

“曹艺舰的迷你语音学实验室”使用说明

中国语言文学系 曹艺舰

2001210879 cyjian@pku.edu.cn

“曹艺舰的迷你语音学实验室”是一个简单的语音分析程序，此程序能完成一些基础的语音学相关的分析工作，如播放语音，切分语音，显示基频、功率谱和共振峰等。

本项目为孔江平老师《语音分析与编程》课期末作业，使用 Matlab R2016b 进行开发，由学生在孔江平老师和梁昌维助教的指导下完成。

一、使用说明

打开 main.m 文件进入程序。

用户可以在 main.fig 对该程序的用户图形界面进行修改。

figure1_WindowButtonDownFcn 和 figure1_WindowButtonUpFcn 函数分别记录了鼠标按下和抬起时指针所在的坐标，进而实现对语音样本指定范围的截取。坐标的获取由函数 `get(axes, 'CurrentPoint')` 实现。

图形界面的主体包括两个图像框，上图用于呈现语音波形，下图用呈现语图和语音的可视化数据。当图像发生更改时，程序调用 `refresh_Callback` 函数，根据新得到的语音信号、坐标轴范围、语音可视化数据的开关，重新绘制波形图和语图，以达到刷新的效果。

接下来，简单介绍菜单栏中各个按钮的功能。工具栏上的按钮的功能都能在菜单栏中找到对应，故不再赘述。

1.1 “文件” 菜单

打开：对应 `open_Callback` 函数，打开一个语音文件，将其信号存放在变量 `handles.sig` 中，同时对各种参数进行初始化定义，如语音可视化数据的开关，语音的时长，文件读取状态等。该功能使用到的核心函数有：`audioread`、`uigetfile`。

保存：对应 `save_Callback` 函数，将程序中的语音文件覆盖保存至原文件。该功能使用到的核心函数有：`audiowrite`。

另存为：对应 `saveas_Callback` 函数，弹出窗口，输入名称，将程序中的语音

文件保存到新的位置。该功能使用到的核心函数有：audiowrite、uiputfile。

退出：对应 exit_Callback 函数，退出程序。

1.2 “查看” 菜单

查看全部：对应 view_all_Callback 函数，在窗口中展示 handles.sig 的全部波形。

查看选中区域：对应 view_selected_Callback 函数，在窗口中展示 handles.sig 被选中区域的波形。该功能通过把选中区域的两端传递到窗口的两端实现。

宽窄带语图切换：对应 switch_band_Callback 函数，在语图窗口中实现宽带语图和窄带语图的切换。切换由开关 handles.bandswitch 控制，开关存放于 refresh_Callback 函数，由 switch...case...语句实现。

语图颜色切换：对应 colormap_Callback 函数，在语图窗口中实现颜色的切换，有三种颜色，分别为 flipud(gray), hsv 和 jet。切换由 handles.colormap 控制，该变量所对应的各种颜色在 refresh_Callback 函数中，由 switch...case...语句实现。

输出波形图：对应 plot_wave_Callback 函数，把当前窗口的波形图输出到一个新的 figure 窗口。

输出语图：对应 plot_spec_Callback 函数，把当前窗口的语图输出到一个新的 figure 窗口。

1.3 “编辑” 菜单

撤销：对应 undo_Callback 函数，每次执行函数，handles.num_time 的值减少 1 个单位，而所有对 handles.sig 的操作都使 handles.num_time 的值加 1 个单位，将每个 num_time 所对应的 sig 存放在新的元胞数组中，refresh_Callback 函数根据当前的 num_time 读取对应的 sig 值。

缺点：只能撤销恢复对 sig 的操作，不能对振幅、基频、共振峰图的删减进行撤销恢复。

恢复：对应 redo_Callback 函数，每次执行函数，handles.num_time 的值增加 1 个单位。

删除（选中）：对应 cut_Callback 函数，删除选中的语音片段，设置两个变量 sig1 与 sig2，分别对应从左到右光标，从右光标到右，然后令 sig=[sig1 sig2]。

