EXPRESSIVE MEANS

Vamp plugins for *Sonic Visualiser* – a brief Reference Manual

"Expressive Means" is a set of Vamp plugins designed for the automated analysis of timing, articulation, pitch vibrato, and portamento in monophonic recordings of musical performances within the open-source, multi-platform audio visualisation software *Sonic Visualiser*.[1] Based on historical thresholds, it identifies and classifies types and relative modulation intensities of such instances – with a primary focus on bowed stringed instruments, but also potentially useful for other instruments and singing.[2] The plugins provide both semantic and "advanced" user interfaces: the former offer a quick start with default parameter settings based on sample tests; the latter allow for manual parameter adjustments, making the plugins adaptive for various recording qualities and future findings on expressive means in musical performances. However, it should be noted that the analyses returned by "Expressive Means" do not lead to direct conclusions about *psychoacoustic* effects or on specific performance styles, as the measurements naturally are subject to interpretation themselves.

"Expressive Means" makes use of Matthias Mauch's "pYIN" code and of Tilo Hähnel's "vibratoanalyse.R" logic.[3] It is open-source and distributed under the GNU General Public License (GPL), funded by the Stuttgart University of Music and the Performing Arts (Germany). You may redistribute and / or modify it under the terms of the GPL (v2.0 or later).

Please mind that "Expressive Means" only works with monophonic audio recordings in a meaningful way (i.e., the voice of interest being the only or at least the 'loudest' part within the signal). If polyphonic and / or strongly reverberating audio is used instead, the analysis may mix up the pitch data of the various voices or miss exact tone durations. However, within certain limits, polyphonic and reverberating recordings may be converted to monophonic, dry ones. To do so, please consider our short video tutorial on how to isolate single voices by means of *Audacity* and / or *Izotope RX*.[4]

Contents of this Reference Manual

1. Quick start (from installation to data export) 2
2. Expressive Means: What it does & which outputs to choose....3
3. Layer overview 7
4. Type Indicators and Inflection Indices 8
5. Semantic Parameters 9
6. "Advanced" Parameters 13
7. Hints & tricks: Using supplemental outputs for parameter and threshold refinements 19

---

[1] See https://www.sonicvisualiser.org/ (2023/05/08); https://www.vamp-plugins.org/ (2023/05/08). Also consider our introductory video tutorial to "Expressive Means", associated with this Reference Manual: https://youtu.be/rkF1VzNEKgw (2023/06/03).

[2] Find a discussion of type classifications and Inflection Indices in Frithjof Vollmer (2023), "Spectrogram-Based Analysis of Expressive Gestures: Classification Approaches for Articulation, Pitch Vibrato, and Portamento", in: Julian Caskel / Frithjof Vollmer / Thomas Wozonig (eds.), *Softwaregestützte Interpretationsforschung. Grundsätze, Desiderate und Grenzen*, Würzburg: Königshausen & Neumann 2023, pp. 197–242. The plugins provide parameter defaults for stringed, keyed, wind and percussion instruments as well as for opera singing; however, only the threshold values for string instruments are supported by studies on historical recordings so far.

[3] Matthias Mauch et al. (2014ff), "pYin. Monophonic pitch and not tracking based on a probabilistic Yin extension", see https://code.soundsoftware.ac.uk/projects/pyin (2023/05/08); Tilo Hähnel (2023), "Automatisierte Vibratoanalyse. Zur Konzeption eines Skripts für R", in: Caskel / Vollmer / Wozonig, *Softwaregestützte Interpretationsforschung* (see note 2), pp. 267–286; and https://www.tilo-haehnel.de/analysen_vibratoanalyse.html (2023/05/08).

[4] See "Separate Voices from Polyphonic Recordings via Audacity or Izotope RX", https://www.youtube.com/watch?v=Yb1jzYHytpA.

1. Quick start (from installation to data export)

Download and install the latest version of *Sonic Visualiser* (SV; the plugin requires multi-line layers provided by v4.5.2 or later) along with the latest release of "Expressive Means" via Github or from the Vamp Plugin homepage.[5] As SV, "Expressive Means" is provided for Windows (64-bit), MacOS, and Linux operating systems – look out for the package including your system's name within its file name. Unzip the package and move the files to your Vamp plugins folder:

Windows (64-bit):	*C:* → *Program Files* → *Vamp Plugins*
MacOS:	*[User / Home]* → *Library* → *Audio* → *Plug-Ins* → *Vamp*[6]
Linux:	*usr* → *local* → *lib* → *vamp*

Start SV. Load an audio file by dropping it into the pane (or via *File* → *Open…*). Add a convenient spectrogram by *Layer* → *Add Spectrogram…* (or press ⇧G).[7] To prompt "Expressive Means", click the *Transform* menu → *Analysis by Plugin Name* → *E - I* → *Expressive Means*. Choose from the various outputs:

- Note onset / note duration detection only: "Onsets" / "Durations"
  (1) Choose a clef, a signal type (Strings / … / Vocal), and the shortest note durations assumed for your recording; deselect "Normalise audio" only if you would like to capture *piano* passages → OK
  (2) Onsets will be shown as time instants, note durations as regions (spans)
  (3) Generally, you may apply different presets to various sections of your audio (e.g., use „short" note durations for fast passages, then „moderate" for subsequent cantilenas: mark region by "Select" tool [cursor] → start plugin → activate "Restrict to selection extents" → choose parameter settings → OK)

- Expressive Means analysis: "Articulation: Summary" / "Pitch Vibrato: Summary" / "Portamento: Summary"
  (1) Repeat the above. Articulation only: choose a sound quality (from 1 = clean to 6 = extremely noisy), and an assumed reverb duration. If the audio contains legato notes, leave "overlap" toggle active.
  (2) Deselect "Normalise audio" only if you would like to capture *piano* passages → OK
  (3) *Types* will be shown as labelled time instants → cf. chapter 3 ("Layer overview") for label explanation
  (4) Note that it is (by nature of Vamp plugins) *not* possible to paste an existing Time Instants layer. Instead, you may use the *Expressive Means: Onset alignment* template for subsequent alignments.

