

=Speech Intonation Assistant= (SIA)

Версия SIA от 2024-05-24

РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ

1. Функциональные характеристики

Программа **Speech Intonation Assistant (SIA)** предназначена для анализа статистических характеристик параметров интонации речи пользователя на достаточно длительных интервалах (порядка 10 и более секунд). На этих интервалах измеряются два параметра, наиболее важных для восприятия:

1. Диапазон изменения ЧОТ – **Pitch Variability [Hz]**;
2. Скорость или темп речи – **Speech Rate [wpm]**.

Первый параметр в значительной степени характеризует **выразительность** речи, а второй – ее **разборчивость**.

На основе расчета этих параметров интонации реализованы три режима работы программы: “**Testing**”, “**Training**”, “**Monitoring**”.

SIA реализован как отдельное **Приложение** для **Windows** и **Linux**.

2. Области применения

Настоящее программное приложение предназначено для *тестирования, тренинга и мониторинга* интонационных (просодических) характеристик устной речи пользователя. При подготовке к публичному выступлению или во время самого выступления пользователь должен придерживаться определенного стиля речи. Каждый стиль устной речи характеризуется своими особенностями просодических параметров устной речи. Различными стилями устной речи являются: официальный, публицистический, художественный, разговорный, учебный, научный и др. Например, востребованным **SIA** может стать для создателей таких видов контента, как:

Аудио-контент:

- *Подкастеры*
- *Радио и телеведущие*

Видеоклипы и YouTube:

- *Блоггеры*
- *Обучающие видео*

Рекламный контент:

- *Пропагандисты*
- *Голосовые актеры*

Учебный контент:

- *Преподаватели и лекторы*
- *Онлайн-курсы и вебинары*

3. Начало работы

Стартовое окно SIA, открывающееся после запуска программы, показано на Рис. 1.

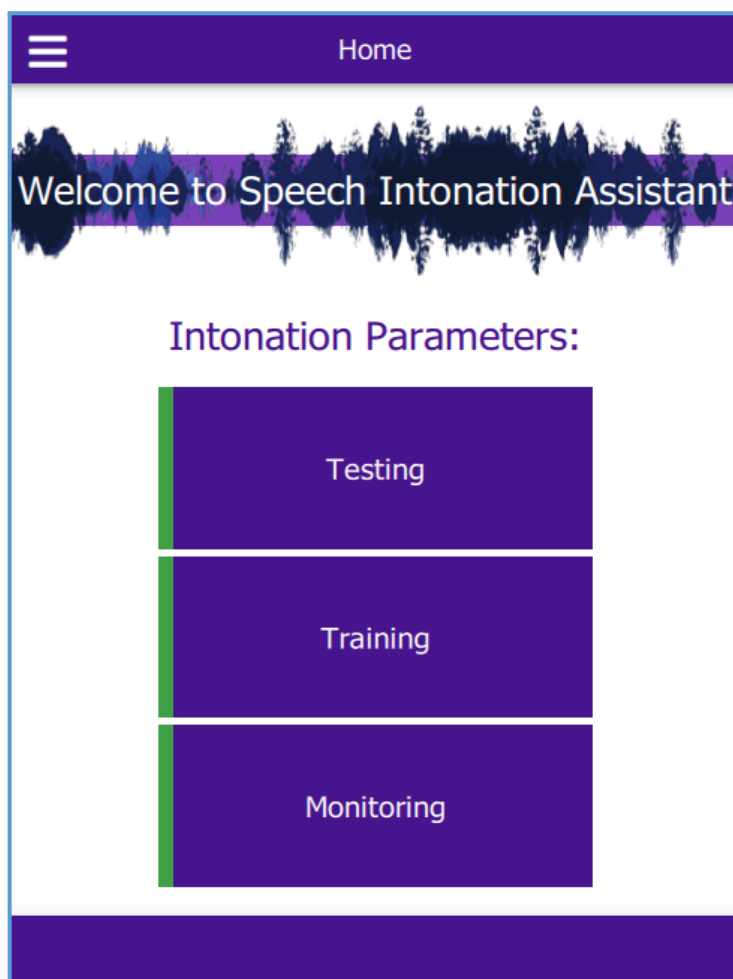


Рис. 1. Стартовое окно SIA

После запуска программы пользователю предоставляются функциональные возможности в 3-х режимах:

1. *Тестирование* устной речи пользователя путем оценки интонационных параметров: диапазона изменения ЧОТ – **Pitch Variability [Hz]** и скорости речи – **Speech Rate [wpm]**.
Первый режим реализуется путём нажатия на значок “**Testing**”.
2. *Практическое обучение навыкам интонирования речи* путем сравнения актуальных интонационных параметров устной речи пользователя с параметрами речи выбранного эталонного диктора и оценки степени их сходства.
Второй режим реализуется путём нажатия на значок “**Training**”.
3. *Мониторинг изменения интонационных параметров* текущей речи в реальном процессе чтения лекции, доклада и т.п.
Третий режим реализуется путём нажатия на значок “**Monitoring**”.

4. Тестирование интонационных параметров говорящего (1-й режим)

При выборе первого режима “**Testing**” открывается рабочее окно (Рис. 2).

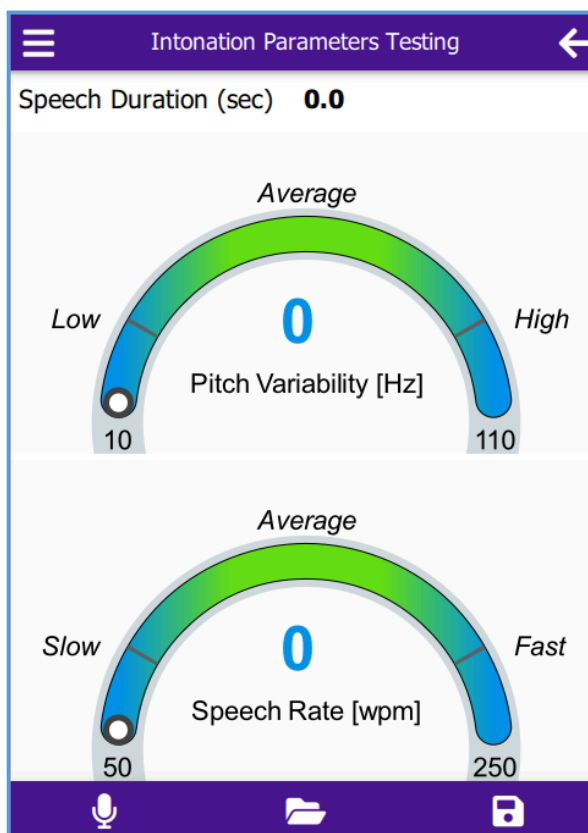





Рис. 2. Рабочее окно первого режима


В центральной части окна расположены две разноцветные дуги, играющие роль шкал значений интонационных параметров **Pitch Variability** и **Speech Rate**. На концах дуг указаны минимальные и максимальные значения этих параметров. Численные значения параметров заданы в разделе настроек и могут быть изменены пользователем (см. Раздел 7).

После открытия рабочего окна пользователю предоставляются следующие возможности:

- 1) Запись через микрофон достаточно длительного отрезка анализируемой речи (желательно более 10 сек);
- 2) Вызов одного из заранее подготовленных тестовых аудио файлов, хранящихся в папке “**data-tests**”.

Первый вариант реализуется путём нажатия на значок микрофона  в левой части нижней фиолетовой полосы. В результате открывается окно со значком , а после его активации появляется динамический значок  и начинается запись произносимого отрезка речи. Для завершения записи необходимо повторное нажатие этого значка.

Произнесенный отрезок речи подготавливается для дальнейшей обработки и отображения результатов интонационного анализа, а также записывается в папку “**data/records**”.

