VOAT: Voice Onset Analysis Tool

Technical manual

Table of contents

Requirements

The current version of VOAT is only supported for **Windows** operative systems.

Supported file formats

- WAV: It can be an audio file contain 1 or more channels. It can also be a file containing audio and EGG in two different channels.
- ADICHT: An ADICHT file is a data file used by LabChart for Windows (version 7 and later).
- CSV: VOAT can read the GAW signals exported from the GAT software. The columns in the CSV file should be separated by ";" (semicolon).

Installation

VOAT is a software written in Python programming language developed for semi-automatic segmentation of voice onset from acoustic, EGG, and airflow signals. VOAT is provided as a standalone executable program, which means that it only requires an executable file that can be run from anywhere in the file system. All files required by VOAT are included in a folder. To run the program, look for the “VOAT.exe” file located in the main folder (Figure 1).

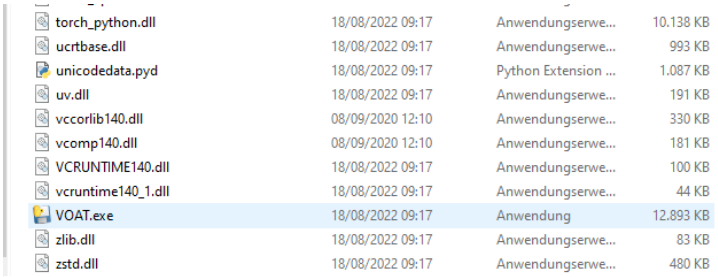


Figure 1. VOAT executable file

Usage

Main menu

When the “VOAT.exe” file is executed, the main menu of VOAT will be displayed in your screen (Figure 2).

There are two options:

- Segmentation: This option is used to segment the voice onset.
- Analysis: This option is still under development and it does not do anything.

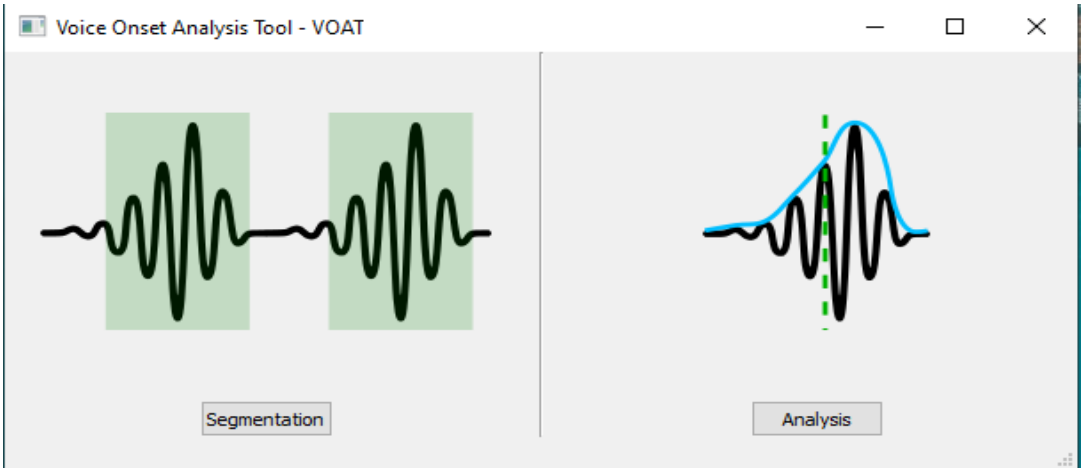


Figure 2. VOAT main menu

Segmentation

When clicked, this option allows to open a second window in the background, which is used for semi-automatic extraction of the voice onset. Figure 3 shows the main components of the ***Segmentation*** window.

Figure 3. Segmentation window.

1. Click on ***File*** (menu bar) to open a new file. The current supported formats are described in Supported file formats.
2. The  button is used to save the results of the voice onset detection on an excel file. The excel file contains four columns:
    a. **ID:** Name of the analyzed file.
    b. **Label:** Name of the segment (can be edited in the software).
    c. **Setting:** The settings used for the segmentation.
    d. **Voice onset [ms]:** The voice onset measured in milliseconds.

    The user always has the option to select the location to save the file. The name will be generated automatically and a message will appear with the location and name of the file. When a new segment is to be saved (in the same location and same excel file), VOAT will add the new measurement in a new row.

3. The ***Record*** selector is used to select the file to be analyzed (only applies for ADICHT files). The ***Channel*** allows to select the channel to be analyzed. Figure 4 shows an
   example for an ADICHT file containing acoustic (Acoustic mic), EGG, and airflow (Airflow-Input A)) signals.

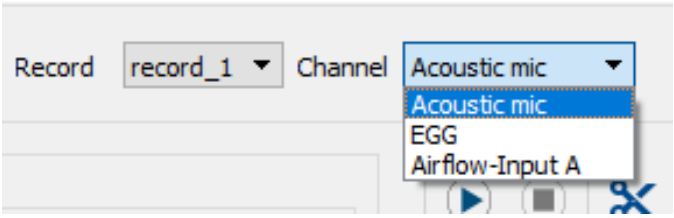


Figure 4. File containing acoustic, EGG, and airflow signals.

4. Space designated to plot the signal to be analyzed (top) and the estimated voice onset (bottom).

5. Buttons used to play and stop ( ) sounds, cut parts of the plotted signal ( ), and zoom in and out parts of the signal ( ). The play, stop, and cut buttons are only enable once a signal is loaded for analysis.

   To Cut out a segment from the signal, you must left-click and hold on the part to be deleted, then select the segment and release the button. The segment will be highlighted in blue and you can press to cut that part of the signal. Figure 5 shows an example.

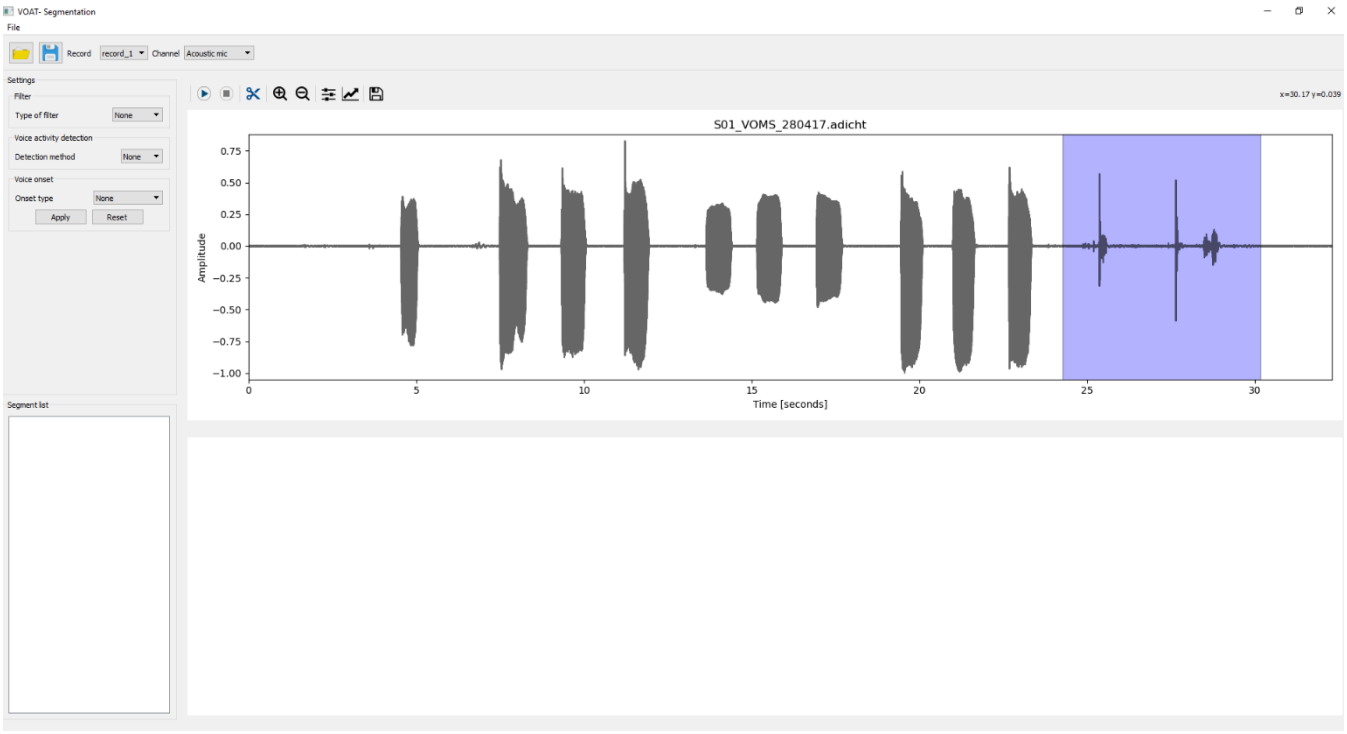


Figure 5. Segment selected from a signal.

