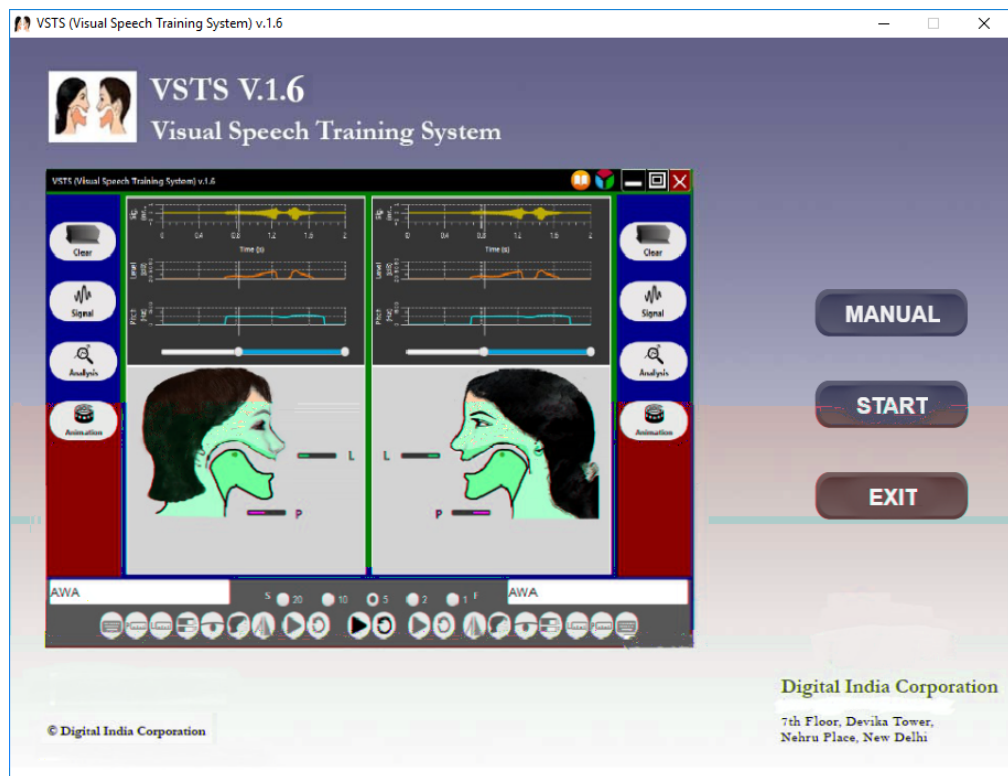


Visual Speech Training System (VSTS v.1.6)

User Manual



Digital India Corporation

(Not-for-Profit Company, Ministry of Electronics and Information Technology, Government of India)

7th Floor, Devika Tower, Nehru Place, New Delhi 110019

Left Blank

Visual Speech Training System (VSTS v.1.6)

User Manual

Abstract

The Visual Speech Training System (VSTS) is a computer-based speech training system which uses information obtained by speech signal analysis to provide a visual feedback of efforts involved in speech production. It has been developed for use as a speech training aid to assist in acquisition of correct articulatory efforts by children with hearing impairments and second language learners. It can also be useful to speech therapists and speech training professionals as an analysis and diagnostic tool. It is an application software which runs on a PC with sound card, without needing any additional hardware. The software has been developed by Digital India Corporation in collaboration with IIT Bombay with support of an R&D grant from the Department of Electronics and Information Technology, Government of India. The user manual provides information on installation and use of the software.

Contents

Section	Page
1. Introduction	4
2. Hardware & Software Requirement	5
3. Installation	6
4. Getting Started	9
5. Signal Acquisition	12
6. Analysis	17
7. Animation	19
8. Color Customization	22
9. VSTS Version History	23

Development Team

Laxita Vyas, Digital India Corporation < laxita@digitalindia.gov.in>
Navin Daniels, Digital India Corporation < ndaniels@digitalindia.gov.in>
K S Nataraj, EE Dept, IIT Bombay <natarajks@ee.iitb.ac.in>
Hirak Dasgupta, EE Dept, IIT Bombay <hirakdgpt@ee.iitb.ac.in>
Vishal Mane, Digital India Corporation < vishal.mane@digitalindia.gov.in>
P C Pandey, EE Dept, IIT Bombay <pcpandey@ee.iitb.ac.in>

1. Introduction

The Visual Speech Training System (VSTS) is a computer-based speech training system which uses information obtained by speech signal analysis to provide a visual feedback of efforts involved in speech production. It has been developed for use as an analysis and diagnostic tool by speech therapists and speech training professionals and as a speech training aid to assist in acquisition of correct articulatory efforts by children with hearing impairments and second language learners. It is in the form of an application software which runs on a PC with sound card, without needing any additional hardware.

The system has the following main features:

- i) Signal acquisition: recording of speech signal as an audio clip of up to 10 s and loading of pre-recorded sounds;
- ii) Signal analysis and display: signal waveform, pitch, energy, spectrogram (2D plot of time-varying magnitude spectrum), and areagram (2D plot of time-varying vocal tract shape);
- iii) Animation of articulatory efforts: display of time-varying vocal tract shape (as obtained from the analysis and displayed in the areagram) with adjustable animation speed along with optional indicators for frame energy, pitch, and place of articulation.

In order to provide corrective feedback with reference to a model speech utterance, the system has two side-by-side display panels. These panels can be used for displaying animation or analysis results for speech signals from the student and a teacher or a reference speaker. It is also possible to use the same signal in both the panels with analysis result in one panel and animation in the second panel.

The color scheme of the overall display can be modified and saved for subsequent use. Animation in each panel can be reconfigured by selecting the face and its orientation. The face can be selected as that of a man, woman, boy, or girl, and its orientation can be selected as left or right facing.

The present version of the system can be used for speech segments containing vowels, semivowels, and diphthongs. Future versions may be useable for speech with other phoneme classes.

2. Hardware & Software Requirement

Computer & Operating System: Windows-based PC with Intel 6th generation processor (Core i3, i5 or i7) or equivalent, sound card with 16-bit resolution and 44.1 kHz sampling frequency, 2 GB memory, hard-disk with 2 GB free space, and Windows 7 or higher.

Microphone & Loudspeaker Configuration: External unidirectional dynamic or condenser microphone and good quality speakers are preferred. Microphone level and boost should be set to record conversational-level speech signal with the microphone at approximately 10 cm from the speaker's lips. The setting steps may vary depending on the system configuration and the OS version. For PCs running Windows 7, the following steps may be followed:

- i) Right-click on the sound icon on the taskbar.
- ii) Select 'Recording devices' from the pop-up menu.
- iii) Select the required input device. The bar on the right indicates the current microphone activity. The number of green levels indicates the current level of sound input as shown in Figure 2.1.
- iv) If there is no activity on the bar, configure the correct input device (by selecting the 'Configure' button) and check the microphone connection.
- v) Click on 'Properties' to adjust the microphone input level as shown in Figure 2.2. To avoid signal distortion, disable any enhancements and noise suppression settings (under enhancement tab).
- vi) Click on the sound icon on the taskbar and set the volume to an appropriate level.