- Add *Indices* as connected points / curve: "Articulation: […] Index" output (repeat above steps)

- If needed, edit the plugin's outputs with SV's edit tools:
  (1) move instances by the cross, delete instances by the rubber, add new instances by the pencil
  (2) change values and label data by the hand → double-click segmentation line (note that values will not change automatically / dynamically when moving a Time Instant!)
  (3) overview & adjust data in a list: right-click layer → *edit layer data*

Export data via *File* → *Export Annotation Layer…* (TXT recommended). The respective output's values will be given together with their time stamps and may be imported, e.g., to a spreadsheet (via copy & paste) for statistical evaluation and / or chart generation.[8]

---

[5] https://github.com/cannam/expressive-means/releases/, or https://www.vamp-plugins.org/download.html?platform=osx&search=expressive+means&go=Go; inclusion to the "Vamp Plugin Pack" is planned: https://www.vamp-plugins.org/pack.html

[6] Before starting SV, you may most certainly need to press *control* + click the "expressive-means.dylib" file, then confirm ("open"), in order to pass MacOS's security barrier. If the *Library* folder is hidden by default, press *Command* + ⇧ + *L*.

[7] For a video tutorial on how to set up spectrograms in SV properly, see https://www.youtube.com/watch?v=EfDfhD3l6T8 (2023/05/08).

[8] Please mind that while SV generally exports layer values by using the decimal point (6.2 Hz), some non-English *Excel* versions (such as the German) operate with the decimal comma instead (6,2 Hz). In this case, all points within the data set must be replaced by commas before pasting them to *Excel* (Windows: *Strg+H*; MacOS: *Command+F*; "Replace all"), otherwise they will be misinterpreted.

2. Expressive Means: What it does & which outputs to choose

"Expressive Means" aims to find and describe soloistic performance practices with respect to the *Type Indicators* and the *Inflection Indices* of Articulation, Pitch Vibrato, and Portamento instances.

Type Indicators consist of strings of characters, each character describing a specific quality of the gestural element in question: e.g., the string *4Fw>* denotes a *continuous* (4), *fast* (F), *wide* (w), *decreasing* (>) *pitch vibrato*. The instrument-specific thresholds in place, largely reflecting recordings of the 20th century, may be adjusted by a number of parameters listed in chapter 4 in order to match other instruments' or vocals' specifications. For a discussion of the violin thresholds see note 2, pp. 200–232.

Inflection Indices (= relative modulation intensities) on the other hand describe *cross-instrumental* degrees of deviation from a steady, not-modulated tone. They are obtained by multiplicative cumulation of all metrical values and standardised on a theoretical mean of 50 points via scaling:[9] E.g., a decreasing (0.9) onset vibrato (0.8) of 7.5 Hz and 50 Cents results in an Index of I_{Vibr} = 0.8 * 7.5 * 50 * 0.9 * 0.2043 (scaling) = 55 points (for the factors in place, see chapter 4). This way, correlations and / or asynchronities of Expressive Means with other musical elements, as well as their developments over time may be detected and described (e.g., by means of "Index" charts).[10]

When prompting the plugins, the user has to choose from two basic modes of parameter settings: Either a 'semantic' mode which asks for rather *musical* qualities of the audio, such as "clef" (e.g., "treble"), "signal type" (e.g., "strings" or "vocal"), or shortest assumed "note durations" (e.g., "short (< 150 ms)"). This mode is designed for a quick start that only requires some basic knowledge about the recording in question, but it may come with significant error rates due to audio qualities that deviate considerably from those of the test samples used to determine the parameter presets.[11] Within SV's *Transform* menu, plugins working in semantic mode are named without extension (e.g., "Expressive Means: Portamento: Summary").

Alternatively, plugins carrying the appendix "(advanced)" offer direct parameter settings (i.e., referring to *metrical* qualities of the audio, as in a "Minimum onset interval" of 150 ms). Their defaults are identical to those used in the 'semantic' modes for solo violin recordings and can be adjusted at will, so high accuracy rates may be reached even for highly idiosyncratic recordings (but the parameter configuration may take some time). See chapter 5 ("Parameters") on how each choice here affects the output's decisions and which defaults were hard-coded for the 'semantic' versions of the plugins.

"Expressive Means" generally performs its analyses in three steps:

1. Determine note onsets, offsets, durations and Inter-Onset Intervals (IOI) based on changes in spectral content, pitch alteration, or raw power (in hierarchical order).
2. Determine pitch data and level development of the audio and attribute them to their respective IOIs.
3. Perform analysis and classification of the data.

Each plugin provides a number of different outputs, of which each represent a specific result of the analysis. The respective "Summary" and "Index" outputs – included to each the Articulation, Pitch Vibrato, and Portamento plugin – will be sufficient in most cases, but the complementary outputs may help with locating specific instances (e.g., the exact moment of a glide) or with optimising the parameter presets (see chapter 6).

---

[9] Obtained by multiplication of average values (see chapter 4); adjustable through the "scaling factor" parameter (see chapter 5).

[10] For an exemplary case study on Beethoven Violin Concert recordings, see pp. 233–238 in the above mentioned discussion (note 2).