删除（未选中）：对应 trim_Callback 函数，删除未选中的语音片段。

上下翻转：对应 invert_Callback 函数，将语音波形上下翻转，令 sig=sig*-1。

前后翻转：对应 `inverse_Callback` 函数，将语音波形前后翻转，该功能使用到的核心函数有：`flipud`。

1.4 “效果” 菜单

播放（全部）：对应 `play_all_Callback` 函数，播放 `sig` 的全部音频，该功能使用到的核心函数有：`sound`。

播放（选中）：对应 `play_selected_Callback` 函数，播放 `sig` 的选中区域音频，该功能使用到的核心函数有：`sound`。

播放（未选中）：对应 `play_unselected_Callback` 函数，播放 `sig` 的未选中区域音频，该功能使用到的核心函数有：`sound`。

语音信号放大：对应 `wave_plus_Callback` 函数，放大波形至原来的 1.2 倍。

语音信号缩小：对应 `wave_minus_Callback` 函数，缩小波形至原来的 0.8 倍。

语音信号归一化：对应 `normalization_Callback` 函数，对波形进行归一化处理，使原波形的最大值定在 0.9，其他值保持相对大小关系。

低通滤波：对应 `low_pass_Callback` 函数，对语音信号进行低通滤波处理。

高通滤波：对应 `hig_pass_Callback` 函数，对语音信号进行高通滤波处理。

带通滤波：对应 `band_pass_Callback` 函数，对语音信号进行带通滤波处理。

滤波功能使用到的核心函数有：`remez`、`conv`。低通滤波 `remez` 函数的参数为`[1 1 0 0]`，高通滤波为`[0 0 1 1]`，带通滤波为`[0 1 1 0]`。

1.5 “音长” 菜单

查看音长（全部）：对应 `duration_all_Callback` 函数，查看整个语音样本的时长。

查看音长（选中）：对应 `duration_selected_Callback` 函数，查看语音样本选中区域的时长。

1.6 “音强” 菜单

显示能量：对应 `power_Callback` 函数，在语图上显示能量数据的折线图。显示由开关 `handles.powerswitch` 控制，开关存放于 `refresh_Callback` 函数。

该回调函数使用外部的 `power_amp` 函数，使用 `enframe` 函数分帧¹。

¹ 所有需要分帧处理的步骤均由 `enframe` 函数实现，下文不再赘述。

显示振幅: 对应 `amplitude_Callback` 函数, 在语图上显示振幅数据的折线图。显示由开关 `handles.amplitudeswitch` 控制, 开关存放于 `refresh_Callback` 函数。

该回调函数使用外部的 `amplitude` 函数, 振幅与能量的关系可表述为 $\text{amplitude} = 10 * \log_{10}(\text{power})$ 。

输出能量图: 对应 `power_output_Callback` 函数, 把能量的折线图打印到新的 figure。

输出振幅图: 对应 `amplitude_output_Callback` 函数, 把能量的折线图打印到新的 figure。

输出能量数据: 对应 `power_output_txt_Callback` 函数, 打印能量数据到 txt 文件, 实现时间与能量的对应。该功能使用到的核心函数有: `table`、`writetable`。

输出振幅数据: 对应 `amplitude_output_txt_Callback` 函数, 打印振幅数据到 txt 文件, 实现时间与振幅的对应。该功能使用到的核心函数有: `table`、`writetable`。

1.7 “音高” 菜单

自相关法: 对应 `f0a_Callback` 函数, 该菜单下的功能均调用外部的 `F0_extraction` 函数, 函数包括低通滤波、分帧、中心削波、三电平处理、计算自相关等步骤。

菜单下的其他功能有: 显示基频、查看全部区域基频、查看选中区域基频、删除选中区域基频、删除未选中区域基频、输出基频数据。

显示基频对应 `f0extraction_Callback` 函数, 在语图上显示基频数据的散点图。显示由开关 `handles.f0switch` 控制, 开关存放于 `refresh_Callback` 函数。

查看全部区域基频对应 `f0_view_all_a_Callback` 函数, 把整个语音信号的基频打印到新的 figure 中。

查看选中区域基频对应 `f0_view_all_a_Callback` 函数, 把选中区域的基频打印到新的 figure 中。

删除选中区域基频对应 `f0_delete_selected_a_Callback` 函数, 把选中区域的基频, 从存放基频数据的数组 `f0` 中删除, 同时在语图上刷新显示。