6. Space designated to list all segments extracted from the plotted signal. Figure 6 shows an example:

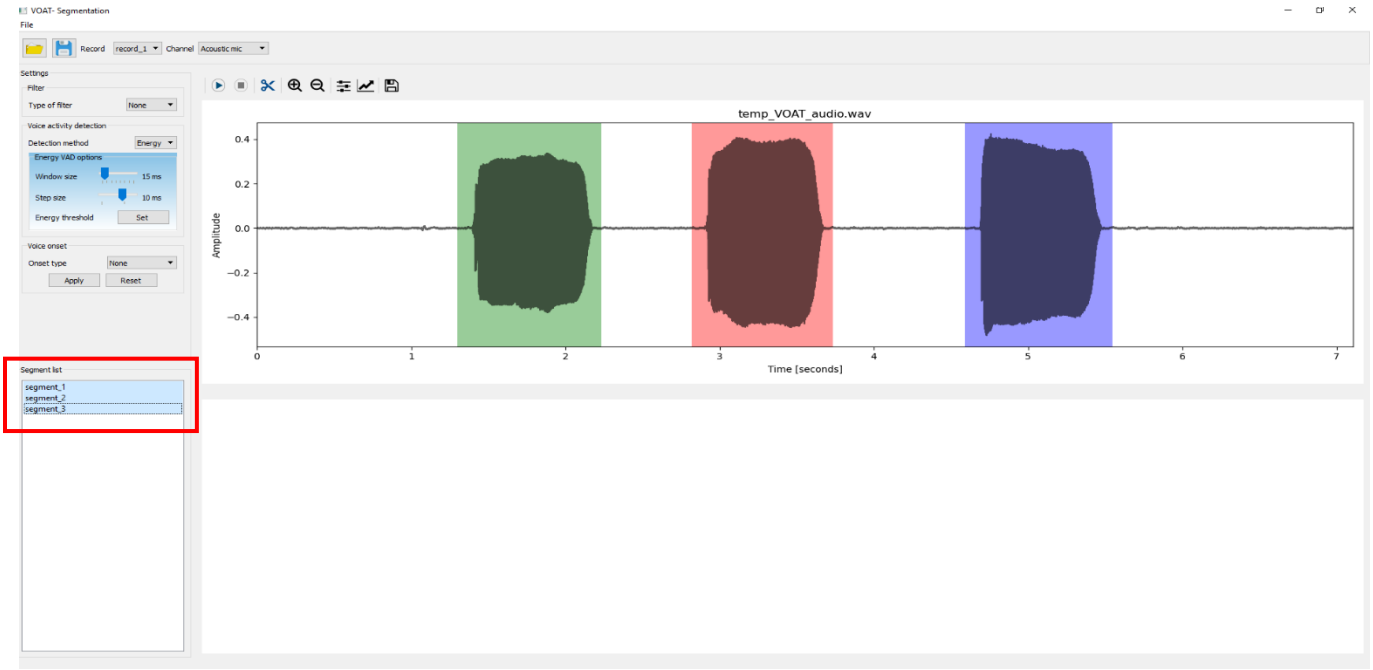


Figure 6. Example of segments selected from an acoustic signal. When the name of one of the segments is clicked, the corresponding region will be highlighted in color in the plot.

7. This is the settings used for automatic detection of the voice onset. The details are explained in the subsection Settings.

Settings

After loading the signal to be analyzed, there are three main settings used for voice onset detection:

- **Filter*:*** Allows to apply a Bandpass, Highpass, or Lowpass filter to the plotted signal. The options are:
    - **None:** If this option is selected, then filtering is NOT applied to the signal.
    - **Bandpass:** Allows to select the low and high cut-off frequencies of the filter. Default values are 75 Hz (low) and 300 Hz (high).
    - **Highpass:** Default value is 300 Hz.
    - **Lowpass:** Default value is 300 Hz.

- **Voice activity detection:** This option allows the user to select segments from a recording to be analyzed. Is useful for cases in which a single recording contains multiple phonations (Figure 7).

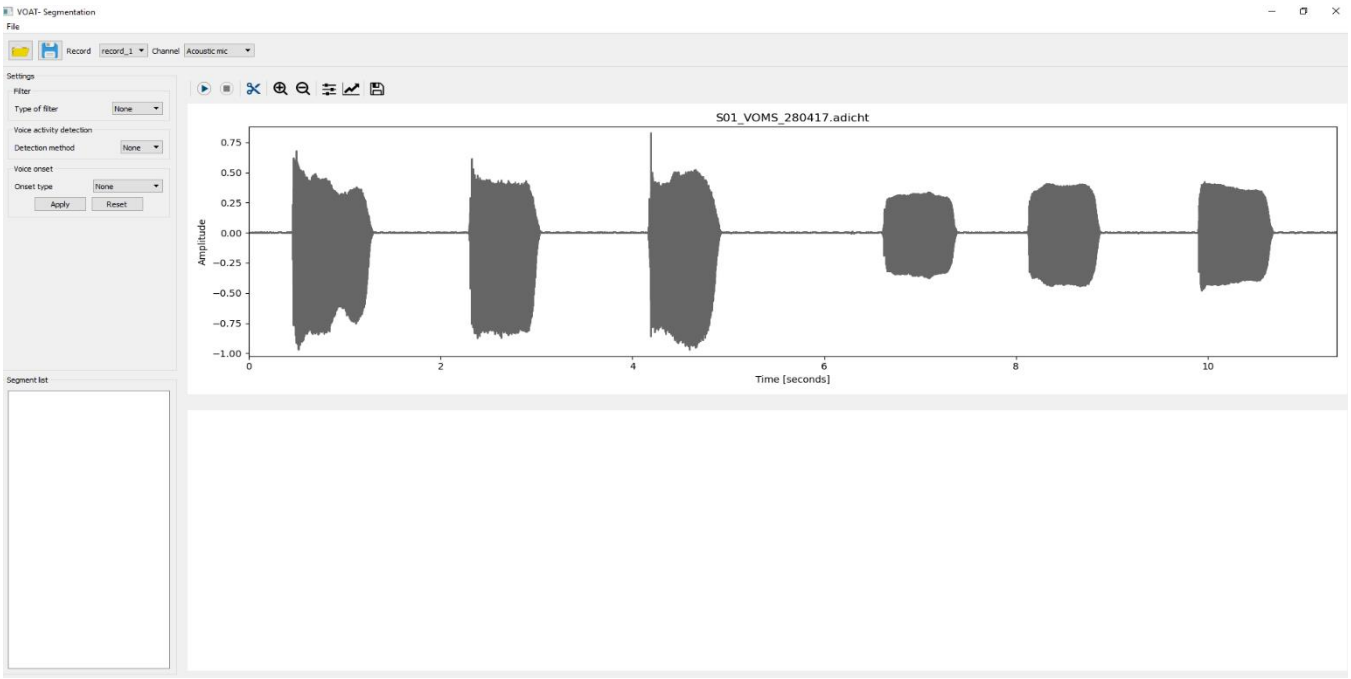


Figure 7. Acoustic signal with multiple sustained phonations.