[11] We used 48 samples of various string, wind, piano, percussion and vocal recordings, dating from 1902 to 2021, representing frequency bandwidths from 150–3,000 to 20–20,000 Hz and Signal-to-Noise Ratios (SNRs) from 42 to 0 dB. However, due to the unique nature of each recording, the parameter settings apt for most of these samples may lead to rather useless results when it comes to other recordings. In our experience, this particularly applies to singing, mainly due to the historically wide span of different vibrato approaches here (which makes it difficult for the plugin to find new onsets). We recommend to try the "advanced" modes in these cases. Across various recordings of solo violin and piano, on the other hand, the 'semantic' presets should work relatively stable.

In addition, "Expressive Means" provides independent and powerful onset and offset detectors, apt for a wide range of instruments and voices, as separate outputs. These detectors proved robust even with audio that show strong vibrato and portamento instances and may be useful for separate measurements of timing.

Expressive Means: onset & offset detectors

standard outputs (semantic version):	*Onsets, Durations*
further outputs available in "advanced" version:	*Offsets, Power, Spectral Rise Onset Detection Function, Pitch Onset Detection Function, Spectral Drop Offset Detection Function*

- *Onsets* are detected either based on a sudden level rise of a certain percentage of FFT bins within a certain range (see "Spectral detection range" parameters) and a certain time ("spectral rise"), *or* based on a change of pitch ("pitch change", determined on base of a moving pitch average's stabilisation below a certain threshold; see "Pitch Onset Detection Function" output), *or* based on a rise of raw power (= overall level of the audio) within a certain span. Since they feature various degrees of sensitivity, the detectors are hierarchically weighted in reverse order of their potential for false-positives ("spectral rise" beats "pitch change" beats "power rise"). Within a certain time span ("Minimum onset interval"), only one onset is accepted at a time. See "onset sensitivity" parameters, chapter 5.
- "Onsets" is given as time instant layer, labelled with the reason ("spectral rise", "pitch change", or "power rise") for onset detection. Comparable to a manual "time instant" annotation, it may be (1) copied & pasted to a "time values" layer and transformed into durations or BPMs this way, or (2) exported as data set containing the time stamp and onset cause for each instance.
- "Offsets" are determined either by a significant drop in raw power or by a level drop below a certain threshold of those frequencies that have been present at the begin of the note's sustain phase (see "offset sensitivity" parameters, chapter 5).
- "Durations" is given as regions layer (creating 'duration bars' from onset to offset of each detected note). Hovering over them shows the note onsets' time stamps. The labels themselves ("1–3") have not further meaning. Exporting this layer as data set will return time stamp, label, and duration of each instance.
- "Spectral Rise Onset Detection Function" returns the ratio of FFT bins above the „Spectral drop floor level" (default at -60 dB) that are experiencing a certain dBFS level change within a certain time. A new onset is defined if a threshold percentage of these bins first exceed, than fall below the level threshold again (see "Onset sensitivity: Noise", "Level", and "Noise time window" parameters, chapter 5).
- "Pitch Onset Detection Function" delivers the difference of the moving pitch average window and the pitch data of the same window's left edge. To define a new pitch onset, the function has to exceed a certain threshold first (see "Onset sensitivity: pitch" parameter, ch. 5) for at least the duration of the "Minimum onset interval" (ibid.) and then fall below this threshold again.
- "Spectral Drop Offset Detection Function" provides the ratio of FFT bins present at the note's sustain phase begin that are still above a certain threshold ("Spectral drop floor level"), indicating an offset as soon as this number falls below the "Spectral drop offset ratio" parameter (ch. 5).
- "Power" delivers a simple raw power curve of the audio (centred).
- "Offsets" is given as time instants layer, "Power" and all onset / offset functions as time values layers.

Expressive Means: Articulation

standard outputs (semantic version):	*Summary*, *Articulation Type*, *Index*, *Noise Type, Volume Development*
further outputs available in "advanced" version:	*Mean Noise Ratio*, *Mean Dynamics*, *Mean Tone Ratio*, *Pitch Track*

- "Summary" combines the results from all analysis steps of the Articulation plugin, given as time instants layer with multi-line labels: (1) time stamp of the IOI's onset and its duration in ms; (2) its Articulation type (as string); (3) its spectrographic noise ratio during its attack & decay phase in %; (4) its volume development (maximum and minimum level differences within sustain phase compared to phase beginning) in dB;

(5) its note duration (as ratio to IOI duration); (6) type's Articulation Index I_{Art} in Index points. Layer export will deliver a data set providing a separate column per quality.[12] Also see chapter 3 ("Label overview").

