызывающий механические колебания частиц упругой среды; с точки зрения физиологии – субъективно воспринимаемые органами слуха человека механические колебания упругой среды, вызывающие у него определенные ощущения.

Звуки речи – минимальные единицы речевой цепи, являющиеся результатом сложной артикуляционной деятельности человека и характеризующиеся определенными акустическими и перцептивными (связанными с восприятием речи) свойствами.

Марка – средство для выделения отдельных областей данных в окне данных.

Метод идентификации диктора по голосу – алгоритм вычисления (выделения) речевых признаков и алгоритм попарного сравнения признаков.

Окно данных – окном данных в программе называется ограниченная рамкой самостоятельная прямоугольная область в пределах центральной рабочей области главного окна программы, в которой отображаются в виде графического изображения определённые данные (осциллограммы, спектрограммы, гистограммы, форманты и т. д.).

Оператор – должностное лицо организации, использующее программу по назначению.

Основной тон (ОТ) – характеристика речевого процесса, описывающая периодическое закрытие голосовых связок гортани; обычно говорят о частоте основного тона – количестве схлопываний голосовых связок в секунду, или о периоде основного тона – времени между ближайшими схлопываниями голосовых связок.

Основная частота – частота колебания сложной формы, равная частоте колебания первой гармоники дискретного спектра; зависит от размеров и быстроты колебаний голосовых связок. Применительно к устной речи характеризует тип голоса (бас, тенор, дискант и т. д.).

Осциллограмма речевого сигнала – изображение амплитуды колебательного процесса в непрерывной временной последовательности, т. е. в координатах амплитуда-время. С помощью осциллограммы оператор выделяет речевые фрагменты, подлежащие дальнейшему исследованию.

Пауза – перерыв в речи, которому акустически соответствует отсутствие звука, а физиологически – остановка в работе речевых органов. При воспроизведении фонограммы – перерыв в воспроизведении.

Помеха – процесс, мешающий слуховому восприятию полезных сигналов при воспроизведении фонограммы, проявляющийся в виде различного рода шумов, фона и других сигналов, не несущих полезной информации.

Смесь гауссовых распределений (СГР) – линейная комбинация гауссовых функций, используемая для аппроксимации различного рода экспериментальных распределений компонент акустического пространства.

Спектр звука – акустическая характеристика сложного звука, дающая информацию о частоте источника, частоте гармоник основного тона и об относительной интенсивности всех его частотных составляющих.

Фонограмма – предварительно записанный в файл речевой сигнал.

Форманта – термин фонетики, обозначающий акустическую характеристику звуков речи (прежде всего гласных), связанную с уровнем частоты голосового тона. Обозначает определённую частотную область, в которой вследствие резонанса усиливается некоторое число гармоник тона, производимого голосовыми связками. Форманта представляет собой резонансный максимум в огибающей спектра речевого сигнала. Ширина резонансной области предполагается в пределах от 300 до 500 Гц. При вычислении положения форманты все более узкие структуры должны игнорироваться.

Фрагмент – выделенная на активной вкладке часть данных.

Приложение Д: Перечень сокращений

В настоящем документе используются следующие сокращения.

DET	(англ. <i>Detection Error Trade-off Plot</i>) – график отношения ошибок классификации, который визуально отображает зависимость значений вероятностей ошибок первого и второго рода.
DRR	(англ. <i>Direct-to-Reverberant Ratio</i>) – отношение интенсивности прямого звука к интенсивности реверберированного звука.
GMM	(англ. <i>Gaussian Mixture Model</i>) – модель гауссовых смесей.
EER	(англ. <i>Equal Error Rate</i>) – уровень равной ошибки или точка совпадения вероятностей ошибок первого и второго рода.
FA	(англ. <i>False Acceptance</i>) – ошибка принятия чужого диктора за своего (ложное опознавание).
FR	(англ. <i>False Rejection</i>) – ошибка принятия своего диктора за чужого (ошибочный пропуск).
LR	(англ. <i>Likelihood Ratio</i>) – оценка отношения функций максимального правдоподобия (максимально-правдоподобная оценка).
JFA	(англ. <i>Joint Factor Analysis</i>) – объединённый факторный анализ.
MFCC	(англ. <i>Mel Fourier Cepstrum coefficients</i>) – кепстральные коэффициенты по шкале мел.
SVM	(англ. <i>Support Vector Machine</i>) – машина опорных векторов.
VAD	(англ. <i>Voice Activity Detection</i> , а также <i>Silence Suppression</i> – подавление тишины) – детектор голосовой активности для отделения активной речи от фонового шума или тишины.
АЧХ	амплитудно-частотная характеристика.
ОСШ	отношение сигнал/шум.
ОТ	основной тон.
ПТ	полутона – используются в качестве единиц оценки скорости изменения тона, а также для дополнительной оценки интервалов.
СГР	смесь гауссовых распределений.
СФ	спектрально-формантный метод идентификации.
ЧОТ	частота основного тона.