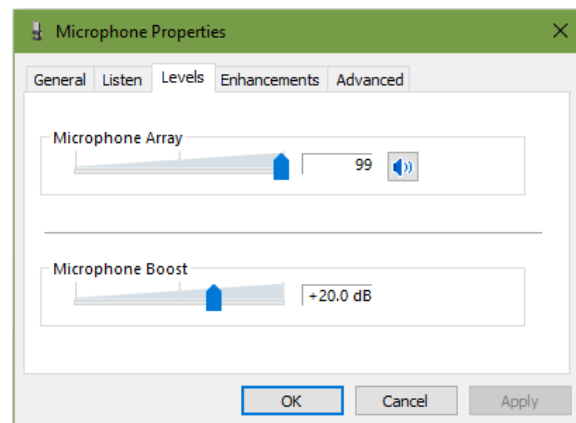
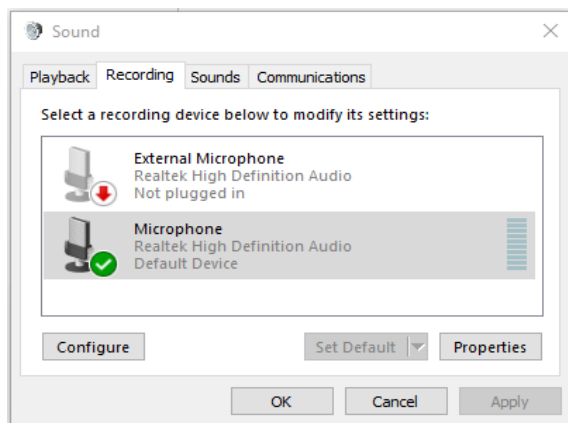


Figure 2.1(a): Current level of sound input indicator. **Figure 2.1(b):** Microphone level & boost controls.

3. Installation

The execution of VSTS requires installation of runtime environments of Java and MATLAB, which should be compatible with the operating system version (32- or 64-bit). There are two distribution folders of 'vsts_v.1.6', one for PC with 32-bit OS and the other one for PC with 64-bit OS. Each folder has corresponding versions of Matlab and Java runtime environment files. Both folders have VSTS application and sound folders.

i) Folder 'vsts_v.1.6_32bit': It has 32-bit versions of runtime environment setup files.

'MCR_R2014b_win32_installer.exe': Matlab runtime environment for 32-bit OS,

'jre-8u111-windows-i586.exe': Java runtime environment for 32-bit OS,

'vsts_v.1.5_setup_32bit.bat': VSTS installation batch file for 32-bit OS,

'VSTS': VSTS application, demo sound folder, recorded sound folders.

ii) Folder 'vsts_v.1.6_64bit': It has 64-bit versions of runtime environment setup files.

'MCR_R2014b_win64_installer.exe': Matlab runtime for 64-bit OS,

'jre-8u111-windows-x64.exe': Java runtime environment setup for 64-bit OS,

'vsts_v.1.6_setup_64bit.bat': VSTS installation batch file for 64-bit OS,

'VSTS': VSTS application, demo sound folder, recorded sound folders.

Download or copy the distribution version folder appropriate for your machine. To install the application, open the folder and double click on the batch file

'vsts_v.1.6_setup_32bit.bat' or 'vsts_v.1.6_setup_64bit.bat' (depending on your version).

The Matlab runtime environment installation process begins. Click on 'Next' button in the installation window as shown in Figure 3.1(a). Select 'Yes' radio button to accept the license agreement as shown in Figure 3.1(b) and click on 'Next' button to continue the installation process. Select the appropriate folder for installation files (default should generally work) as shown in Figure 3.1(c) and click on 'Next'. After the next page appears as shown in Figure 3.1(d), click on 'Install' button. The installation progress gets displayed as in Figure 3.1(e). Complete the installation process by clicking on 'Finish' button as shown in Figure 3.1(f). Next the Java runtime environment installation process begins. Click on 'Install' button as shown in Figure 3.2(a). The installation progress gets displayed as shown in Figure 3.2(b). When the next page appears, complete the installation process by clicking on 'Finish'.

After installation of MCR and JRE, the batch file creates a shortcut of the application on desktop with the name 'vsts_v.1.6'. The batch file creates 'VSTS' folder in the C drive which contains the executable application 'vsts_v.1.6.exe', and folders of 'demo_sounds' and 'recorded_sounds'. The 'demo_sounds' folder has subfolders with '.wav' sound files provided as part of the distribution of the application. The 'recorded_sounds' folder is used for storing

the sounds recorded as ‘.wav’ files by the application in the subfolder linked to date (ddmmyy) and time (hhmmss) of their creation.

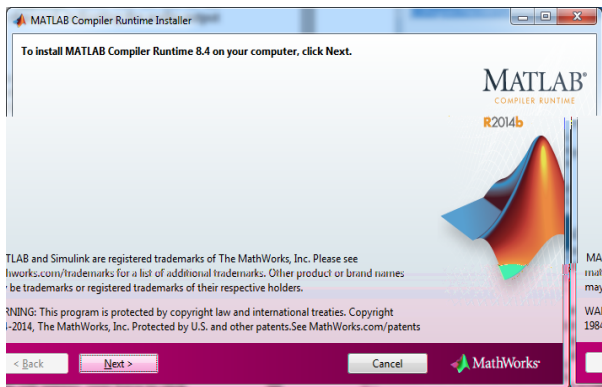


Figure 3.1(a): First page of installation of MATLAB runtime environment.

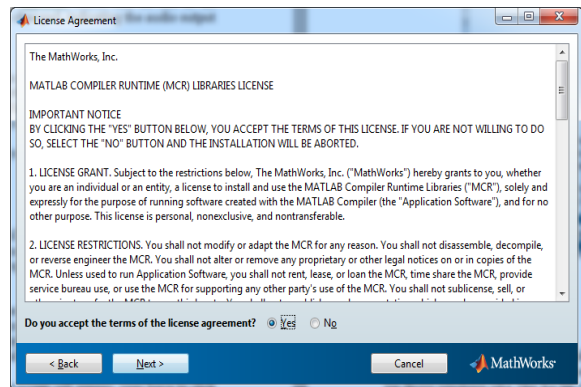


Figure 3.1(b): Second page of installation of MATLAB runtime environment.

