=Speech Auto Tuner=

(SAT)

Version from 2025-04-12

User Guide

1. Functional characteristics

The Speech Auto Tuner (SAT) program is designed for automatic adjustment to normalize the acoustic characteristics of a microphone signal during the creation of speech audio content. The use of normalization tools enables standardization of the speech signal level (volume) and its frequency characteristics (spectrum), which are essential for the most comfortable and accurate auditory perception of spoken language.

SIA is implemented as a separate Application for Windows and Linux

2. Applications

This software application could become highly sought-after in the preparation of speech content for various online applications, such as:

- News and entertainment content.
- TV interviews and conversations.
- Lectures and educational content.
- Webinars and online conferences.
- Podcasts, video blogs, streaming services, social networks, and much more.

Problems with the perception of speech content, related to the variation in microphone acoustic parameters and its placement, are common among most bloggers. These issues often lead to listeners avoiding further engagement due to "muffled" speech (high-frequency attenuation) or "whistling" speech (low-frequency attenuation). The Speech Auto Tuner (SAT) program ensures automatic normalization (standardization) of the microphone signal's acoustic characteristics during the creation of speech audio content, namely: volume level and frequency spectrum.

3. Getting started

The SAT start window, which opens after launching the program, is shown in Fig.1.

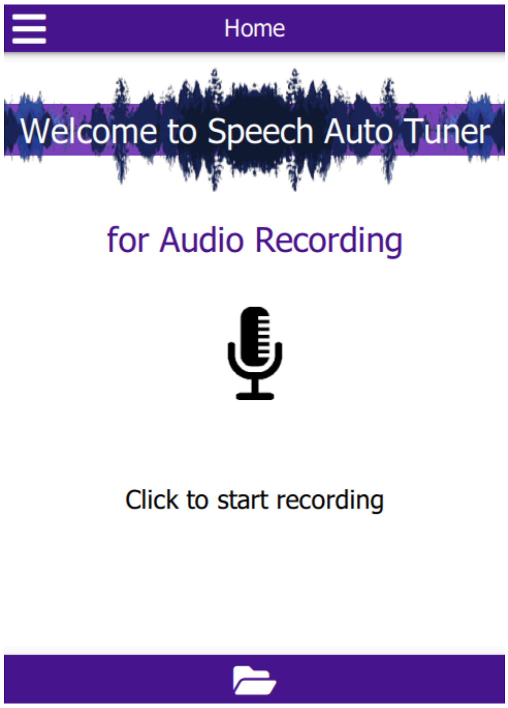


Fig. 1. Start window

After launching the program, the user is provided with the option to record a new speech signal using their existing microphone or to retrieve a previously recorded speech signal (SS) from another speech source, using the corresponding icons: , or . As a result of recording the SS, the working window opens (see Fig. 2).

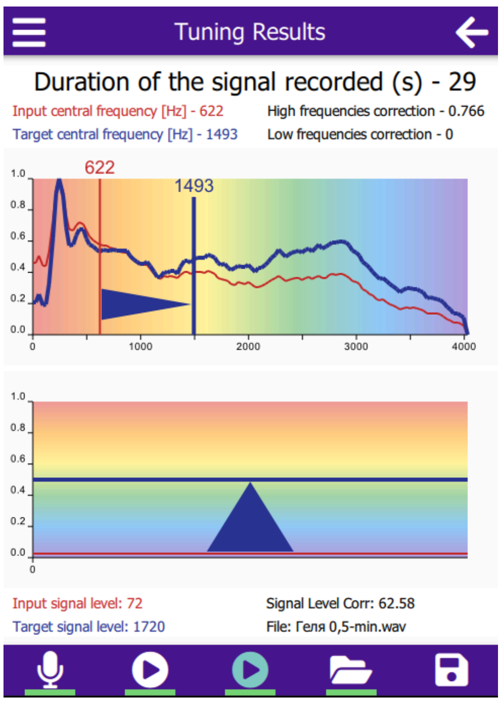


Fig. 2. Resulting view of the working window

In the upper graph (see Fig. 2), the frequency response of the original signal (red line) and the normalized signal (blue line) are displayed, along with vertical lines (red and blue) showing the position of the center frequency (CF) of the frequency response before and after spectrum normalization. In the lower graph, the volume levels for the original low-level speech signal (red horizontal line) and the normalized speech signal (blue line) are shown. The standard value for the normalized speech signal level is chosen to be 0.5 of the maximum allowable level of the speech signal (see Table 1).