The ***Voice activity detection*** has four options:

- **None:** No segment is detected.
- **Manual:** The segment(s) must be selected manually from the plot. For this, the user must press and hold the left click and drag the mouse pointer to select the region of interest. Then, click on the  icon and the segment

will be added to the ***Segment list*** and colored in green in the plot***.*** Figures Figure *9* and Figure *8* show the selection and inclusion of the voice

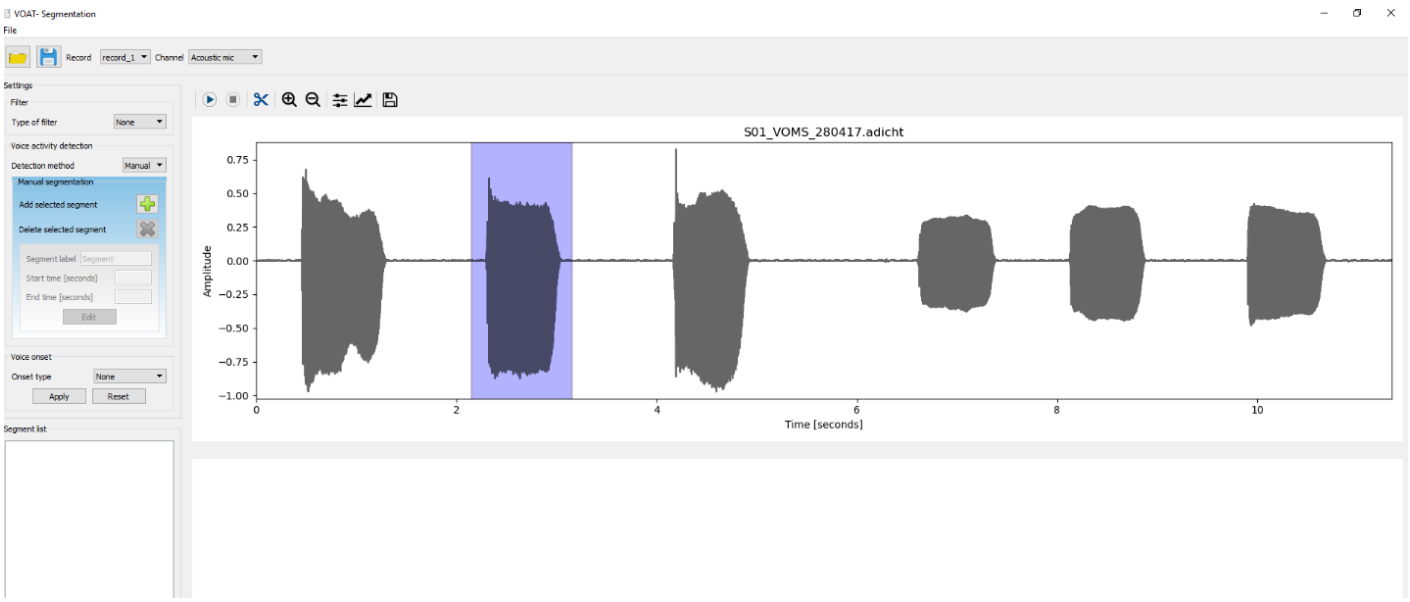


Figure 8. Selection of segment.

segment.

The  button allows to delete the selected segment in ***Segment list***. It is also possible to change the name of the selected segment by writing the name in the field ***Segment label*** (Figure 8) and clicking on the ***Edit*** button. It is also possible to change the starting and ending times (***Start time [seconds]l End time [seconds]***).

- **Energy:** This option will automatically select the segments from the plot based on the energy contour of the signal. The options ***Window size*** and ***Step size*** are used to control the precision of the selection e.g., high values

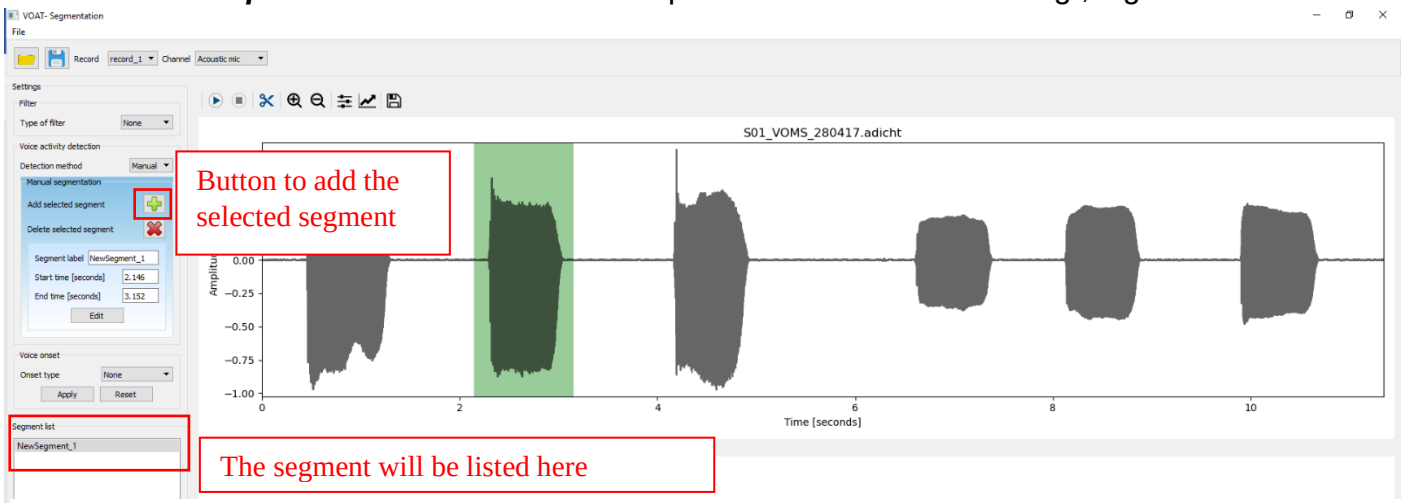


Figure 9. Adding the segment to the Segment list.

result in large segments. It is also possible to select an ***Energy Threshold*** by clicking on the ***Set*** button. Then, you must select the energy level on the plot by moving the pointer (a horizontal red line will be displayed in plot) and then left-clicking on the desired threshold. Figure 10 shows an example.

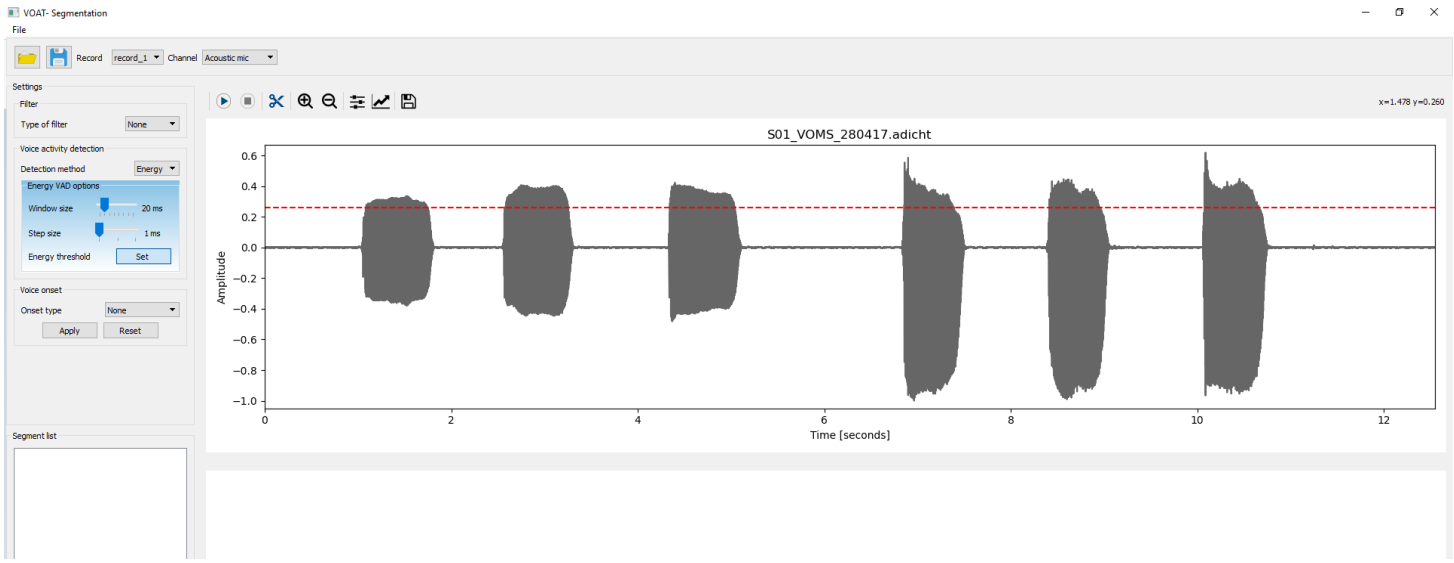


Figure 10. Selection of energy threshold.

