Tutorial on VoiceSauce - A program for voice analysis

Yuan Chai
University of California Los Angeles
yuanchai@g.ucla.edu

02/10/2023

Goal of the workshop

- Have a basic understanding of the rationale and usage of VoiceSauce;
- Get hands-on experience of using VoiceSauce to process audio files;
- Visualize and interpret data in Excel
- (Try some visualization using R code!)

What is VoiceSauce

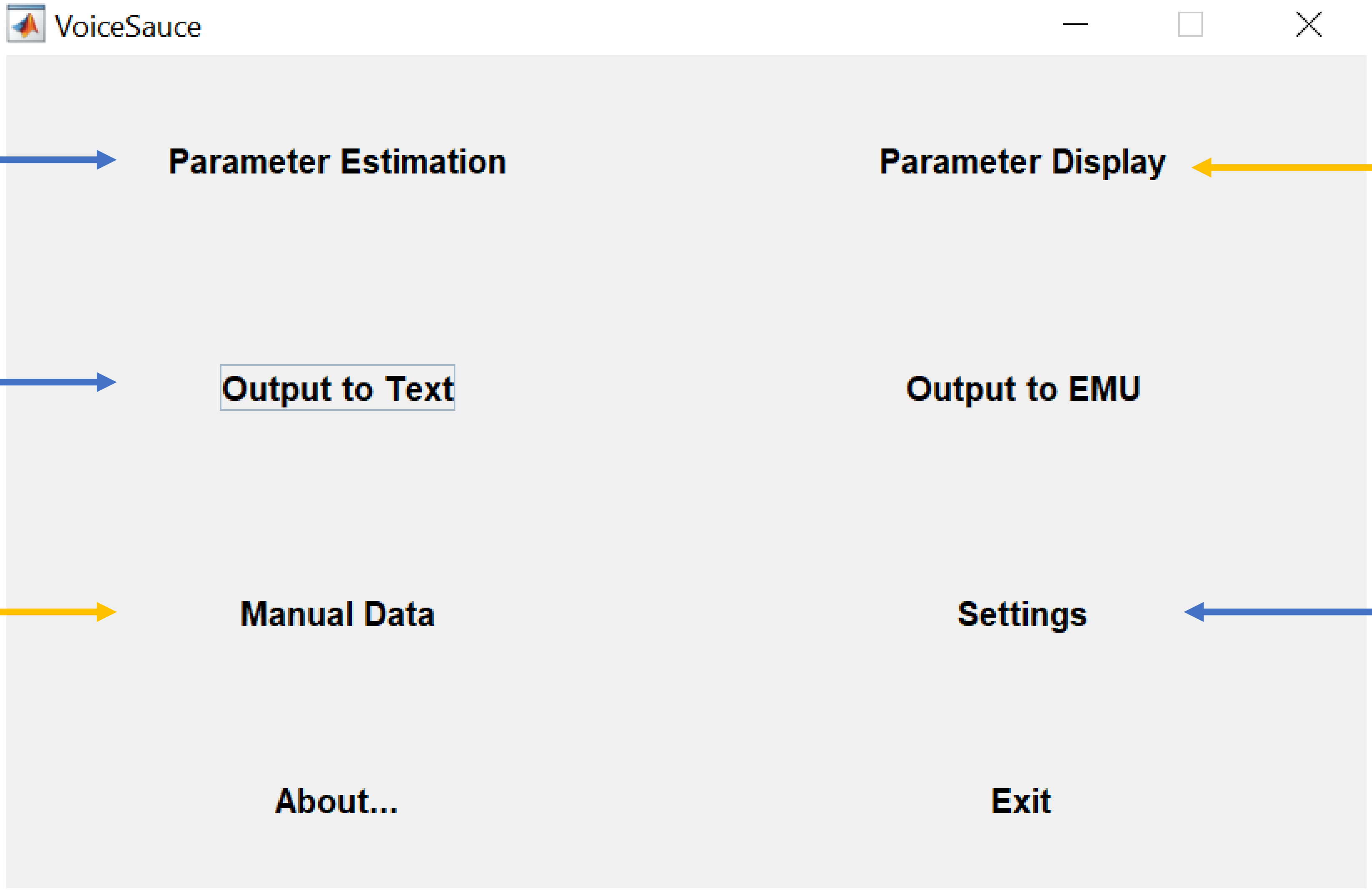
- VoiceSauce is a software that can be used to analyze acoustic measurements related to **voicing**.
- Compared to Praat, VoiceSauce specializes in measuring parameters related to **voice quality**:
  - **Spectral tilt** (H1-H2, H2-H4), **noise** (HNR), **voicing amplitude** (SoE). Those measures indicate whether there is glottal constriction or F0 irregularity in the voicing.
- VoiceSauce can also calculate measures that Praat can calculate:
  - Pitch (F0), vowel formant, duration, intensity (RMS Energy)

What does VoiceSauce look like

Step 2

Step 3

If you spot errors in the output and want to manually add data



If you want to visualize the data in VoiceSauce

Step 1

Output from VoiceSauce

- VoiceSauce output one datapoint every 1 millisecond.
- VoiceSauce can also calculate mean
  - Either the overall mean,
  - Or you can specify how many intervals you want to divide a sound file into, and calculate the mean of each interval.

All data points:

Filename	Label	seg_Start	seg_End	t_ms	H1c	H2c	H4c	A1c	A2c
Gai.mat	1-a-short-	154.993	241.778	155	NaN	NaN	NaN	NaN	NaN
Gai.mat	1-a-short-	154.993	241.778	156	NaN	NaN	NaN	NaN	NaN
Gai.mat	1-a-short-	154.993	241.778	157	11.07	5.879	-7.695	-26.586	-25.205
Gai.mat	1-a-short-	154.993	241.778	158	12.481	7.54	-6.098	-23.711	-22.406
Gai.mat	1-a-short-	154.993	241.778	159	13.906	9.01	-4.08	-21.1	-19.914
Gai.mat	1-a-short-	154.993	241.778	160	15.442	10.657	-2.1	-18.369	-17.362
Gai.mat	1-a-short-	154.993	241.778	161	17.096	12.307	0.13	-15.384	-14.507
Gai.mat	1-a-short-	154.993	241.778	162	18.805	14.063	2.319	-12.217	-11.768
Gai.mat	1-a-short-	154.993	241.778	163	20.362	15.38	3.559	-10.39	-10.283
Gai.mat	1-a-short-	154.993	241.778	164	21.752	16.67	4.531	-8.742	-9.111
Gai.mat	1-a-short-	154.993	241.778	165	22.642	17.403	5.192	-7.341	-8.114
Gai.mat	1-a-short-	154.993	241.778	166	23.054	17.767	5.711	-6.216	-7.292
Gai.mat	1-a-short-	154.993	241.778	167	23.415	17.92	6.016	-5.299	-6.53
Gai.mat	1-a-short-	154.993	241.778	168	23.648	18.216	6.227	-4.607	-6.111
Gai.mat	1-a-short-	154.993	241.778	169	23.86	18.556	6.455	-4.022	-5.6
Gai.mat	1-a-short-	154.993	241.778	170	24.064	18.842	6.813	-3.511	-5.181
Gai.mat	1-a-short-	154.993	241.778	171	24.169	19.034	7.078	-2.989	-4.792
Gai.mat	1-a-short-	154.993	241.778	172	24.248	19.223	7.149	-2.445	-4.473
Gai.mat	1-a-short-	154.993	241.778	173	24.349	19.363	7.138	-1.952	-4.186
Gai.mat	1-a-short-	154.993	241.778	174	24.385	19.488	7.194	-1.543	-4.028

Just the mean

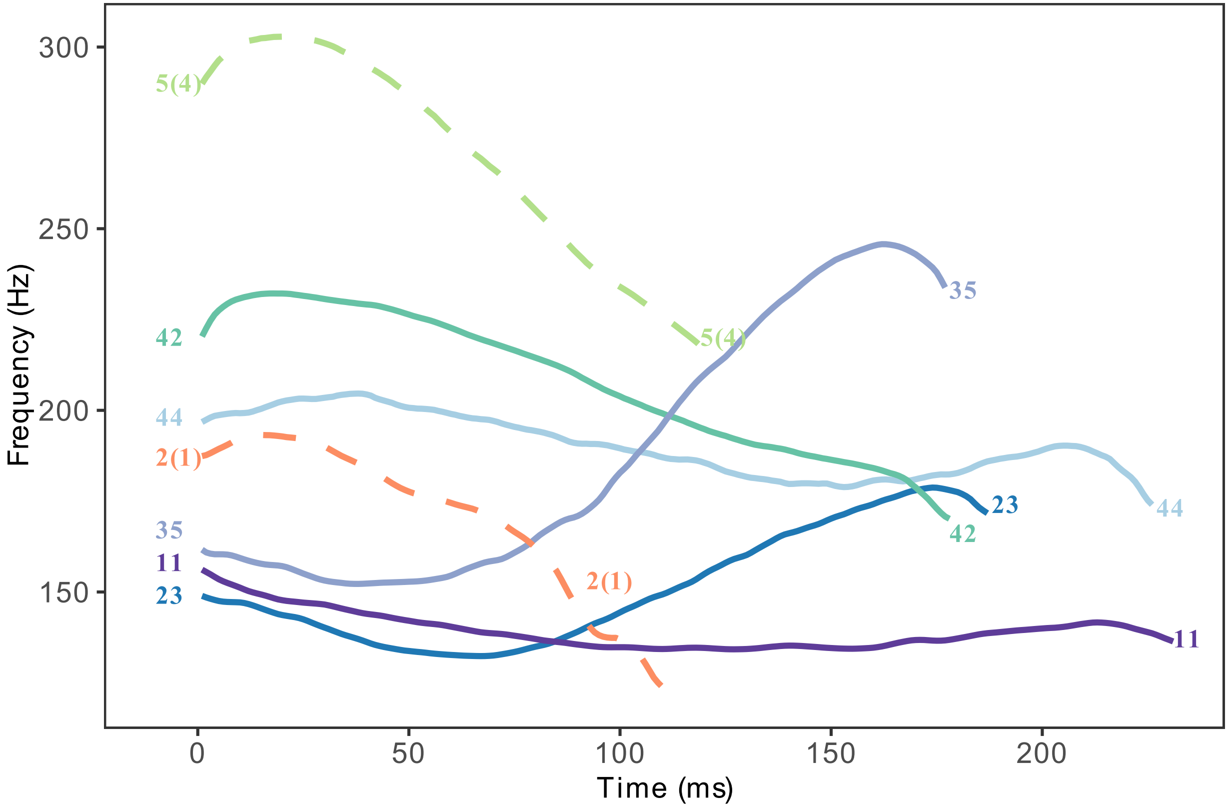
Filename	positi	vowel	length	phonation	seg_Start	seg_End	H1c_mean	H1H2c_mean	CPP_mean	Energy_mean
aGa.mat	1	a	short	glottal	127.354	215.402	15.279	1.946	17.511	3
aGa.mat	2	a	short	glottal	242.999	353.387	8.207	-3.91	16.243	0.973
aka.mat	1	a	short	modal	327.73	381.61	9.135	-0.786	16.362	0.477
aka.mat	2	a	short	modal	514.339	637.869	7.967	-2.142	16.766	0.365
kaGa.mat	1	a	short	glottal	110.185	174.578	16.543	8.428	17.414	1.696
kaGa.mat	2	a	short	glottal	235.029	323.077	11.401	-0.247	16.171	1.09
koGu.mat	1	o	short	glottal	167.14	223.641	15.259	8.602	16.71	8.866
koGu.mat	2	u	short	glottal	284.793	396.801	17.068	3.731	15.888	3.101
kou.mat	1	o	short	modal	258.052	498.542	13.007	-2.459	18.71	20.469
kou.mat	2	u	short	modal	498.542	660.182	10.456	-1.433	15.491	1.007
kouL.mat	1	o	short	modal	183.771	371.84	15.715	0.582	17.661	10.577
kouL.mat	2	u	long	modal	371.84	664.019	17.66	-0.226	17.621	5.994
noGu.mat	1	o	short	glottal	203.659	267.468	23.627	14.075	18.396	6.587
noGu.mat	2	u	short	glottal	334.636	465.612	15.828	7.704	15.669	1.454
noLu.mat	1	o	long	modal	206.389	444.834	19.19	4.425	19.165	7.607
noLu.mat	2	u	short	modal	444.834	595.961	11.085	-4.378	15.915	1.871
nou.mat	1	o	short	modal	177.954	374.419	10.721	0.382	19.414	2.846
nou.mat	2	u	short	modal	374.419	471.812	5.618	2.438	15.364	0.199

