

# 语音分析程序说明

Phonetics Lab ver 1.0, Jan 25<sup>th</sup> 2025

by Dat-Danny Ta

## 目录

- 一、 程序简介
- 二、 语音播放与查看
- 三、 时域编辑
- 四、 录音
- 五、 逆滤波
- 六、 振幅
- 七、 语图
- 八、 基频
- 九、 共振峰
- 十、 FFT 快速傅立叶变换
- 十一、 LPC 分析
- 十二、 语音合成

\*\*\*

## 一、 程序简介

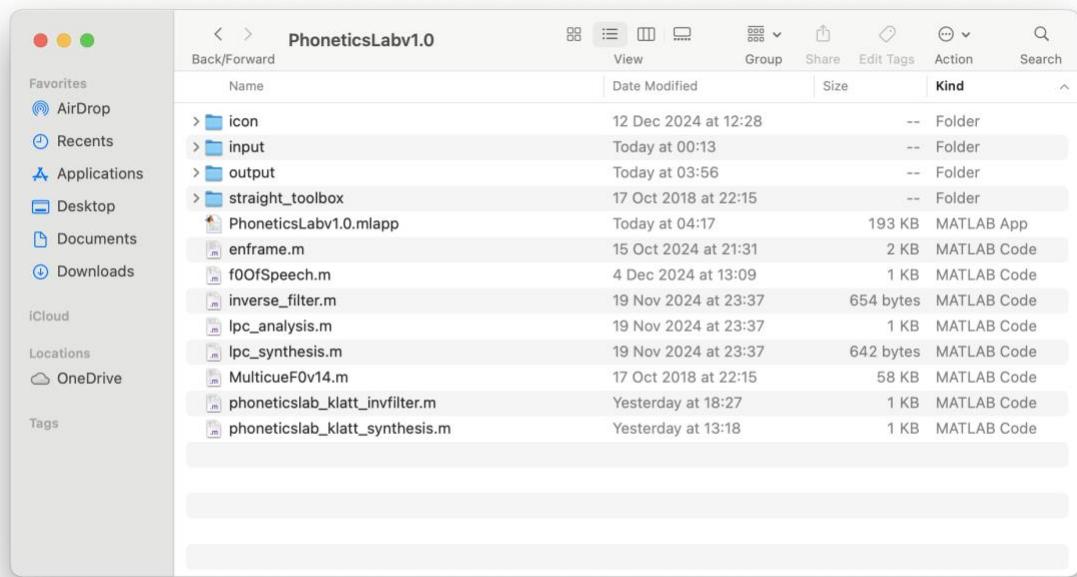
本程序在 MATLAB 中运行，能够完成较为全面的语音分析工作，其功能比较齐全，包括录音、语音播放、时域编辑、频域分析以及语音合成器等。

在此文件夹中，icon 为程序内 button 的标志，input 和 output 分别为数据输入和输出文件夹。您在程序内部提取出来的参数将保存至 output 文件夹里。

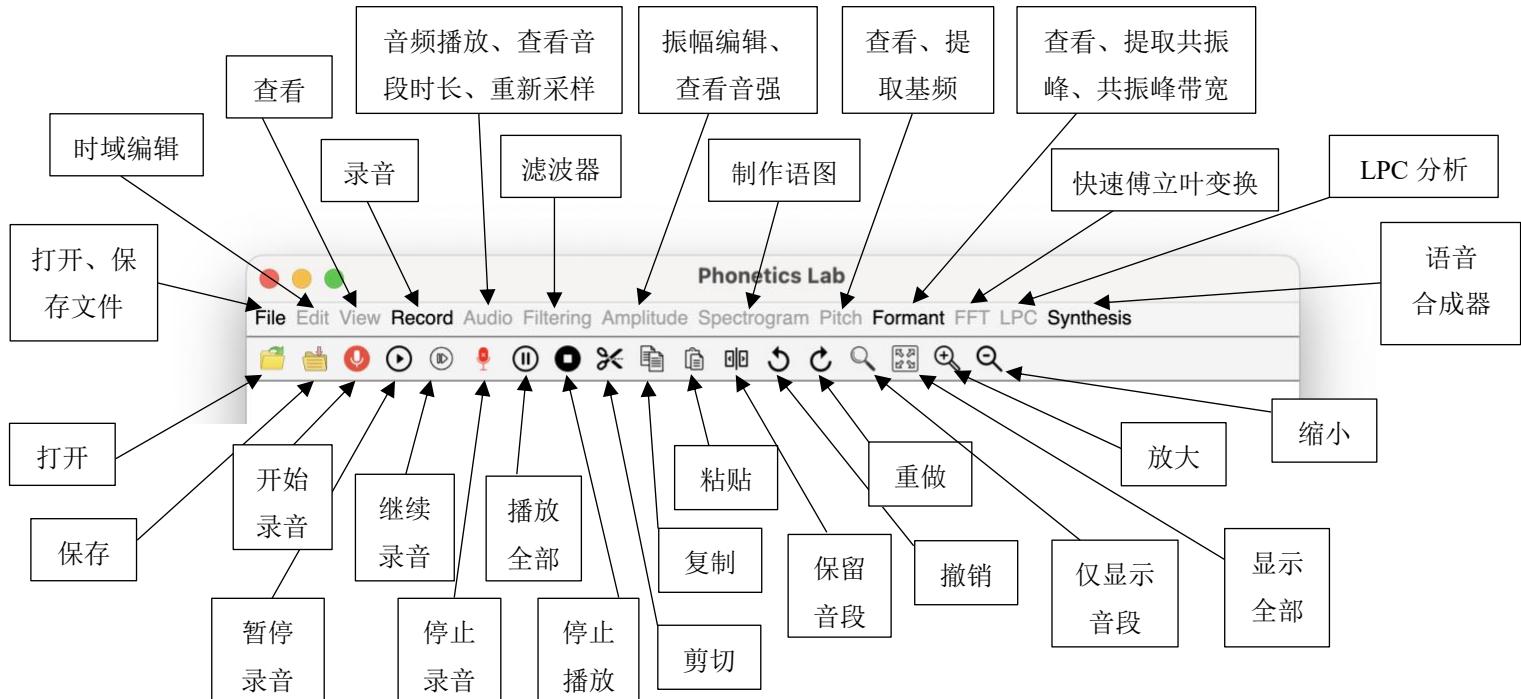
本程序的 straight\_toolbox 参考 Kawahara, H. 等人（1999）的研究成果<sup>1</sup>。enframe.m、f0OfSpeech.m、inverse\_filter.m、lpc\_analysis.m、lpc\_synthesis.m、phoneticslab\_klatt\_invfilter.m、phoneticslab\_klatt\_synthesis.m 等函数由北京大学孔江平教授和梁昌维老师提供。

---

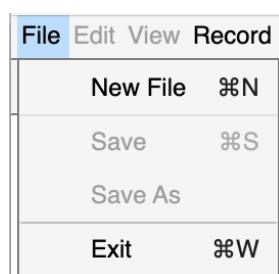
<sup>1</sup> Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, 27(3-4), 187-207. [https://doi.org/10.1016/S0167-6393\(99\)00063-3](https://doi.org/10.1016/S0167-6393(99)00063-3)



## 菜单说明：



## File 选项：



- New File:** 打开新语音文件
- Save:** 将目前语音文件保存到目前工作路径，默认格式为 'recording\_**目前时间**.wav'。
- Save As:** 选择工作路径，输入文件名，保存目前语音文件。
- Exit:** 退出本程序。退出程序之前需要确认是否要关闭程序。



**快捷键:** Windows 用户请将 Cmd 替换为 Ctrl

Cmd + N = New (打开新的语音文件)

Cmd + S = Save (保存目前的语音文件)

Cmd + W = Exit (退出程序)

Cmd + X = Cut (剪切)

Cmd + C = Copy (复制)

Cmd + V = Paste (粘贴)

Cmd + T = Trim (仅保留光标内的音段)

Cmd + R = Reverse (倒放光标内的语音信号)

Cmd + Z = Undo (撤销, 最多可撤销 5 次)

Cmd + Y = Redo (重做)

Cmd + I = Zoom In (放大)

Cmd + O = Zoom Out (缩小)

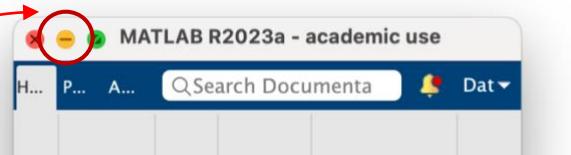
Cmd + P = Play Selected (播放光标内的音段)

Cmd + U = Play Unselected (播放光标外的音段)

**光标:** 先按下, 滑动后放手即可。若需要查看某一点的参数, 只需在目标位置点击一下就好。

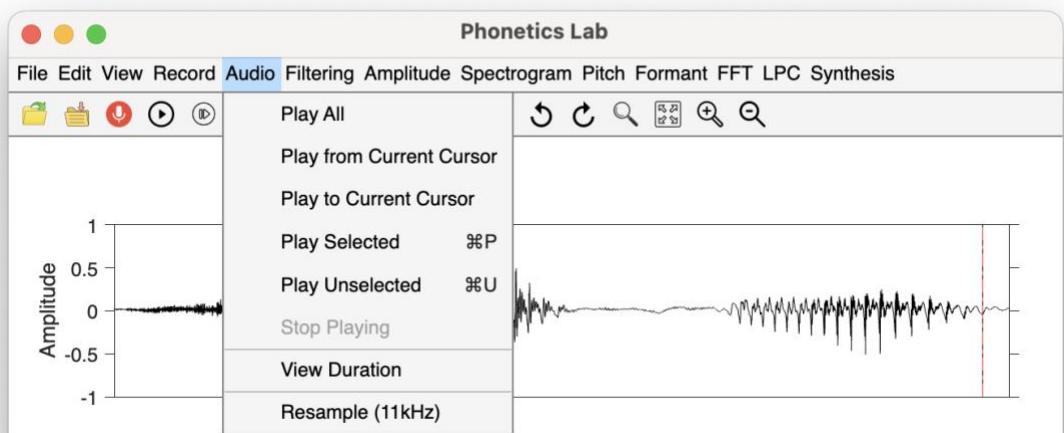
Tips for MacOS users: 点击 minimize

MATLAB 的窗口, 这样程序运行的时候就不要重新找 Phonetics Lab 窗口了!



## 二、 语音播放与查看

Audio 菜单功能:



**Play All:** 播放已打开的语音文件;

**Play from Current Cursor:** 从光标位置开始播放;

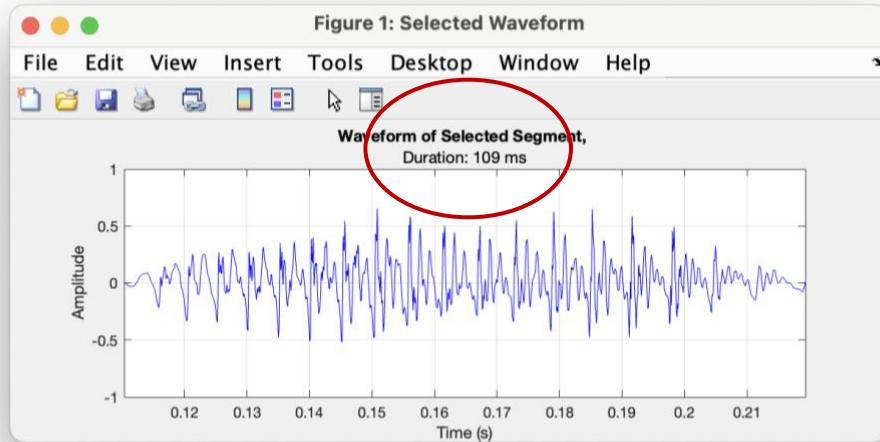
**Play to Current Cursor:** 从头播放到光标位置;