- "Type" only returns the sounds' specific articulation qualities as labels per onset in a time instants layer (e.g., "p<L" means a "plosive, increasingly loud, long sound"). As a premise, each IOI is assumed to be somehow articulated, so each onset comes with an Articulation type. Exporting this layer will deliver time stamps alongside type indicators. Also see chapter 4 (type indicators) and chapter 5 (thresholds).
- "Index" only returns Articulation Index values (cf. chapter 4), given by default as connected points in a time values layer. Export will deliver time stamps and Index values.
- "Noise type" denotes the degrees of noisiness during the IOIs' attack & decay phases. These degrees are determined based on the percentage of FFT bins within the "Spectral detection range" that exceed the "Spectral drop floor level" within a floating "Noise time window" (NTW). Example: With default parameter settings, if 87% FFT bins exceed -70 dBFS within the first half of NTW *and* 42% FFT bins do so within the second half, the IOI is classified as "affricative". "Noise type" delivers the first character within the "Articulation: Type" labels: "sonorous" (s), "fricative" (f), "plosive" (p), or "affricative" (a). Values are given as connected points in a time values layer. Thresholds may be adjusted by changing the "Impulse noise ratio" parameters in the "advanced" plugin mode (see chapter 5).
- "Volume development" denotes dBFS level developments during the tones' sustain phases (i.e., starting an adjustable 50 ms after onset). They correspond to the second character of the "Type" string: "constant" (=, development does not exceed threshold), "increasing" (<), "decreasing" (>), "in- and decreasing" (:). Values are given as duration bars in a regions layer, indicating the spans of the tones' sustain phases. Also see "Sustain phase" and "Volume development" parameters in "advanced" plugin mode.
- *Duration* (last "Type" character) is determined based on the ratio of tone duration (onset–offset) and IOI.
- The "Mean […]" outputs return averages of the corresponding values in the "Summary" output, given as global Notes layers. They may be helpful for purposeful adjustments of the "Type" and "Index" thresholds.
- "Pitch track" returns the raw data gathered from Matthias Mauch's "pYin" algorithm as time values layer.

Expressive Means: Pitch Vibrato

standard outputs (semantic version):	*Summary*, *Vibrato Type*, *Vibrato Index*, *Vibrato-Only Pitch Track*
further outputs available in "advanced" version:	*Mean Duration*, *Mean Rate*, *Mean Maximum Range*, *Pitch Track*

- "Summary" combines the results from all analysis steps of the Vibrato plugin, given as time instants layer with multi-line labels: (1) time stamp of the IOI's onset and its duration in ms; (2) its Vibrato type (as string); (3) duration of detected "vibrato elements" (see below, "Vibrato-Only Pitch Track") within the pitch data as ratio to the duration of the full *tone* (not IOI) in %; (4) mean vibrato rate (speed) in Hertz (cycles per second); (5) maximum vibrato range in Cents (100 Cents equal a semitone); (6) time stamp of the range maximum & duration of the full tone; (7) type's Vibrato Index I_{Vibr} in Index points.
- "Type" only returns the tones' specific pitch vibrato qualities as labels per onset in a time instants layer (e.g., "3Mn>" means a "moderately fast, narrow onset vibrato, decreasing in range"). Types are determined based on the position, rate, range, and range development of "vibrato elements" within their respective Inter-Onset Intervals (IOIs). Exporting this layer will deliver time stamps alongside type indicators. Also see chapter 4 (type indicators) and chapter 5 (thresholds).
- "Index" only returns the Vibrato Index values (cf. chapter 4), given by default as connected points in a time values layer. Export will deliver time stamps and index values.
- "Vibrato-Only Pitch Track" returns only those parts of the "pYin" pitch raw data that qualified as "vibrato elements" after Tilo Hähnel's method (see note 3): pitch data developments will be considered as vibrato cycles if they correlate to a certain degree ("Vibrato shape: correlation threshold") with a cosine-like development *and* if they meet the rate and range thresholds (see "Vibrato" parameters, chapter 5). Consecutive cycles will be considered as coherent vibratos per Inter-Onset Interval. Compensating for pitch data losses due to note overlaps, a "Section duration threshold" parameter interpolates these elements' characteris-

[12] As of SV v4.5.2, this feature still appears faulty, as multi-line labels aren't transformed correctly to columns. We're working on it.

tics towards the respective tone's onset and offset (which is the reason why a tone with 40% vibrato elements still may be classified as carrying a "continuous" vibrato, "4"). A "Note segmentation" parameter offers four options to deal with glides and pitch steps (from "None" to "Without glides and Segmented", the latter ruling out potential glides *and* pitch steps as being part of a vibrato; cf. chapter 5).

- The "Mean […]" outputs return averages of the corresponding values in the "Summary" output, given as global Notes layers. They may be helpful for purposeful adjustments of the "Type" and "Index" thresholds.
- "Pitch track" returns the raw data gathered from Matthias Mauch's "pYin" algorithm as time values layer.

Expressive Means: Portamento

standard outputs (semantic version):	*Summary*, *Portamento Type*, *Portamento Index*, *Significant Points*
further outputs available in "advanced" version:	*Glide Direction, Glide Dynamic, Glide Link, Glide-Only Pitch Track, Mean Range, Mean Duration, Mean Dynamics, Pitch Track*