Второй вариант реализуется путём нажатия на значок  в средней части нижней фиолетовой полосы. Он предназначен для вызова из папки “**data/tests**” заранее подготовленных аудио файлов – фрагментов речи создателей различного вида контента. В папке “**tests**” в качестве примеров размещены различные образцы речи для мужских и женских голосов (см. рис. 3).

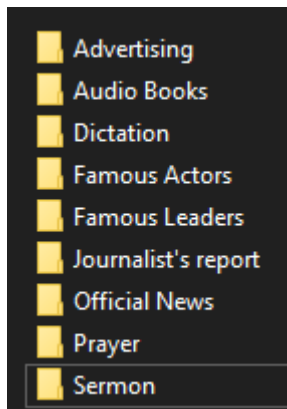


Рис. 3. Содержимое папки “**tests**”

В девяти папках размещены образцы различных стилей речи в девяти тематических папках:

- Звуковые рекламные ролики;
- Примеры чтения аудио книг;
- Примеры стиля «Диктант»;
- Выступления известных актеров;
- Выступления известных лидеров;
- Репортажи радио и ТВ журналистов;
- Чтение дикторами официальных новостей;
- Примеры религиозных молитв;
- Примеры религиозных проповедей.

Примечание. Содержимое папки “**tests**” может пополняться дополнительными примерами с использованием имеющихся средств (например, **Sound Forge**). Следует иметь в виду, что в настоящей версии **SIA** предусмотрена возможность анализа только для сигналов в формате: **wav, mono, 8 kHz**. Любые другие форматы следует адаптировать под этот формат, используя подходящие программные средства.

После окончания записи через микрофон или после вызова тестового аудио файла открывается окно с результатами расчета интонационных параметров: диапазона изменения ЧОТ – **Pitch Variability [Hz]** и скорости речи – **Speech Rate [wpm]** (см. Рис. 4).

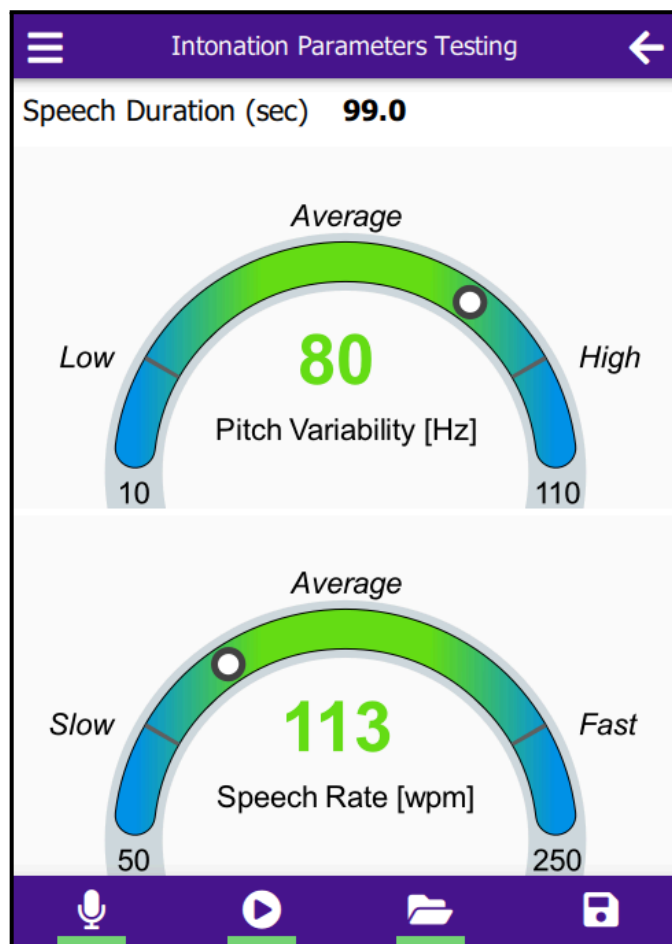




Рис. 4. Окно результатов в первом режиме

В верхней части окна отображается длительность анализируемого фрагмента речи – **Speech Duration (sec)**. Ниже расположены 2 шкалы:

- верхняя отображает графически (белым кружком) и численно (под дугой) результаты измерения диапазона изменения ЧОТ – **Pitch Variability [Hz]**,
- нижняя – результаты измерения скорости речи – **Speech Rate [wpm]**.