Table 1. LUFS dB values for permissible loudness levels

№	Relative loudness in LUFS	Relative in ratios	Fractions of 2^{16} S(t)
1	-10	3,16	20739
2	-11	3,55	18461
3	-12	3,98	16466
4	-13	4,47	14661
5	-14	5,01	13081
6	-15	5,62	11661
7	-16	6,31	10386
8	-17	7,08	9256
9	-18	7,94	8254
10	-19	8,91	7355
11	-20	10,0	6554

Duration of the signal recorded (s) - 29

Input central frequency [Hz] - 957 High frequencies correction - 0.508

Target central frequency [Hz] - 1408 Low frequencies correction - 0

At the top of the working window, information about the duration of the speech signal (SS) and the data used for correcting its frequency spectrum are displayed.

Input signal level: 4840 Signal Level Corr: 0.44

Target signal level: 2130 File: Speech Level-Max.wav

In the lower part of the working window, data necessary for adjusting the volume level of the speech signal (SS) are displayed, as well as the name of the audio file used during testing.

After the completion of each speech fragment spoken into the microphone, it is recorded into the folder **data/records** as a digital signal with a sampling rate of 8 kHz, with the date and time of the recording indicated.

The first icon on the left is used for playback of the recorded speech signal (SS) for review, while the second icon allows for playback of the normalized SS. The normalization of the frequency response (FR) and volume level of the SS is performed automatically after the recording of each speech fragment. Upon activation of the second icon, the normalized SS is saved in the **data/records** folder with the added name **Corrected**.

14.04.2025.11.38.18 - Corrected

14.04.2025.11.38.18

In the **data/tests** folder, test examples of speech signal (SS) recordings with varying volume levels and spectral characteristics are stored.

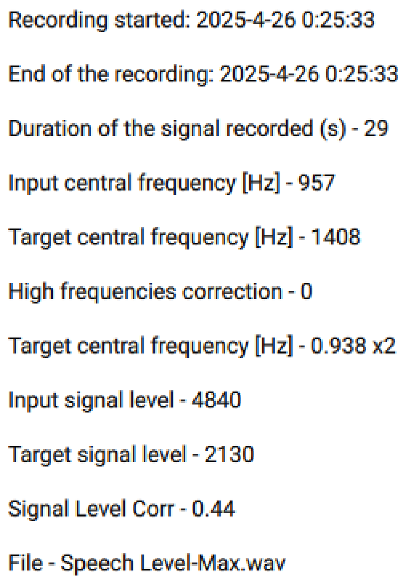
4. Helper icons

In the upper bar on the left of the working window of each mode there is an icon ☰, which, when clicked, opens an additional information window (see Fig. 3).

Fig. 3. Information window

The settings icon Settings allows the user to change application properties. Internal information, called up by the mark in the upper right square Show advanced, is intended for the developer only. Changing its contents by the user is highly undesirable and can lead to incorrect operation of the program.

By selecting the icon: Results History the contents of the “results” file are called up**,** that store signal and processing data. Such as: file name, recording date, its length, frequency and level values of input and output signals, processing parameters.



Recording started: 2025-4-26 0:25:33

End of the recording: 2025-4-26 0:25:33

Duration of the signal recorded (s) - 29

Input central frequency [Hz] - 957

Target central frequency [Hz] - 1408

High frequencies correction - 0

Target central frequency [Hz] - 0.938 x2

Input signal level - 4840

Target signal level - 2130

Signal Level Corr - 0.44

File - Speech Level-Max.wav

CONCLUDING REMARKS

We encourage users to check our website from time to time for useful SAT updates.
To learn more about the theoretical background and code for developing SAT: see additional information posted on this site