- **Voice onset*:*** This option allows semi-automatic detection of voice onset using five different options that can be change by the user.
    - **Method:** Defines the method used to computed the envelope of the signal. Options are Hilbert and Kunduk (please refer to Section ***Voice onset computation -> Vocal Rise Time*** for more information).
    - **Smoothing factor:** Allows to select 6 levels of smoothing for the amplitude envelope. A low value results in an amplitude envelope that is highly adapted to the shape of the original signal. A high value results in an amplitude envelope that describes the general shape of the signal. Default is 3.
    - **Segment:** Allows to select the duration of the signal where the algorithm should look for the voice onset. Default is 400 ms.
    - **Onset threshold [%]:** Allows to select the starting point of the voice onset. Default is 10% of the maximum value of the amplitude envelope.
    - **Saturation point [%]:** Allows to select the end point of the voice onset. Default is 90% of the maximum value of the amplitude envelope.

  Figure 11 shows an example of the voice onset computed on three different phonations contained in a single recording.

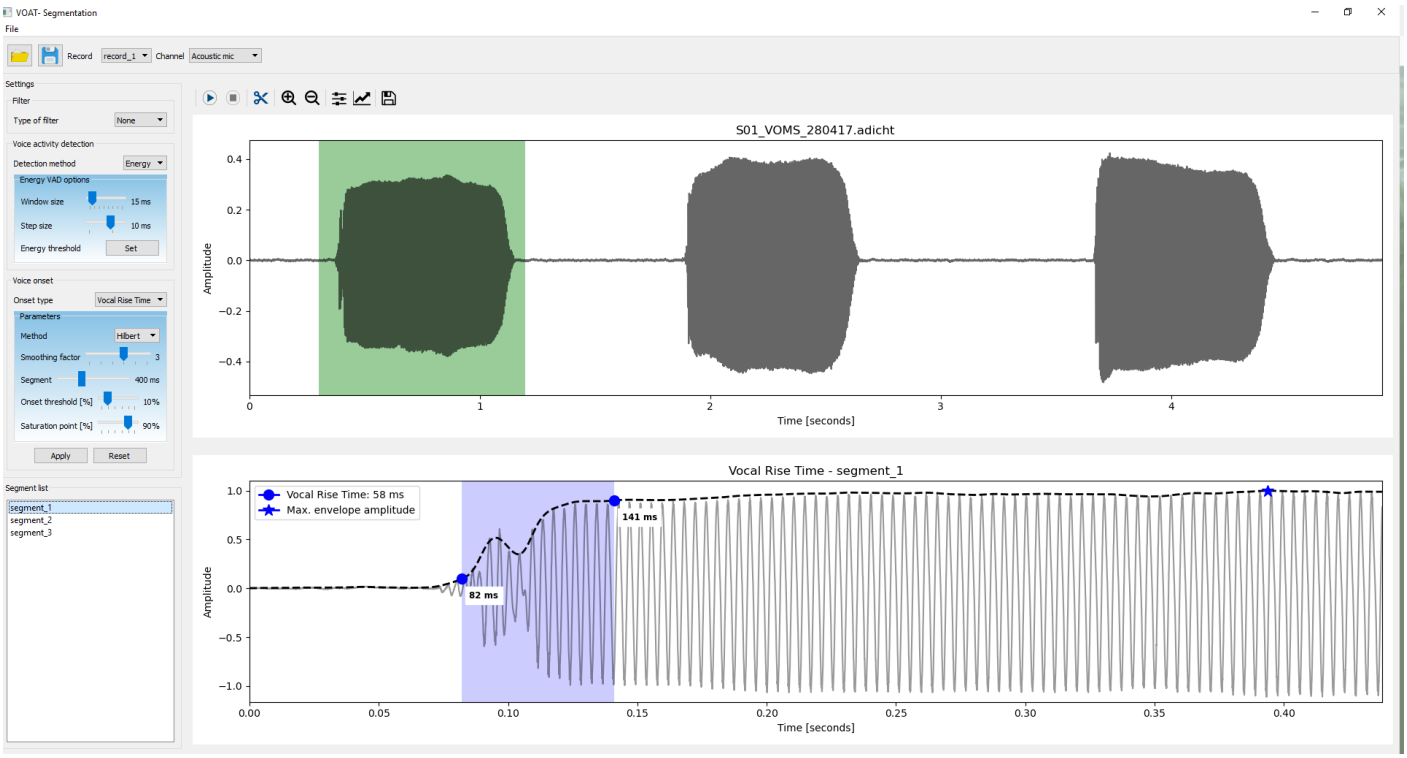


Figure 11. Voice onset detection on three different segments. The top shows the original signal and the botton shows the voice onset of the selected segment. The blue region shows the voice onset, the blue points show the start and end points, and the black dotted line represent the amplitude envelope computed with the Hilbert method.

Voice onset computation

Vocal Rise Time (VRT)

The VRT is used to estimate the voice onset and is defined as the time needed for the amplitude envelope of the signal to go from 10% up to 90% of the maximum value of the amplitude envelope. The VRT can be computed not only on acoustic signals, but also on EGG, airflow, glottal area waveforms, among others. In VOAT, the percentages of the initial and final points can be selected by the user. The algorithm implemented in VOAT to compute the voice onset is as follows:

1. Eliminate the DC level of the signal and re-scale the amplitude values in the range between -1 and 1.

2. Add 50 ms of silence at the beginning of the signal to ensure that there is a pause before the onset.

3. Compute the amplitude envelope of the signal, defined as a smooth curve outlining its extremes. VOAT allows the computation using two approaches **Hilbert** and **Kunduk**.

a. **Hilbert:** Applies the Hilbert transform to the signal to obtain an analytic version of it (real and imaginary components):

$$s(t) = x(t) + iy(t)$$

Where $s(t)$ is the analytic signal, $x(t)$ is the real component (original acoustic/EGG/airflow signal), and $y(t)$ is the imaginary component (Hilbert transform of $x(t)$). The amplitude envelope is computed as the magnitude of the resulting analytic signal:

$$e(t) = \sqrt{x(t)^2 + y(t)^2}$$

The resulting envelope $e(t)$ is convolved with a Gaussian window. The size of the window can be 20 ms, 40 ms, 60 ms, 80 ms or 100 ms, depending on the ***Smoothing factor*** (Section Voice onset). For instance, if the ***Smoothing factor*** is 3, then the size of the window is 60 ms.

b. **Kunduk:** The amplitude envelope is computed by first finding all of the amplitude peaks of the signal, i.e., the local maxima. In this case, the ***Smoothing factor*** controls the distance between neighbor maxima. The resulting peak points are then interpolated using a B-spline.

4. After computing the envelope, the maximum value of it is computed. Then, the starting voice onset point is located by looking for the value of the amplitude envelope equal to the 10% of the maximum value.

5. Since the voice onset is located at the start of the signal, only a segment of the envelope is considered. The length of the segment is selected by the user with the ***Segment*** option (Section ***Usage->Segmentation->Settings-> Voice onset***).

   The final length of the signal depends on the point of the starting voice onset point, found in step 4. For instance, if ***Segment*** is 400 ms and the starting point (10% of the maximum value of the envelope) is located 20 ms **after** the beginning of the signal, then the final length of the signal is 420 ms.

   The reason for this is as follows: If the user has a signal with 500 ms of silence at the beginning of the signal and it does **not** cut it, then VOAT has to ensure that the voice onset is at least in the last 400 ms (depends on ***Segment***) of the segment.

6. Re-scale the amplitude envelope to have values between -1 and 1 by dividing each point of the amplitude envelope by the maximum value of the envelope. Since the envelope was segmented in the previous step, we must ensure that the maximum value is 1 to simplify the localization of the voice onset.

7. Locate the point of the envelope which is greater or equal to ***Onset threshold [%]*** (default is 10%) of the maximum. For instance, since the maximum value of the envelope is 1 (Step 6), then the 10% is 0.1.