- "Summary" combines the results from all analysis steps of the Portamento plugin, given as time instants layer with multi-line labels: (1) time stamp of the IOI's onset and its duration in ms; (2) its Portamento type (as string); (3) start pitch (F_0) / end pitch in Hertz (Hz) & range in Cents (c); (4) portamento's start & end time stamp, portamento duration in Milliseconds (ms); (5) maximum dB deviation of the glides' maximum / minimum levels from their start & end pitches; (6) type's Portamento Index I_{Port} in Index points. Also see chapter 3 ("Label overview").
- "Type" only returns the tones' specific pitch vibrato qualities as labels per onset in a time instants layer (e.g., "/3Sm+" means an "ascending, small, moderately long, loud target tone portamento"). Exporting this layer will deliver time stamps alongside type indicators. See chapter 4 ("type indicators") as well as chapter 5 on how to manipulate the threshold values in place.
- "Index" only returns Portamento Index values (cf. chapter 4), given by default as connected points in a time values layer. Export will deliver time stamps and index values.
- "Significant Points" shows the glides' onsets and offsets in a time values layer. Glides are detected as soon as the pitch starts to constantly move forward in one direction for at least a "Minimum duration" *and* if within this time, the difference of two consecutive hops' pitch track values exceeds the "Minimum hop difference" (see chapter 5, "Glide detection" parameters). Glide ends as soon as the "Pitch Onset Detection Function" (see onset outputs) falls below the "Minimum hop difference" again. If within the span of the "minimum duration" any hop's pitch value deviates more than "Maximum hop difference" from the preceding hop, movements are ruled out as being a glide, for they are considered as rather being a pitch step. Moreover, movements' spans (pitch difference between start and end) have to meet the "Minimum pitch range" to be considered as being glides.
- "Glide Direction" returns the directions of glides by subtracting their end pitch values from their start pitch values. If the difference is positive, the glide is descending; if it is negative, it is ascending. The output corresponds to the first character within the "Portamento: Type" labels: "ascending" (/) or "descending" (\).
- "Glide Dynamic" denotes the glides' dB developments by comparing their respective start and end tone dBFS levels with the glides' level maxima and minima. The output corresponds to the last character within the "Portamento: Type" labels: "louder" (+, meaning higher levels of the glide than of both its start and end pitch), "stable" (=, meaning glide levels in between the start and end pitch levels), or "quieter" (-, meaning lower glide levels than both start and end pitch levels).
- "Glide Link" gives the glides' types of connection to their start and end notes, corresponding to the second character within the "Portamento: Type" labels: "Starting" (1, meaning a glide beginning at start note pitch and leaving a gap towards target note), "Interconnecting" (2, meaning a nearly uninterrupted glide), "Targeting" (3, pitch gap between start note and glide). "Links" are detected by comparing the glides' pitch data to the pitch median of their respective start and end tones.
- "Glide-Only Pitch Track" extracts the pitch data of glides from the "pYin" raw data.
- The "Mean […]" outputs return averages of the corresponding values in the "Summary" output, given as global Notes layers. They may be helpful for purposeful adjustments of the "Type" and "Index" thresholds.
- "Pitch Track" returns the pitch track raw data gathered from the "pYin" algorithm as time values layer.

3. Layer overview

Figure 3.1 shows an exemplary *Sonic Visualiser* pane with layers of three different "Expressive Means: Pitch Vibrato" outputs: *Summary*, *Vibrato-Only Pitch Track*, and *Vibrato Index* (the latter determining y-axis units). Designs of "Articulation" and "Portamento" plugins feature the same basic structure. For their specific values (particularly regarding the "Summary" labels – i.e., "type codes + type values"), see previous chapter.

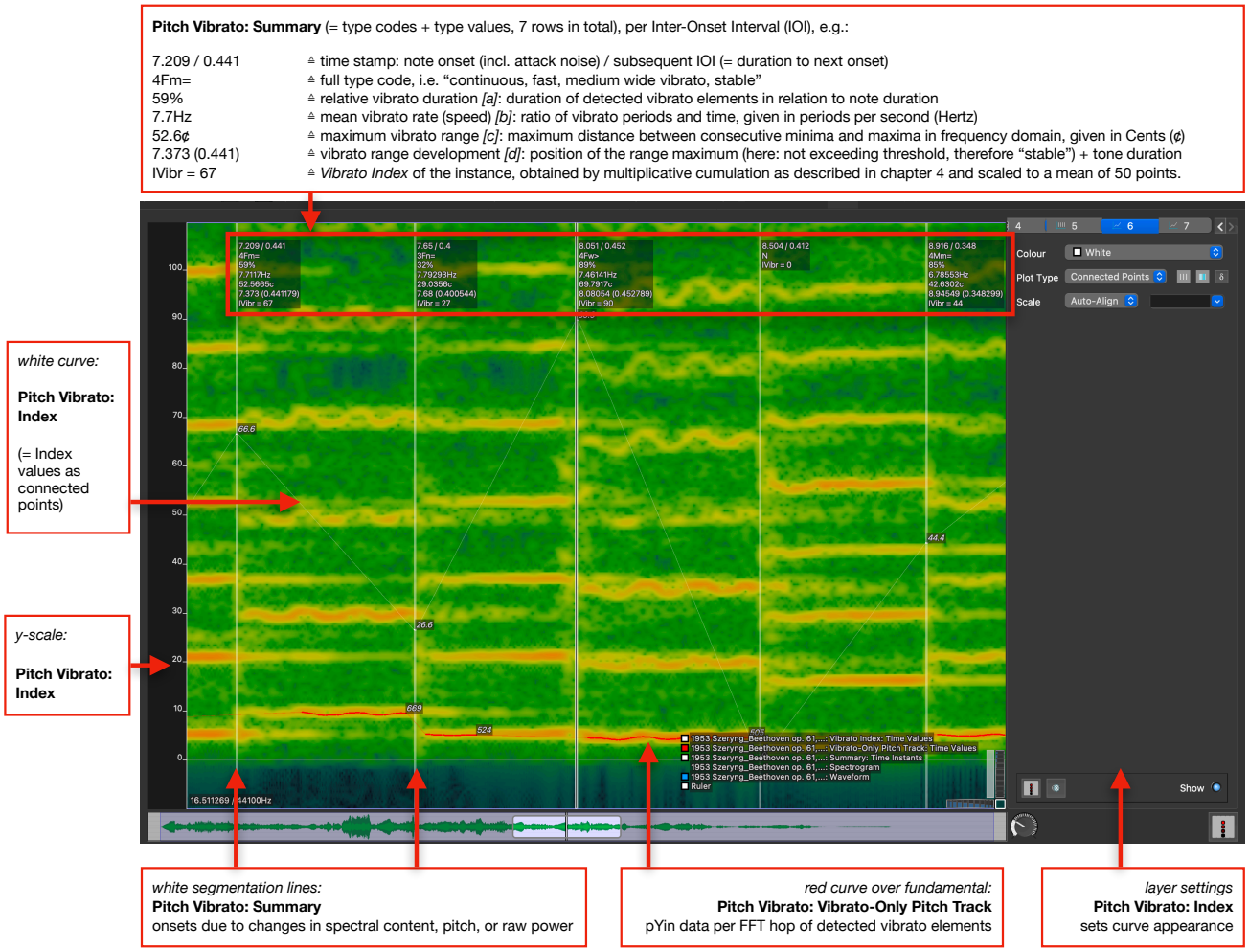

Figure 3.1: Exemplary pane view with three "Expressive Means: Pitch Vibrato" output layers over a 1953 Szeryng recording of Beethoven’s Violin Concerto