**Play Selected:** 播放光标内的音段;

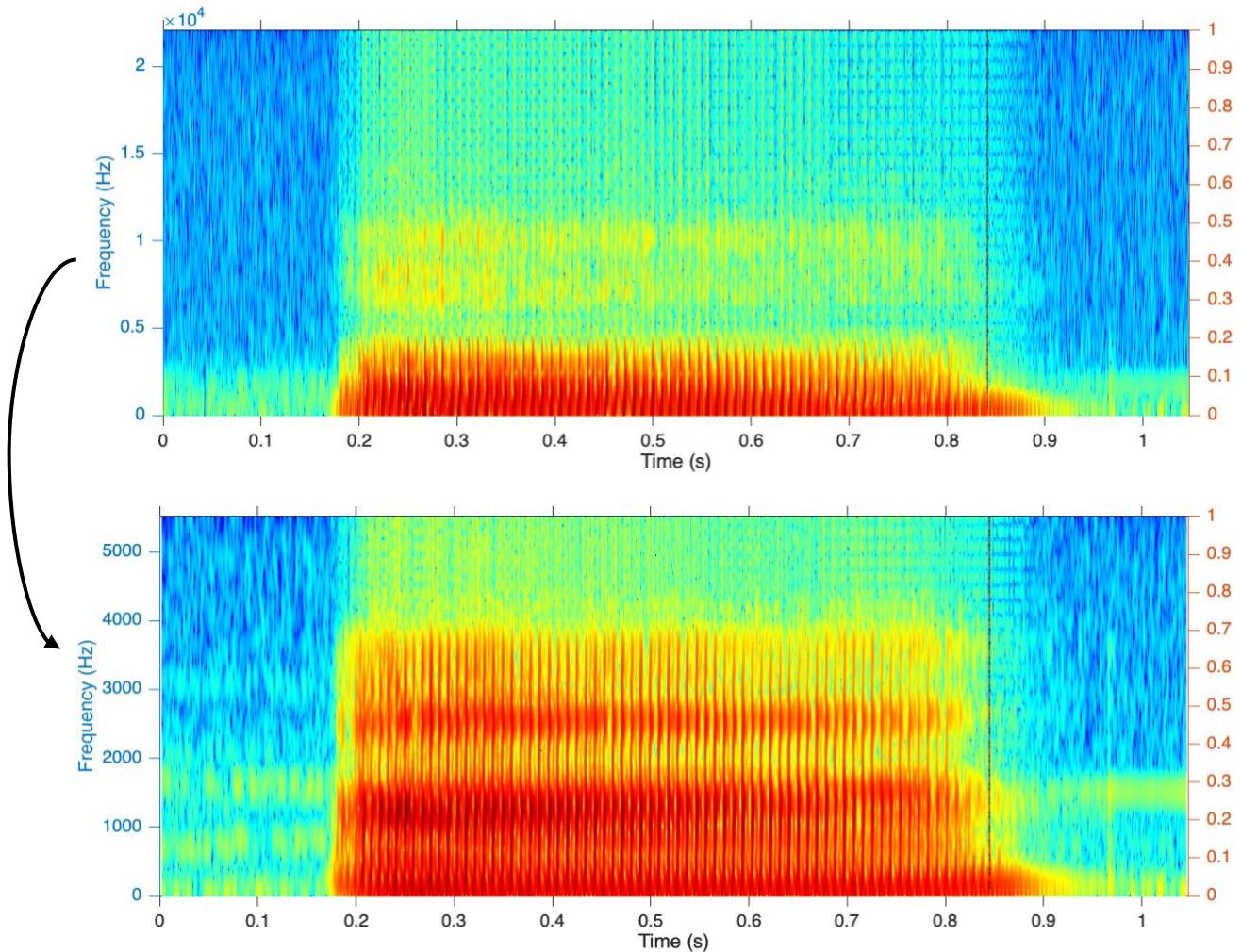
**Play Unselected:** 播放光标外的音段；

**Stop Playing:** 停止播放；

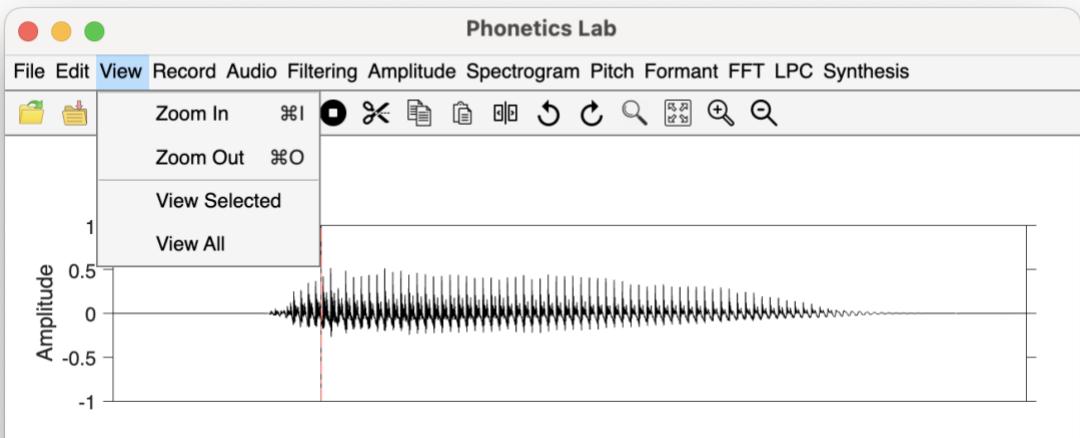
**View Duration:** 查看光标内音段的时长。选取音段，点击 View Duration 后弹出如下窗口：



**Resample (11kHz):** 重新采样到 11kHz



View 菜单功能：

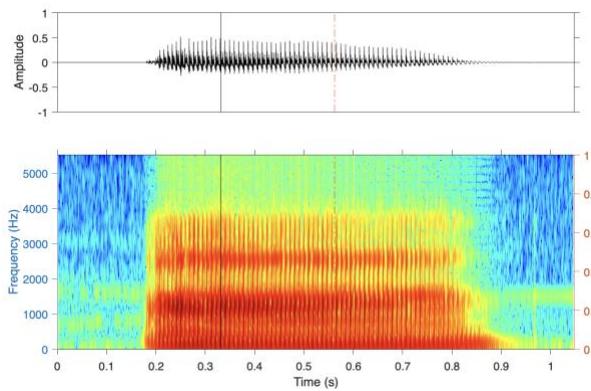


**Zoom In:** 放大；

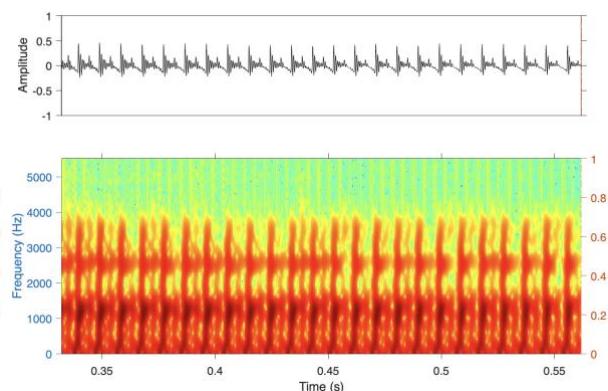
**Zoom Out:** 缩小；

**View Selected:** 仅显示光标内的音段；

先选取音段：

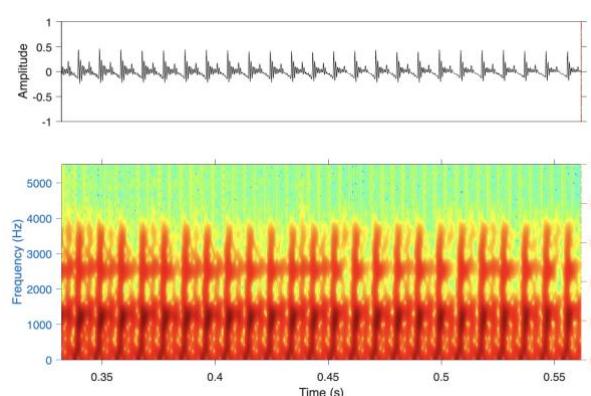


点击 View Selected 查看光标内的音段：

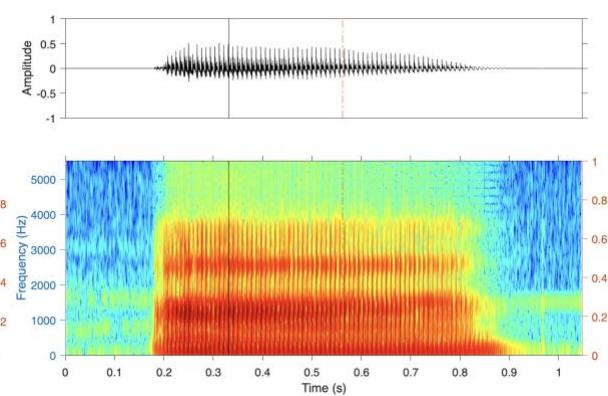


**View All:** 显示全部。

查看光标内的音段：

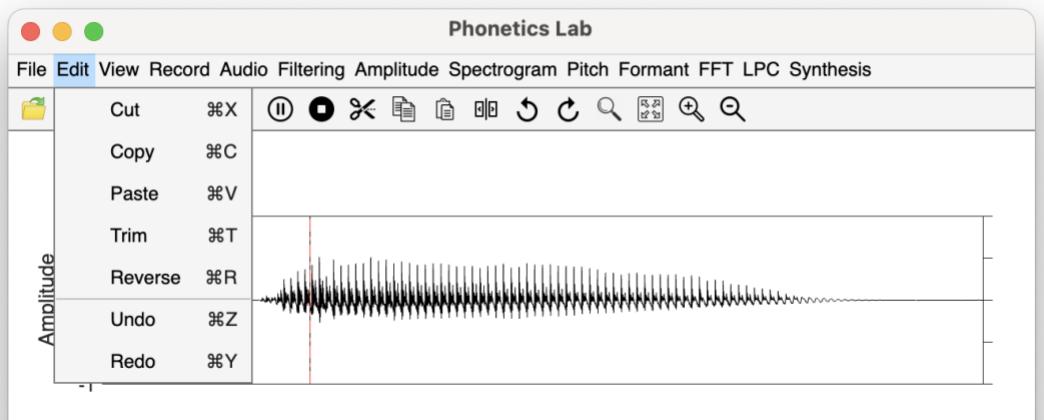


点击 View All 查看全部：



### 三、 时域编辑

Edit 菜单功能:

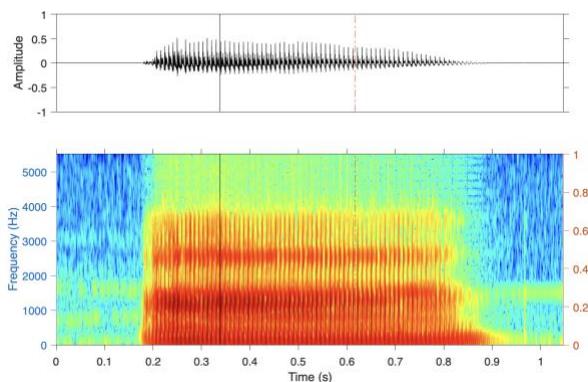


**Cut、Copy、Paste** 功能分别为剪切、复制、粘贴。

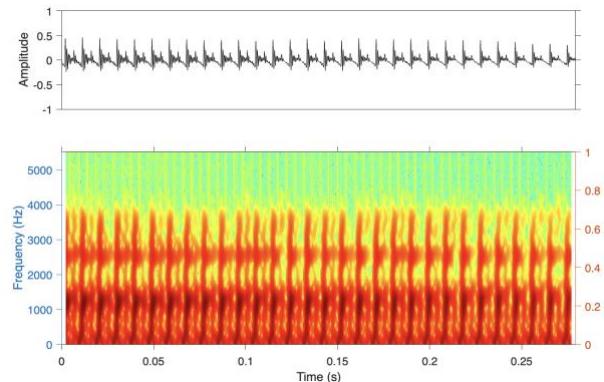
**Undo、Redo** 分别为撤销、重做。请注意，本系统最多可撤销 5 次。

**Trim:** 保留光标内的音段;

先选取音段:

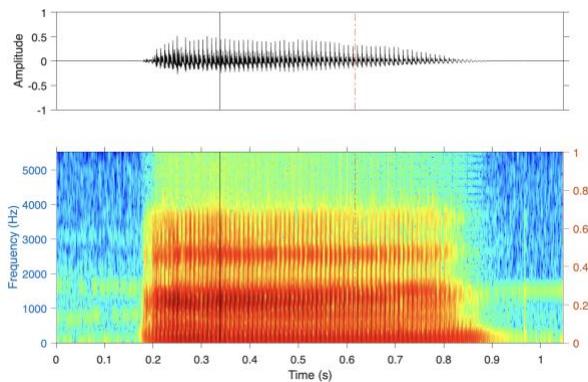


点击 Trim 仅保留光标内的音段:

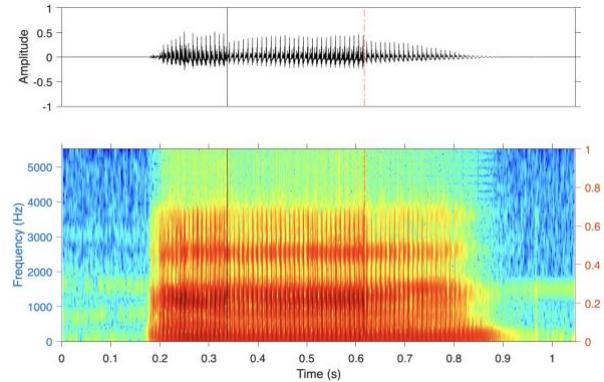


**Reverse:** 倒放光标内的音段;

先选取音段:



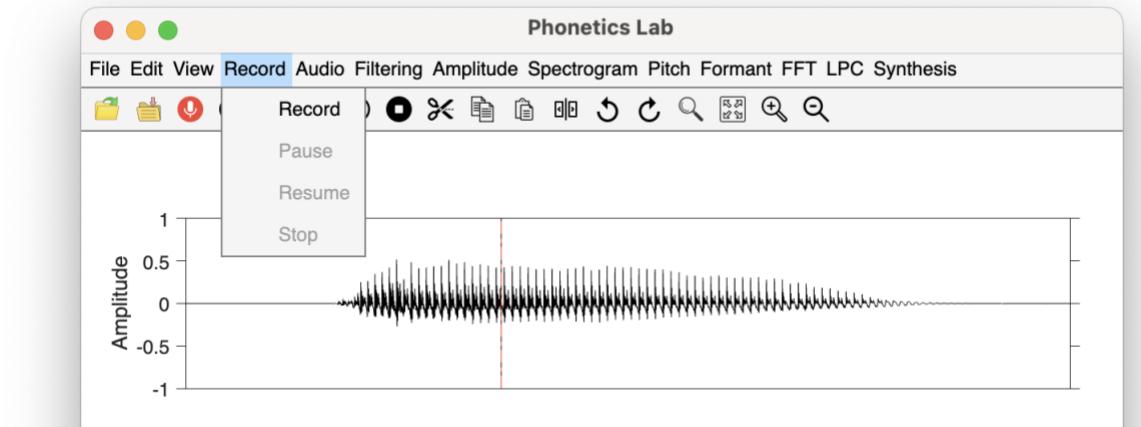
点击 Reverse 倒放光标内的音段:



想要倒放全部的话，将鼠标从头滑动到音频最后，选取全部音段即可。

## 四、录音

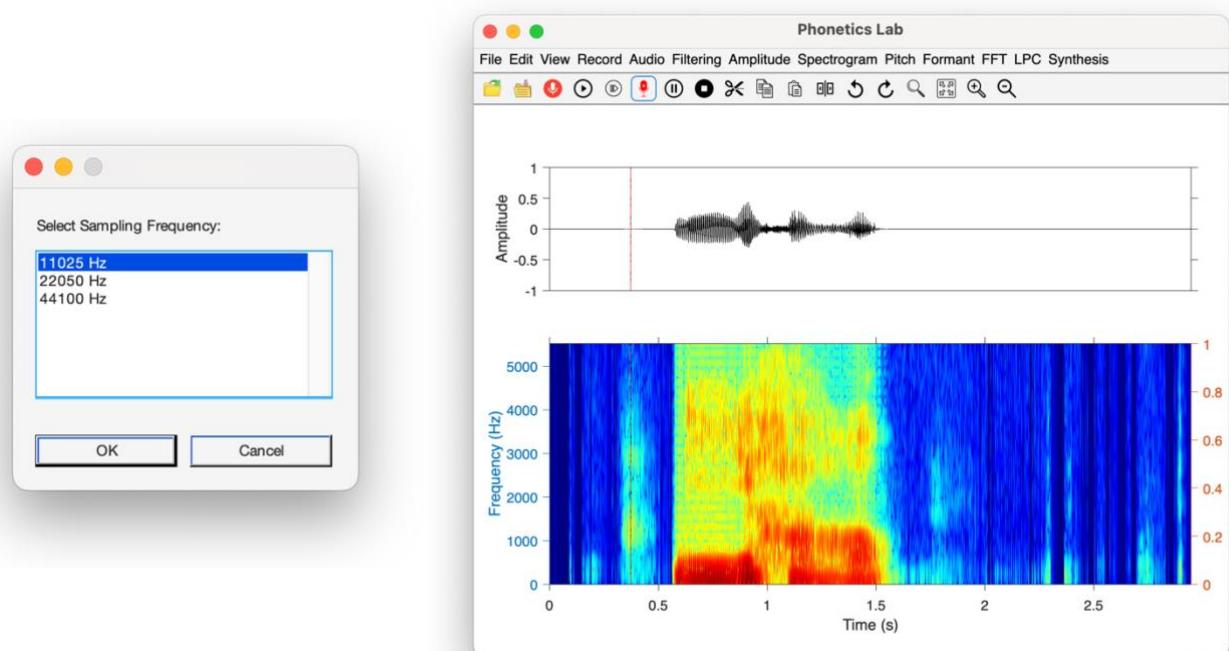
Record 菜单功能：



**Record、Pause、Resume、Stop** 分别是开始、暂停、继续、停止录音。

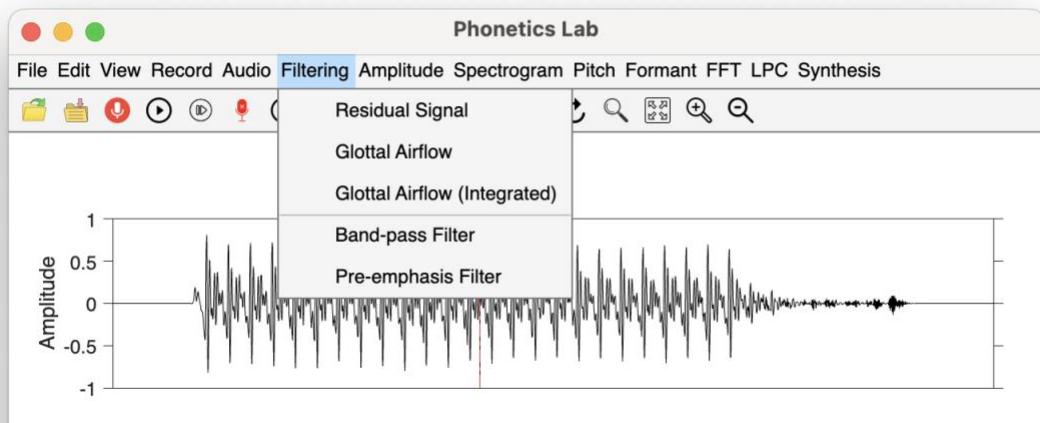
点击 Record 之后，请先选择采样频率。

点击 OK 确认后，系统开始录音，您可随时暂停、继续录音。停止录音后将自动制作波形图和语图：

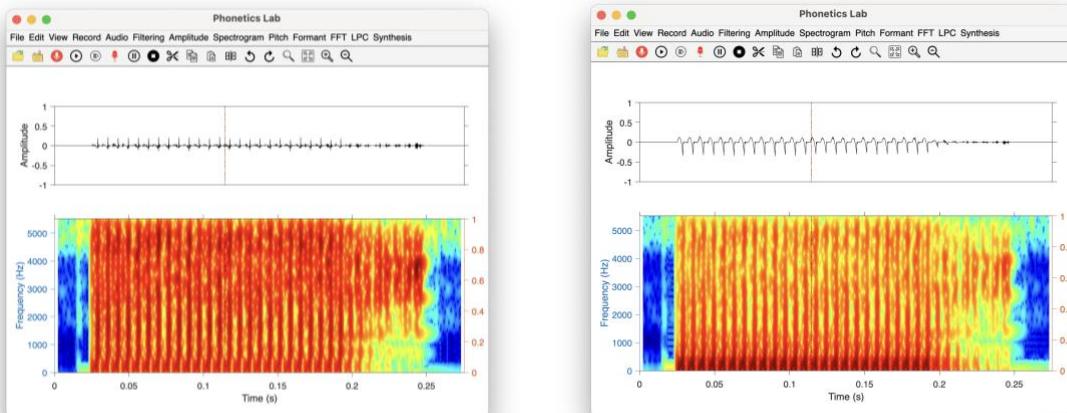


## 五、逆滤波

Filtering 菜单功能:

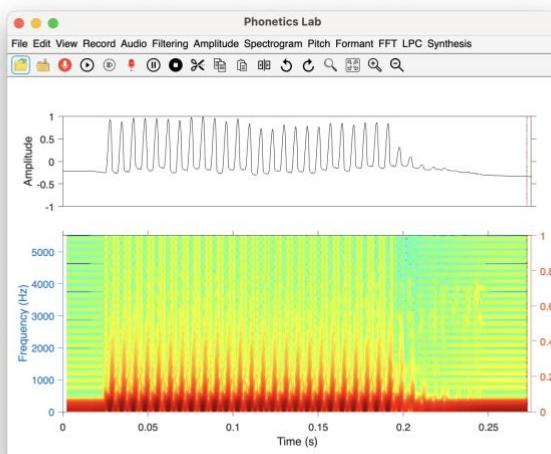


**Residual Signal:** 光标内（如有）或全部音频的残差信号（左图）；

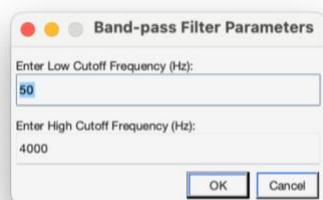


**Glottal Airflow:** 光标内（如有）或全部音频的声门气流信号（右图）；

**Glottal Airflow (Integrated):** 光标内（如有）或全部音频的声门气流信号（积分）；



**Band-pass Filter** (带通滤波器): 允许某个频率范围内的信号通过，同时衰减低于下限频率和高于上限频率的信号，用于提取特定频率范围的信号，例如语音信号的特定共振峰。

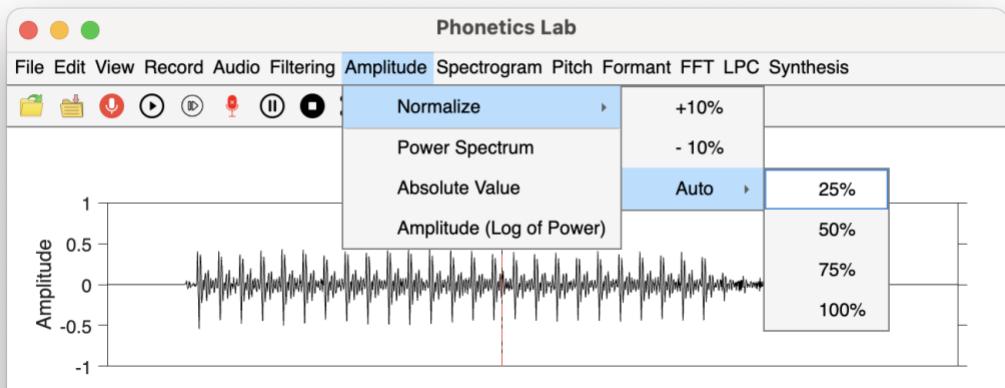


**Pre-emphasis Filter** (预加重滤波器): 增强信号的高频成分(通常通过一阶高通滤波器实现)，以补偿高频成分在传输过程中的衰减。若元音共振峰提取不准确，可尝试先针对音频进行预加重滤波器处理。

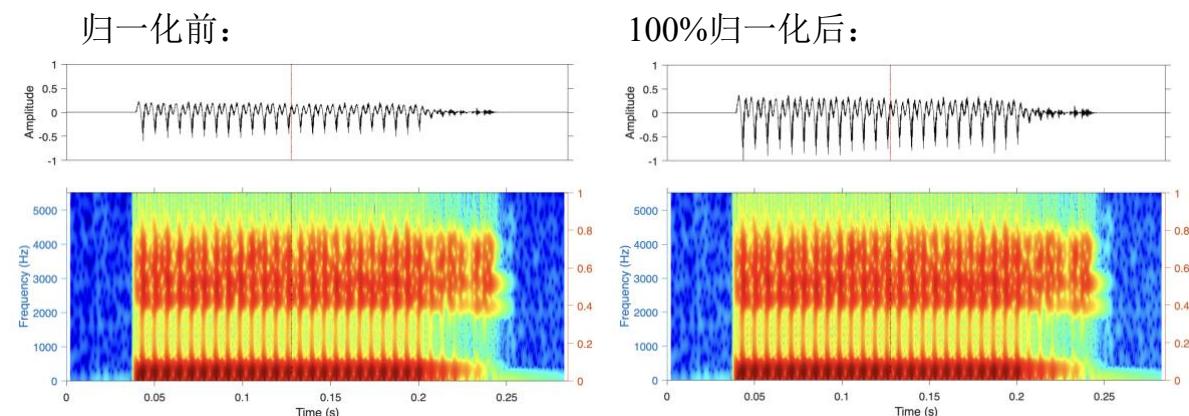


## 六、 振幅

Amplitude 菜单功能：



**Normalize** (归一化): 对振幅进行归一化。可手±10%，或直接选取归一 25%、50%、75%、100%。

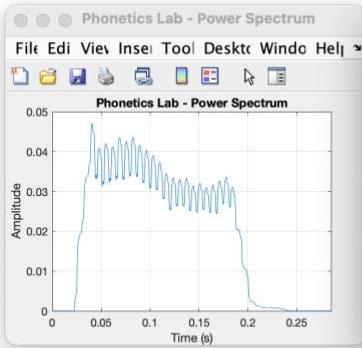


**Power Spectrum** (功率谱): 查看音频的功率谱；

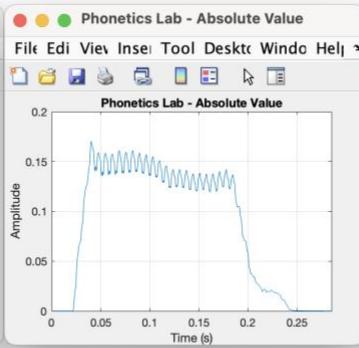
**Absolute Value** (绝对值): 查看音频的音强绝对值图形；

**Amplitude (Log of Power)**: 查看音频的音强图。

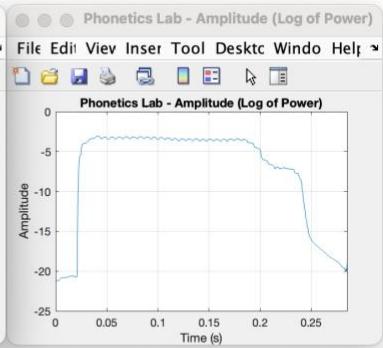
Power Spectrum



Absolute Value

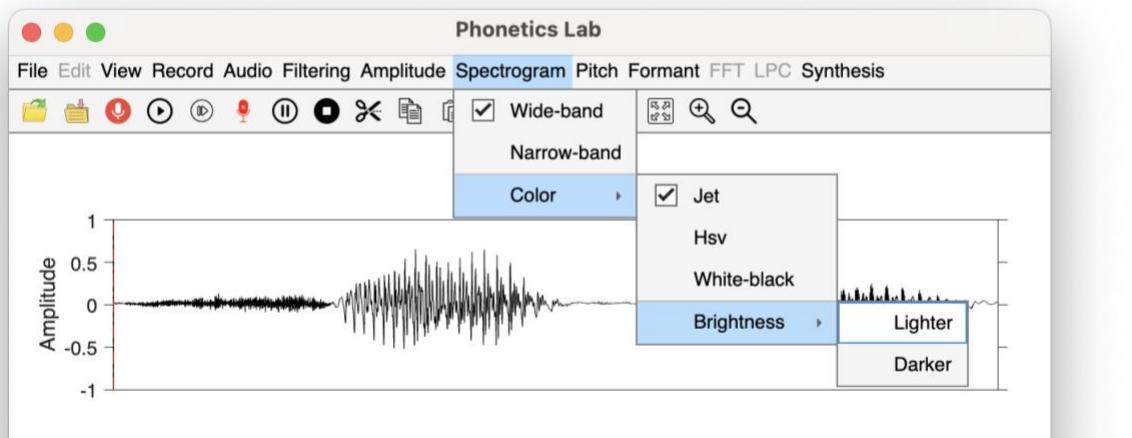


Amplitude (Log of Power)