8. Locate the point of the envelope which is greater or equal to ***Saturation point [%]*** (default is 90%) of the maximum. For instance, since the maximum value of the envelope is 1 (Step 6), then the 90% is 0.9.

9. Subtract 50 ms from the onset and saturation points. This 50 ms correspond to the silence added in step 2.

10. Use the resulting onset and saturation points to measure the voice onset and to locate its position in the original signal.

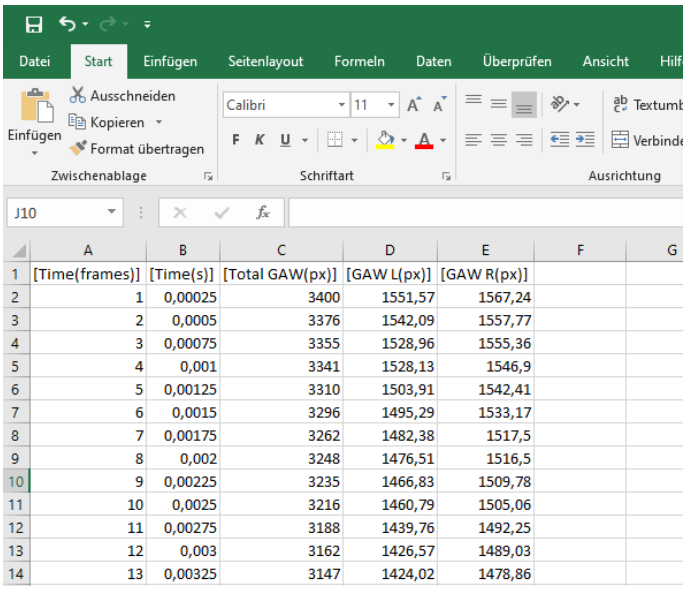
Glottal Analysis Tool (GAT)

VOAT can also compute the voice onset from Glottal Area Waveform (GAW) and trajectory files obtained with the software **GAT,** a software developed at the University Hospital Erlangen used to analyze high-speed videoendoscopy data[1].

Estimation of the voice onset from a GAW file

File format: GAT offers the possibility to create GAW files in the **.csv** format. To read such a file in VOAT, it is necessary that the columns [Time(s)], [Total GAW(px)], [GAW L(px)], and [GAW R(px)] are contained. This is because, internally, VOAT looks for the column [Time(s)] to compute the sampling frequency of the GAW signal. Then, it is assumed that the following three columns correspond to the total GAW ([Total GAW(px)]), the GAW from the left vocal fold ([GAW L(px)]), and , the GAW from the right vocal fold ([GAW R(px)]), respectively. Figure 12 shows an example of a GAW file (.csv) generated with GAT.

[1] https://www.hno-klinik.uk-erlangen.de/phoniatrie/forschung/computational-medicine/gat-software/



	A	B	C	D	E
1	[Time(frames)]	[Time(s)]	[Total GAW(px)]	[GAW L(px)]	[GAW R(px)]
2	1	0,00025	3400	1551,57	1567,24
3	2	0,0005	3376	1542,09	1557,77
4	3	0,00075	3355	1528,96	1555,36
5	4	0,001	3341	1528,13	1546,9
6	5	0,00125	3310	1503,91	1542,41
7	6	0,0015	3296	1495,29	1533,17
8	7	0,00175	3262	1482,38	1517,5
9	8	0,002	3248	1476,51	1516,5
10	9	0,00225	3235	1466,83	1509,78
11	10	0,0025	3216	1460,79	1505,06
12	11	0,00275	3188	1439,76	1492,25
13	12	0,003	3162	1426,57	1489,03
14	13	0,00325	3147	1424,02	1478,86

Figure 12. Information contained in a GAW file generated with the software GAT

Once the csv file the GAW has been created, then it can be read by VOAT by clicking on ***File->Load GAT file ->Glottal Area Waveform*** (Figure 13).

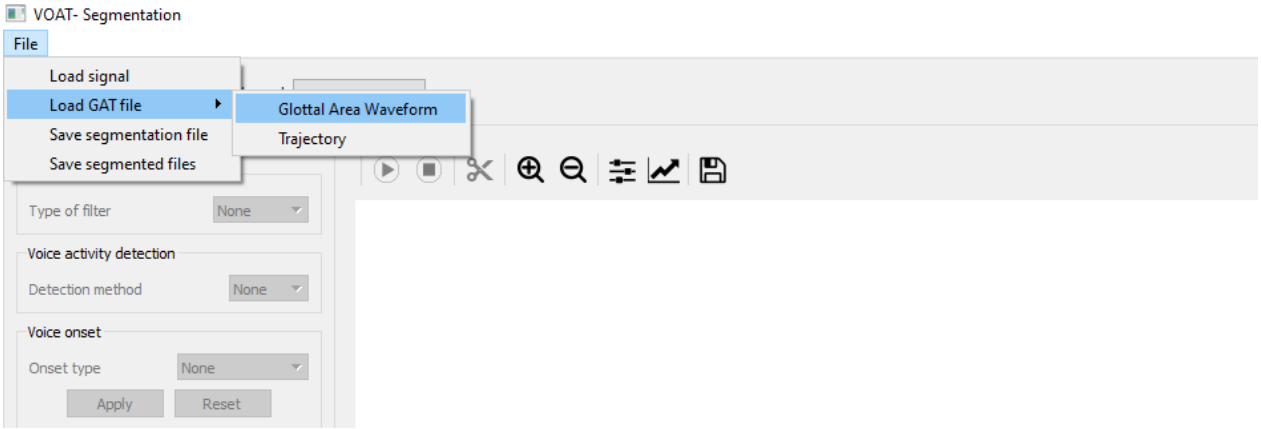


Figure 13. Opening a GAW file generated with GAT.

Then, look for the GAW file generated with GAT and open it. As shown in Figure 14, the onset of a GAW signal starts from a high value which decreases as the vocal folds start oscillating. In order to compute the voice onset with VOAT, it is necessary to filter the signal (as indicate in Settings) to “flatten” the beginning of the phonation. Figure 15 shows the resulting GAW signal after applying a bandpass filter between 75Hz and 300Hz.

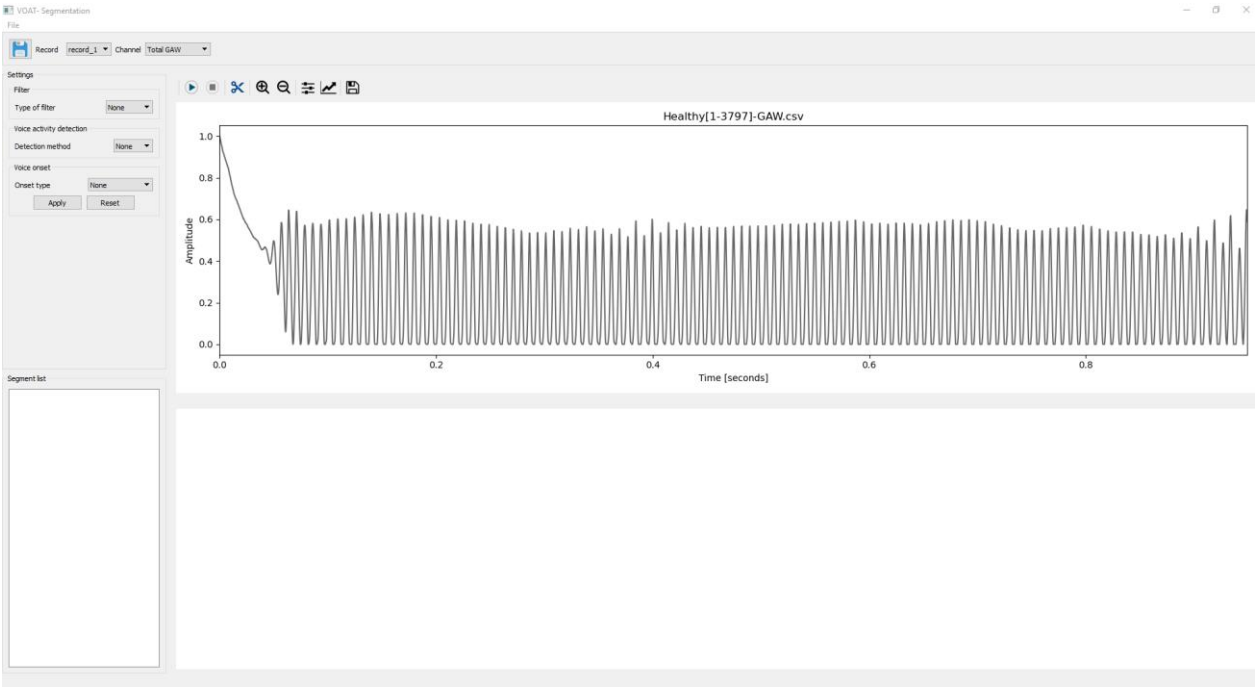


Figure 14. GAW signal generated with GAT.