Приложение Е. Таблица весов по типам и подтипам мелодического контура

НАЗВАНИЕ ПАРАМЕТРА	Весовой коэффициент при совпадении	Весовой коэффициент при различии
Имя подтипа. Ядерный слог: восходящий	0.13	
Начальная частота	0.15	0.15
Конечная частота	0.07	0.07
Максимум	0.08	0.08
Время максимума	0.05	0.05
Минимум	0.13	0.13
Время минимума	0.05	0.05
Интервал-Гц	0.05	0.05
Интервал-ПТ	0.07	0.07
Время половинной частоты	0.05	0.05
Средняя частота	0.15	0.15
Скорость изменения тона	0.07	0.07
Длительность	0.01	0.01
Имя подтипа. Ядерный слог: нисходящий	0.06	
Начальная частота	0.08	0.08
Конечная частота	0.14	0.14
Максимум	0.08	0.08
Время максимума	0.01	0.01
Минимум	0.17	0.17
Интервал-Гц	0.07	0.07
Интервал-ПТ	0.08	0.08
Время половинной частоты	0.06	0.06
Средняя частота	0.18	0.18
Скорость изменения тона	0.06	0.06
Длительность	0.01	0.01
Имя подтипа. Ядерный слог: восходяще-нисходящий	0.01	
Начальная частота	0.06	0.06
Конечная частота	0.06	0.06
Максимум	0.12	0.12
Время максимума	0.06	0.06
Минимум	0.15	0.15
Время минимума	0.05	0.05
Интервал-Гц	0.08	0.08
Интервал-ПТ	0.03	0.03
Время половинной частоты	0.02	0.02
Средняя частота	0.13	0.13
Скорость изменения тона	0.01	0.01
Экссесс	0.06	0.06
Скос	0.06	0.06

НАЗВАНИЕ ПАРАМЕТРА	Весовой коэффициент при совпадении	Весовой коэффициент при различии
Длительность	0.04	0.04
Имя подтипа. Ядерно-заядерный участок: восходящий	0.12	
Начальная частота	0.11	0.11
Конечная частота	0.06	0.064
Максимум	0.10	0.10
Время максимума	0.03	0.03
Минимум	0.12	0.12
Время минимума	0.02	0.02
Интервал-Гц	0.06	0.06
Интервал-ПТ	0.09	0.09
Время половинной частоты	0.03	0.03
Средняя частота	0.19	0.19
Скорость изменения тона	0.06	0.06
Эксцесс	0.01	0.01
Скос	0.02	0.02
Изрезанность	0.02	0.02
Имя подтипа. Ядерно-заядерный участок: нисходящий	0.04	
Начальная частота	0.09	0.09
Конечная частота	0.08	0.08
Максимум	0.05	0.05
Время максимума	0.05	0.05
Минимум	0.17	0.17
Интервал-Гц	0.04	0.04
Интервал-ПТ	0.10	0.10
Время половинной частоты	0.08	0.08
Средняя частота	0.17	0.17
Скорость изменения тона	0.04	0.04
Эксцесс	0.01	0.01
Скос	0.03	0.03
Изрезанность	0.03	0.03
Имя подтипа. Ядерно-заядерный участок: восходяще-нисходящий	0.04	
Начальная частота	0.05	0.05
Конечная частота	0.13	0.13
Максимум	0.12	0.12
Время максимума	0.01	0.01
Минимум	0.15	0.15
Интервал-Гц	0.08	0.08
Интервал-ПТ	0.12	0.12
Средняя частота	0.13	0.13
Скорость изменения тона	0.05	0.05