Means of three equal-timed intervals for each file

Filename	Label	seg_Start	seg_End	H1c_mean	H1c_means001	H1c_means002	H1c_means003
Gai.mat	1-a-short-	154.993	241.778	19.524	21.76	21.788	15.265
Gai.mat	2-i-short-g	241.778	447.722	14.792	14.646	17.97	11.827
aGa.mat	1-a-short-	127.354	215.402	15.279	14.356	19.158	12.472
aGa.mat	2-a-short-	242.999	353.387	8.207	4.842	13.055	6.819
ai.mat	1-a-short-	109.075	316.273	19.063	12.878	21.267	23.019
ai.mat	2-i-short-r	316.273	513.833	18.351	21.449	18.616	14.915
aka.mat	1-a-short-	327.73	381.61	9.135	10	10.052	7.632
aka.mat	2-a-short-	514.339	637.869	7.967	9.353	8.235	6.361
kaGa.mat	1-a-short-	110.185	174.578	16.543	16.443	17.589	15.722
kaGa.mat	2-a-short-	235.029	323.077	11.401	9.54	8.756	15.049
koGu.mat	1-o-short-	167.14	223.641	15.259	17.046	15.767	13.061
koGu.mat	2-u-short-	284.793	396.801	17.068	13.588	17.119	20.063
kou.mat	1-o-short-	258.052	498.542	13.007	17.423	13.044	8.675
kou.mat	2-u-short-	498.542	660.182	10.456	10.635	13.418	7.448
kouL.mat	1-o-short-	183.771	371.84	15.715	16.504	17.35	13.4
kouL.mat	2-u-long-n	371.84	664.019	17.66	17.874	20.16	14.953

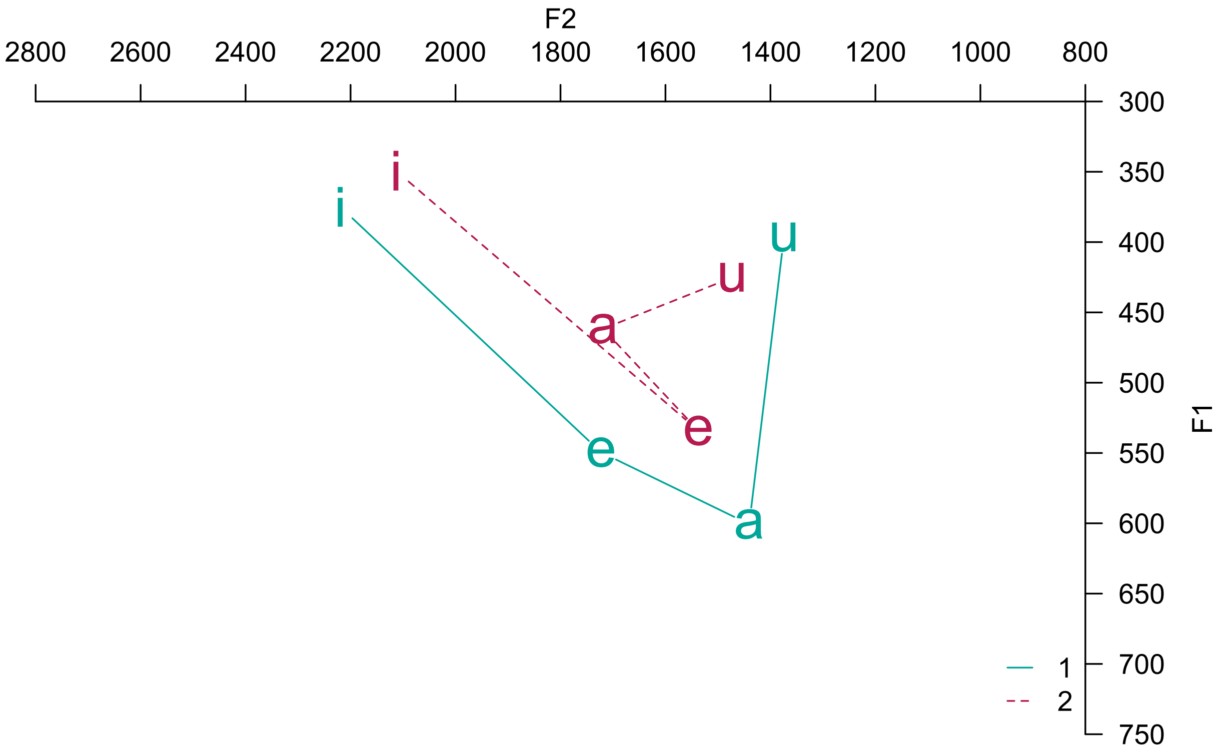
What can you draw/analyze using output from VoiceSauce

Pitch track (F0 track of the seven tones in Xiapu Min)



What can you draw/analyze using output from VoiceSauce

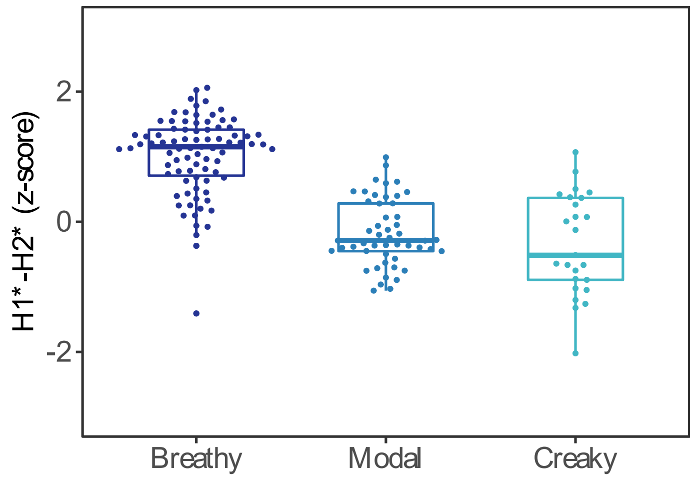
Vowel chart (stressed and unstressed vowels in Cahuilla)



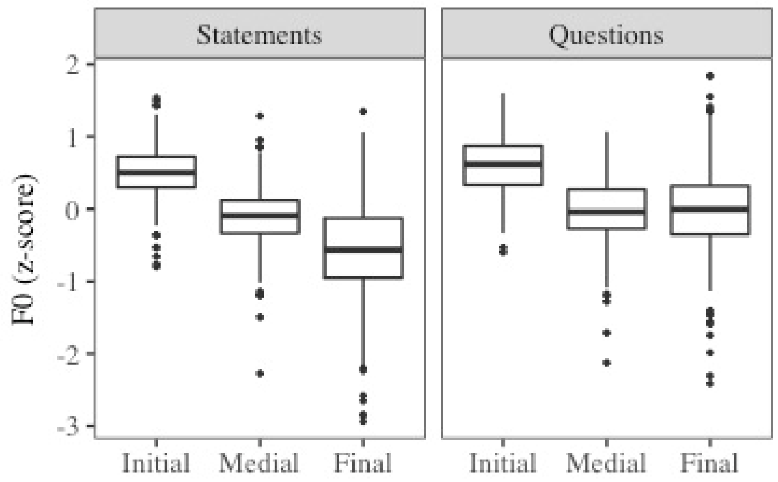
What can you draw/analyze using output from VoiceSauce

Boxplots of various measures

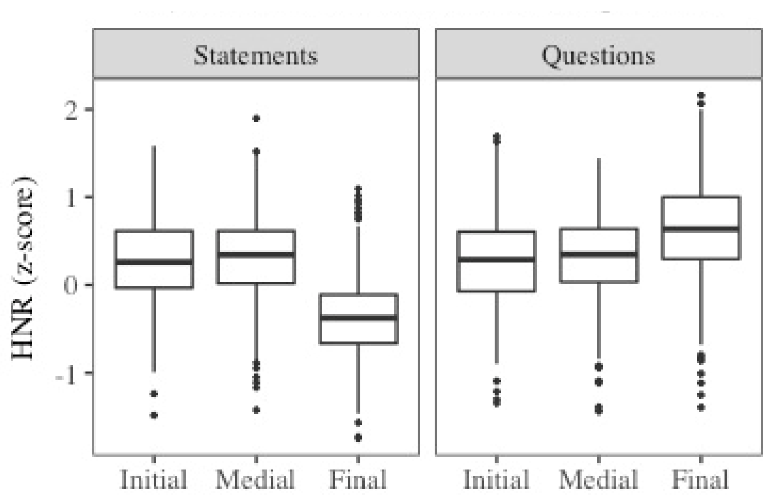
H1*—H2*



F0



Harmonic-to-noise ratio



Summary of the parameters

- F0: F0 from Straight (strF0), Snack (sF0), Praat (pF0), Subharmonic to harmonic ratio F0 (shrF0)
- Formant: Formant frequencies and bandwidths by Snack (sF1, sF2) and by Praat (pF1, pF2)
- Spectral measures: H1, H2, H1H2c, H2H4c
- Energy: Energy (overall); Strength of Excitation (SoE)
- Noise: Cepstral Peak Prominence (CPP); Harmonic to noise ratios: HNR05 (0-500Hz), HNR15 (0-1500Hz), HNR25 (0-2500Hz), HNR35 (0-3500Hz), Subharmonic to harmonic ratio: SHR

Sample research questions:

- Is the consonant pre-glottalized or post-glottalized?
- Are vowels following ejectives more glottalized than vowels following non-ejectives?
- Do implosive have stronger voicing than non-implosive?
- Does vowel quality differ between stressed and unstressed syllables?
- Do vowels after voiceless stops have a higher F0 than vowels after voiced stops?
- What is the F0 contour and shape of the tones in the language?
- Do vowels following aspirated stops have a breathy voice quality?