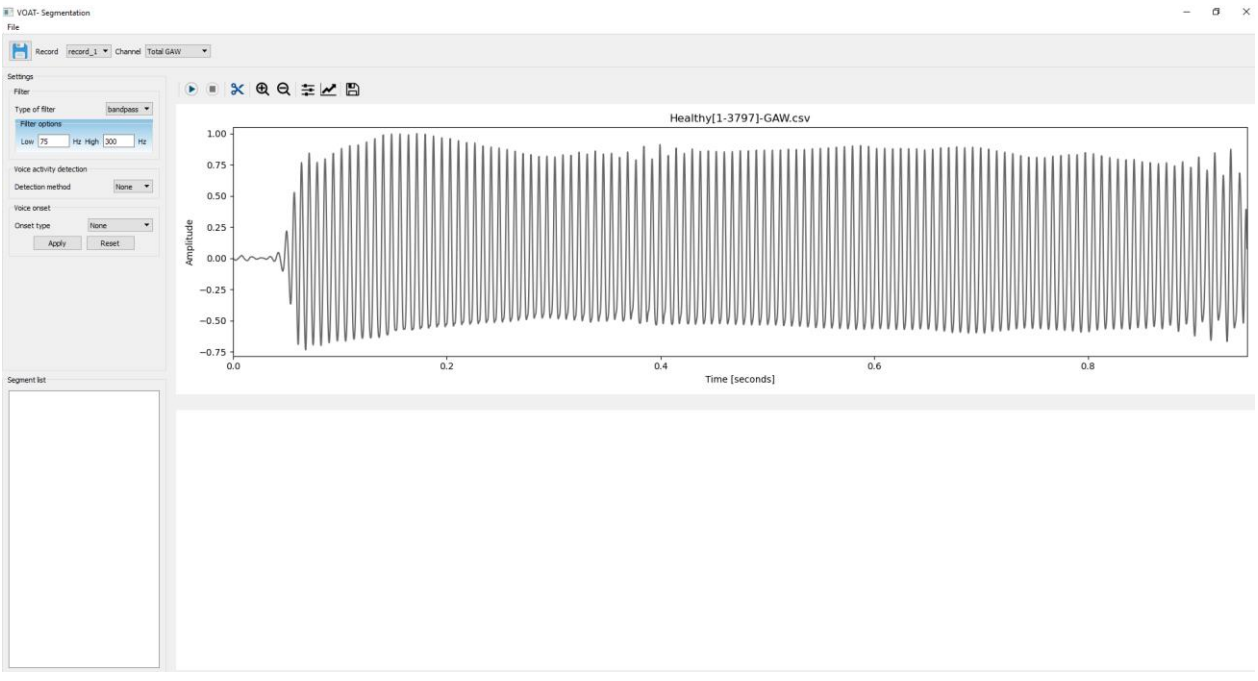
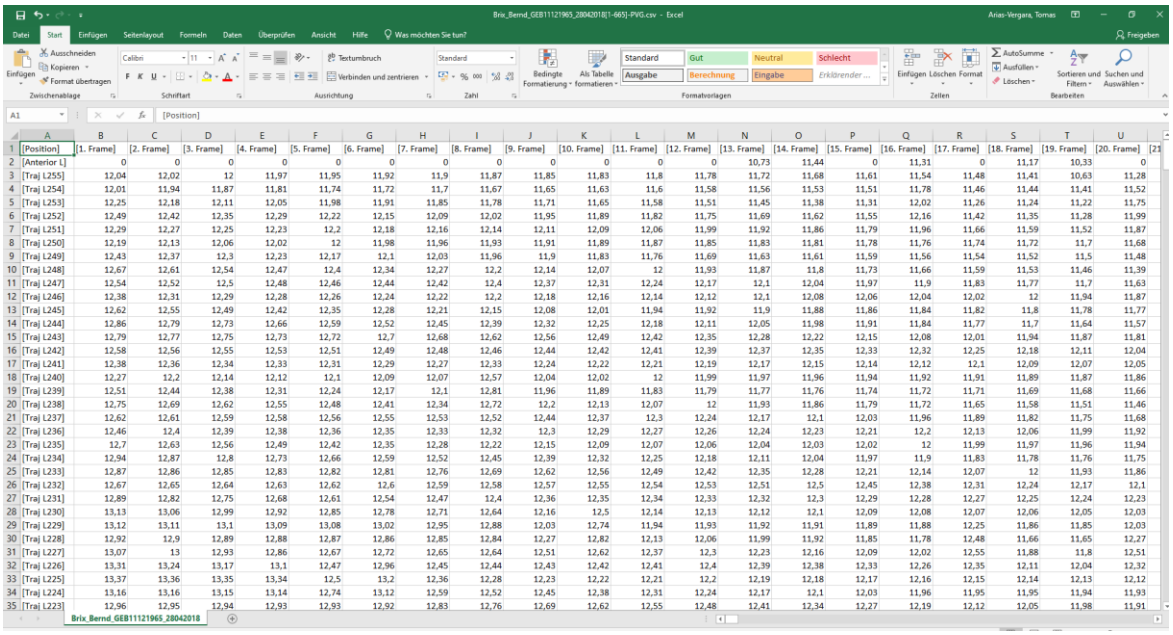


Figure 15. GAW signal after filtering.

Estimation of the voice onset from a Trajectory file (PVG)