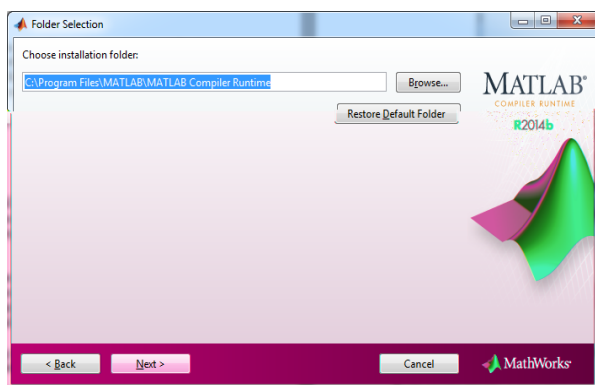


Figure 3.1(c): Third page of the installation of MATLAB runtime environment

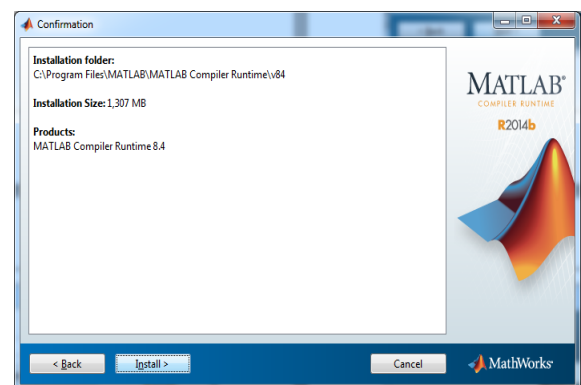


Figure 3.1(d): Fourth page of the installation of MATLAB runtime environment

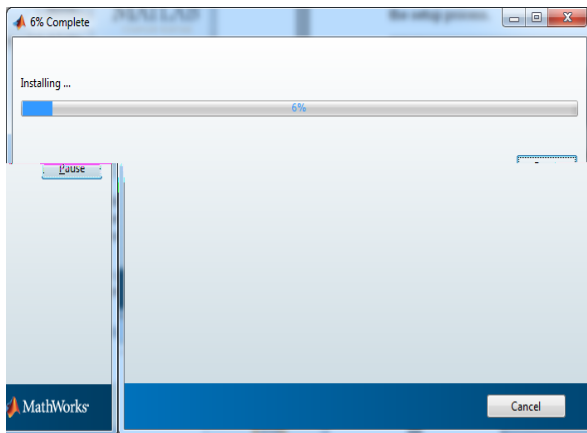


Figure 3.1(e): Fifth page of installation of MATLAB runtime environment.



Figure 3.1(f): Sixth page of installation of MATLAB runtime environment.

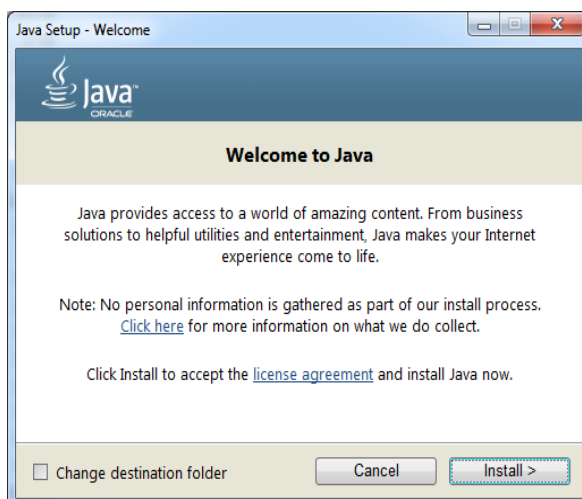


Figure 3.2 (a): First page of installation of Java runtime environment.



Figure 3.2 (b): Second page of installation of Java runtime environment.

4. Getting Started

Double click on the 'vsts_v.1.6' icon on the desktop. A welcome screen appears as shown in Figure 4.1, with the following three buttons:

Manual: VSTS user manual,

Start: Navigate to the main screen,

Exit: Close the application.

After clicking on the 'Start' button, the main screen appears as shown in Figure 4.2. Use 'maximize' button for better visualization of the graphics.

The screen is horizontally split into two panels: one for the reference or model speech (from the teacher or pre-recorded) and the other for the learner's speech. Each panel has a vertical icon bar. One of the following functions can be selected by clicking on the corresponding icon (if indicated as active by being highlighted):

Clear: Clears any previously generated display.

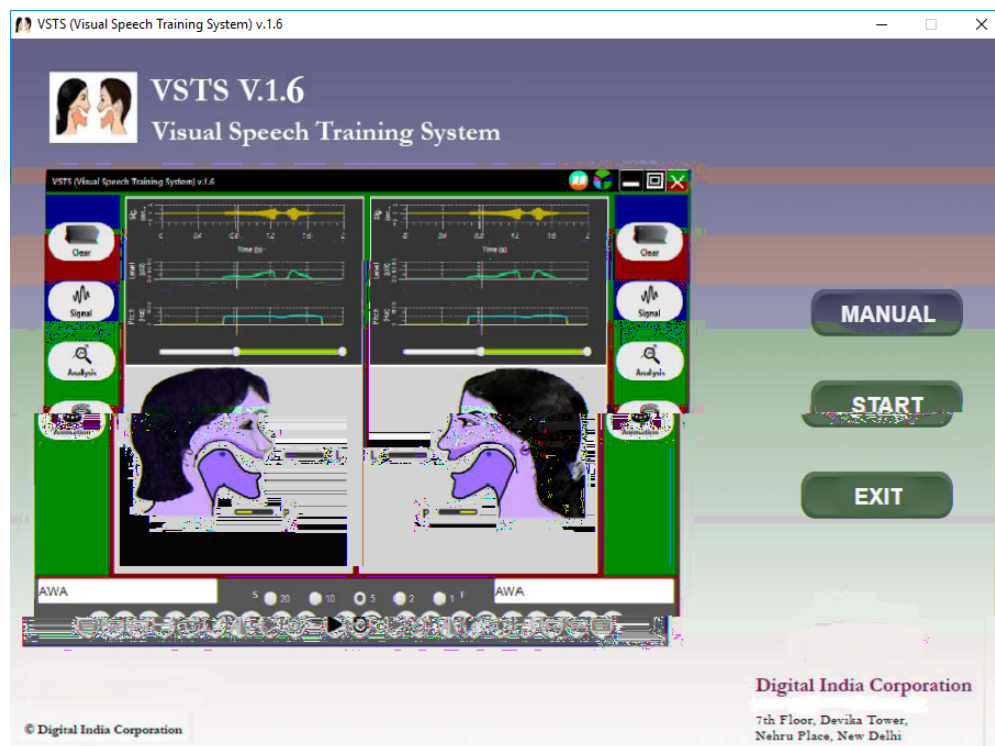


Figure 4.1: Welcome screen.