Зеленым цветом отмечены диапазоны изменения среднестатистических значений этих параметров, полученные на основе примеров речи создателей различного вида контента (Рис. 3), фиолетовыми – отклонения от них.

Кроме того, в нижней фиолетовой полоске появляется дополнительный значок , обеспечивающий прослушивание анализируемого фрагмента речи. В правой части этой полоски имеется значок , при нажатии на который осуществляется фиксация полученных результатов (см. раздел 7).

5. Обучение навыкам интонирования речи (2-й режим)

При выборе второго режима “**Training**” открывается рабочее окно (Рис. 5).

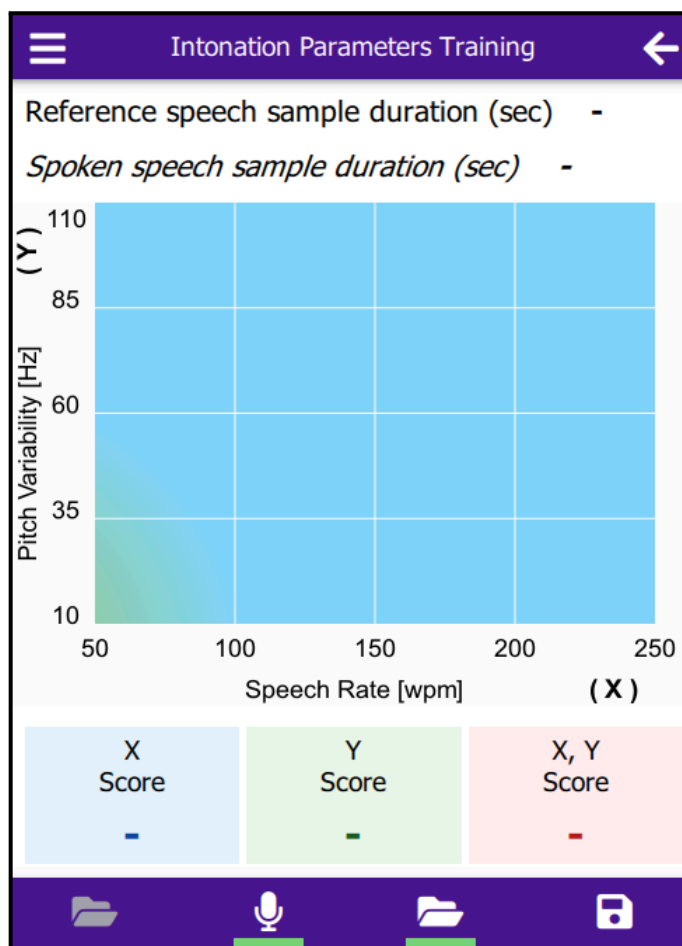



Рис. 5. Рабочее окно второго режима

В рабочем окне отображается тренировочная мишень в виде плоскости с координатами (**X**, **Y**), которые соответствуют двум измеряемым интонационным параметрам, проградуированным в численных значениях – [**wpm**] и [**Hz**].

После открытия рабочего окна пользователю предоставляется возможность вызова из папки “**data-tests**” одного из заранее подготовленных эталонных образцов устной речи путём нажатия на значок  в левой части нижней фиолетовой полосы. При этом открывается папка “**data/tests**”, в подпапках которой (см. рис. 3) размещены аудио файлы различных образцов устной речи. Эти образцы речи – фрагменты речи создателей различного вида Интернет-контента – могут быть выбраны пользователем, как подходящие образцы для тренинга собственной речи.

После выбора подходящего образца устной речи на тренировочной мишени рабочего окна появляется соответствующий значок в виде черно-белого кольца (см. Рис. 6). Координаты его на плоскости соответствует измеренным значениям интонационных параметров.