VOAT allows reading a trajectory file (**csv format**) generated using GAT. This file is created by saving the Phonovibrogram (PVG) as a csv file in GAT. The format of the file is shown in Figure 16. The csv file must contain a column called **[Position]** used by VOAT to know which section correspond to the left (top 256 rows in the csv file) or right (bottom 256 rows) trajectory.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	[Position]	[1. Frame]	[2. Frame]	[3. Frame]	[4. Frame]	[5. Frame]	[6. Frame]	[7. Frame]	[8. Frame]	[9. Frame]	[10. Frame]	[11. Frame]	[12. Frame]	[13. Frame]	[14. Frame]	[15. Frame]	[16. Frame]	[17. Frame]	[18. Frame]	[19. Frame]	[20. Frame]
2	[Anterior L]	0	0	0	0	0	0	0	0	0	0	0	0	10,73	11,44	0	11,31	0	11,17	10,33	0
3	[Traj L255]	12,04	12,02	12	11,97	11,95	11,92	11,9	11,87	11,85	11,83	11,8	11,78	11,72	11,68	11,61	11,54	11,48	11,41	10,63	11,28
4	[Traj L254]	12,01	11,94	11,87	11,81	11,74	11,72	11,7	11,67	11,65	11,63	11,6	11,58	11,56	11,53	11,51	11,78	11,46	11,44	11,41	11,52
5	[Traj L253]	12,25	12,18	12,11	12,05	11,98	11,91	11,85	11,78	11,71	11,65	11,58	11,51	11,45	11,38	11,31	12,02	11,26	11,24	11,22	11,75
6	[Traj L252]	12,49	12,42	12,35	12,29	12,22	12,15	12,09	12,02	11,95	11,89	11,82	11,75	11,69	11,62	11,55	12,16	11,42	11,35	11,28	11,99
7	[Traj L251]	12,29	12,27	12,25	12,23	12,2	12,18	12,16	12,14	12,11	12,09	12,06	11,99	11,92	11,86	11,79	11,96	11,66	11,59	11,52	11,87
8	[Traj L250]	12,19	12,13	12,06	12,02	12	11,98	11,96	11,93	11,91	11,89	11,87	11,85	11,83	11,81	11,78	11,76	11,74	11,72	11,7	11,68
9	[Traj L249]	12,43	12,37	12,3	12,23	12,17	12,1	12,03	11,96	11,9	11,83	11,76	11,69	11,63	11,61	11,59	11,56	11,54	11,52	11,5	11,48
10	[Traj L248]	12,67	12,61	12,54	12,47	12,4	12,34	12,27	12,2	12,14	12,07	12	11,93	11,87	11,8	11,73	11,66	11,59	11,53	11,46	11,39
11	[Traj L247]	12,54	12,52	12,5	12,48	12,46	12,44	12,42	12,4	12,37	12,31	12,24	12,17	12,1	12,04	11,97	11,9	11,83	11,77	11,7	11,63
12	[Traj L246]	12,38	12,31	12,29	12,28	12,26	12,24	12,22	12,2	12,18	12,16	12,14	12,12	12,1	12,08	12,06	12,04	12,02	12	11,94	11,87
13	[Traj L245]	12,62	12,55	12,49	12,42	12,35	12,28	12,21	12,15	12,08	12,01	11,94	11,92	11,9	11,88	11,86	11,84	11,82	11,8	11,78	11,77
14	[Traj L244]	12,86	12,79	12,73	12,66	12,59	12,52	12,45	12,39	12,32	12,25	12,18	12,11	12,05	11,98	11,91	11,84	11,77	11,7	11,64	11,57
15	[Traj L243]	12,79	12,77	12,75	12,73	12,72	12,7	12,68	12,62	12,56	12,49	12,42	12,35	12,28	12,22	12,15	12,08	12,01	11,94	11,87	11,81
16	[Traj L242]	12,58	12,56	12,55	12,53	12,51	12,49	12,48	12,46	12,44	12,42	12,41	12,39	12,37	12,35	12,33	12,32	12,25	12,18	12,11	12,04
17	[Traj L241]	12,38	12,36	12,34	12,33	12,31	12,29	12,27	12,33	12,24	12,22	12,21	12,19	12,17	12,15	12,14	12,12	12,1	12,09	12,07	12,05
18	[Traj L240]	12,27	12,2	12,14	12,12	12,1	12,09	12,07	12,57	12,04	12,02	12	11,99	11,97	11,96	11,94	11,92	11,91	11,89	11,87	11,86
19	[Traj L239]	12,51	12,44	12,38	12,31	12,24	12,17	12,1	12,81	11,96	11,89	11,83	11,79	11,77	11,76	11,74	11,72	11,71	11,69	11,68	11,66
20	[Traj L238]	12,75	12,69	12,62	12,55	12,48	12,41	12,34	12,72	12,2	12,13	12,07	12	11,93	11,86	11,79	11,72	11,65	11,58	11,51	11,46
21	[Traj L237]	12,62	12,61	12,59	12,58	12,56	12,55	12,53	12,52	12,44	12,37	12,3	12,24	12,17	12,1	12,03	11,96	11,89	11,82	11,75	11,68
22	[Traj L236]	12,46	12,4	12,39	12,38	12,36	12,35	12,33	12,32	12,3	12,29	12,27	12,26	12,24	12,23	12,21	12,2	12,13	12,06	11,99	11,92
23	[Traj L235]	12,7	12,63	12,56	12,49	12,42	12,35	12,28	12,22	12,15	12,09	12,07	12,06	12,04	12,03	12,02	12	11,99	11,97	11,96	11,94
24	[Traj L234]	12,94	12,87	12,8	12,73	12,66	12,59	12,52	12,45	12,39	12,32	12,25	12,18	12,11	12,04	11,97	11,9	11,83	11,78	11,76	11,75
25	[Traj L233]	12,87	12,86	12,85	12,83	12,82	12,81	12,76	12,69	12,62	12,56	12,49	12,42	12,35	12,28	12,21	12,14	12,07	12	11,93	11,86
26	[Traj L232]	12,67	12,65	12,64	12,63	12,62	12,6	12,59	12,58	12,57	12,55	12,54	12,53	12,51	12,5	12,45	12,38	12,31	12,24	12,17	12,1
27	[Traj L231]	12,89	12,82	12,75	12,68	12,61	12,54	12,47	12,4	12,36	12,35	12,34	12,33	12,32	12,3	12,29	12,28	12,27	12,25	12,24	12,23
28	[Traj L230]	13,13	13,06	12,99	12,92	12,85	12,78	12,71	12,64	12,16	12,5	12,14	12,13	12,12	12,1	12,09	12,08	12,07	12,06	12,05	12,03
29	[Traj L229]	13,12	13,11	13,1	13,09	13,08	13,02	12,95	12,88	12,03	12,74	11,94	11,93	11,92	11,91	11,89	11,88	12,25	11,86	11,85	12,03
30	[Traj L228]	12,92	12,9	12,89	12,88	12,87	12,86	12,85	12,84	12,27	12,82	12,13	12,06	11,99	11,92	11,85	11,78	12,48	11,66	11,65	12,27
31	[Traj L227]	13,07	13	12,93	12,86	12,67	12,72	12,65	12,64	12,51	12,62	12,37	12,3	12,23	12,16	12,09	12,02	12,55	11,88	11,8	12,51
32	[Traj L226]	13,31	13,24	13,17	13,1	12,47	12,96	12,45	12,44	12,43	12,42	12,41	12,4	12,39	12,38	12,33	12,26	12,35	12,11	12,04	12,32
33	[Traj L225]	13,37	13,36	13,35	13,34	12,5	13,2	12,36	12,28	12,23	12,22	12,21	12,2	12,19	12,18	12,17	12,16	12,15	12,14	12,13	12,12
34	[Traj L224]	13,16	13,16	13,15	13,14	12,74	13,12	12,59	12,52	12,45	12,38	12,31	12,24	12,17	12,1	12,03	11,96	11,95	11,95	11,94	11,93
35	[Traj L223]	12,96	12,95	12,94	12,93	12,93	12,92	12,83	12,76	12,69	12,62	12,55	12,48	12,41	12,34	12,27	12,19	12,12	12,05	11,98	11,91

Figure 16. Trajectory file generated with GAT.

To read a trajectory file, you must select the option ***File->Load GAT file->Trajectory*** in the menu bar (Figure 17). Then, a window will pop-up (Figure 18), asking the trajectory position (see the GAT manual for more information) and sampling frequency (default is 4000).

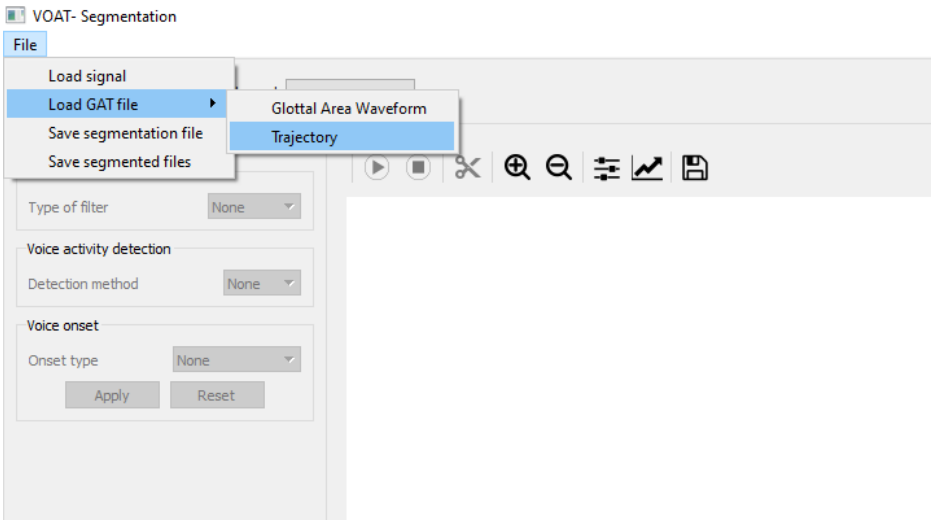


Figure 17. Selecting a trajectory file.