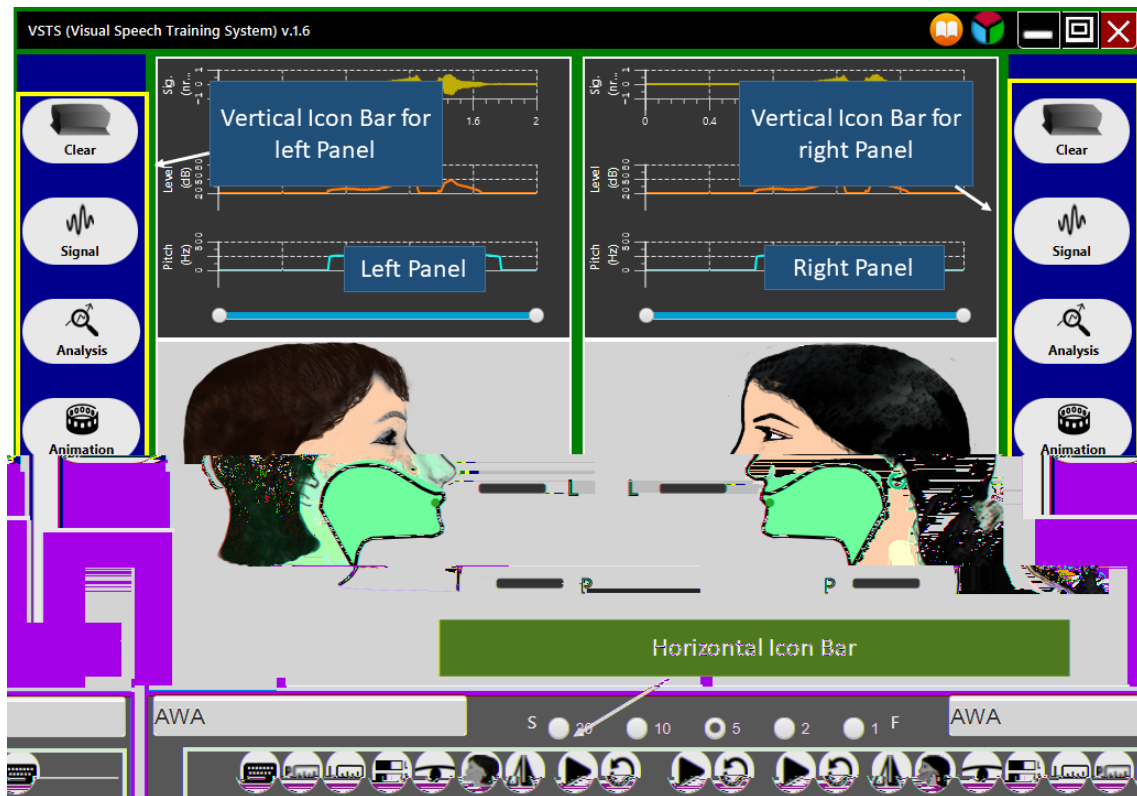


Figure 4.2: Main screen.

Signal: Speech signal can be recorded or previously stored audio files can be browsed and the selected segment of the audio file can be stored and processed for analysis and animation.











Analysis: The signal is analyzed for display of the analysis parameters and use in animation. The icon is active only if an audio segment has been selected for analysis and animation.

Animation: Information as obtained by analysis of the signal is used for animation. The icon is active only if an audio segment has been selected for analysis and animation.

These functions are described further in the subsequent sections.

Below the two panels, there is a horizontal icon bar, which has icons for display control and configuration during analysis and animation. These controls and functions, as listed in Table 4.1, can be used in different combinations.

Table 4.1: Icons in the bottom horizontal icon bar along with corresponding function and brief description.

Sr.	Icon	Function	Description
1.		Play	Starts the animation. The left dot on the progress bar moves towards right, indicating the position of the animation frame.
2.		Pause	Pauses the animation at the current position.
3.		Reset	Resets the left dot to the start of the signal.
4.		Pitch	Toggles the pitch scale as 0-500 or 0-250 Hz.
5.		Level	Toggles the level scale as 20-80 or 20-60 dB.
6.		Face	Selects the face cyclically as man, woman, boy, or girl.
7.		Mirror	Toggles the animation as left or right facing.
8.		POA	Toggles red dot indicator of POA (place of articulation) as on/off.
9.		L/P	Selects the level and pitch indicator bars as vertical, horizontal, or off.
10.		Message	Opens message window to enter text or graphical symbols.

5. Signal Acquisition

A speech signal can be recorded and stored or a pre-recorded signal can be loaded for display, analysis, and animation on either panel. The duration of the recorded or loaded signal can be up to 10 s, from which a segment of duration 0.3 – 4.8 s can be selected. The selected segment gets appended with preceding and succeeding silences of 0.1 s, resulting in segment of duration 0.5 – 5.0 s for processing.

Both recording and loading operations are available under the 'Signal' icon. After 'Signal' is clicked, a dialog box appears giving options for recording and loading along with options for selecting a segment, resetting, playing, saving, and processing. The dialog box has fields for entering the name and age of the speaker, and these along with the time information are used for forming the name of the audio file for saving the selected segment.

Recording the Speech Signal

A speech signal can be recorded using a microphone for analysis and animation on either panel by clicking the 'Record' icon, after the 'Signal' icon has been clicked. The recording is controlled by 'Start' and 'Stop' icons. A progress bar is displayed to indicate the duration for which recording has already taken place.



Figure 5.1: Window for signal acquisition (recording and loading).

The sounds recorded using the application are stored as '.wav' files in a subfolder inside 'recorded_sounds'. Names of the subfolder and the files inside it are associated with the date (ddmmyy) and time (hhmmss) of their creation. After the application is started, it creates a subfolder named as 'rec_ddmmyy_hhmmss' and this subfolder is subsequently used for recording the sounds in files named as 'name_age_ddmmyy_hhmmss.wav'. The sound files may be renamed or copied to another folder as required.

For recording, click on the 'Signal' icon of the desired panel (left/right). The Signal Acquisition window with options for 'Record' and 'Load' appears, as shown in Figure 5.1. Click on the 'Record' icon. A box appears with the 'Start' and 'Stop' icons along with the progress bar. Recording is started after the 'Start' icon is clicked. It is stopped if the 'Stop' icon is clicked or if the duration reaches 10 s, whichever happens earlier. After recording, a dialogue box appears which shows the recorded waveform as two graphs along with five icons of 'Select', 'Reset', 'Play', 'Save' and 'Proceed' at the bottom of the box as shown in Figure 5.2. A specific segment of the signal can be selected by dragging the two cursors below the upper graph, as shown in Figure 5.3. On clicking the 'Select' icon, the selected segment is displayed in the lower graph. The 'Reset' icon can be used to move the cursor positions to their default positions. If 'Play' is clicked, the selected signal is played. Playback of the selected segment (multiple times if needed) may be needed in deciding to save it, to select another segment, or to go to another recording or loading session. If the 'Save' icon is clicked, the selected segment is stored with the file name obtained using current date and time, as

'name_age_ddmmyy_hhmmss.wav'

in the subfolder as created at starting of the application. If 'Proceed' is clicked, the selected segment is processed, with the processing progress indicated by the hour glass image. After processing, the Signal Acquisition window is closed and the Animation window is opened with 'Analysis' and 'Animation' icons activation on the corresponding panel.