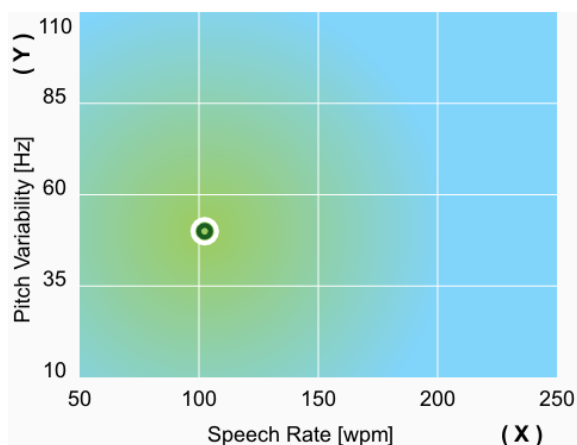


Рис. 6. Тренировочная мишень рабочего окна

В процессе тренинга производится запись пользователем своей речи для её сравнения с желаемым образцом. Запись осуществляется через микрофон в соответствии с процедурой, описанной в п. 4. Произнесенный отрезок речи подготавливается для дальнейшей обработки и отображения результатов интонационного анализа, а также записывается в папку **“data/records”**.

После произнесения каждого из фрагментов анализируемой речи осуществляется отображение значений её интонационных параметров в виде точек на тренировочной плоскости. В результате рабочее окно второго режима приобретает вид, показанный на Рис. 7, где небольшими точками с белыми кружками отображается результат каждого тренировочного произнесения фрагмента речи. Последнее (наилучшее) произнесение фрагмента отображается белым кружком с контрастной черной точкой в его центре.

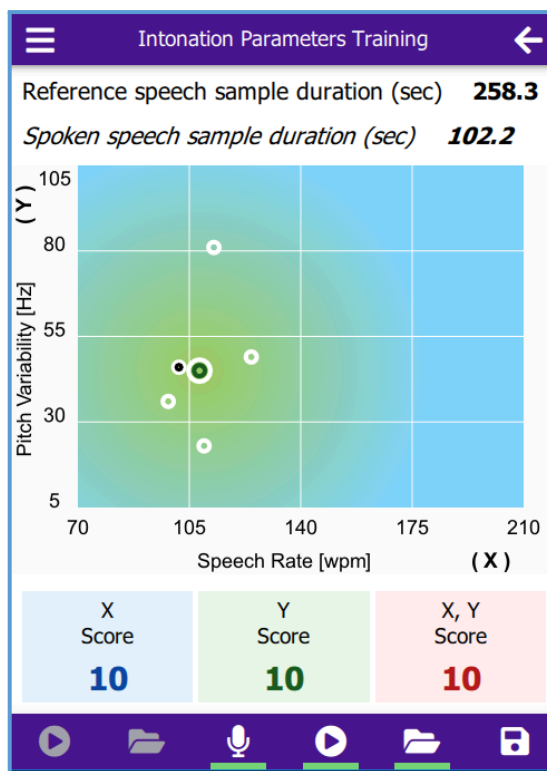





Рис. 7. Результирующий вид рабочего окна

После произнесения фрагмента анализируемой речи, показавшего наибольшую близость к избранному образцу, результаты сравнения с образцом отображаются (см. Рис. 7) следующим образом:

- Под тренировочной плоскостью приводятся численные оценки близости по 10 бальной шкале отдельно для каждого (X) и (Y) параметра, а также обобщенная оценка (X, Y).
- Над тренировочной плоскостью показываются длительности в секундах эталонного образца и образца произнесенной речи.

Значок  в нижней фиолетовой полоске предназначен для прослушивания эталонного образца устной речи, а значок  – для прослушивания анализируемого фрагмента речи. При нажатии значок  осуществляется фиксация полученного результата (см. раздел 7).

6. Мониторинг интонационных параметров текущей речи (3-й режим)

При выборе третьего режима - **“Monitoring”** - открывается первый (начальный) вид рабочего окна (Рис. 8).