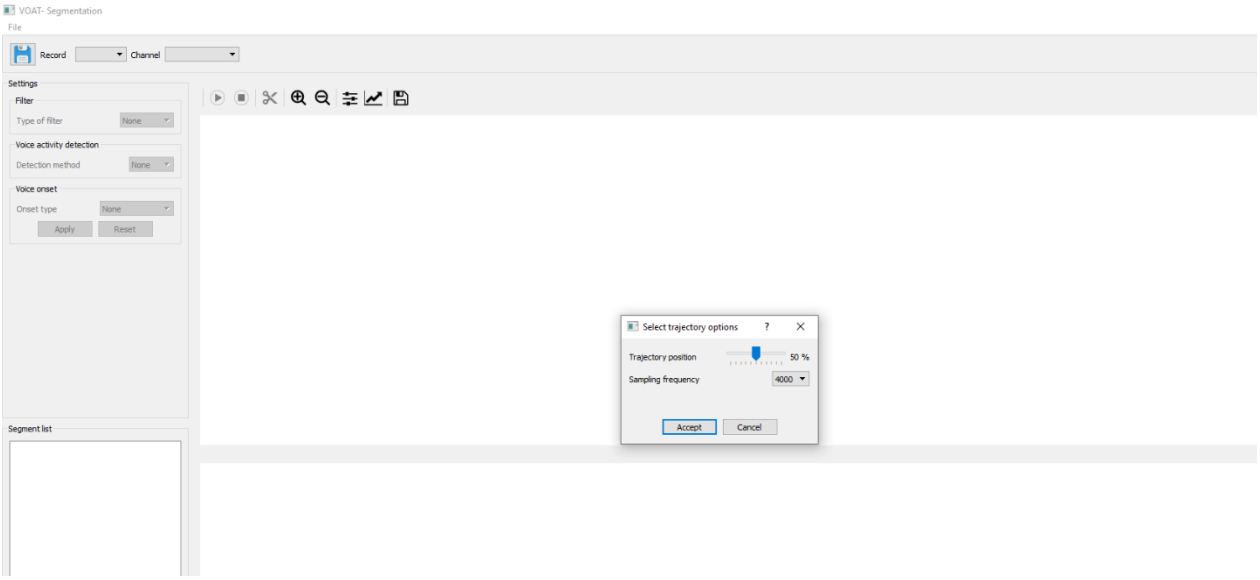


Figure 18. Selecting the region of the vocal folds (base on the trajectory) to be analyze.

Once the position and sampling frequency are confirmed, VOAT will display the selected trajectory (Figure 19).

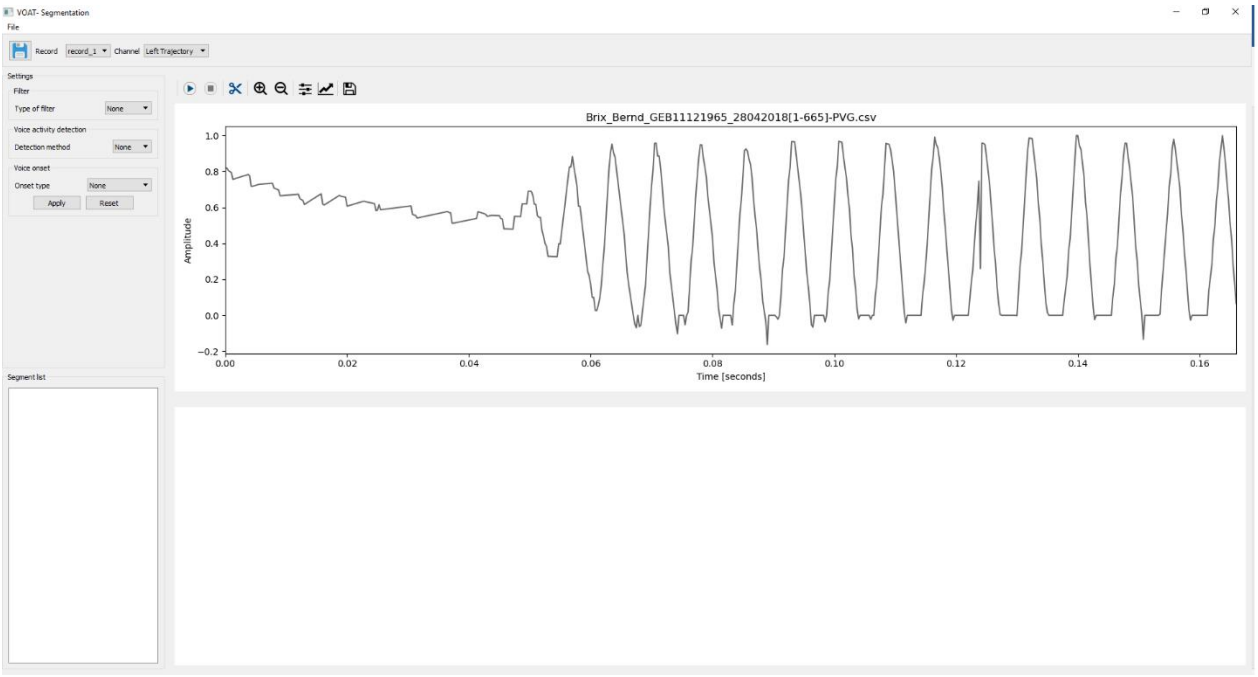


Figure 19. Trajectory at 50% showing the onset of phonation.

Similar to the GAW signal, the trajectory should also be filtered to visualize the onset properly (Figure 20).

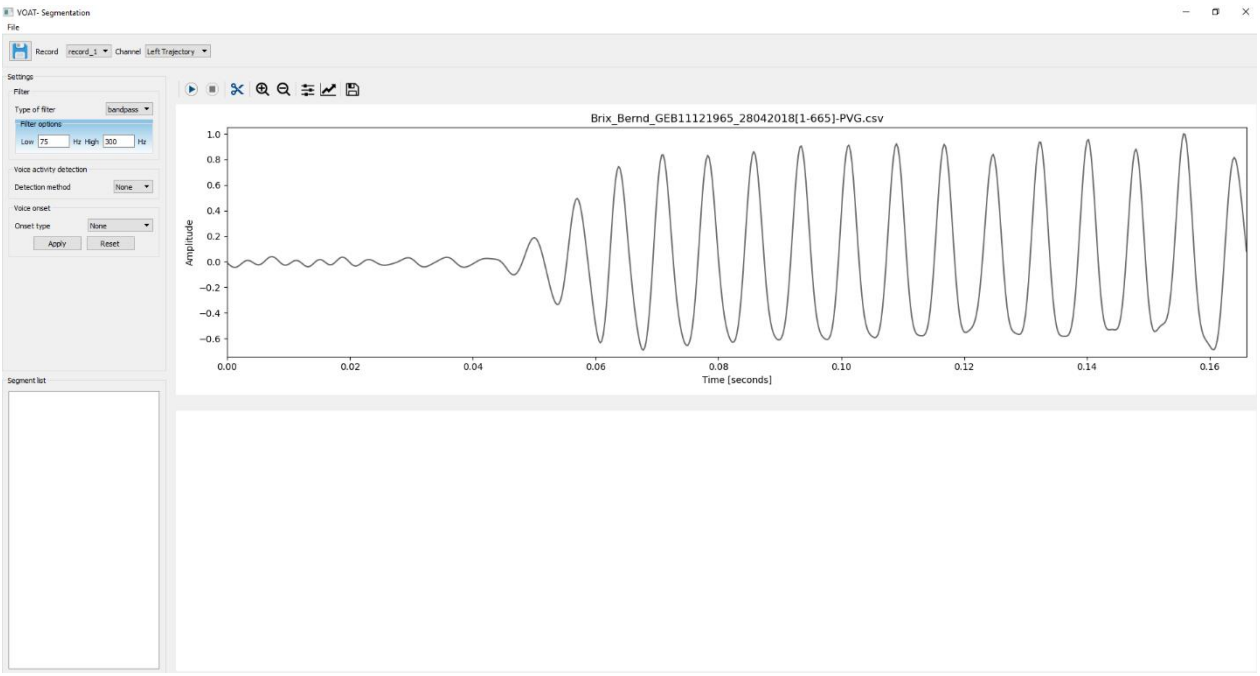


Figure 20. Filtered trajectory signal (bandpass filter with cut-off frequencies at 75Hz and 300Hz