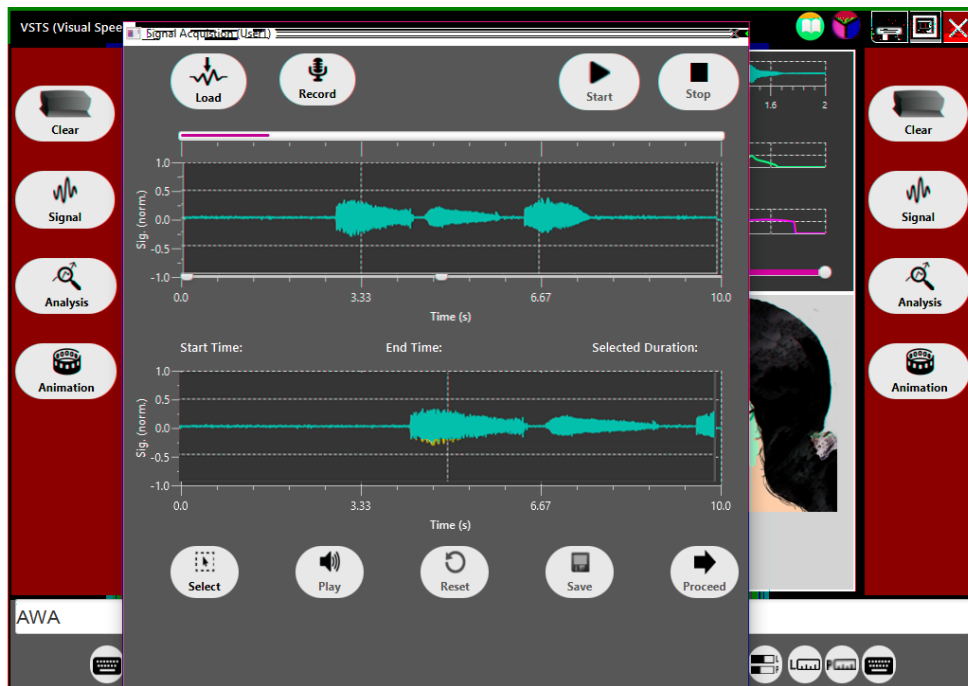


Figure 5.2: Window showing recorded sound signal.

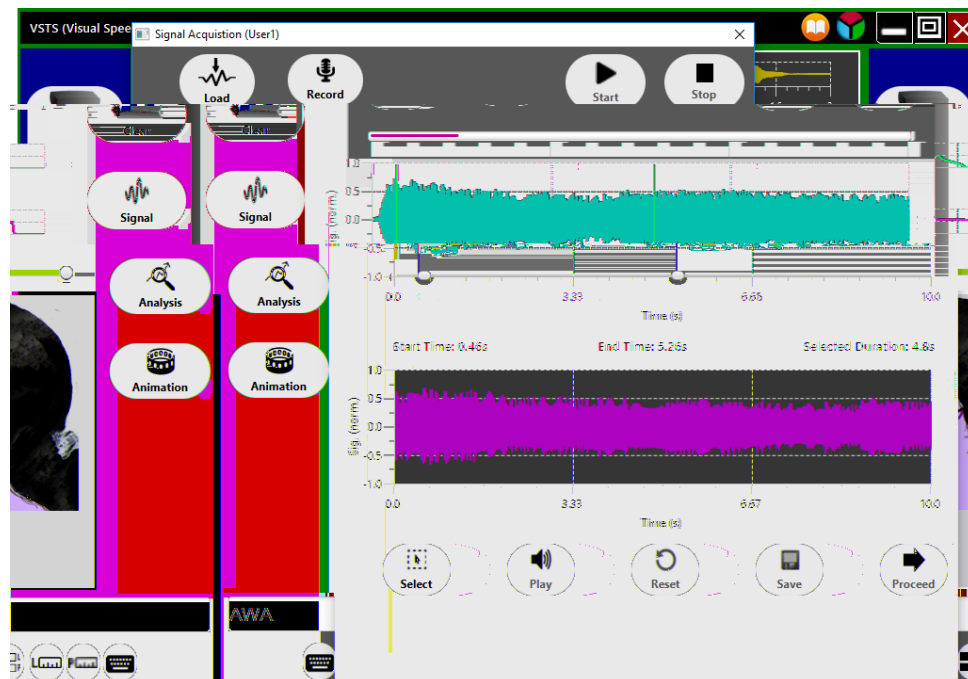


Figure 5.3: Window for selection of a segment of the recorded signal.

Loading the Speech Signal

A pre-recorded sound file can be loaded on either panel of the display and used for analysis and animation. Click on the 'Signal' icon on the desired panel (left/right). The Signal Acquisition window with options for 'Record' and 'Load' appears. Click on the 'Load' icon. A box appears to browse the pre-recorded files as shown in Figure 5.4. A set of demo sound files are included in the package in the folder 'demo_sounds'. Select the desired audio file from the folder 'demo_sounds' or 'recorded_sounds'.

After loading, a dialogue box appears which shows the loaded waveform as two graphs along with five icons of 'Select', 'Reset', 'Play', 'Save' and 'Proceed' as shown in Figure 5.5. These icons have the same functions as described earlier for recording the speech signal.

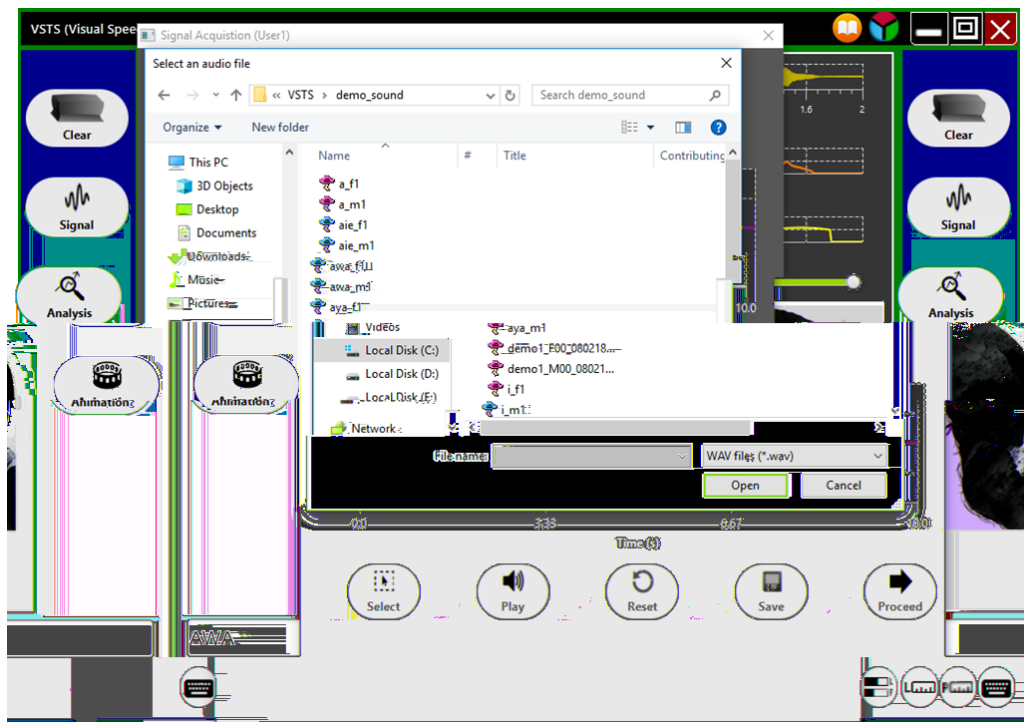


Figure 5.4: Window for browsing pre-recorded sound file.