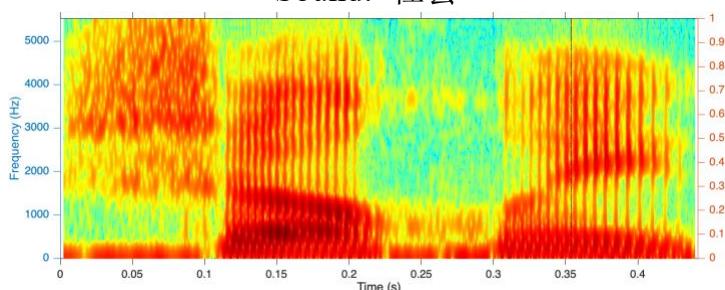
## 七、语图

Spectrogram 菜单功能：



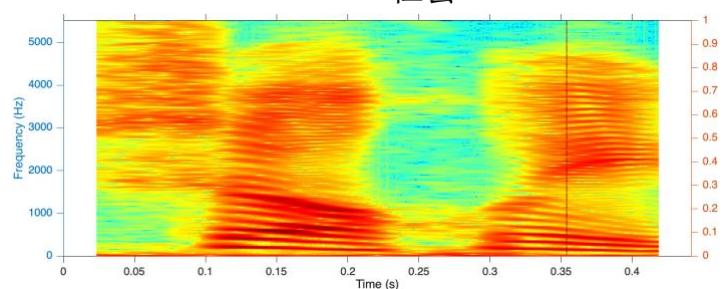
**Wide-band:** 制作宽带语图；

Sound: 社会



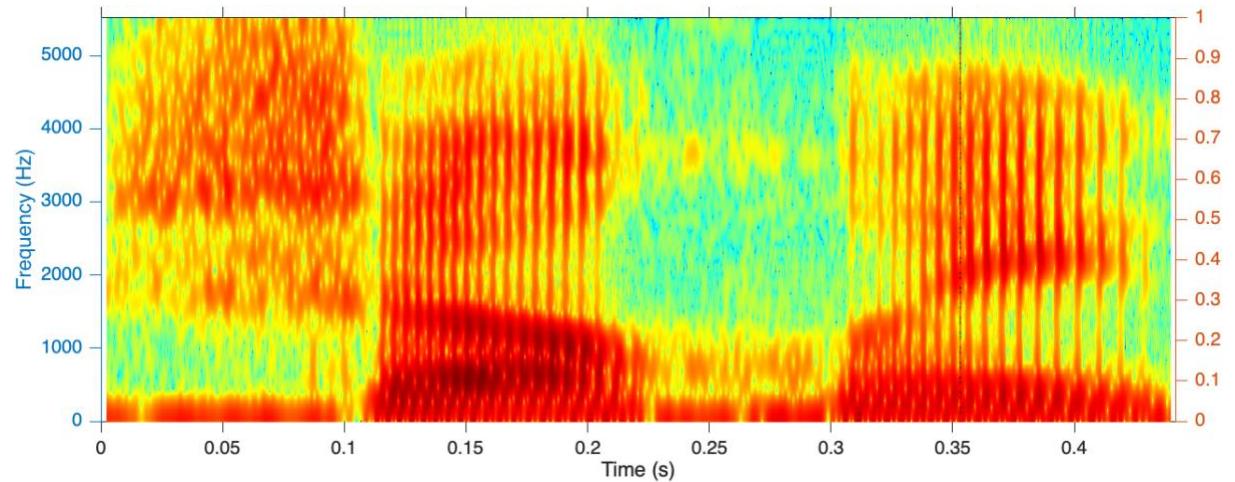
**Narrow-band:** 制作窄带语图；

Sound: 社会

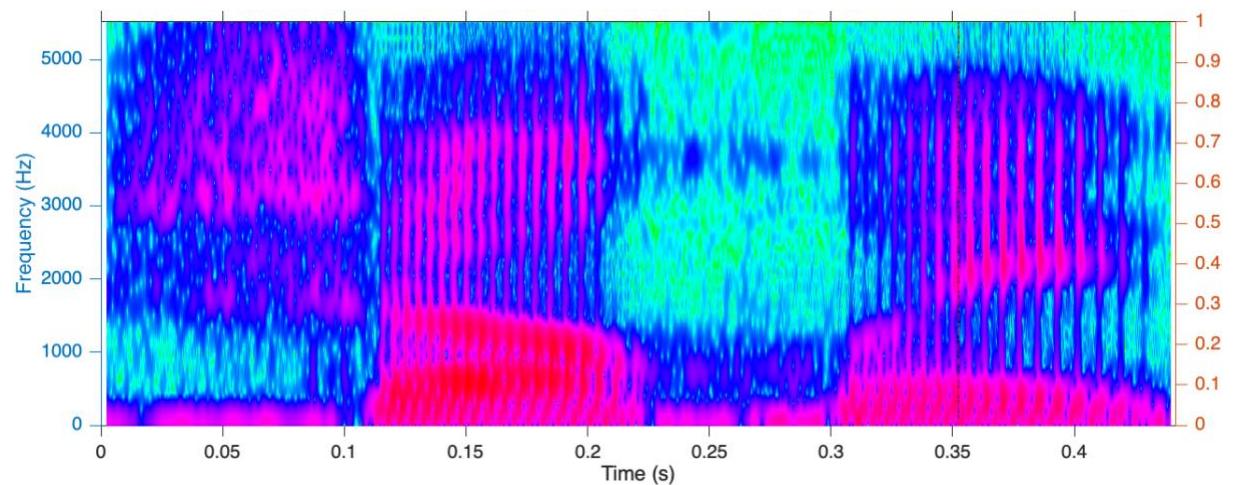


**Color:** 更换语图颜色, 目前支持 Jet、Hsv、White-black 三种颜色 (如下图);  
**Brightness:** 调整语图亮度。

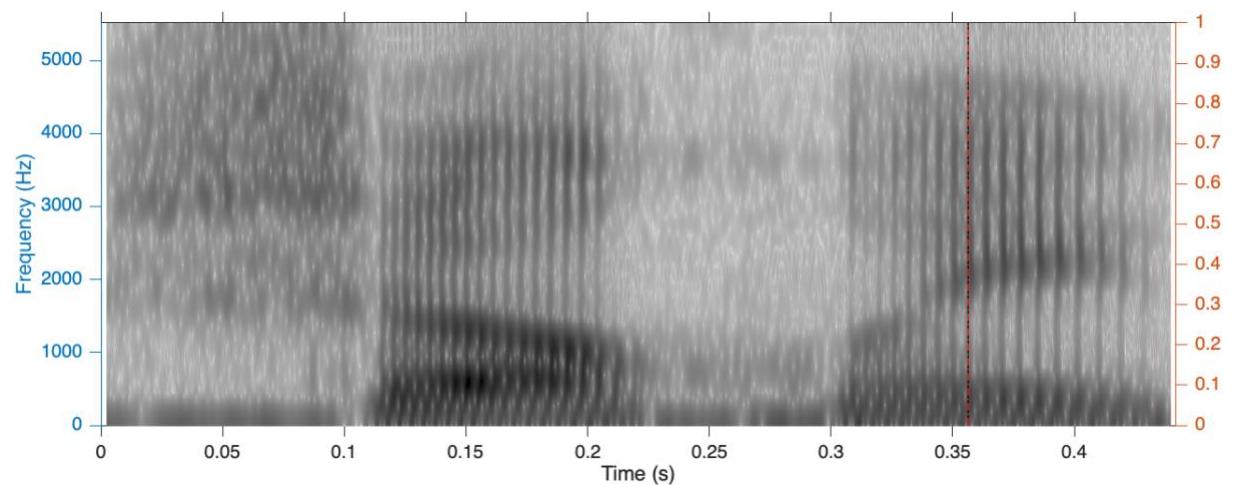
Sound: 社会; color: Jet



Sound: 社会; color: Hsv

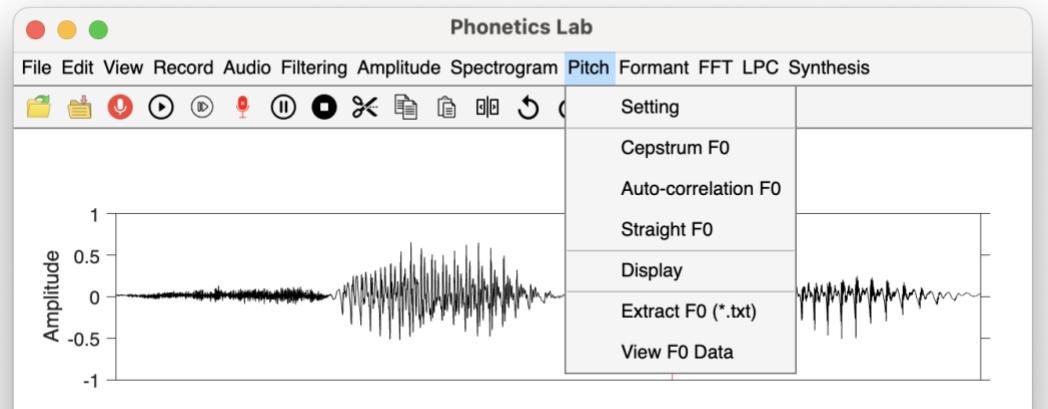


Sound: 社会; color: White-black



## 八、 基频

Pitch 菜单功能：



**Setting** (设置): 调整基频上下限、ZCR (过零率, Zero-Crossing Rate)、STE (短时能量, Short-Time Energy) 等参数。默认值如右图所示:



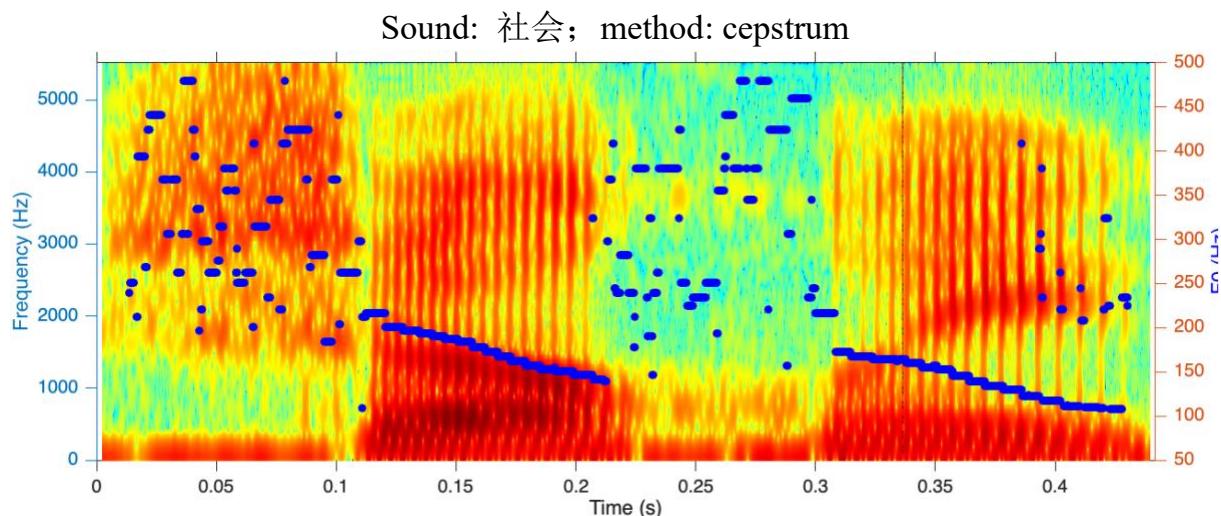
**Cepstrum F0:** 利用倒谱方法计算基频;

**Auto-correlation F0:** 利用自相关方法计算基频;

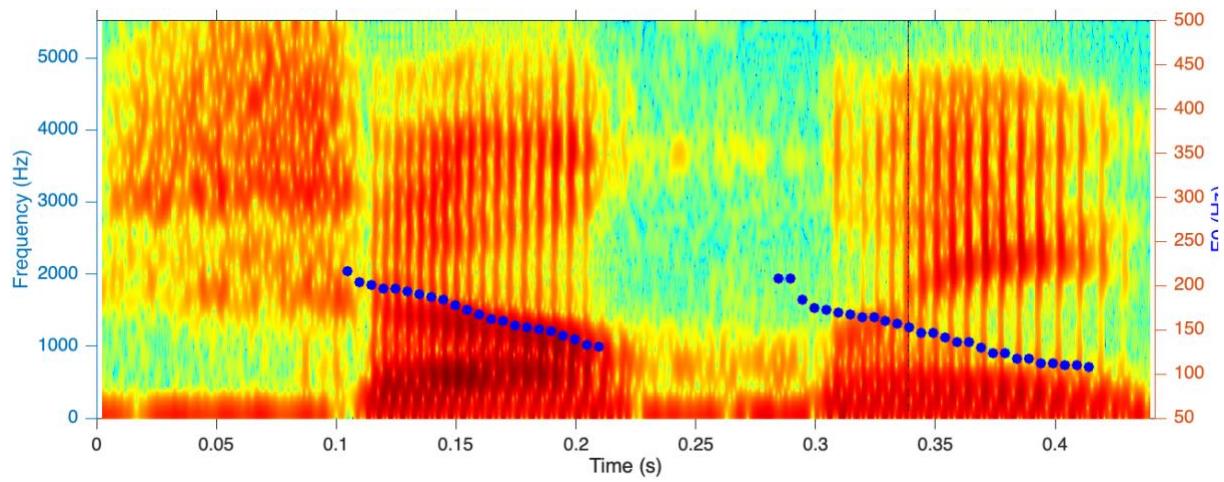
**Straight F0:** 利用 Straight 方法 (Kawahara 等人, 1999) 计算基频;

**Display:** 取消显示基频曲线;

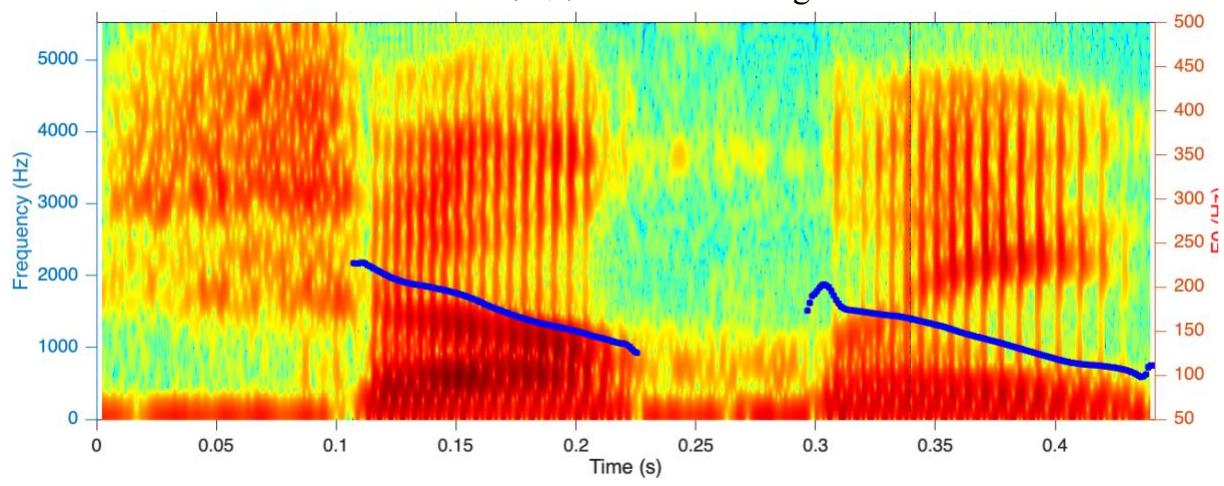
三种基频算法比较:



Sound: 社会; method: auto-correlation

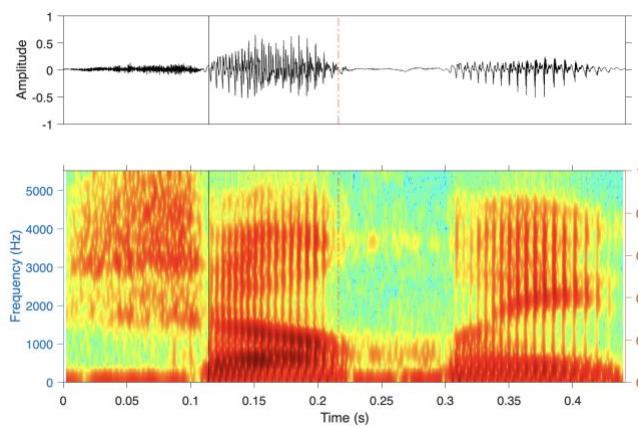


Sound: 社会; method: straight

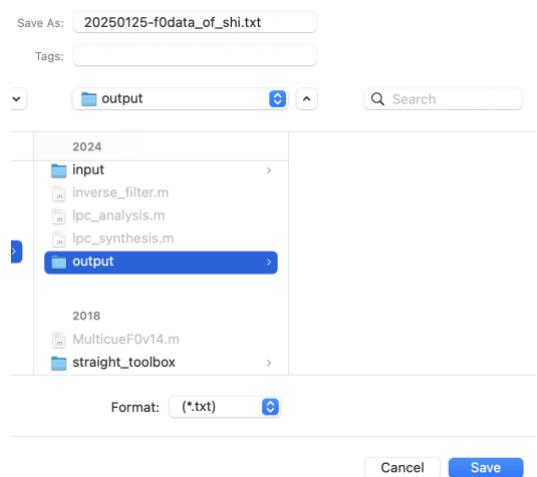


**Extract F0 Data (\*.txt):** 将光标内音段的基频保存为.txt文件（光标内 11 个等距点的数据，含时间点和对应该时间点的基频值，采用 Straight 方法）。

先选取音段：



输入文件名、选择工作路径：

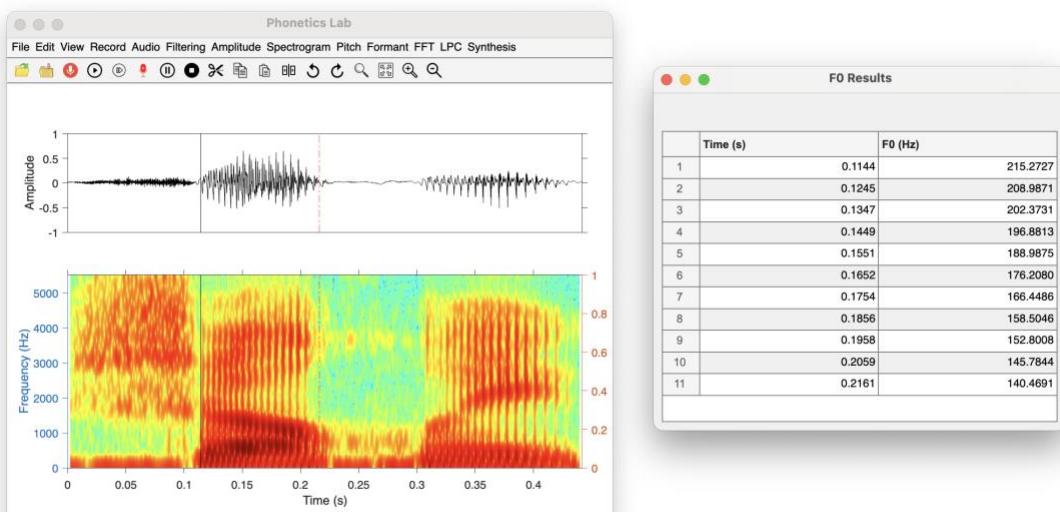




您的数据文件格式如下，将数据导入 Excel、R 等程序作图即可！

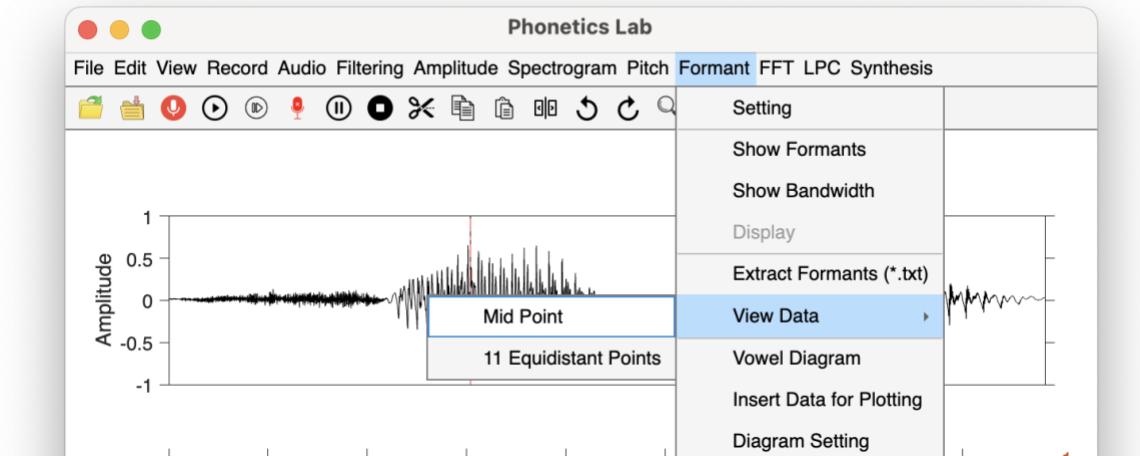
20250125-f0data_of_shi.txt	
Time (s)	F0 (Hz)
0.114359	215.272721
0.124533	208.987091
0.134707	202.373064
0.144881	196.881338
0.155056	188.987484
0.165230	177.336591
0.175404	166.448591
0.185578	158.504649
0.195752	152.800844
0.205926	145.784352
0.216100	140.469148