How to download and use VoiceSauce

- Windows users: Standalone .exe file
- Mac users: Install Matlab and run the scripts in Matlab
- Refer to https://yuanchaiyc.github.io/website/subpages/VS-tutorial.html for detailed installation instructions

Case study for today

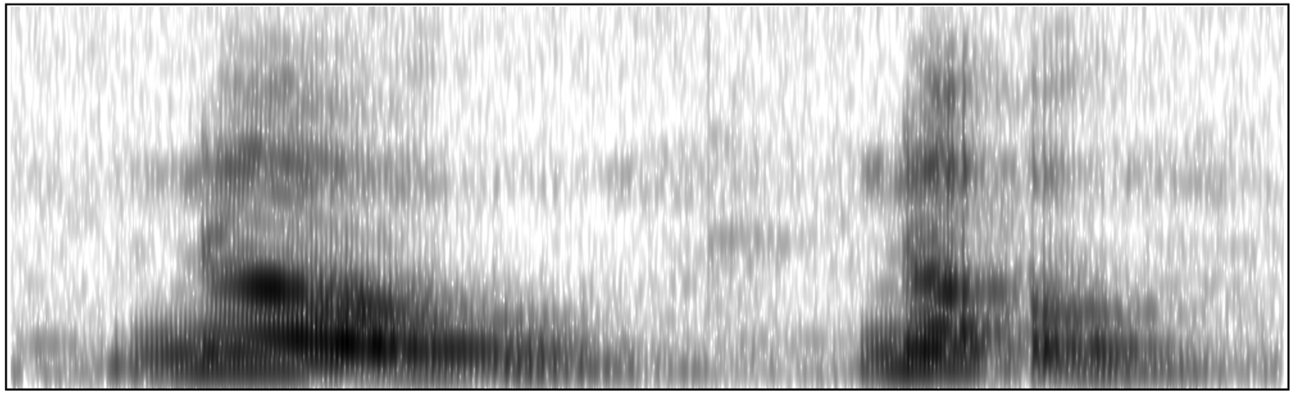
- The acoustics of **V** and **VʔV** in Hawaiian
- Hawaiian has phonemic glottal stop:

aha	"what"
ʔaha	"line, life"
noːu	"yours"
noʔu	"mine"

(Data and recordings from the UCLA Phonetics Lab Archive
http://archive.phonetics.ucla.edu/Language/HAW/haw_word-list_1973_01.html#1)

Case study for today

- Research question: Are the vowels surrounding the glottal stop creakier than the plain vowels?
  - noːu vs. noʔu



	oː	u		o		u	
noːu				noʔu			
yours				mine			

Case study for today

- Word list:

word	gloss	word	gloss	word	gloss
pe	"thus"	peːpeː	"crushed"	peʔe	"to hide (oneself)"
nou	"to throw"	noːu	"yours"	noʔu	"mine"
kou	"yours"	kouː	"moist"	koʔu	"mine"
aka	"shadows"			aʔa	"roots"
				kaʔa	"to roll"

Case study for today

- Parameters of interest:
  - F0
  - Harmonic amplitude: H1, H1—H2
    - The lower the harmonic amplitude, the more glottal constriction
  - Noise: Harmonic-to-noise ratio (HNR05, meaning between 0 to 500 Hz)
    - The lower the HNR, the noisier the signal
  - Amplitude of voicing: Strength of excitation (SoE)
    - Glottalization tends to results in lower amplitude in voicing (SoE)

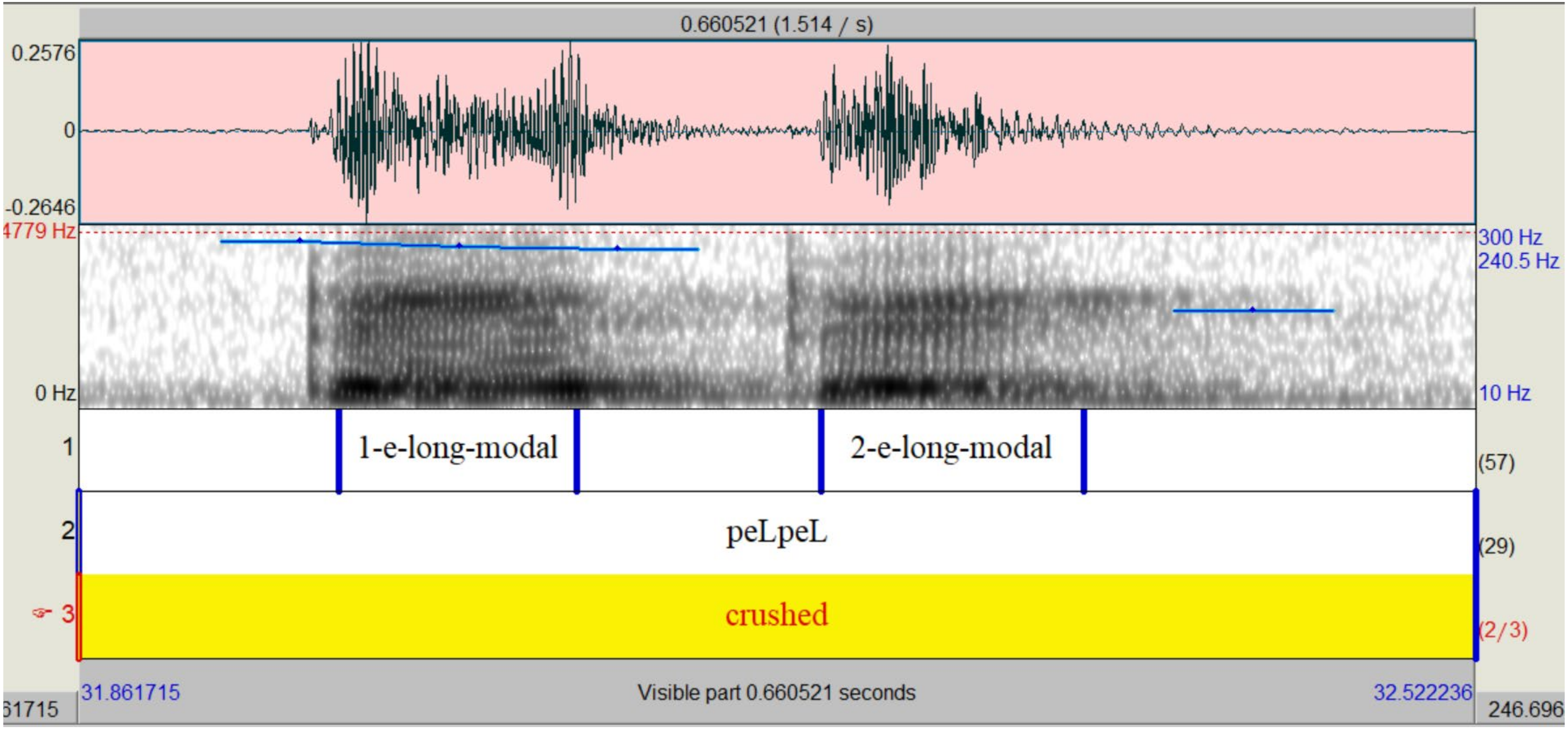
Getting started

- Prepare data in Praat
  - Create a Textgrid
  - Segment and annotate the target segment
  - Save the Textgrid
    - either as for the whole recording
    - or split the recording into individual target words – RECOMMENDED
    - You can use **Praat scripts** or **Praat plugins** to chop a long recordings into smaller chunks. Come talk to me if you want to know more about the tools!

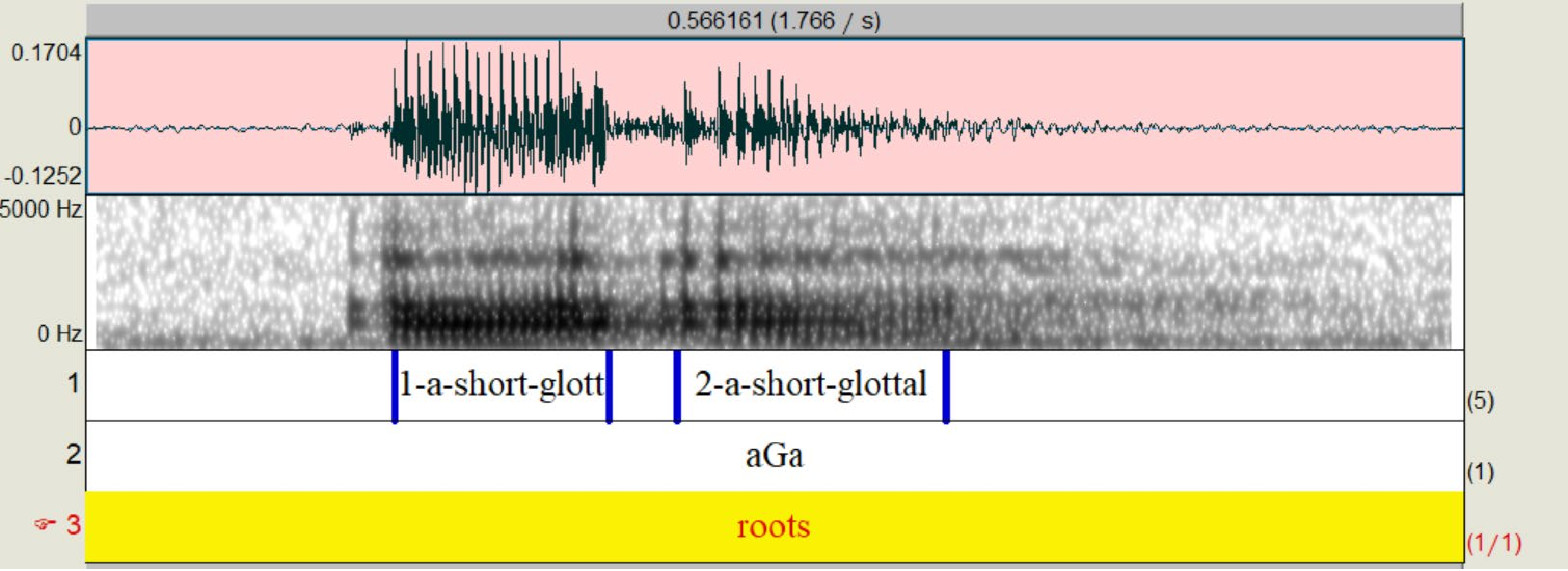
Getting started

- Annotation strategy for the current task
  - Segment out each vowel (monophthongs or nuclei and glide in dipthongs)
    - e.g. [peːpeː] → Segment out two [eː]s
    - e.g. [nou] → Segment out [o] and [u]
  - Assign label at the word level and the segment level
    - Word level: peLpeL (use "L" to replace diacritic [ː] because VS does not allow special symbols)
    - Segment level:
      - 1-e-long-modal
      - position-vowel-length-phonation