删除未选中区域基频对应 `f0_delete_unselected_a_Callback` 函数, 把选中区域的基频, 从存放基频数据的数组 `f0` 中删除, 同时在语图上刷新显示。

输出基频数据对应 `f0_output_original_a_Callback` 函数, 打印基频数据到 txt 文件, 该功能使用到的核心函数有: `table`、`writetable`。

倒谱法: 对应 `f0b_Callback` 函数, 该菜单下的功能均调用外部的 `F0_extraction_cep` 函数, 使用倒谱提取基频。

菜单下的其他功能与上文自相关法类似, 不再赘述。

Kawahara 法：对应 `f0c_Callback` 函数，该菜单下的功能均调用外部的 `MulticueF0v14` 函数²。

菜单下的其他功能与上文自相关法类似，不再赘述。

1.8 “傅立叶变换”菜单

功率谱（单点）：对应 `power_spec_Callback` 函数，把单点的功率谱打印到新的 figure 上。

该回调函数使用外部的 `fft_single` 函数，在选中的信号点左右共选取 512 个点，执行快速傅立叶变换，该功能使用到的核心函数有：`blackman`、`fft`。

功率谱（选中区域）：对应 `power_spec_selected_Callback` 函数，把选中区域的功率谱打印到新的 figure 上。

该回调函数使用外部的 `fft_all` 函数，对选中的信号范围执行快速傅立叶变换，该功能使用到的核心函数有：`blackman`、`fft`。

谱包络（单点）：对应 `spec_envelope_Callback` 函数，把单点的谱包络打印到新的 figure 上。

该回调函数使用外部的 `cep_env` 函数，在 `fft` 的基础上计算谱包络。该功能使用到的核心函数有：`dct`、`idct`。

谱包络（选中区域）：对应 `spec_envelope_selected_Callback` 函数，把选中区域的谱包络打印到新的 figure 上。

该回调函数使用外部的 `cep_env_all` 函数，在 `fft` 的基础上计算谱包络。该功能使用到的核心函数有：`dct`、`idct`。

功率谱和谱包络（单点）：对应 `power_spec_envelope_Callback` 函数，把单点的功率谱和谱包络打印到新的 figure 上，使用 `hold on` 功能，把上文两个相关功能实现在同一个 figure 中。

功率谱和谱包络（选中区域）：对应 `power_spec_envelope_selected_Callback` 函数，把选中区域的功率谱和谱包络打印到新的 figure 上，使用 `hold on` 功能，把上文两个相关功能实现在同一个 figure 中。

倒谱（单点）：对应 `cepstrum_Callback` 函数，把单点的倒谱打印到新的 figure 上，并把该点的基频值标注在图中。

该回调函数使用外部的 `cepstrum` 函数，在 `fft` 的基础上计算倒谱。该功能使用到的核心函数有：`dct`。

² 该函数的注释中提到了函数的设计者：Designed and coded by Hideki Kawahara, 31/August/2004 first conceiled version, 30/June/2016 refactored for Octave compatibility。

1.9 “线性预测分析” 菜单

LPC(单点): 对应 `lpc_single_Callback` 函数, 把单点的 LPC 打印到新的 figure 上。

该回调函数使用外部的 `lpc_single` 函数, 该功能使用到的核心函数有: `lpc`、`fft`。

LPC(选中区域): 对应 `lpc_all_Callback` 函数, 把选中区域的 LPC 打印到新的 figure 上。

该回调函数使用外部的 `lpc_all` 函数, 该功能使用到的核心函数有: `lpc`、`fft`。

逆滤波: 对应 `lpc_inversefilter_Callback` 函数, 对语音信号 `sig` 进行逆滤波处理。

该回调函数使用外部的 `lpc_inversefilter` 函数³, 该功能使用到的核心函数有: `lpc`、`real`、`filter`。

高频提升逆滤波: 对应 `lpc_preinversefilter_Callback` 函数, 对语音信号进行高频提升逆滤波处理。

该回调函数使用外部的 `lpc_pre_inversefilter` 函数⁴, 该功能使用到的核心函数有: `lpc`、`real`、`filter`。和逆滤波相比, 多了一步 `filter([1 -0.98],...)` 的高频提升。

