Speech Stimulus Tool

A MATLAB GUI for Segment-Level Signal Processing

Documentation and Technical Overview

Ravi Umadi

Abstract

The *Speech Stimulus Tool* is an interactive MATLAB GUI designed for recording, loading, segmenting, processing, visualising, and exporting speech or bioacoustic signals. The system provides experimenters with a unified interface to apply standardised manipulations such as fading, band-pass filtering, and frequency-domain normalisation, while maintaining full control over time-domain shape, spectral envelope, RMS level, and reproducibility of processing settings. This document describes all features, processing steps, GUI components, mathematical methods, and file export mechanisms used in the tool.

Contents

1 Overview

The Speech Stimulus Tool provides a complete workflow for creating experiment-ready acoustic stimuli. A user can:

- Record audio directly from any input device.

- Load an existing WAV file.
- Visualise the full waveform, spectrum, and spectrogram.
- Select arbitrary time segments via graphical interaction.
- Apply a range of processing operations:
  - onset/offset fades,
  - band-pass filtering,
  - frequency-domain envelope shifting (spectral normalisation),
  - RMS-preserving level adjustments.
- Compare original and processed waveforms and spectra.
- Save processed segments together with all associated metadata:
  - processed audio,
  - waveform, spectrum, and spectrogram images,
  - YAML-formatted parameter configuration,
  - user notes.

The interface is designed for controlled stimulus generation, such as in psychophysical, perceptual, and neuroethological experiments.

2 GUI Structure

The window is divided vertically into two halves:

1. **Left: Visualisation Panel** (waveform, spectrum, spectrogram)
2. **Right: Control Panels**

   The right-side controls are further subdivided into:

   1. **INPUT** panel
   2. **SEGMENT** panel
   3. **PROCESS** panel
   4. **NOTES** panel

   Each panel and its components are described below.

3 The INPUT Panel

This is where the user configures audio input, recording, and file loading.

Device Selection

- A dropdown lists all available input devices returned by `audiodevinfo`.
- The special item "Default device" corresponds to MATLAB's default input.

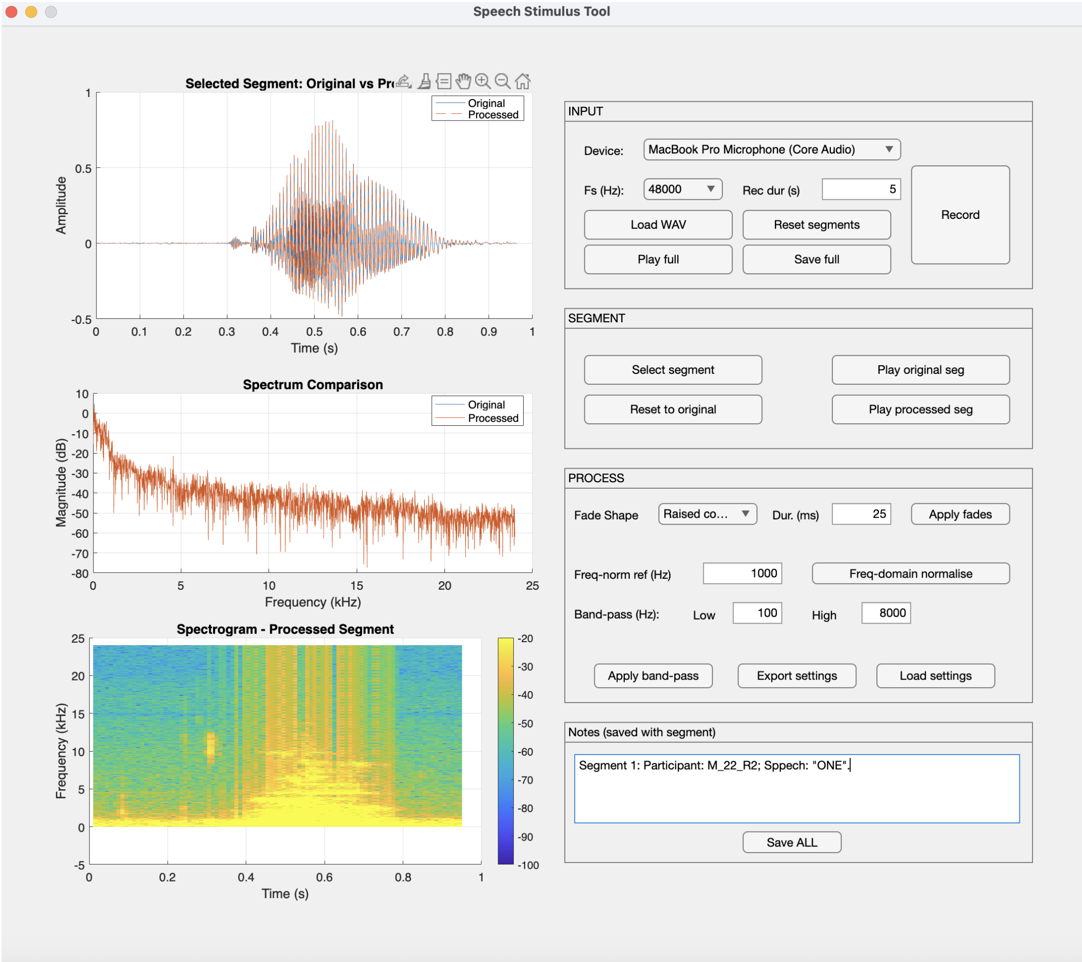

Figure 1: Graphical user interface of the *Speech Stimulus Tool*, showing the time-domain waveform, spectral views, and control panels for input, segment selection, and processing.