**View F0 Data:** 查看基频值（光标内 11 个等距点的数据，您也可以点击在语图上的某一个点，查看那一点的数据，含时间点和对应该时间点的基频值，Straight 方法）。F0 Results 窗口的数据可以复制粘贴哦！



## 九、 共振峰

Formant 菜单功能:

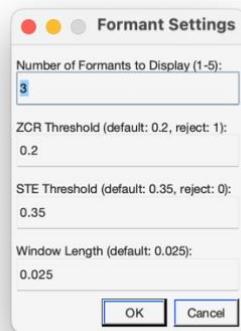


**Setting** (设置): 调整显示的共振峰数量、ZCR (过零率, Zero-Crossing Rate)、STE(短时能量, Short-Time Energy) 和窗长。默认值如右图所示:

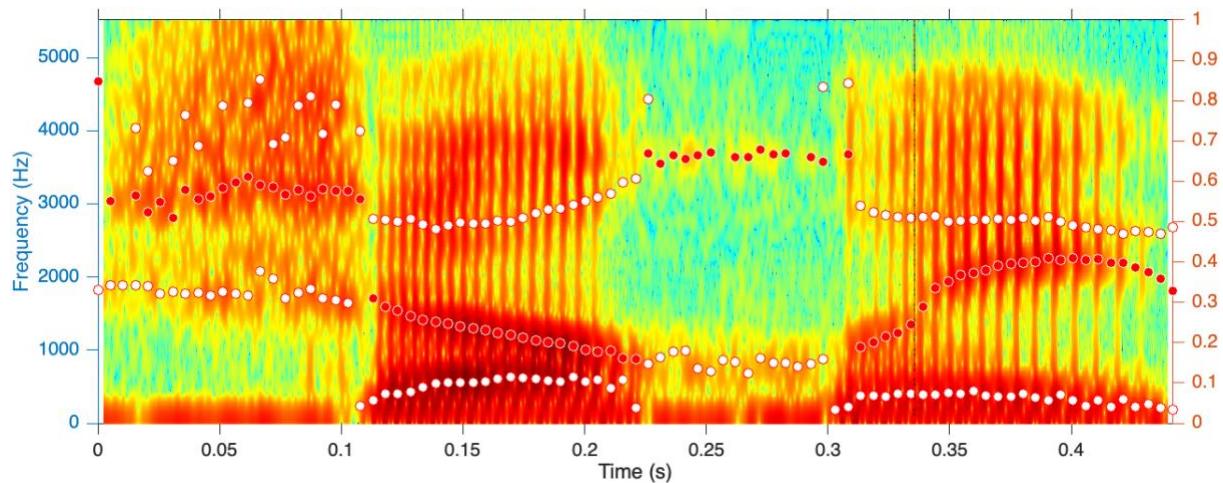
**Show Formants:** 显示共振峰;

**Show Bandwidth:** 显示共振峰及其带宽;

**Display:** 取消显示共振峰。

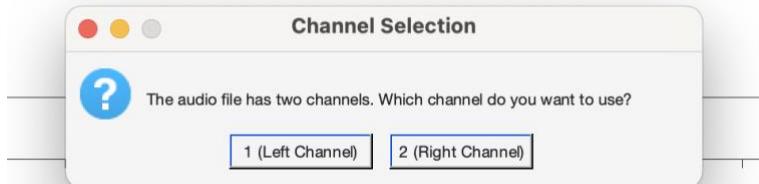


Sound: 社会

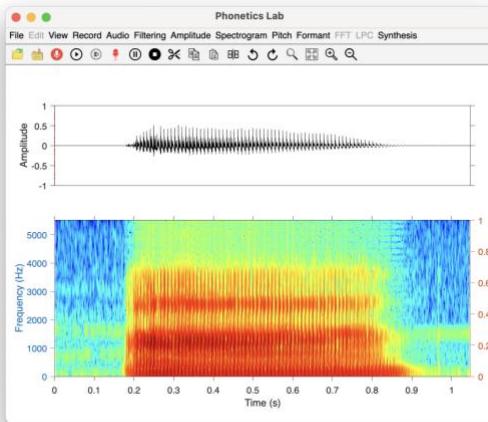


如果程序所提取的共振峰不太理想，可事先对语音信号进行滤波器处理（如 Pre-emphasis Filter），这样可提高提取效果。下面的语图为元音/a/，采样频率 44kHz，16bit，双声道，录音设备：FIFINE USB T669。

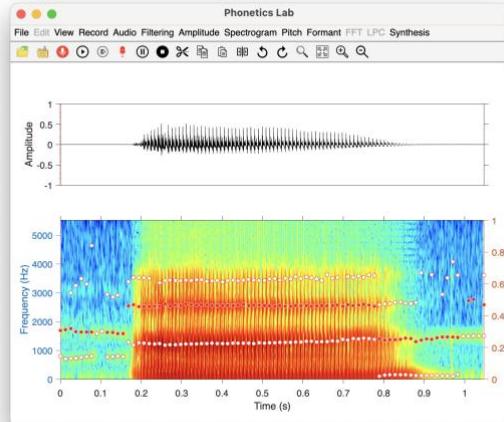
先选取需要分析的声道<sup>2</sup>:



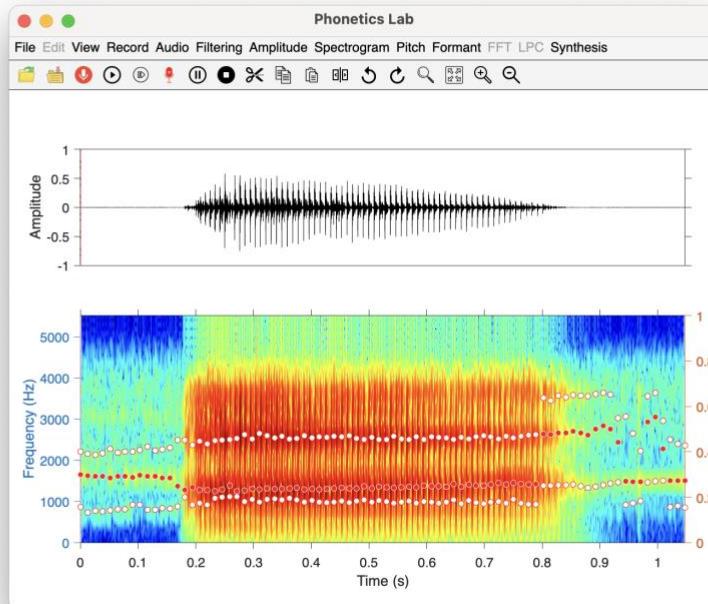
确认声道后重新采样到 11kHz:



首次提取共振峰不太理想:



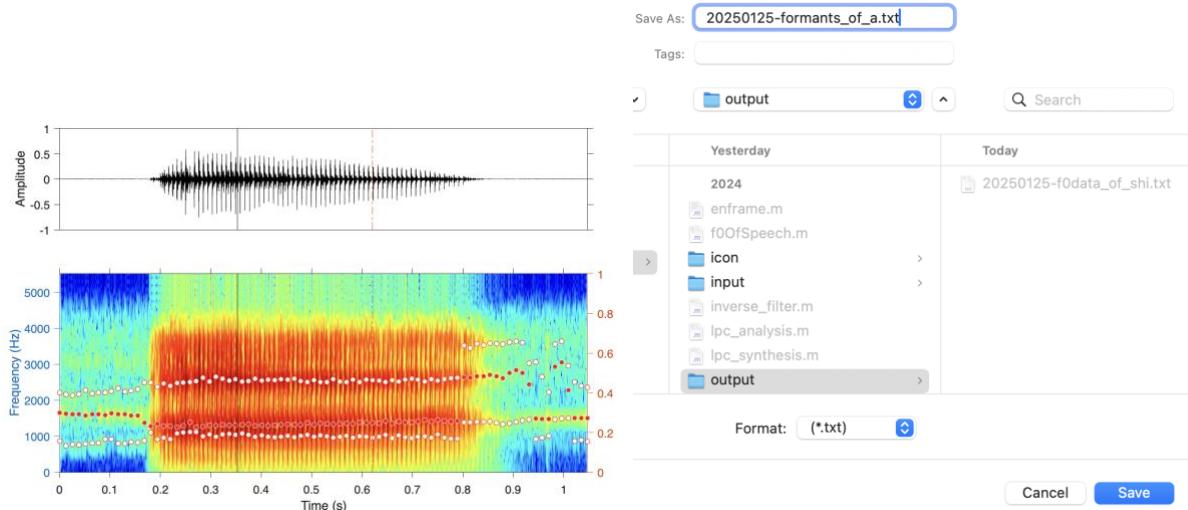
对语音信号进行 Band-pass Filter, 2~3 次 Pre-emphasis Filter, 以及振幅归一化之后, 效果明显提升:



<sup>2</sup> 如果您读入双声道信号, 程序将要求您选择所需的声音。

**Extract Formants (\*.txt):** 将光标内音段的基频保存为.txt 文件（光标内 11 个等距点的数据，含时间点和对应该时间点的基频值）。

先选取音段：



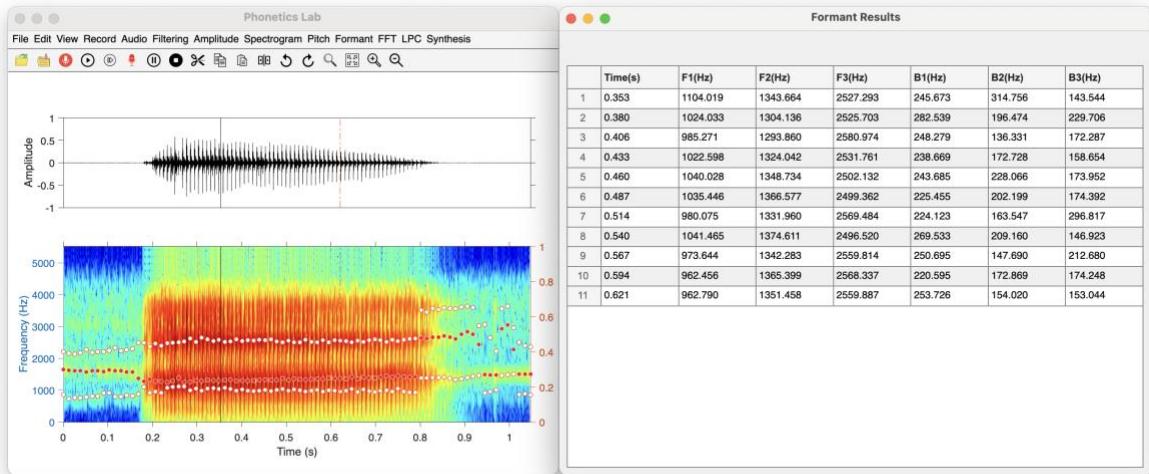
输入文件名、选择工作路径：



您的数据文件格式如下，将数据导入 Excel、R 等程序进行分析即可！

Time (s)	F1 (Hz)	F2 (Hz)	F3 (Hz)	B1 (Hz)	B2 (Hz)	B3 (Hz)
0.352878	1104.019458	1343.663659	2527.293458	245.673245	314.755720	143.544248
0.379664	1024.032551	1304.136380	2525.703053	282.539095	196.474049	229.706133
0.406451	985.271327	1293.859753	2580.974303	248.278684	136.331455	172.286862
0.433237	1022.597761	1324.042286	2531.760575	238.669285	172.728472	158.653787
0.460023	1040.027811	1348.733962	2502.132473	243.684834	228.065532	173.951853
0.486810	1035.445919	1366.576591	2499.362262	225.454852	202.198942	174.391981
0.513596	980.075097	1331.960432	2569.483628	224.122572	163.546696	296.816679
0.540383	1041.465350	1374.611297	2496.520214	269.533412	209.159591	146.922547
0.567169	973.643693	1342.283362	2559.814418	250.695128	147.689726	212.679839
0.593955	962.455612	1365.398763	2568.337011	220.594988	172.869338	174.247775
0.620742	962.789631	1351.457897	2559.886700	253.725502	154.020103	153.043608