4. Type Indicators and Inflection Indices

Figures 4.1 to 4.3 show classifications of expressive means in spectrogram view, illustrating Type codes and Index calculation as returned by "Expressive Means" (based on pp. 208–213 of the discussion in note 2; please mind the updated Index calculation method as well as the updated scaling factors).

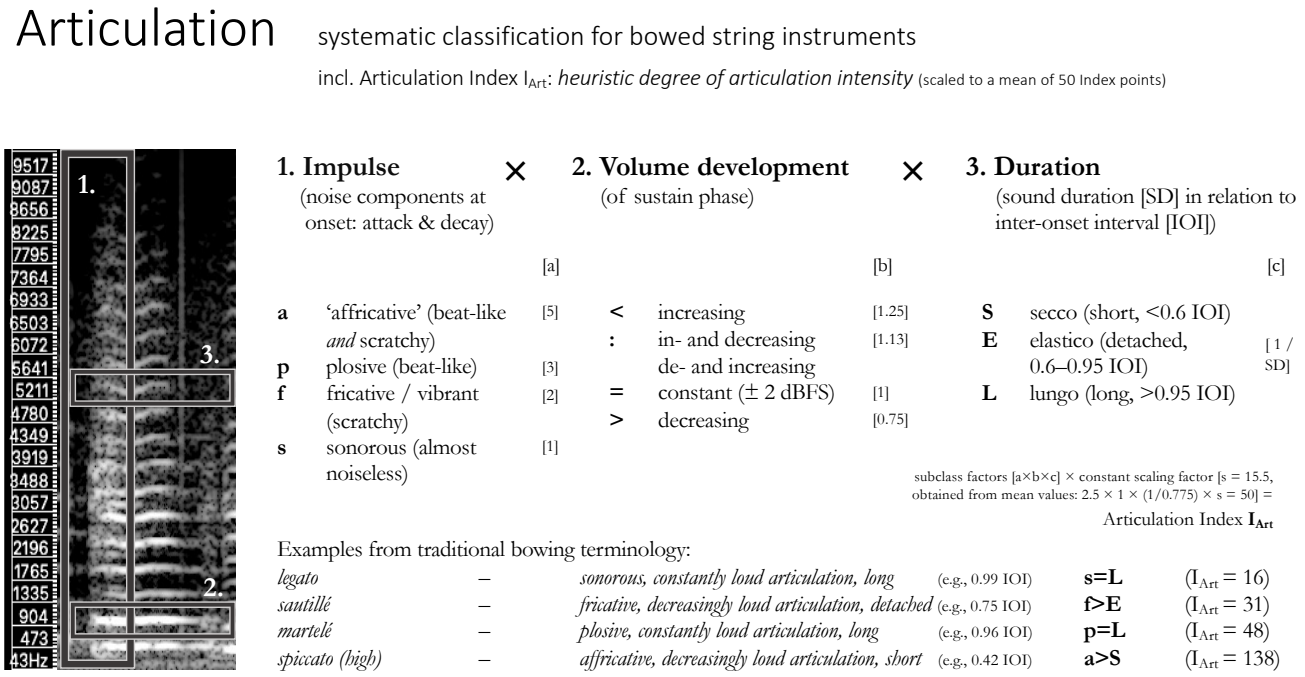

Figure 4.1: Spectrogram-based classification for Articulation in bowed string instrument performance