Sampling Frequency Selection

A dropdown provides standard FS options:

$$\{8000, 16000, 22050, 44100, 48000, 96000\} \text{ Hz}.$$

Recording Duration

A numeric field defines the recording length in seconds.

Buttons

Load WAV Opens a file dialogue and loads a mono WAV file. Multi-channel files are down-mixed to mono.

Record Starts a timed recording with:

- non-blocking acquisition,
- a live countdown shown inside the button,
- automatic termination at the specified duration,
- full replacement of previously loaded data.

Play full Plays the entire loaded/recorded signal.

Save full Saves the current full signal. Output is peak-normalised to 0.99.

Reset segments Clears any selected segment and returns plots to full-signal mode.

4 The SEGMENT Panel

This panel allows extraction and playback of specific regions of interest.

Buttons

Select segment Prompts the user to draw a rectangle on the waveform axis. The x-limits determine the time interval. The corresponding samples are extracted into:

`segOrig` (original)

`segProc` (processed copy)

Reset to original Restores `segProc = segOrig`.

Play original seg Plays the unmodified extracted segment.

Play processed seg Plays the processed segment.

5 The PROCESS Panel

This is the computational core of the tool. It contains three categories:

5.1 Fade Processing

Supported shapes:

- Linear
- Raised cosine (Hann)
- Gaussian

Let $x[n]$ be the input segment and N_f the number of fade samples. The tool constructs an envelope:

$$e[n] = \begin{cases} e_{\text{fade-in}}[n], & n < N_f, \\ 1, & N_f \leq n < L - N_f, \\ e_{\text{fade-out}}[n], & n \geq L - N_f, \end{cases}$$

and computes:

$$y[n] = x[n] \cdot e[n].$$

To ensure perceptual loudness consistency:

$$\mathrm{RMS}(y) = \mathrm{RMS}(x) \quad \Rightarrow \quad y \leftarrow y \cdot \frac{\mathrm{RMS}(x)}{\mathrm{RMS}(y)}.$$

5.2 Band-pass Filtering

A 4th-order Butterworth filter is constructed:

$$[b, a] = \texttt{butter}(4,\ [f_{\text{low}}, f_{\text{high}}]/\frac{F_s}{2}).$$

Zero-phase filtering is done via:

$$y = \texttt{filtfilt}(b, a, x).$$

As before, the RMS after filtering is normalised back to the original RMS level.

5.3 Frequency-Domain Normalisation

The purpose of this method is to:

1. Estimate and smooth the spectral envelope.
2. Detect its dominant peak frequency.
3. Shift the envelope so that the peak occurs at a user-selected reference frequency f_{ref}.
4. Preserve:
   - overall spectral shape,
   - all phase relationships,
   - the time-domain structure,
   - RMS level.

Method Summary

Given signal x, compute its FFT:

$$X[k] = \mathcal{F}\{x[n]\}.$$

Extract magnitude and phase:

$$|X[k]|, \quad \phi[k] = \arg X[k].$$

Apply Gaussian smoothing in the frequency domain to obtain a clean spectral envelope:

$$E[k] = |X[k]| * G_\sigma[k].$$

Find the index of the peak:

$$k_{\text{peak}} = \arg\max_k E[k].$$

Shift the envelope by Δk bins such that the new peak aligns at f_{ref}:

$$\Delta k = \text{round}\left(\frac{f_{\text{ref}} - f_{\text{peak}}}{F_s/N}\right).$$

Construct a multiplicative shaping ratio:

$$R[k] = \frac{E_{\text{shift}}[k]}{E[k] + \varepsilon},$$

clipped to prevent extreme amplification.

Apply shaping:

$$|X_{\text{new}}[k]| = |X[k]| \cdot R[k].$$

Recombine magnitude and phase, then apply IFFT and truncate:

$$y[n] = \Re\{\mathcal{F}^{-1}(X_{\text{new}})\}.$$

Finally, RMS-match:

$$y \leftarrow y \cdot \frac{\text{RMS}(x)}{\text{RMS}(y)}.$$

This procedure shifts the *dominant spectral peak* without introducing tonal artefacts and without altering the microstructure of the waveform.

6 Visualisation

The left panel dynamically updates:

6.1 Waveform Plot

Shows:

- Full signal, or
- Segment (original and processed).

6.2 Spectrum Plot

Two FFT spectra (original vs processed) are shown in dB scale.

6.3 Spectrogram

A log-magnitude spectrogram of the processed segment is plotted using:

- window length $0.03F_s$,
- overlap $0.02F_s$,
- 2048-point FFT,
- dB colour scale.

7 Notes and Saving

The **Save ALL** button automatically creates a folder and writes:

1. Processed segment (.wav)
2. Waveform.png
3. Spectrum.png
4. Spectrogram.png
5. settings.yml (all user-selected parameters)
6. notes.txt (user annotations)

This ensures perfect reproducibility.

8 Conclusion

The Speech Stimulus Tool provides a controlled and transparent workflow for preparing behavioural and perceptual stimulus material. All major steps—recording, segmentation, processing, visualisation, and saving—are implemented as distinct, well-defined, and reproducible operations. The tool is suitable for experiments involving speech, animal vocalisations, or general acoustic stimuli requiring precise manipulation of spectral and temporal structure.