**View Data:** 查看基频值（光标中间（Mid Point 选项）或光标内 11 个等距点的数据（11 Equidistant Points 选项），含时间点和对应该时间点的前三个共振峰及其带宽）。Formant Results 窗口的数据可以复制粘贴哦！



## ～声学元音图～

**Diagram Setting** (元音图设置): 设置元音图的 x、y 轴范围，  
默认值如右图所示。

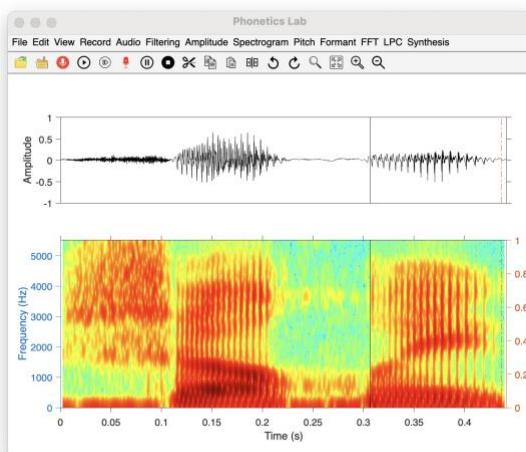
**Vowel Diagram:** 制作光标内音段的元音声学空间。如果您希望将其他音段的数据叠加在已经制作的元音图上，请勿关闭元音图的窗口。

第一步：选取所需的音段；

第二部：点击 Vowel Diagram;

第三部：输入元音符号，点击 OK；

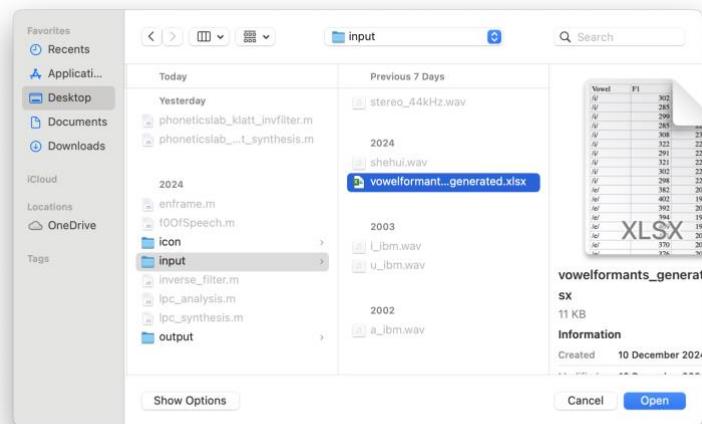
第四部：制作其他元音的声学空间（可打开新的语音文件），作图步骤如上。制作结果如下图所示：



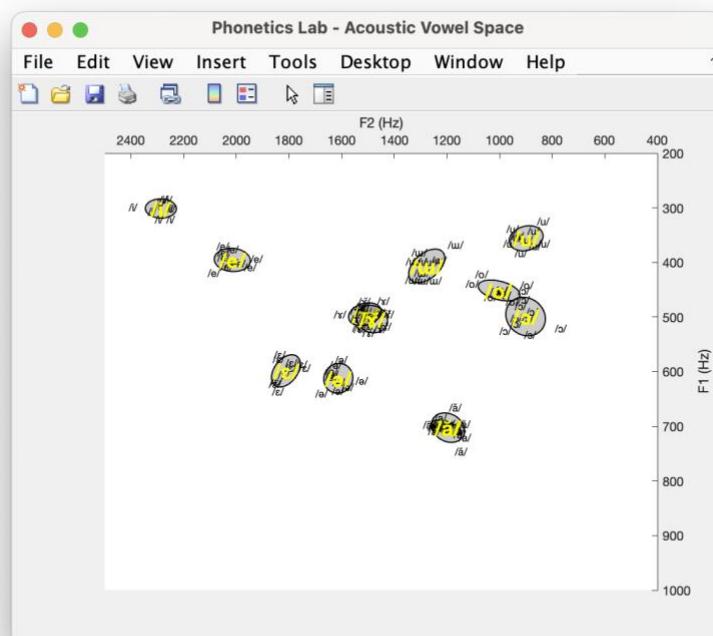
**Insert Data for Plotting:** 读入数据，制作声学元音图。要求 excel 文件由 Vowel、F1、F2 三列组成。请务必确认 excel 文件里的列名是否准确！

第一步：Insert Data for Plotting，读入数据文件。

例如 input 文件夹里的 vowelformants\_generated.xlsx 文件

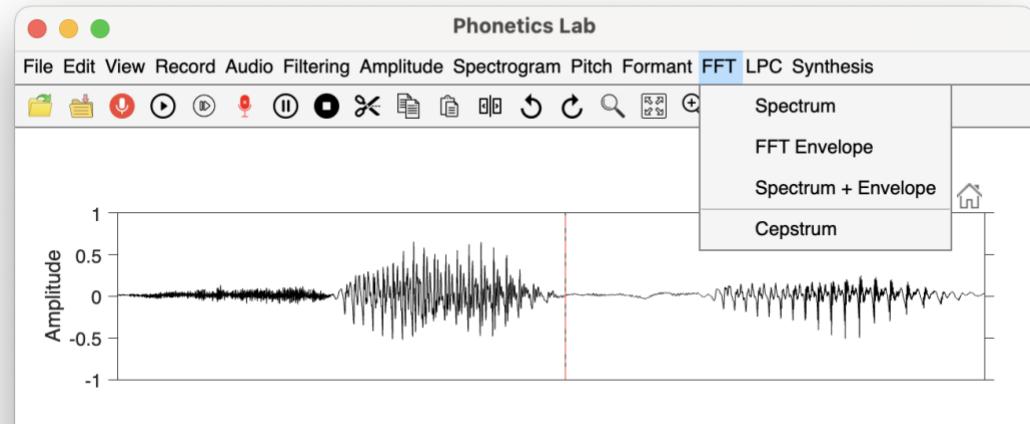


第二部：没有第二部了，程序直接弹出元音图窗口！！

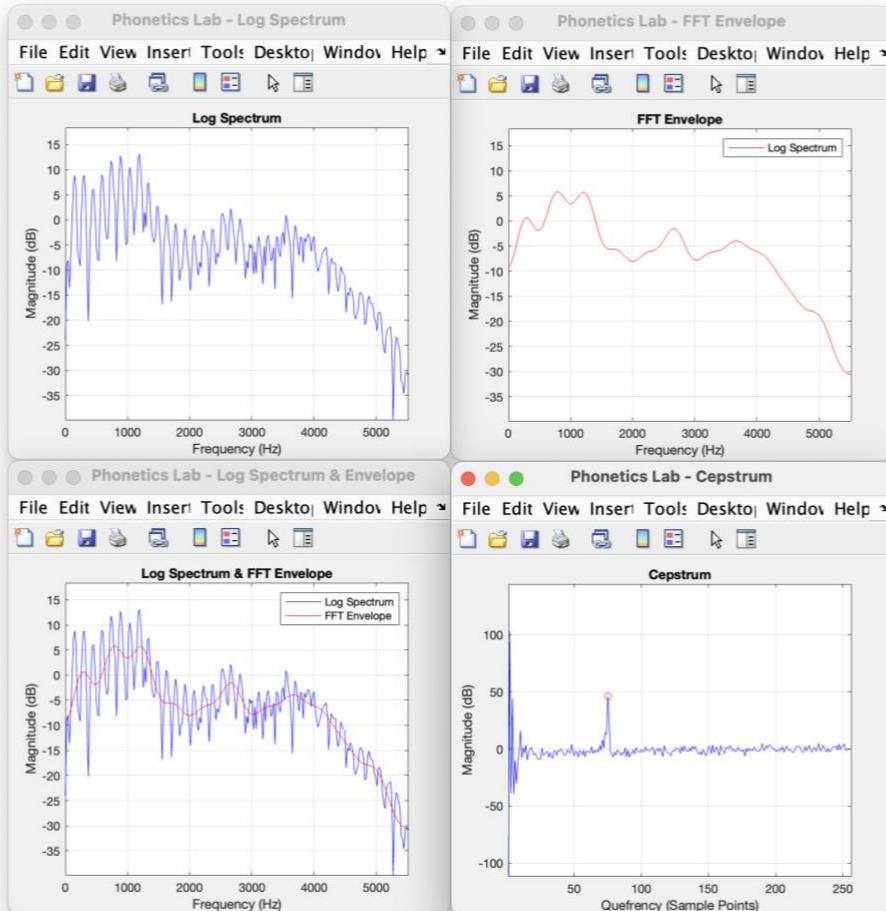


## 十、 FFT 快速傅立叶变换

FFT 菜单功能：

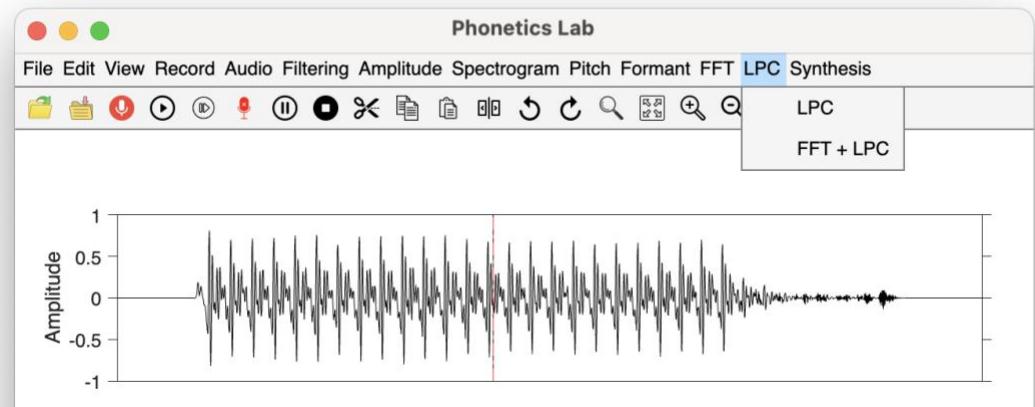


**Spectrum**、**FFT Envelope**、**Spectrum + Envelope**、**Cepstrum** 分别是频谱图(Log)、FFT 谱包络、频谱图+谱包络、倒谱。程序对光标所在位置（某一点）或光标内（某一段音频）进行傅立叶变换，并制作相应的图形。

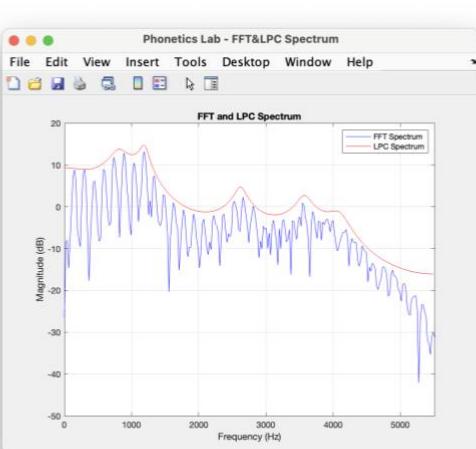
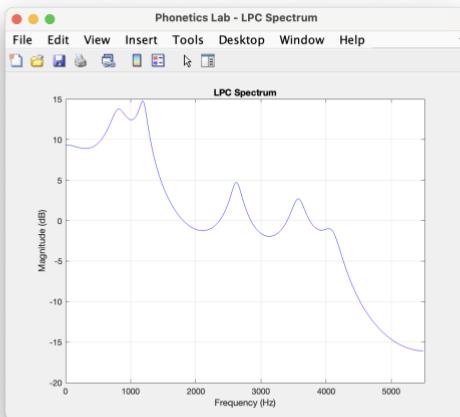