Getting started



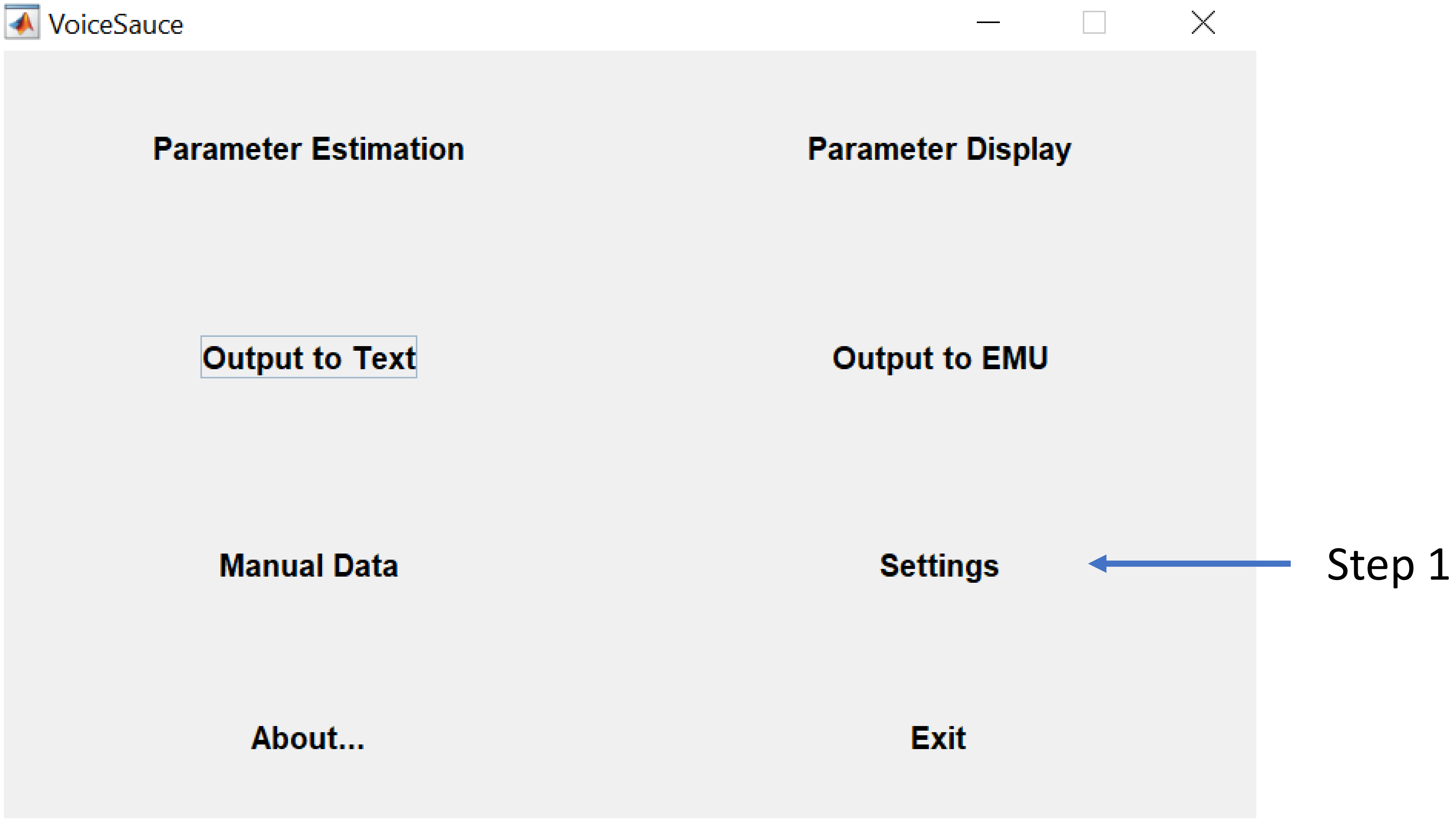
Getting started

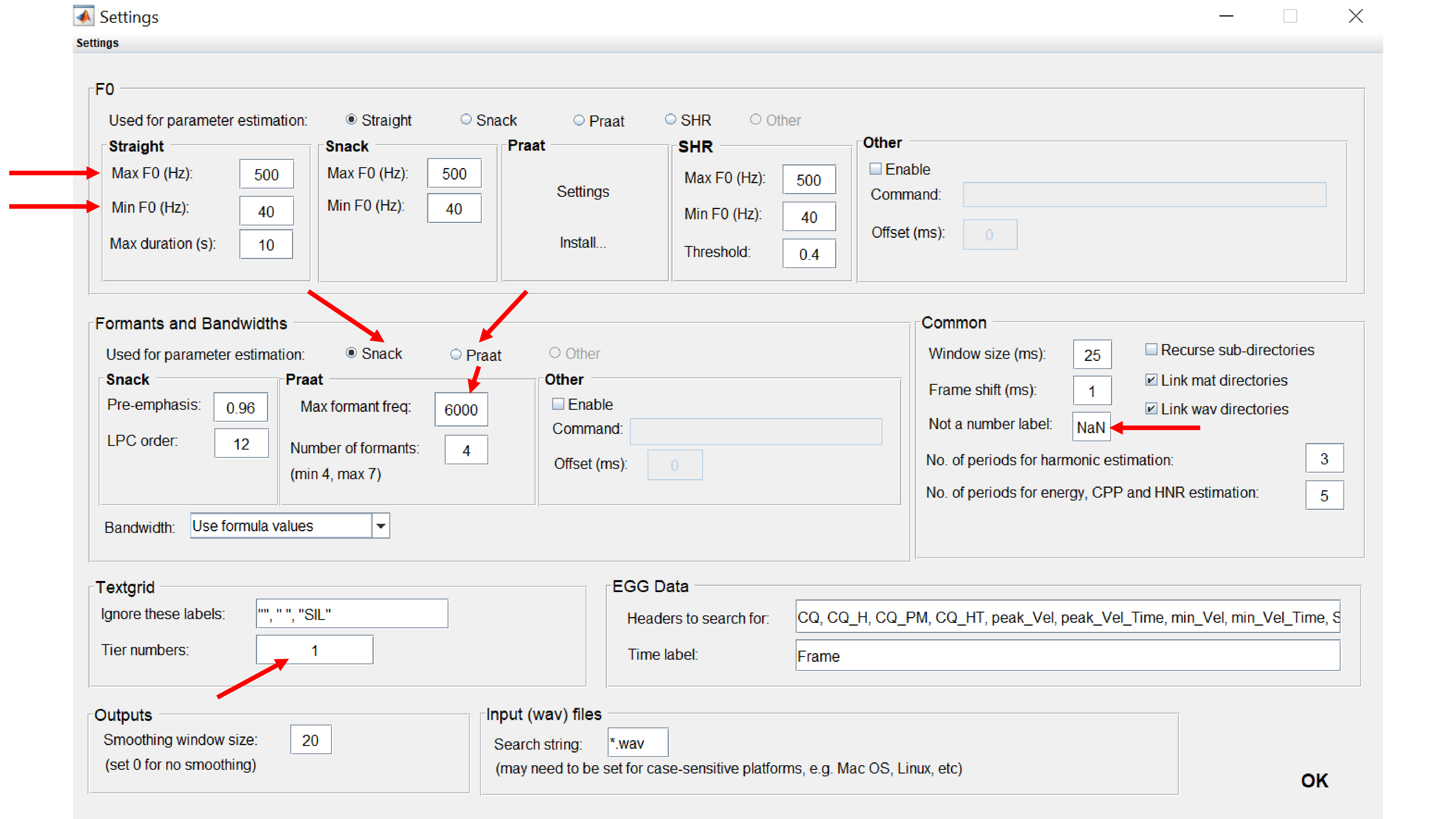


Getting started

- Download the preprocessed data here:
- https://yuanchaiyc.github.io/website/subpages/sample/Hawaiian_data.zip

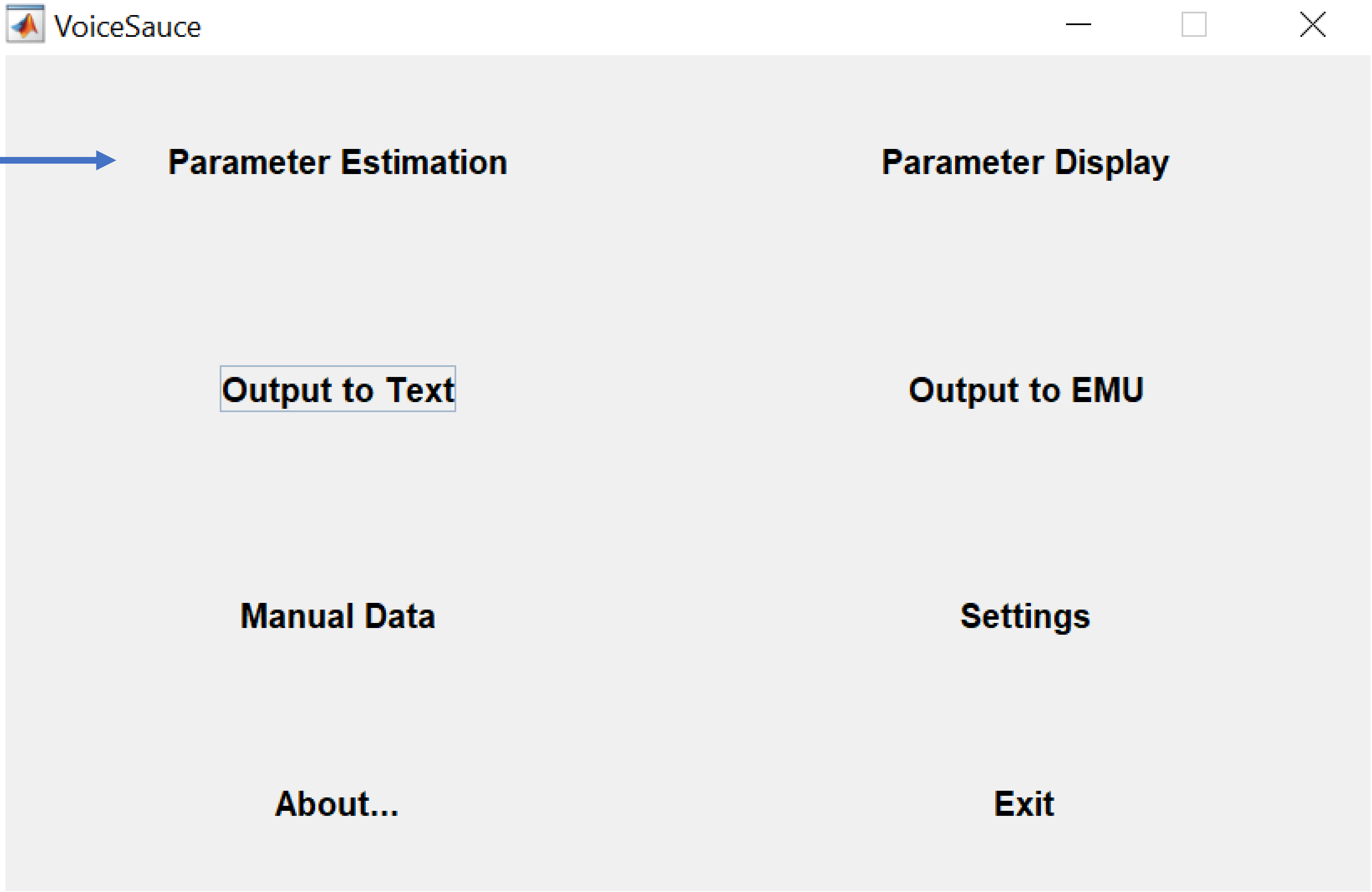
Pass on the .wav and .Textgrid to VoiceSauce