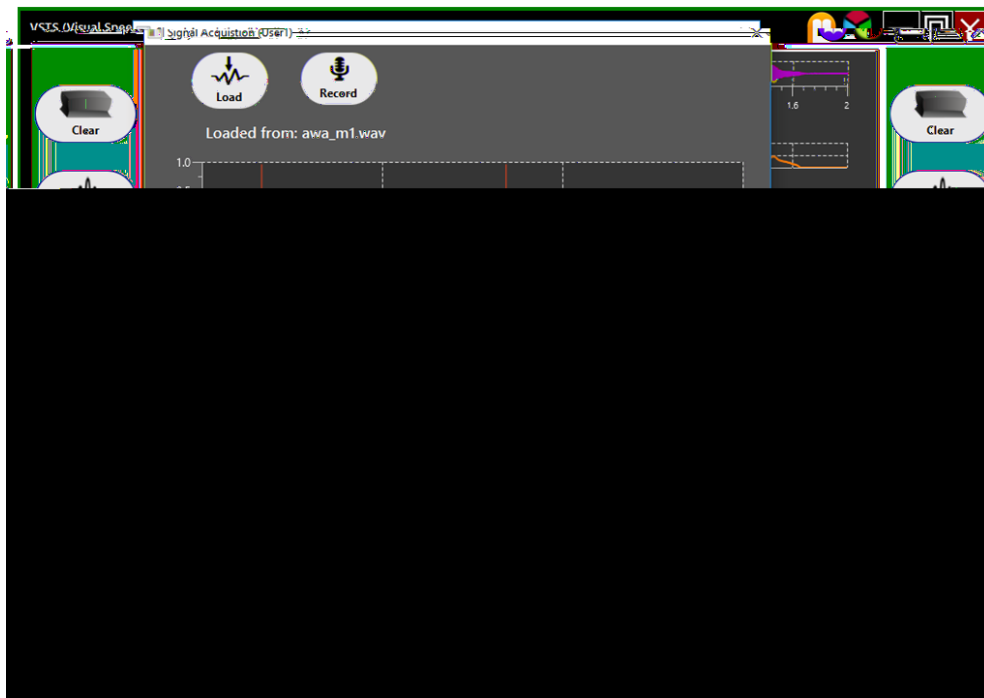


Figure 5.5: Window for display of pre-recorded sound.

6. Analysis

Analysis is used to display the information obtained by processing the recorded or loaded signal. It provides information on the speech signal in the form of waveform, level, pitch, spectrogram, and areagram. This mode is useful for objective measurements, while the animation mode may be used for dynamic visualization and speech training.

After a signal has been recorded or loaded, it is processed and the 'Analysis' and 'Animation' icons of the corresponding panel get activated. Clicking on the 'Analysis' icon displays the signal waveform, level, pitch, spectrogram, and areagram, as shown in Figure 6.1. The signal waveform is normalized to the range $[-1, +1]$. Level indicates time-varying signal level (short-time energy calculated over a duration equal two average pitch periods and converted to dB scale). Pitch is the rate of vibration of the vocal chords and it is obtained by short-time analysis of the speech signal. The level and pitch are plotted in accordance with the scale as selected by the corresponding icons on the bottom bar. Clicking on the 'L' icon may be used to toggle the level scale between 20 – 80 dB and 20 – 60 dB and that on the 'P' icon may be used to toggle the pitch scale between 0 – 500 Hz and 0 – 250 Hz. Spectrogram is two-dimensional plot of the short-time spectrum, showing spectrum level displayed using gray scale as a function of frequency and time, with frequency along the vertical axis and time along the horizontal axis. Areagram is a two-dimensional plot of vocal tract area as a function of distance from the lips and time. The spectrogram and areagram are obtained by short-time analysis of the signal and are displayed using fixed scales. All plots share the same time scale. Measurements on all four plots can be made using the cursor. As the cursor is moved, the values related to its position are displayed in the black bar above the spectrogram.

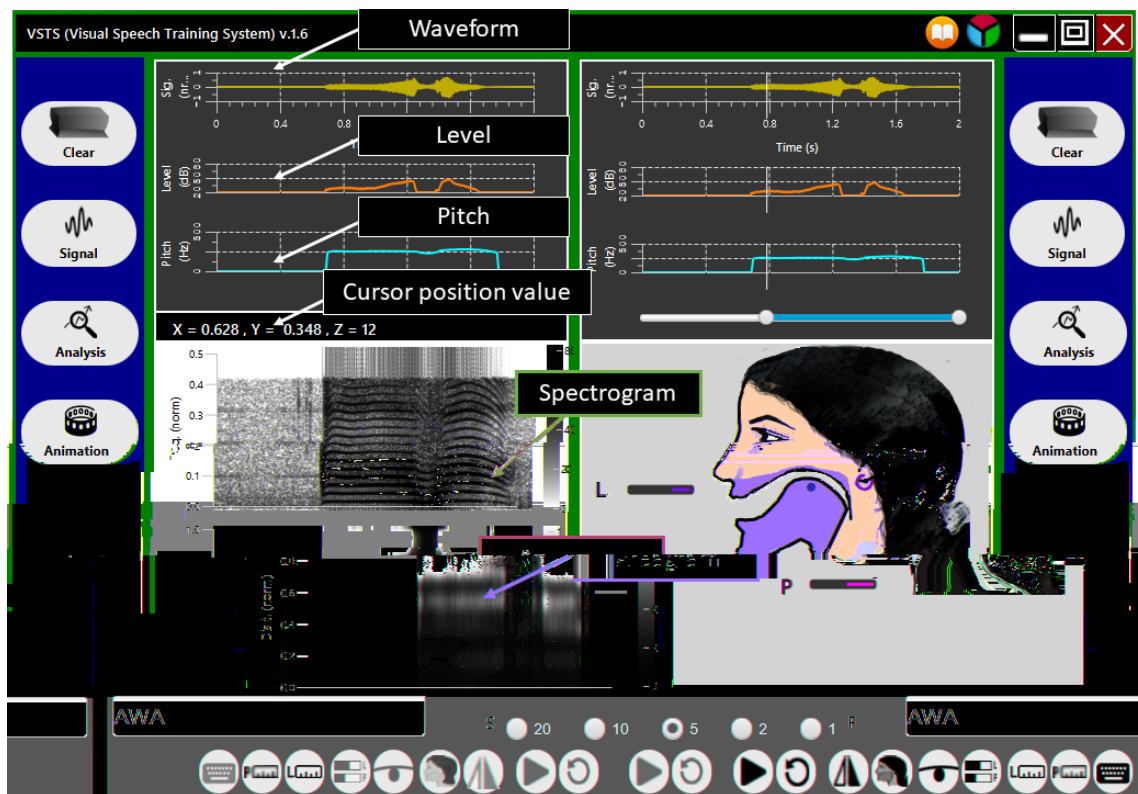


Figure 6.1: Window for display of analysis parameters.