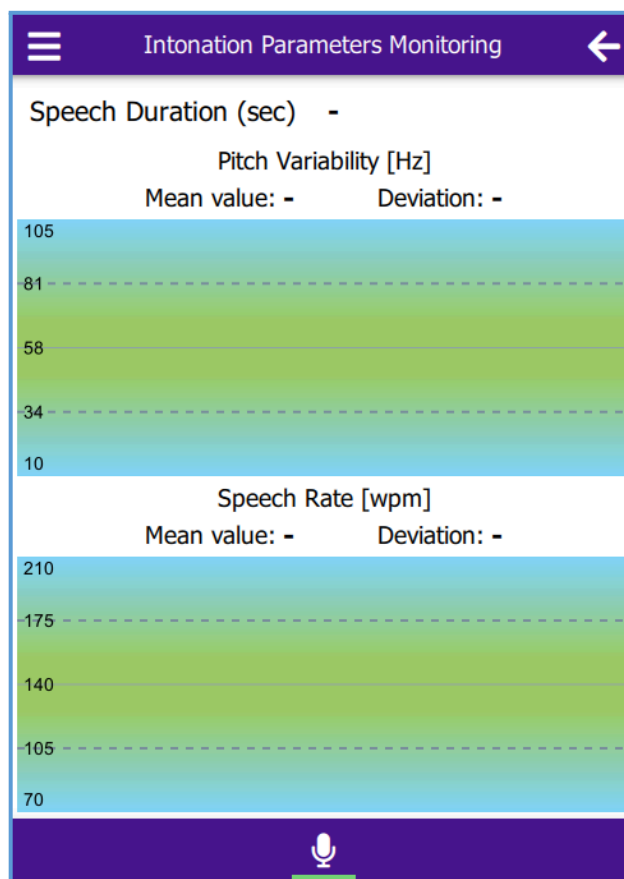



Рис. 8 Начальный вид рабочего окна **“Monitoring”**

В рабочем окне представлены два отдельных поля, предназначенные для отображения динамики развития интонационных параметров устной речи:

- верхняя – результаты измерения диапазона изменения частоты основного тона речи – **Pitch Variability [Hz]**,
- нижняя – результаты измерения скорости речи – **Speech Rate [wpm]**.

Отображение динамики развития интонационных параметров осуществляется в виде точечной кривой на полях рабочего окна в реальном времени произнесения пользователем различного рода аудио контента.

В левой части полей показаны шкалы численных значений параметров и диапазоны их изменения. Штриховыми линиями показана условные зоны средних значений, заданные разработчиком и помечены зеленым цветом. Интервал дискретизации и диапазоны изменения интонационных параметров могут быть изменены пользователем в разделе настроек (см. Раздел 7).

Запись через микрофон анализируемой речи реализуется путём нажатия на значок микрофона . После нажатия на этот значок в рабочем окне (Рис. 9) и начинается отображение текущих значений интонационных параметров анализируемого сигнала речи.