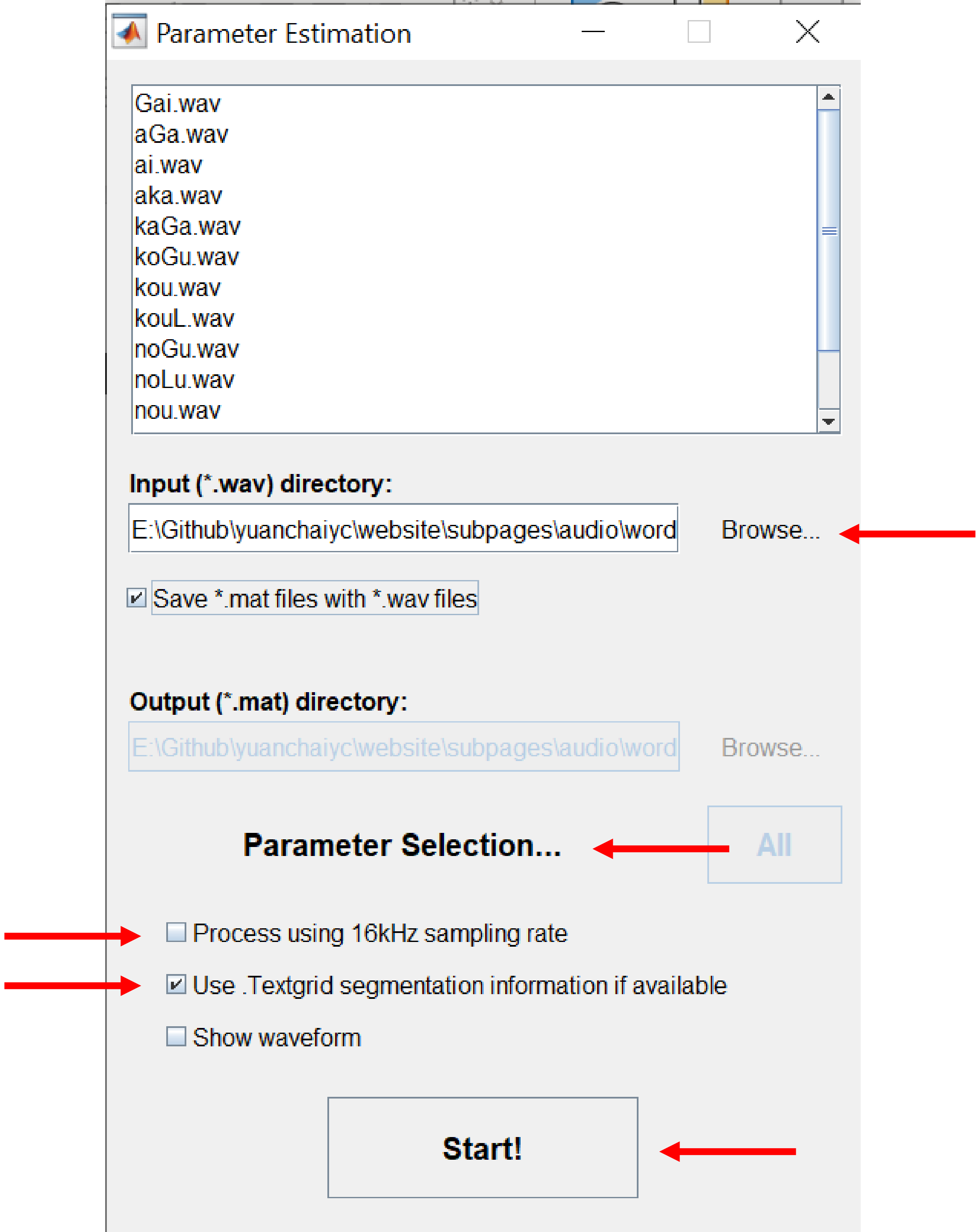



Settings
Settings
F0
Used for parameter estimation:
Straight
Snack
Praat
SHR
Other
Straight
Max F0 (Hz): 500
Min F0 (Hz): 40
Max duration (s): 10
Snack
Max F0 (Hz): 500
Min F0 (Hz): 40
Praat
Settings
Install...
SHR
Max F0 (Hz): 500
Min F0 (Hz): 40
Threshold: 0.4
Other
Enable
Command:
Offset (ms): 0
Formants and Bandwidths
Used for parameter estimation:
Snack
Praat
Other
Snack
Pre-emphasis: 0.96
LPC order: 12
Praat
Max formant freq: 6000
Number of formants: 4
(min 4, max 7)
Other
Enable
Command:
Offset (ms): 0
Bandwidth: Use formula values
Common
Window size (ms): 25
Frame shift (ms): 1
Not a number label: NaN
Recurse sub-directories
Link mat directories
Link wav directories
No. of periods for harmonic estimation: 3
No. of periods for energy, CPP and HNR estimation: 5
Textgrid
Ignore these labels: "", " ", "SIL"
Tier numbers: 1
EGG Data
Headers to search for: CQ, CQ_H, CQ_PM, CQ_HT, peak_Vel, peak_Vel_Time, min_Vel, min_Vel_Time, S
Time label: Frame
Outputs
Smoothing window size: 20
(set 0 for no smoothing)
Input (wav) files
Search string: *.wav
(may need to be set for case-sensitive platforms, e.g. Mac OS, Linux, etc)
OK