## 十一、LPC 分析

LPC 菜单功能：

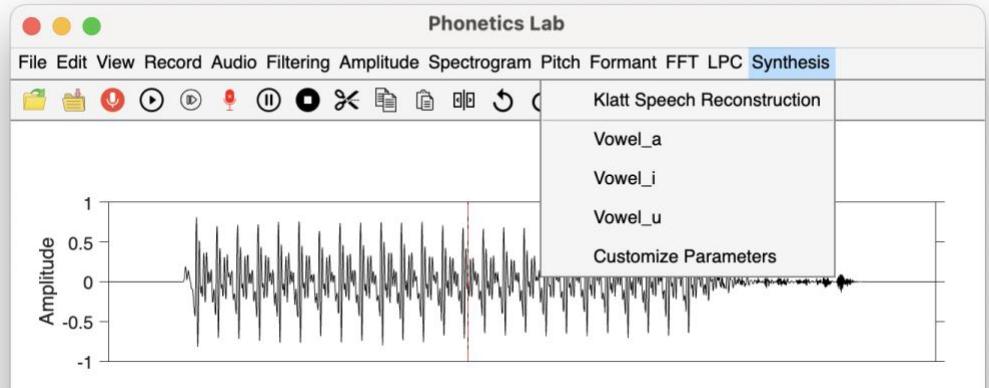


**LPC**、**FFT + LPC** 分别是频谱图 LPC 谱与 FFT&LPC 叠加谱。程序对光标所在位置（某一点）或光标内（某一段音频）进行 LPC 分析，并制作相应的图形。



## 十二、语音合成

Synthesis 菜单功能：



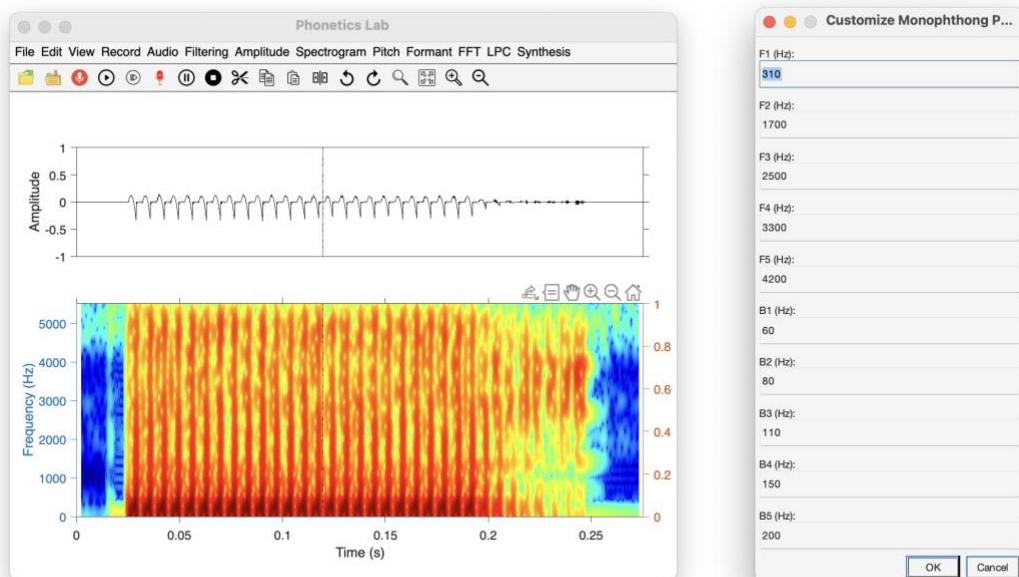
**Vowel\_a:** 元音 a 共振峰与共振峰带宽参数；

**Vowel\_i:** 元音 i 共振峰与共振峰带宽参数；

**Vowel\_u:** 元音 u 共振峰与共振峰带宽参数；

**Customize Parameters:** 自定义参数（窗口如下右图所示，默认值为元音/y/）。

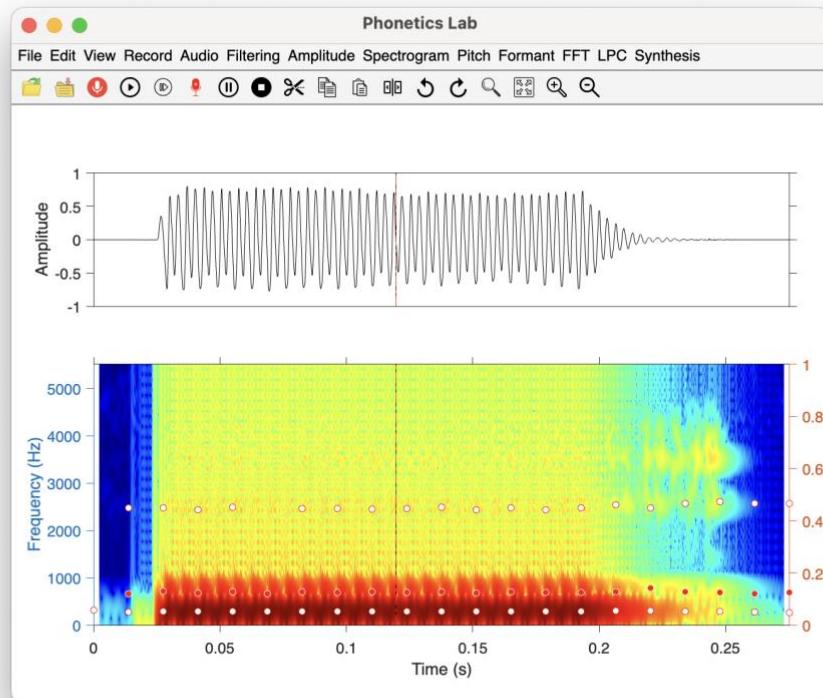
Sound: a (inverse filtered)



第一步：对语音信号进行逆滤波处理，提取 Glottal Airflow 信号 (Filtering 菜单)，如上左图所示。

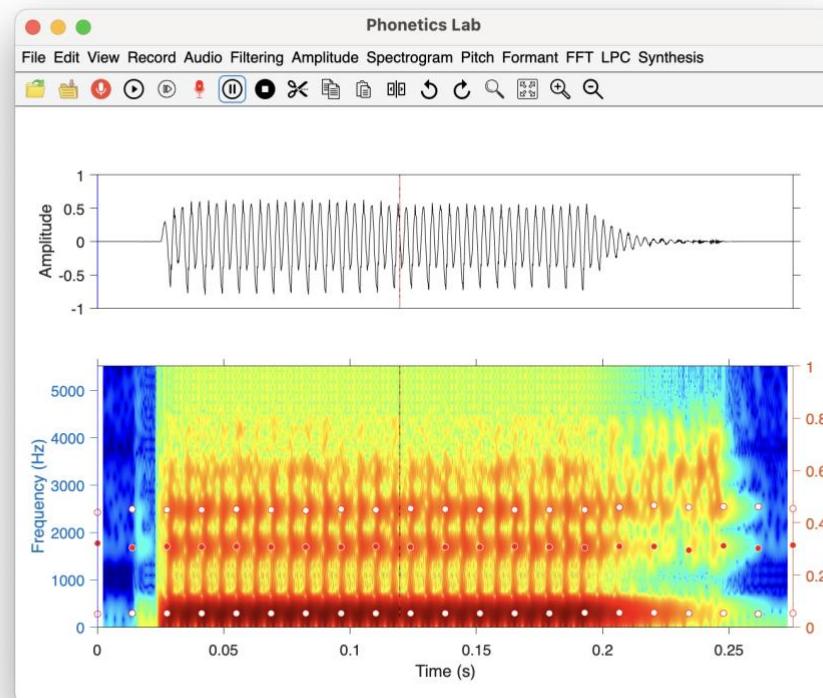
第二部：点击需要合成的元音 (Vowel\_a、Vowel\_i 或 Vowel\_u)。

第三部：点击 Synthesis 菜单的 Klatt Speech Reconstruction，结果如下图所示（合成出来的元音 u）：



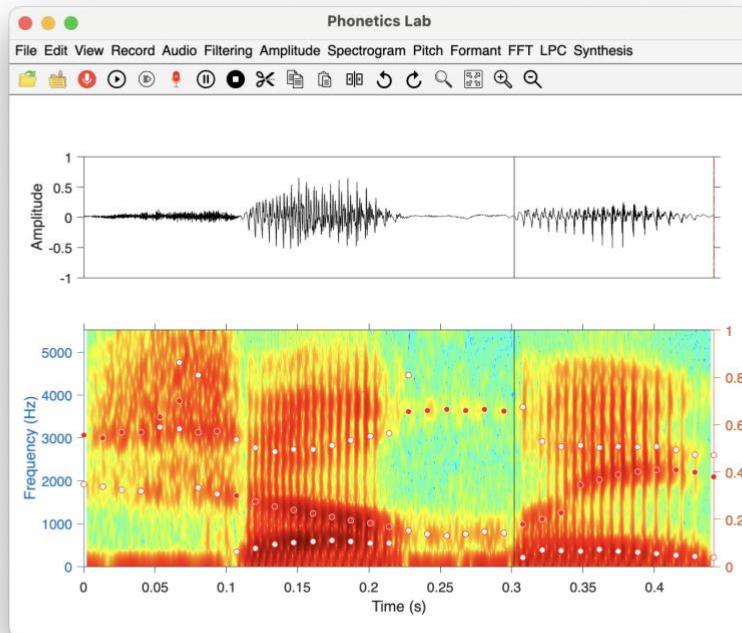
## 自定义参数

第一步：对语音信号进行逆滤波处理，提取 Glottal Airflow 信号（Filtering 菜单）。  
 第二步：点击自定义参数 Customize Parameters，自行调整参数（默认值为元音/y/），  
 调整好后点击 OK 即可。下图为合成出来的元音/y/。

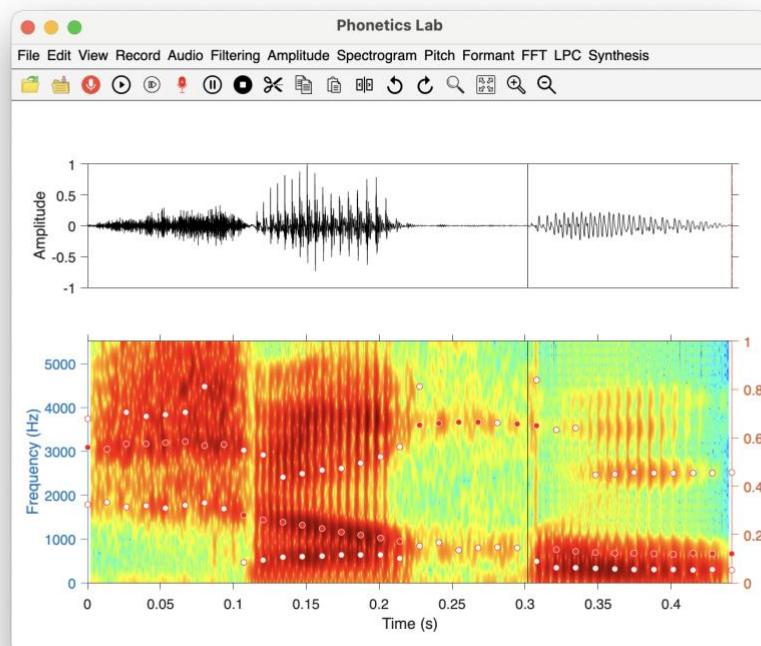


本程序支持仅合成光标内的音段：

Sound: 社会 shèhuì



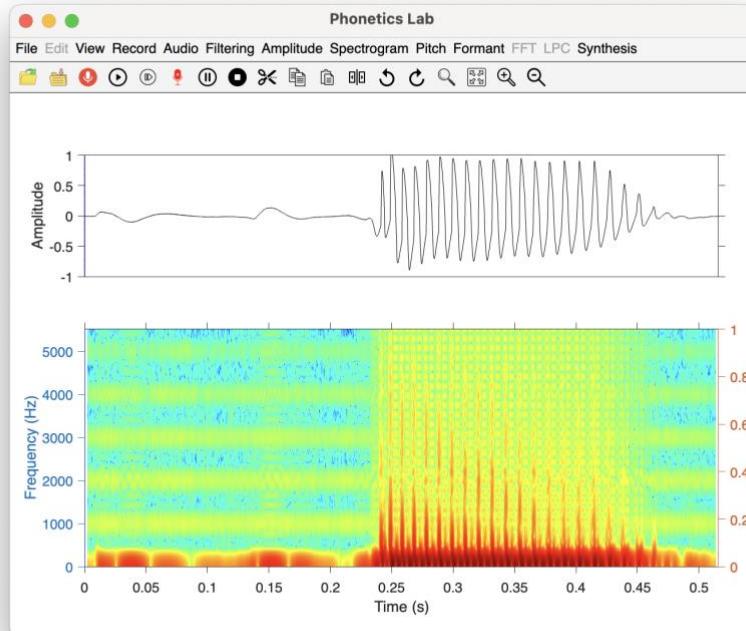
Resynthesized sound: shè hù



EGG 信号能很好地合成元音，下面以 input 文件夹里的 EGG\_channel2.wav 为例。

第一步：读入 input 里的 EGG\_channel2.wav；

第二部：选择 EGG 信号所在的声音(Right Channel)，重新采样 11kHz(optional)；  
Sound: 越南语 C1 调，EGG 信号



第三部：选择或自行输入元音参数，合成元音。

Sound: 由 EGG 信号合成的元音/a/，越南语 C1 调