7. Animation

Animation provides a dynamic display of vocal tract shape, level, and pitch as obtained by processing the recorded or loaded signal. It can be used for actual time-scaled display with audio playback or slow-motion display without audio playback. There are several options for configuring these displays as described later. The slow-motion display can be carried out on either panel or simultaneously on both.

After the 'Animation' icon of either panel is clicked, the corresponding panel is displayed with two sub-panels. In the upper sub-panel, the signal waveform is displayed on normalized $[-1, +1]$ scale, along with plots of level and pitch in accordance with the scales as selected using the corresponding icons. The lower sub-panel is used for slow-motion animation of vocal tract shape along with dynamic display of the level and pitch values using bars. The sub-panels along with the control icons and indicators are shown in Figure 7.1. The animation can be configured and controlled by the set of icons in the horizontal icon bar at the bottom of the display, and described in Table 4.1. During reconfiguration and animation, the relevant icons are visible as active.

The 'Play'/'Pause' icon is used for slow-motion animation of the vocal tract with audio playback, with its label toggling between 'Play' and 'Pause'. The animation speed is selected by clicking on one of the five icons corresponding to the slow-down factors of 1, 2, 5, 10, and 20. In the upper sub-panel, there is a horizontal 'play progress' bar, located below the three plots. The left and right dots are used to select the start and end points for slow-motion animation. During animation, the left dot moves showing the current position. The two dots can be shifted, even during animation, using cursor control. The 'Reset' icon can be used to reset the left dot to the start of the signal. The animations can be configured using 'Face' and 'Mirror' icons. The 'Face' icon can be used to cyclically select the animation face as that of a man, woman, boy, or girl. The 'Mirror' icon can be used to toggle the face as left or right facing. The 'POA' icon can be used to indicate the place of articulation as a red dot at the position of maximum constriction along the length of the vocal tract. The level and pitch values for the current animation frame can be displayed as bars, and mode of this display can be controlled by clicking on 'L/P' icon (shown as two bars) as vertical, horizontal, or off. These icons get disabled during animation.

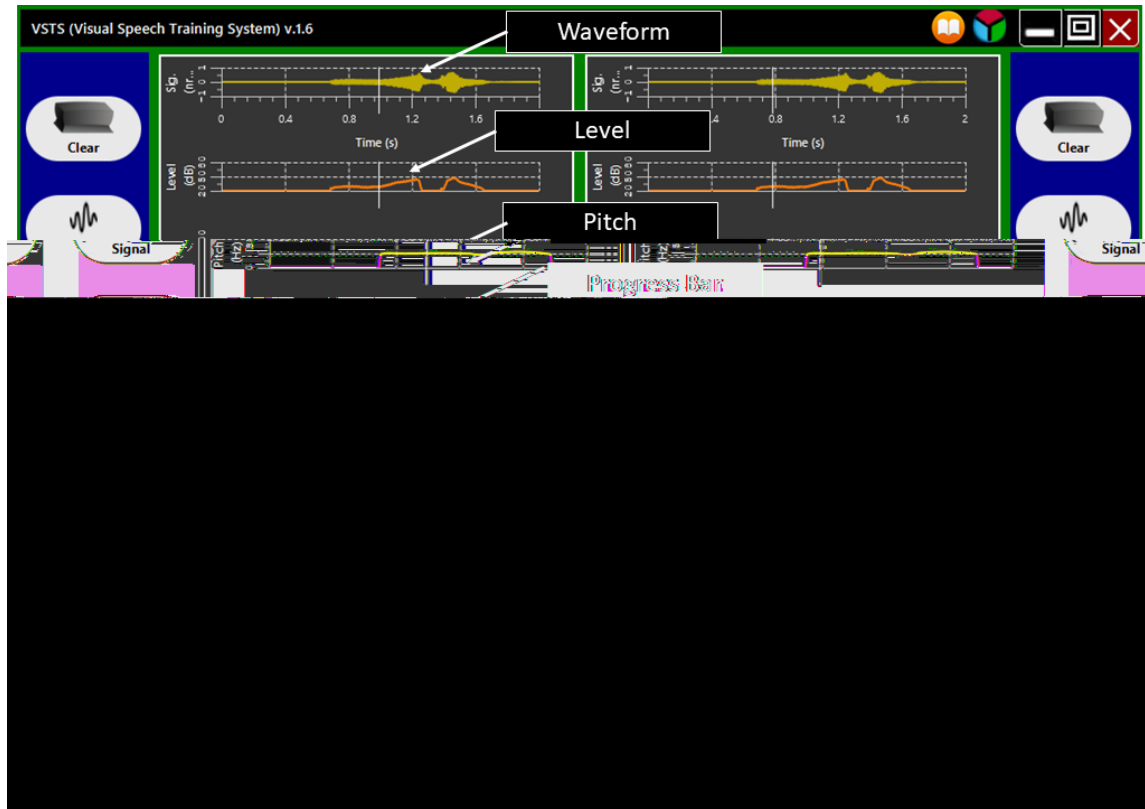


Figure 7.1: Window for display of animation and controls.

The 'Speed' buttons with slow-down factors of 1, 2, 5, 10 and 20 are located at the center of the horizontal bar and are common for controlling the animation speed on both panels. The 'Play'/'Pause' and 'Reset' icons below these speed control buttons are for simultaneous animation on the two panels. Clicking on the common 'Play' disables all the configuration icons on both panels and these become active after animation is paused or completed. Common Play shows the animation in both panels without audio, The common 'Reset' resets the left dot on the animation progress bar in both the panels.