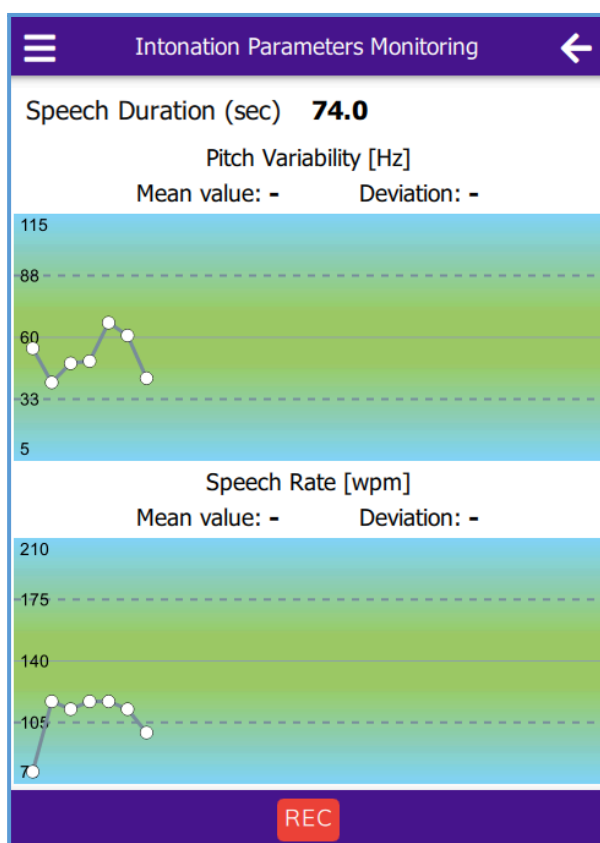



Рис. 9. Отображение текущих значений интонационных параметров

После завершения речи и нажатия на значок  заканчивается запись анализируемого сигнала речи и открывается третий (финальный) вид рабочего окна режима “**Monitoring**” (Рис. 10).

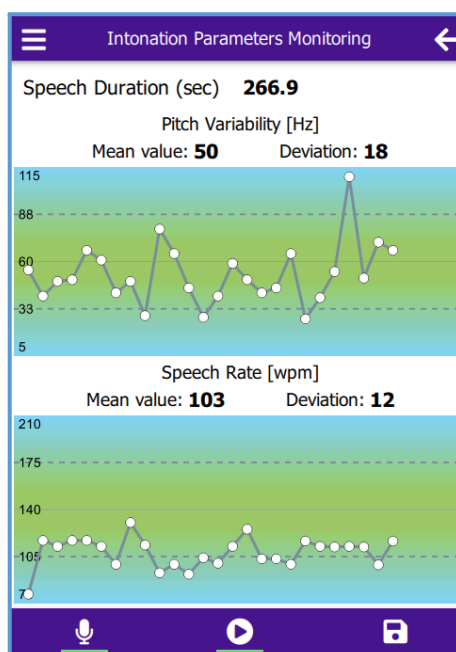




Рис. 10. Финальный вид рабочего окна режима “Monitoring”

В верхней части рабочего окна показана длительность произнесенного речевого сигнала — **Speech Duration (sec)**. На нижнем и верхнем полях показаны траектории изменения соответствующих интонационных параметров, а над каждым полем: средние их значения – **Mean Value** – и отклонения от среднего – **Deviation** – за указанное время.

Значок  в нижней фиолетовой полоске предназначен для прослушивания анализируемого фрагмента речи. При нажатии значок  осуществляется фиксация полученного результата (см. раздел 7).

На Рис. 11 в качестве примера приказаны результаты измерения интонационных параметров 2-х докладчиков на конференции за одинаковое время выступления. Очевидны различия в динамике изменения этих параметров в процессе речи, а также в оценках **Mean Value** and **Deviation** за одинаковый пятиминутный отрезок времени.