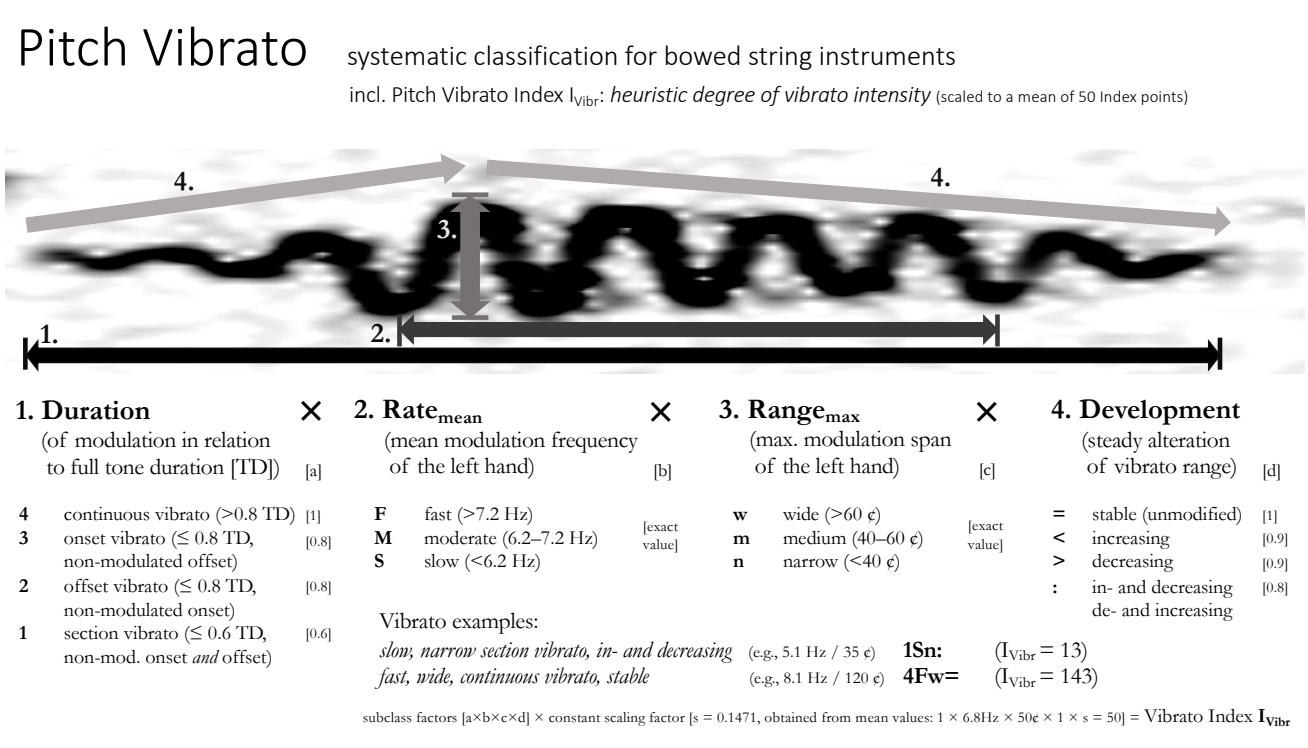

Figure 4.2: Spectrogram-based classification for Pitch Vibrato in bowed string instrument performance[13]

[13] See chapter 5.5 (overview on 'semantic' parameter settings) and note 14 on "Type" thresholds for other instruments and for vocals.

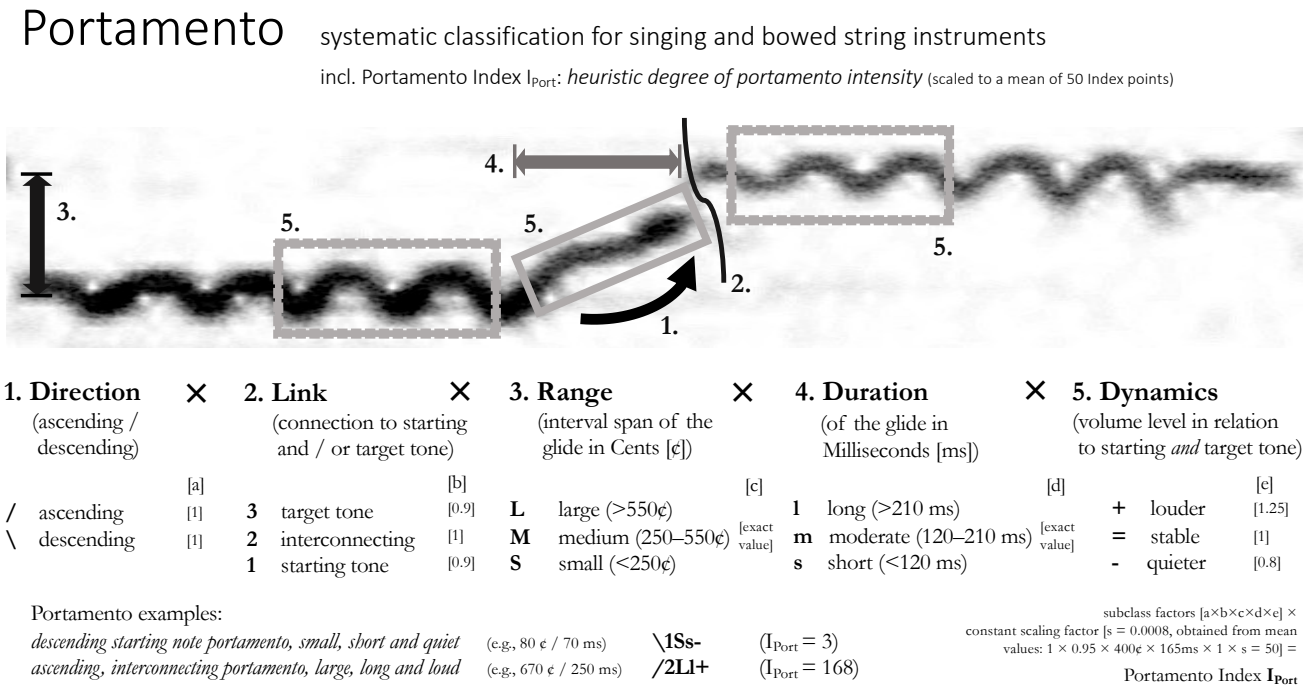

Figure 4.3: Spectrogram-based classification for Portamento in bowed string instrument and vocal performance[14]

5. Semantic Parameters[15]

! Advice: You may apply different presets to various sections of your audio (e.g., use „short“ note durations for fast passages, then „moderate“ for subsequent cantilenas: mark region by SV's "Select" tool [the 'cursor'] → start plugin → activate "Restrict to selection extents" → choose parameter settings → OK)

5.1) Parameter names, functions, and which "advanced" (= metric) parameters they affect