Each panel has a facility for on-screen message. It can be used by the teacher for displaying a text or graphical prompt for speaking or for providing feedback. After clicking on the 'Message' icon, an onscreen keypad appears with a display bar and keys for Roman and Devanagari characters and graphical symbols, as shown in Figure 7.2. The message entered using the keypad is displayed in the horizontal bar above the control icons

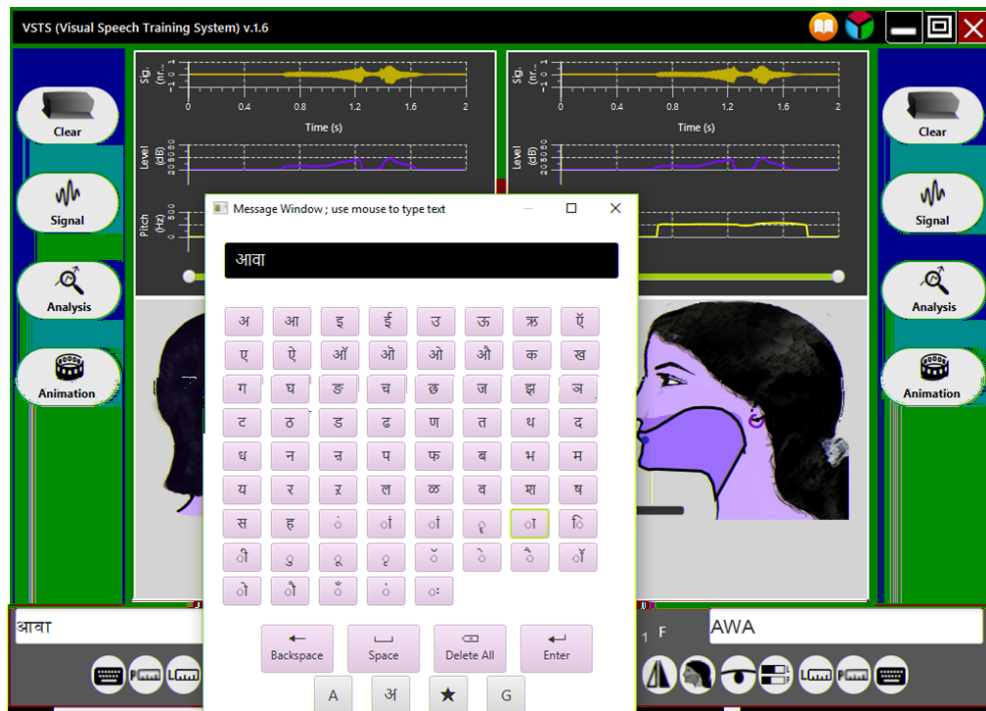


Figure 7.2: Message window for displaying prompt or feedback.

8. Color Customization

The color of different sections of the main screen can be customized by clicking on the 'Color' icon in the top-right corner of title bar. Effect of changing the color on different sections can be visualized by choosing the color from color pallet located in the front of the section as shown in the Figure 8.1. Clicking on the 'OK' button will accept the colors for subsequent use, while clicking on the 'Cancel' button will revert back to the previously selected colors.

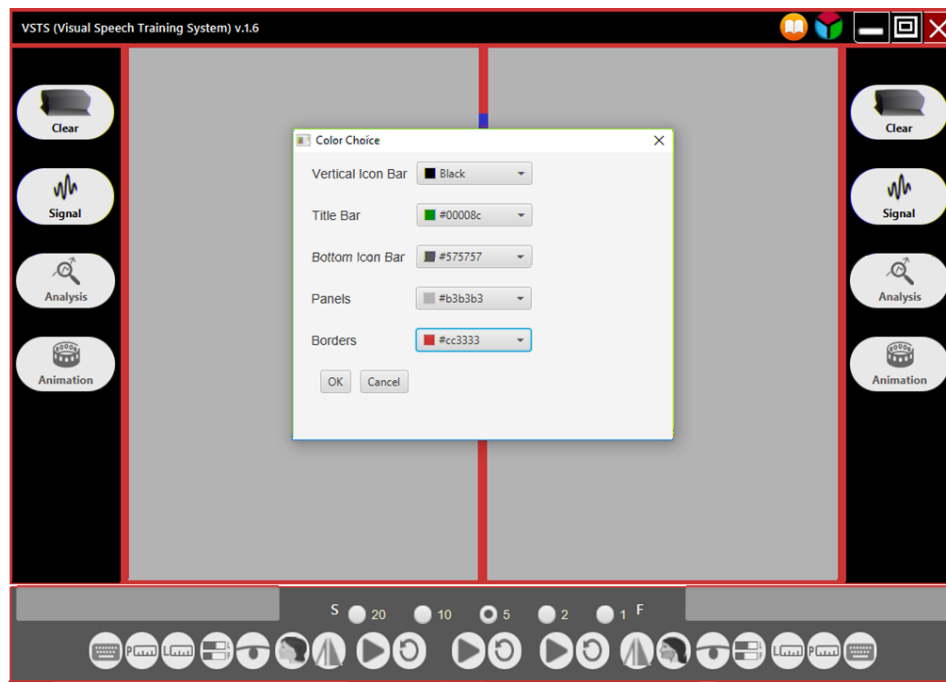


Figure 8.1: Window for selecting color.

9. VSTS Version History

VSTS v.1.x is the system for vowels and vowel-like sounds, with ‘x’ referring to different versions with significant changes incorporated to address the issues as identified during development and user interactions.

Sr.	Version	Functionality and Features
1.	VSTS v.1.1 (January 2017)	<ul style="list-style-type: none">• Integration of Matlab-based back-end signal processing engine & Java-based user interface.• Graphics framework with inputs from speech therapists and the animation experts.• Polygon-based animation interface in Java from the area value matrix from the Matlab-based back-end engine.• Incorporation of estimation & display of level & pitch.• Cursor-based measurement of analysis parameters.• Icon-based controls for analysis & animation.• Animation features: Segment selection, Play / Pause / Move / Reset, POA emphaziser, Level & Pitch bars, Face mirroring, Face select, Animation progress indicator.
2.	VSTS v.1.2 (March 2017)	<ul style="list-style-type: none">• Batch script for automatic installation (runtime environment of Java & Matlab, and Demo & Recorded Sound folders).• Configuration functionality for selecting colors of different user interface sections.• Executable compatible with 32/64-bit OS.• ‘Log’ file for runtime exceptions.
3.	VSTS v.1.3 (June 2017)	<ul style="list-style-type: none">• Icons on the vertical and horizontal bars revised for color-independent visibility & simplified functionality.• ‘Signal’ module for integrating recording and loading functions: Load, Record, Select, Reset, Play, Save & Proceed.• Display of Place of Articulation ('POA') revised as a rolling ball.• Zooming of spectrogram and areagram in separate window.• Subfolder for recorded sounds as a session, associated with date and time inside ‘recorded sounds’ folder.
4.	VSTS v.1.4 (August 2017)	<ul style="list-style-type: none">• Recording of speech signal up to 10 s duration, segment selection from 0.3 s to 4.8 s for analysis and animation.• Introduction of actual-time animation with audio playback.• Slow-motion animation with slow-down factors of 1, 2, 5, 10, 20.
5.	VSTS v.1.5 (October 2017)	<ul style="list-style-type: none">• Message window for prompt and feedback.• Redesigning of icons.• Rationalization of icon functionalities and features.
6.	VSTS v.1.6 (April 2018)	<ul style="list-style-type: none">• Signal Acquisition window functionality customized.• Slow Motion Audio Playback alongwith animation.• Redesigning of some icons.