1.10 “共振峰” 菜单

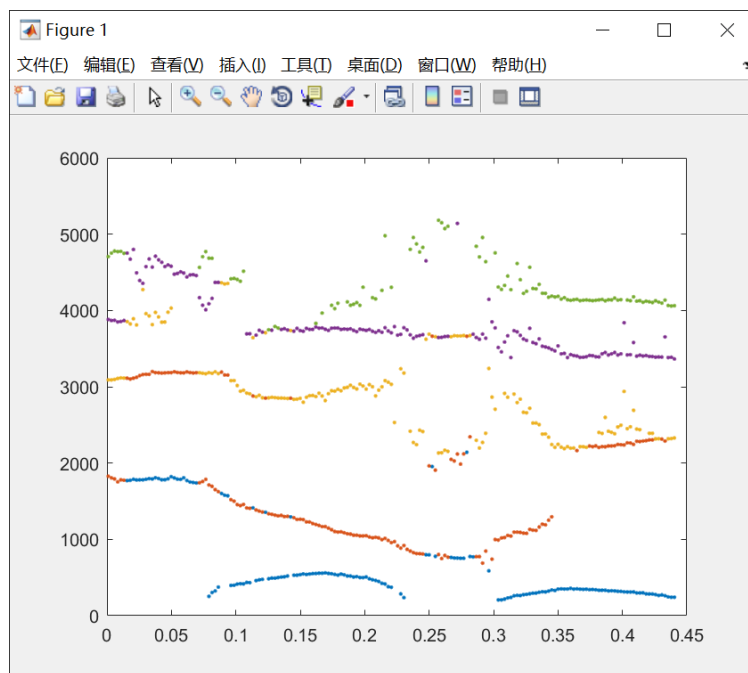
显示共振峰: 对应 `formant_display_Callback` 函数, 在语图上显示共振峰数据的散点图。显示由开关 `handles.formantswitch` 控制, 开关存放于 `refresh_Callback` 函数。

共振峰的提取调用了外部的 `formant_display` 函数, 该函数使用 LPC 求根法对共振峰进行计算。共振峰的频率要介于 150Hz 和采样频率的一半之间, 且带宽要小于 700Hz, 这是数据能确实成为共振峰的必要条件(宋知用 2013, 267)。学生把这个条件加到了 `formant_display` 函数中进行判断, 不符合要求的数据均用 NaN 替代。

但实际上, 共振峰的提取并不是特别理想, 以 `shehui.wav` 为例, 第二共振峰和第三共振峰会出现错乱, 如图所示:

³ 该函数由孔江平老师编写。

⁴ 同上。



输出共振峰图：对应 `formant_output_Callback` 函数，把整个语音信号的共振峰数据打印到新的 figure 中。

输出共振峰数据：对应 `formant_output_txt_Callback` 函数，打印共振峰数据到 txt 文件，前五个共振峰分布在表格的五行。该功能使用到的核心函数有：`table`、`writetable`。

二、结语与不足

这学期的学习让学生收获很多，在今后的田野调查中，学生可以试着用自己编写的程序去解决一些实际问题，如更高效地提取基频，并直接在 Matlab 内部完成数据的输出和统计，这会大大提升研究生的工作效率。

当然，不足也是存在的：程序方面，所有的功能目前都在采样频率为 11025Hz 的 `shehui.wav` 上完成，当采样率改变后，函数内部的一些参数可能也会改变。把语音数据投射到语图的时间轴上时，对应关系可能会有小数的偏差，使得信号点和时间并不总是一一对应。此外，上文提到过的撤销恢复问题、共振峰问题，都可以进一步改进。

学生个人方面，我对许多语音学概念的数学和物理意义还不是很了解，仍需进一步学习。

参考文献

- [1]. 宋知用. MATLAB 在语音信号分析与合成中的应用[M]. 北京航空航天大学出版社, 2013