Step 2: Parameter estimation

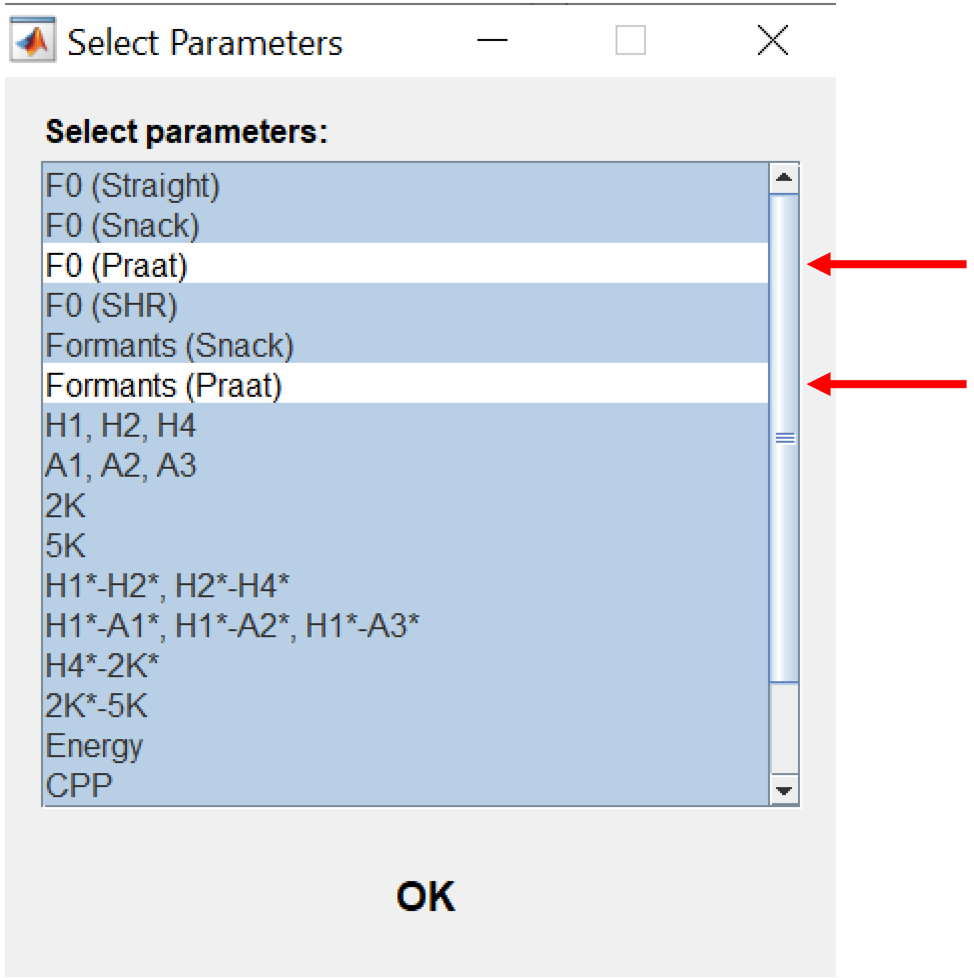
Step 2



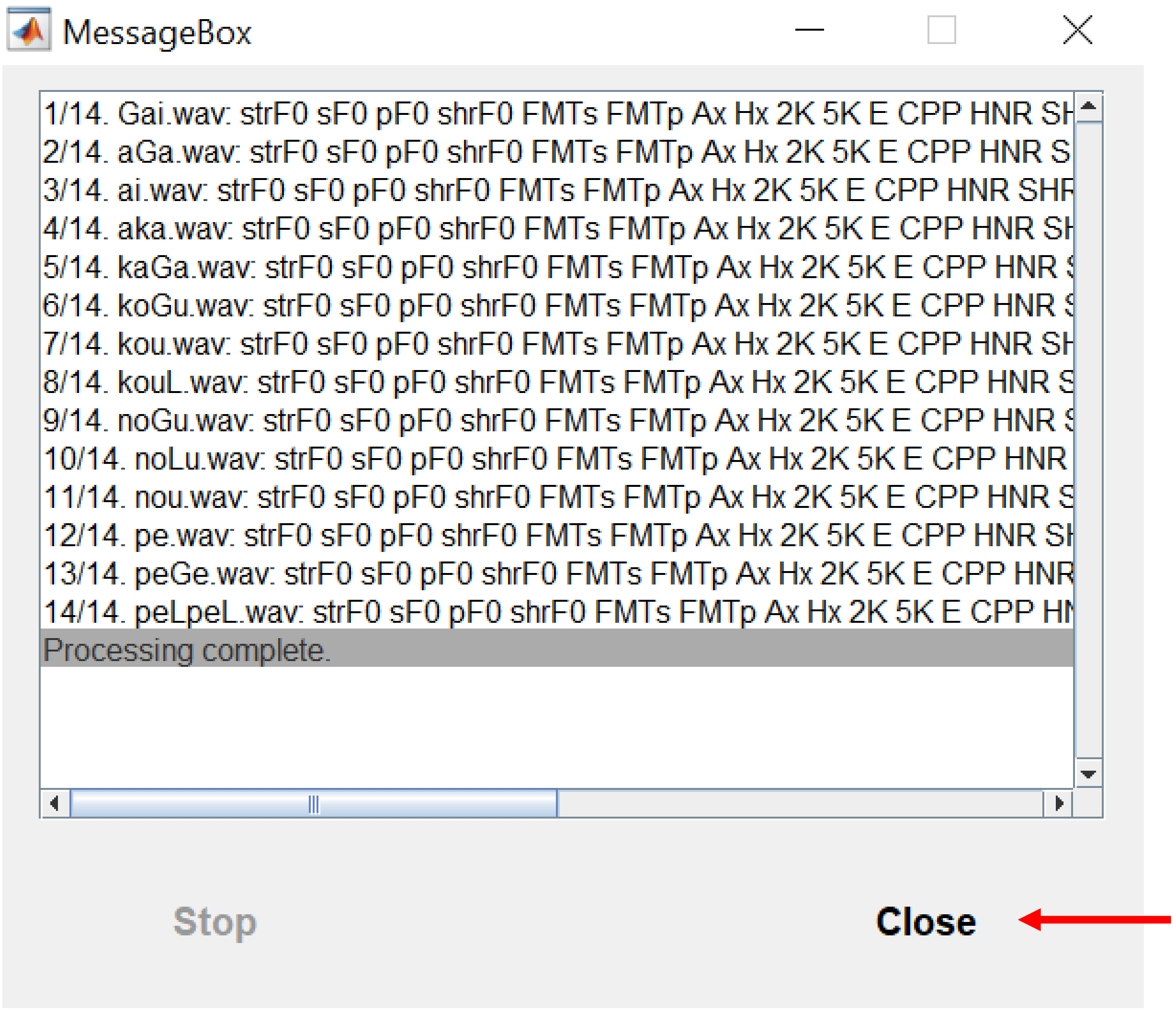

Parameter Estimation
Gai.wav
aGa.wav
ai.wav
aka.wav
kaGa.wav
koGu.wav
kou.wav
kouL.wav
noGu.wav
noLu.wav
nou.wav
Input (*.wav) directory:
E:\Github\yuanchaiyc\website\subpages\audio\word
Browse...
Save *.mat files with *.wav files
Output (*.mat) directory:
E:\Github\yuanchaiyc\website\subpages\audio\word
Browse...
Parameter Selection...
All
Process using 16kHz sampling rate
Use .Textgrid segmentation information if available
Show waveform
Start!

Step 2: Parameter estimation

If you are using Matlab **online**, make sure you **deselect** all the measures involving Praat.