НАЗВАНИЕ ПАРАМЕТРА	Весовой коэффициент при совпадении	Весовой коэффициент при различии
Экссесс	0.03	0.03
Скос	0.05	0.05
Изрезанность	0.03	0.03
Имя подтипа. Шкала: нисходящая	0.05	
Начальная частота	0.08	0.08
Конечная частота	0.06	0.06
Максимум	0.25	0.25
Время максимума	0.06	0.06
Минимум	0.02	0.02
Время минимума	0.06	0.06
Интервал-Гц	0.26	0.26
Интервал-ПТ	0.07	0.07
Изрезанность	0.10	0.10
Имя подтипа. Шкала: восходящая	0	
Начальная частота	0.04	0.04
Конечная частота	0.04	0.04
Максимум	0.19	0.19
Время максимума	0.03	0.03
Минимум	0.03	0.03
Время минимума	0.03	0.03
Интервал-Гц	0.20	0.20
Интервал-ПТ	0.04	0.04
Средняя частота	0.31	0.31
Изрезанность	0.04	0.04
Имя подтипа. Предшкала: высокая	0.01	
Максимум	0.23	0.23
Средняя частота	0.76	0.76
Имя подтипа. Предшкала: низкая	0.03	
Максимум	0.31	0.31
Минимум	0.36	0.36
Средняя частота	0.32	0.32
Имя подтипа. Синтагма: нисходящее завершение	0.01	
Конечная частота	0.16	0.16
Максимум	0.18	0.18
Время максимума	0.11	0.11
Минимум	0.16	0.16
Время минимума	0.11	0.11

НАЗВАНИЕ ПАРАМЕТРА	Весовой коэффициент при совпадении	Весовой коэффициент при различии
Интервал-ПТ	0.10	0.10
Изрезанность	0.14	0.14
Имя подтипа. Синтагма: восходящее завершение	0.01	
Конечная частота	0.03	0.03
Максимум	0.21	0.21
Время максимума	0.12	0.12
Минимум	0.21	0.21
Время минимума	0.12	0.12
Интервал-ПТ	0.16	0.16
Изрезанность	0.12	0.12
Имя типа. Длительный речевой фрагмент	0.42	
Максимум	0.01	0.07
Минимум	0.21	0.23
Интервал-Гц	0.11	0.09
Интервал-ПТ	0.12	0.07
Средняя частота	0.19	0.20
Скорость изменения тона	0.14	0.14
Экссесс	0.02	0.03
Скос	0.05	0.05
Изрезанность	0.11	0.07
Имя типа. Пауза хезитации	0.06	
Максимум	0.30	0.30
Минимум	0.34	0.34
Средняя частота	0.35	0.35
Имя подтипа. Предъядерный слог + ядро: восходящий	0.01	
Начальная частота	0.02	0.02
Конечная частота	0.008	0.008
Максимум	0.03	0.03
Время максимума	0.17	0.17
Минимум	0.01	0.01
Время минимума	0.03	0.03
Интервал-Гц	0.02	0.02
Интервал-ПТ	0.02	0.02
Время половинной частоты	0.02	0.02
Средняя частота	0.02	0.02
Экссесс	0.004	0.004
Скос	0.19	0.19
Изрезанность	0.19	0.19

НАЗВАНИЕ ПАРАМЕТРА	Весовой коэффициент при совпадении	Весовой коэффициент при различии
Длительность	0.19	0.19
Имя подтипа. Предъядерный слог + ядро: нисходящий	0.06	
Начальная частота	0.04	0.04
Конечная частота	0.03	0.03
Максимум	0.04	0.04
Время максимума	0.03	0.03
Время минимума	0.02	0.02
Интервал-Гц	0.02	0.02
Интервал-ПТ	0.07	0.07
Средняя частота	0.01	0.01
Экссесс	0.03	0.03
Скос	0.22	0.22
Изрезанность	0.22	0.22
Длительность	0.22	0.22

008-180614