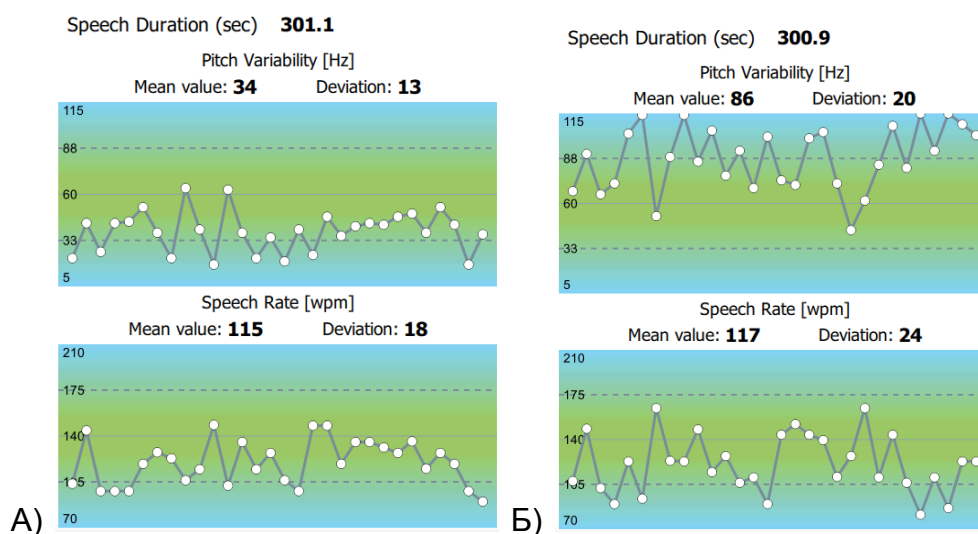



Рис. 11. Результаты измерений для докладчиков А) и Б)

7. Вспомогательные значки

В верхней полоске слева рабочего окна каждого из режимов находится значок , при нажатии на который открывается дополнительное информационное окно (см. рис. 10),

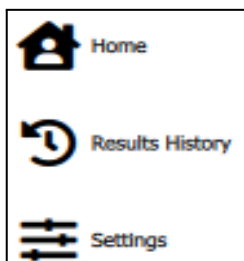



Рис. 12. Информационное окно

Значок настроек  Settings позволяет пользователю изменить рекомендованные разработчиком предельно возможные значения интонационных параметров (см. Рис. 13), одновременно для всех режимов.

Pitch Variation Maximum

115

Maximum Value [Hz]

Pitch Variation Minimum

5

Minimum Value [Hz]

Speech Rate Maximum

210

Maximum Value [WpM]

Speech Rate Minimum

70

Minimum Value [WpM]

Рис. 13. Информационное окно

Дополнительно, только для третьего режима, имеется возможность изменения пользователем двух характеристик процесса записи и отображения параметров (см. Рис. 14).

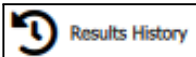
Visible monitoring points number

30

Sliding monitoring window in Sec

10

Рис. 14. Информационное окно

Путём выбора значка:  вызывается содержимое файла “results” (см. рис. 15А для первого режима, Рис. 15Б для 2-го режима и Рис. 15В – для третьего режима).

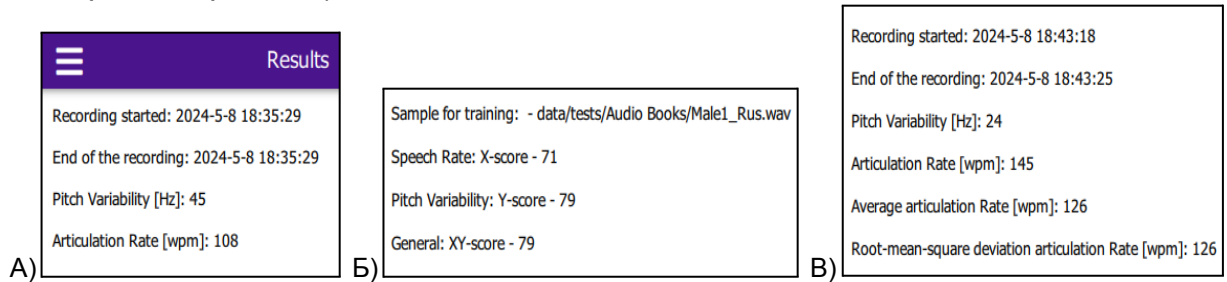
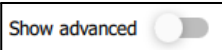


Рис. 15. Содержимое файла “results”

Данные сохраняются также в файле **results.csv** в папке с программой SIA.

ВНИМАНИЕ! Внутренняя информация, вызываемая пометкой в правом верхнем квадрате , предназначена только для разработчика. Изменения её содержимого пользователем крайне нежелательно и может привести к неправильной работе программы.

ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Мы рекомендуем пользователям время от времени заглядывать на наш сайт, в котором могут появляться полезные обновления SIA. Узнать больше о теоретических основах и программном коде разработки SIA: см. дополнительную информацию, размещаемую на этом сайте.