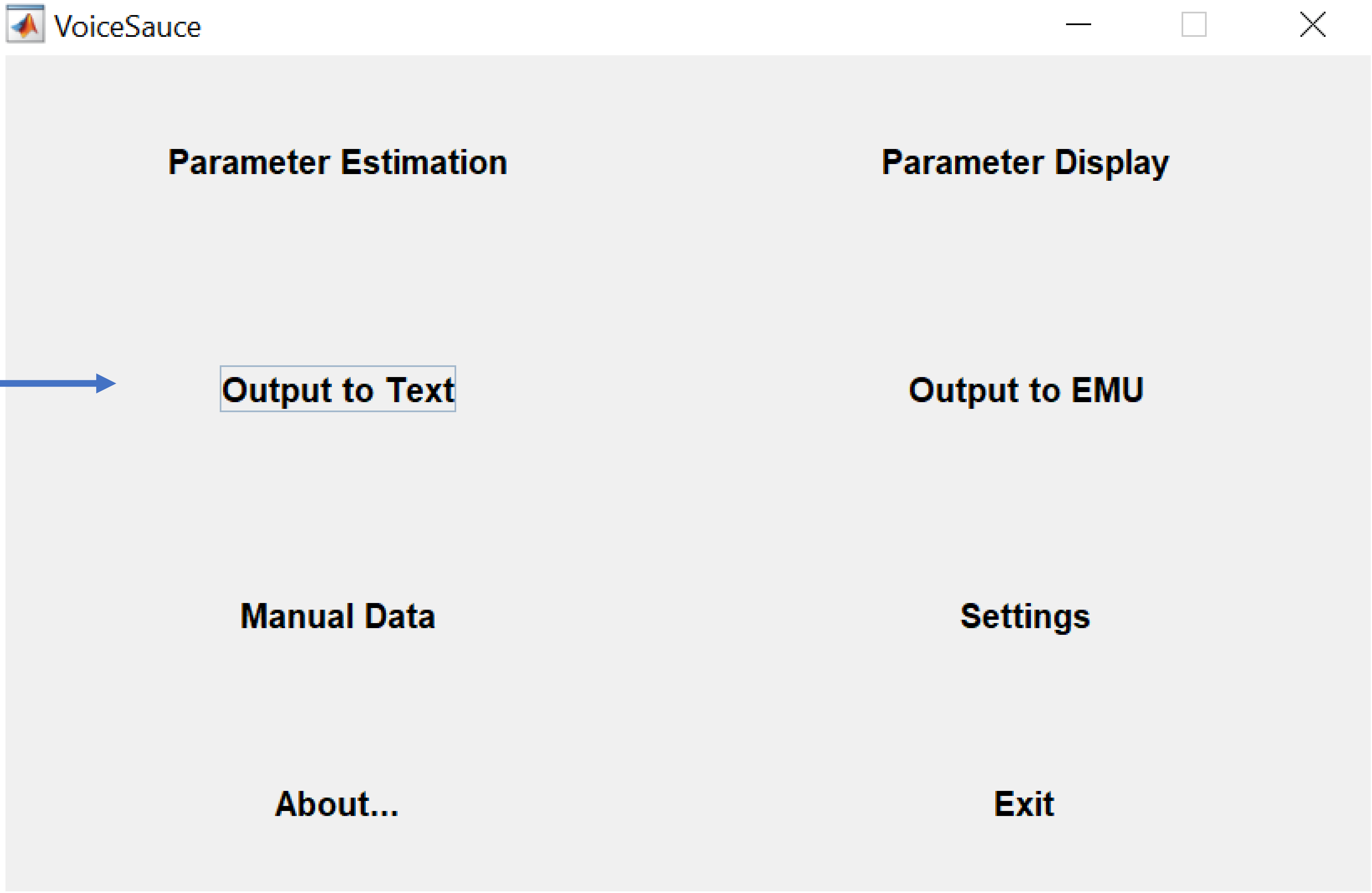


Step 2: Parameter estimation



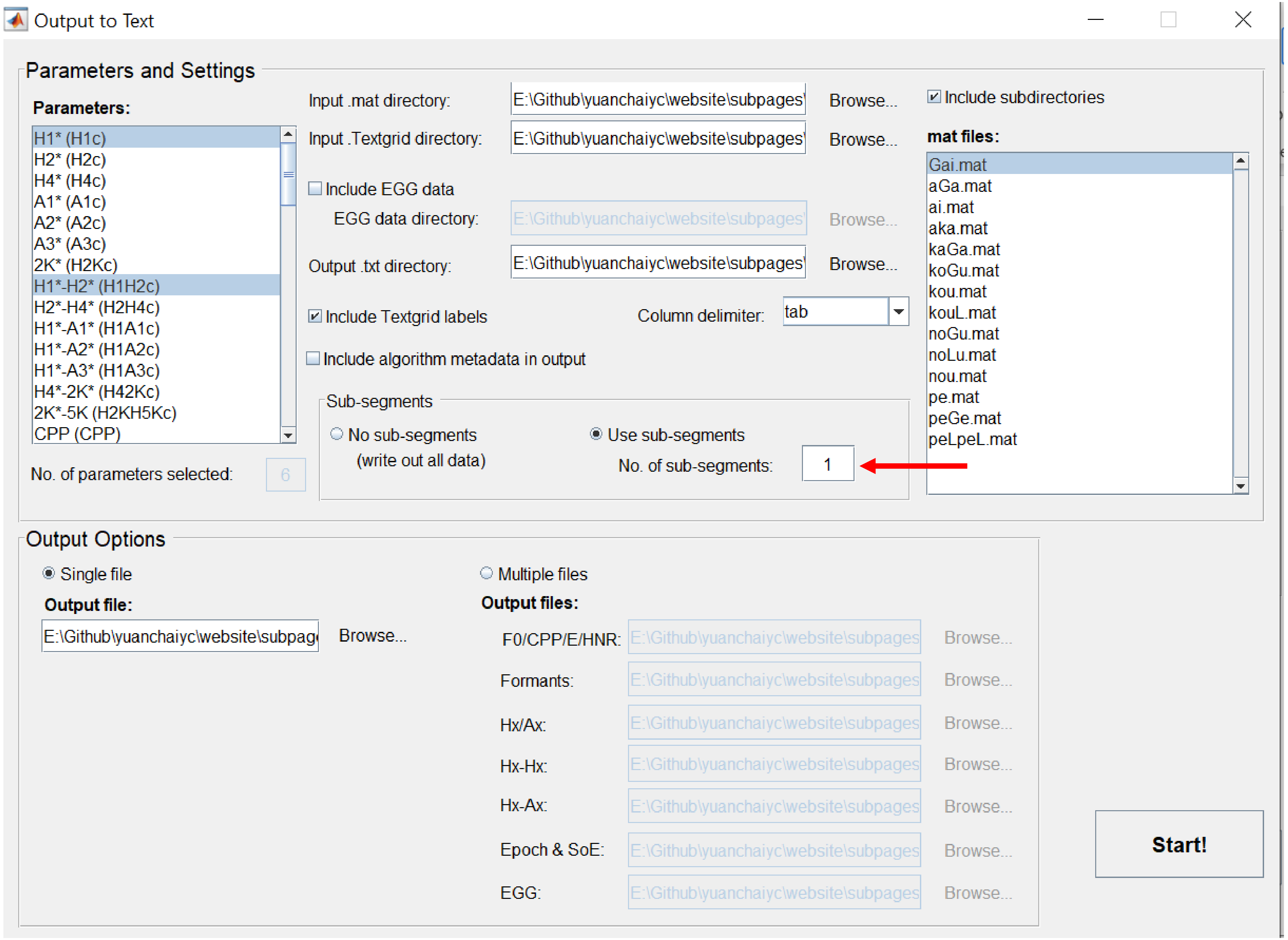
Step 3: Output to text

Step 3



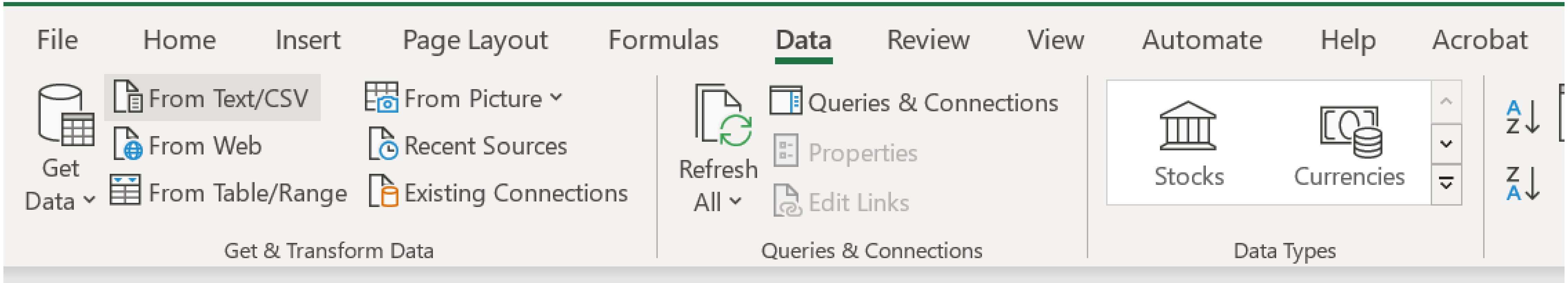
Parameters to select for today:

H1*
H1*-H2*
HNR05
strF0
sF1
sF2
SoE



Step 4: Visualize the output in Excel

- Open output.txt in Excel:
  - Open Excel → Data → From Text/CSV



output_mean_selected.txt

File Origin: 65001: Unicode (UTF-8)

Delimiter: Tab

Data Type Detection: Based on first 200 rows

Filename	Label	seg_Start	seg_End	H1c_mean	H1H2c_mean	CPP_mean	Energy_mean	HNR05_mean	strF0_mean
Gai.mat	1-a-short-glottal	154.993	241.778	19.524	0.259	19.521	8.869	2.782	190.1
Gai.mat	2-i-short-glottal	241.778	447.722	14.792	7.24	16.564	2.978	3.201	167.7
aGa.mat	1-a-short-glottal	127.354	215.402	15.279	1.946	17.511	3	6.043	201.5
aGa.mat	2-a-short-glottal	242.999	353.387	8.207	-3.91	16.243	0.973	2.858	132.5
ai.mat	1-a-short-modal	109.075	316.273	19.063	10.303	16.203	0.749	7.723	167.6
ai.mat	2-i-short-modal	316.273	513.833	18.351	13.719	15.878	0.566	6.278	185.6
aka.mat	1-a-short-modal	327.73	381.61	9.135	-0.786	16.362	0.477	6.822	187.6
aka.mat	2-a-short-modal	514.339	637.869	7.967	-2.142	16.766	0.365	10.015	217.4
kaGa.mat	1-a-short-glottal	110.185	174.578	16.543	8.428	17.414	1.696	4.561	215.0
kaGa.mat	2-a-short-glottal	235.029	323.077	11.401	-0.247	16.171	1.09	-1.107	132.0
koGu.mat	1-o-short-glottal	167.14	223.641	15.259	8.602	16.71	8.866	3.881	262.7
koGu.mat	2-u-short-glottal	284.793	396.801	17.068	3.731	15.888	3.101	-0.308	165.3
kou.mat	1-o-short-modal	258.052	498.542	13.007	-2.459	18.71	20.469	8.858	188.
kou.mat	2-u-short-modal	498.542	660.182	10.456	-1.433	15.491	1.007	3.138	228.9
kouL.mat	1-o-short-modal	183.771	371.84	15.715	0.582	17.661	10.577	4.053	203.4
kouL.mat	2-u-long-modal	371.84	664.019	17.66	-0.226	17.621	5.994	5.656	246.7
noGu.mat	1-o-short-glottal	203.659	267.468	23.627	14.075	18.396	6.587	4.782	236.6
noGu.mat	2-u-short-glottal	334.636	465.612	15.828	7.704	15.669	1.454	-1.101	160.0
noLu.mat	1-o-long-modal	206.389	444.834	19.19	4.425	19.165	7.607	9.403	194.0
noLu.mat	2-u-short-modal	444.834	595.961	11.085	-4.378	15.915	1.871	-1.672	110.1
nou.mat	1-o-short-modal	177.954	374.419	10.721	0.382	19.414	2.846	12.269	180.

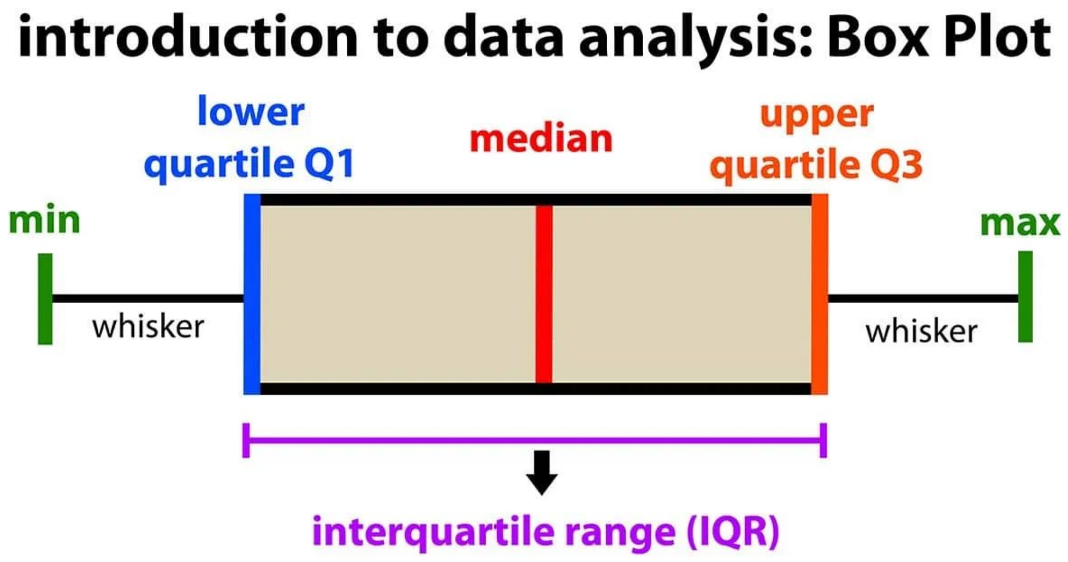
Load | Transform Data | Cancel

Step 4: Visualize the output in Excel

- Open output.txt in Excel:
  - Open Excel → Data → From Text/CSV;
  - Load the data;
  - Save the data file as a .xlsx file.

Step 4: Visualize the output in Excel

- Draw boxplots:
  - Boxplots present the median, first and third quantile, and the minimum and maximum of the data.



Picture from https://www.simplypsychology.org/boxplots.html

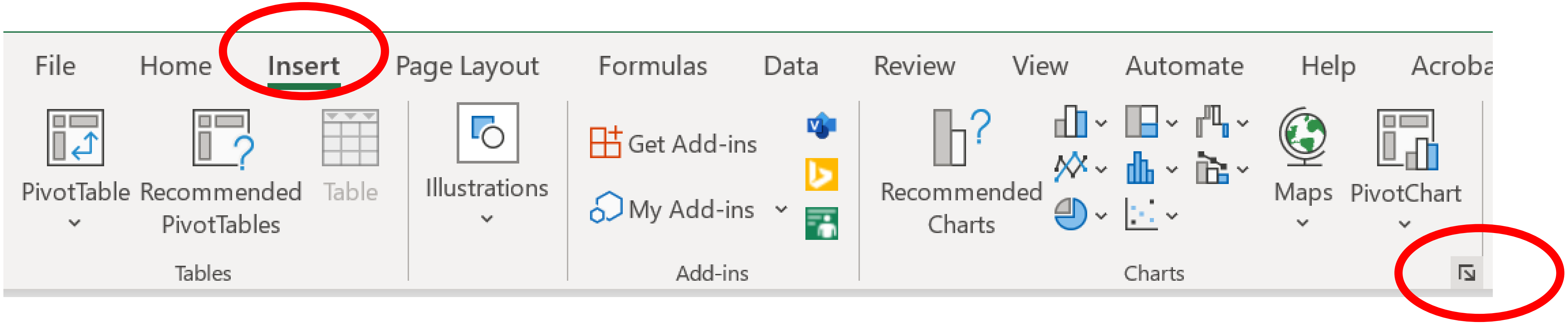
Step 4: Visualize the output in Excel

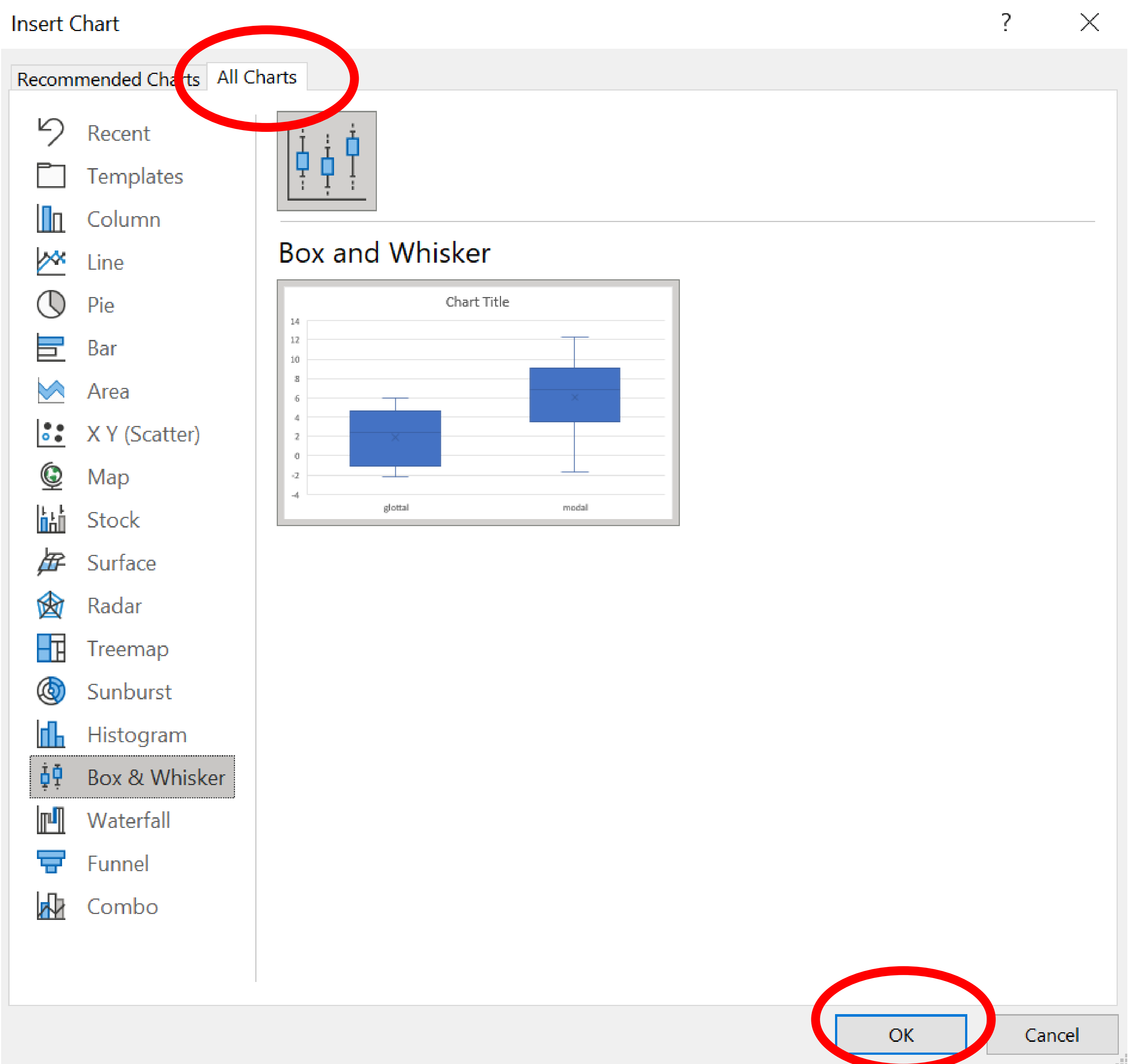
- Draw boxplots: H1-H2 distribution of modal vs. glottalized phonation
  - Select the column of “phonation”; Press “ctrl” on the keyboard, and Select the column of “H1H2c_mean”

A	B	C	D	E	F	G	H	I	J	K	L
Filename	positi	vowel	length	phonation	seg_Start	seg_End	H1c_mean	H1H2c_mean	CPP_mean	Energy_mean	HNR05_mean
aGa.mat	1	a	short	glottal	127.354	215.402	15.279	1.946	17.511	3	6.043
aGa.mat	2	a	short	glottal	242.999	353.387	8.207	-3.91	16.243	0.973	2.858
aka.mat	1	a	short	modal	327.73	381.61	9.135	-0.786	16.362	0.477	6.822
aka.mat	2	a	short	modal	514.339	637.869	7.967	-2.142	16.766	0.365	10.015
kaGa.mat	1	a	short	glottal	110.185	174.578	16.543	8.428	17.414	1.696	4.561
kaGa.mat	2	a	short	glottal	235.029	323.077	11.401	-0.247	16.171	1.09	-1.107
koGu.mat	1	o	short	glottal	167.14	223.641	15.259	8.602	16.71	8.866	3.881
koGu.mat	2	u	short	glottal	284.793	396.801	17.068	3.731	15.888	3.101	-0.308
kou.mat	1	o	short	modal	258.052	498.542	13.007	-2.459	18.71	20.469	8.858

Step 4: Visualize the output in Excel

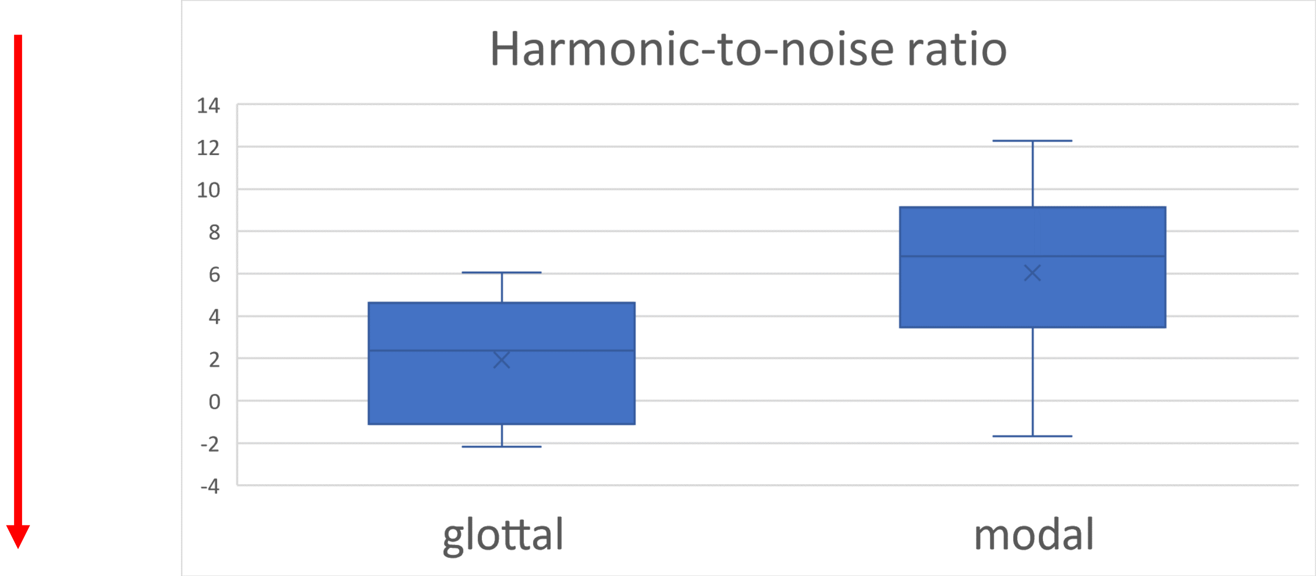
- Draw boxplots: HNR distribution of modal vs. glottalized phonation
  - Select the column of "phonation"; Press "ctrl" on the keyboard, and Select the column of "HNR05_mean"
  - Go to Insert → Charts → All charts → Box & Whisker → Press "OK"




Insert Chart
?
Recommended Charts
All Charts
Recent
Templates
Column
Line
Pie
Bar
Area
X Y (Scatter)
Map
Stock
Surface
Radar
Treemap
Sunburst
Histogram
Box & Whisker
Waterfall
Funnel
Combo
Box and Whisker
Chart Title
14
12
10
8
6
4
2
0
-2
-4
glottal
modal
OK
Cancel

Step 4: Visualize the output in Excel

- Draw boxplots: HNR distribution of modal vs. glottalized phonation



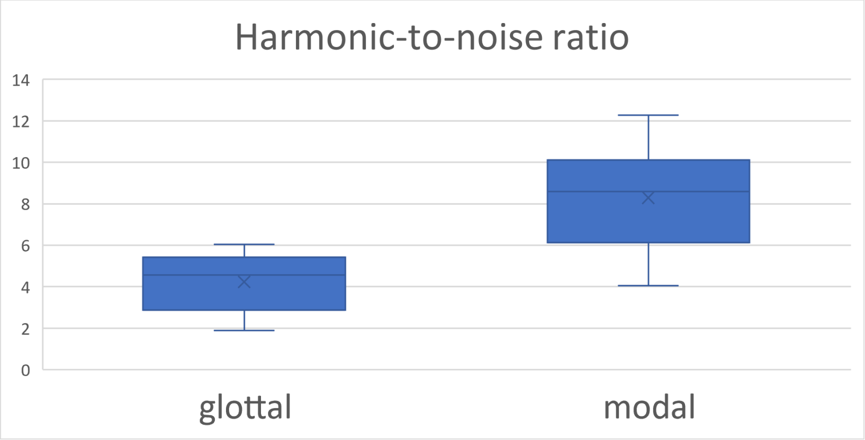
More glottalized

We see that vowels surrounding glottal stops have lower HNR than vowels that do not. This indicates that vowels in VʔV words are more glottalized than vowels in V or VV words.

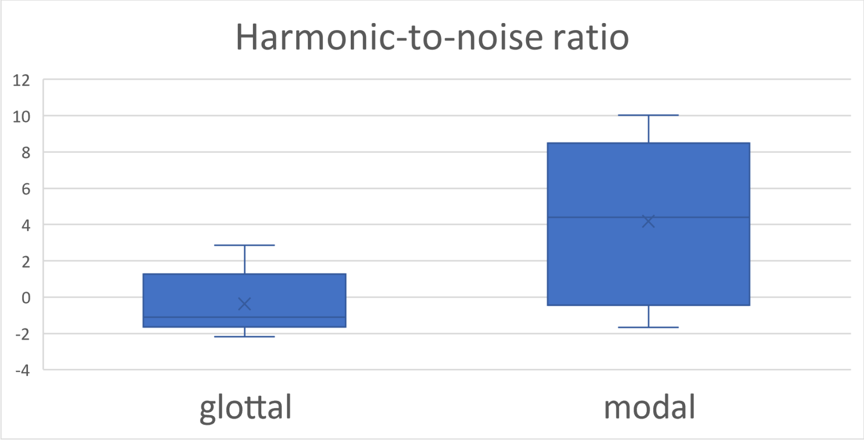
Step 4: Visualize the output in Excel

- Draw boxplots: HNR distribution of modal vs. glottalized phonation
  - You can filter the data and see how the plot changes.
    - Filter the “position” column by only selecting “1”
    - Then filter the “position” column by only selecting “2”

Position 1



Position 2



More glottalized

Step 4: Visualize the output in Excel

- Let’s try more graphs!
  - Draw boxplots for H1H2c_mean (H1—H2), soe_mean (Strength of Excitation), strF0 (F0 using “straight” algorithm)

Step 4: Visualize the output in Excel

- Let’s try drawing graphs in R
- R studio online: https://posit.cloud/content/5398051
- R script offline: https://yuanchaiyc.github.io/website/subpages/VS-tutorial.Rmd

Take-home message

- VoiceSauce is a tool for analyzing acoustics of sound signals;
- Its advantage are:
  - Able to process a large batch of sound files in one sitting;
  - Able to calculate parameters relating to voice quality;
  - Able to compare different algorithms for one measure (e.g. F0, formants);
  - The output is in a tab-delimited format and is ready to be passed on to statistical tests and data visualization.