Parameter name	**Choose from…** (see below for specifications)	**Available in following plugins…**
Clef	Treble, Alto, Bass	[Onsets], [Articulation], [Pitch Vibrato], [Portamento]
	• choose according to which comes closest to the instrument's / voice's clef in your audio • affects the following parameters: Spectral detection range, Volume development	
Signal Type	Bowed Strings, Vocal (Classical), Vocal (Jazz & Pop), Piano / Plugged Strings, Piano / Plugged Strings (historical), Wind / Organ, Percussion	[Onsets], [Articulation]
	Instrumental, Vocal (Classical), Vocal (Jazz & Pop)	[Pitch Vibrato], [Portamento]
	• choose according to your instrument's / voice's general family • affects: Pitch toggle, all Onset sensitivity parameters (Pitch, Noise, Level, Noise time, Power rise thresholds), Sustain phase begin	

[14] For vocals, all "Type" thresholds are currently set solely on base of a 1902 Caruso recording ("Vesti la giubba", Gramophone matr. no. 2875 R) and – regarding vibrato rate – on Tilo Hähnel (2020), "Über die Quantifizierung des Heldentenors. Vibrato, Ornamentik, Glissando, Tempo und Register in akustischen Tonaufnahmen zwischen 1900 und 1930", in: *Musik in Konfrontation und Vermittlung: Beiträge zur Jahrestagung der Gesellschaft für Musikforschung 2018 in Osnabrück*, ed. by Dietrich Helms et al., Osnabrück: epOs, pp. 369–386, Table 5. Caruso, as it becomes clear from Hähnel's study, may be regarded as representing a good average when it comes to singers' vibrato extents between 1900 and 1930. Because of the lack of any further detailed studies on the topic so far, the thresholds for all other instruments currently are set to those of violin performance (even though it seemed from our samples as if wind instruments' vibratos and portamentos, for instance, tend to be generally smaller as opposed to those in violin performance).

[15] When choosing the plugins in SV's "Transform" menu, semantic plugins are identified by the missing appendix "advanced" in name.

Parameter name	**Choose from…** (see below for specifications)	**Available in following plugins…**
Note durations	Short (< 150 ms), Moderate (150–300 ms), Long (> 300 ms) • choose according to the (assumed) **shortest** note durations within your audio • affects the following parameters: Minimum onset interval, Moving pitch average window	[Onsets], [Articulation], [Pitch Vibrato], [Portamento]
Sound quality	1 [= 'clean'], 2, 3, 4, 5, 6 [= 'extremely noisy'] • estimate according to the sound quality / amount of technical artefacts within your audio • affects: Impulse noise parameters ("Plosive" and "Fricative" thresholds)	[Onsets], [Articulation]
Reverb duration	Small studio (<150 ms), Large studio (150–600 ms), Concert hall (600–1500 ms), Church (> 1500 ms) • estimate according to the sound quality / amount of technical artefacts within your audio • **factorises** the Impulse noise parameters: (= makes "Plosive" and "Fricative" thresholds more robust against overlaps / ringing tones due to reverb) Small studio: *1 Large studio: *1.5 Concert hall: *2.25 Church: *3.375	[Onsets], [Articulation]
Overlap compensation	On, Off • only switch off if no "lungo" note durations ('legato' notes, > 0.95 IOI) are expected • **further factorises** the "Fricative" parameter only: (= makes "Fricative" threshold more robust against overlapping notes) On: *1.6 Off: *1	[Articulation]
Normalise audio	On, Off • normalises signal to an amplitude maximum close to 1.0 (= 0 dBFS; prevents too 'quiet' recordings from failing pitch track and noise analysis) • only toggle off if your recording is extracted from an already normalised audio, or if purposefully is a very quiet one (i.e., the intend is to analyse a *piano* passage)	[Onsets], [Articulation], [Pitch Vibrato], [Portamento]
pYin: non-standard precise timing	On, Off • aligns pitch analysis to precise time raster. If switched off, pitches will appear a few hops delayed. • switching on will deliver exact time-pitch alignments, but will take about five times longer for calculation • we recommend to leave this option off, as the delay usually amounts to a few milliseconds (10–20ms) only. It will save time. Switch on if you're looking for the exact beginning of, e.g., a vibrato.	[Onsets], [Articulation], [Pitch Vibrato], [Portamento]

5.2) Numeric values prompted by semantic parameters (see chapter 6, "'Advanced' parameters", for parameter specifications):

"Advanced" parameter name	**Onsets / Durations**	**Articulation**	**Pitch Vibrato**	**Portamento**
Normalise audio [toggle]	[on]			
pYin: Yin threshold distribution	Beta (mean 0.15)			
pYin: Suppress low amplitude pitch estimates	0.1			
Spectral detection range minimum frequency	[Clef: Treble]: [Clef: Alto]: [Clef: Bass]:	4,000 Hz 3,000 Hz 2,000 Hz		
Spectral detection range maximum frequency	[Clef: Treble]: [Clef: Alto]: [Clef: Bass]:	100 Hz 100 Hz 50 Hz		

"Advanced" parameter name	Onsets / Durations	Articulation	Pitch Vibrato	Portamento
Minimum onset interval	[Note durations: Long (> 300 ms)]: 280 ms [Note durations: Moderate (150–300 ms)]: 150 ms [Note durations: Short (< 150 ms)]: 50 ms			
Moving pitch average window	[Note durations: Long (> 300 ms)]: 200 ms [Note durations: Moderate (150–300 ms)]: 150 ms [Note durations: Short (< 150 ms)]: 50 ms			
Onset: Use Pitch	[Signal type: Piano / Pl. Str. (both) / Percussion]: [off] [Signal type: (other)]: [on]		[Signal type: Piano / Pl. Str. (both) / Percussion]: [off] [Signal type: (other)]: [on]	
Onset sensitivity: Pitch	[Signal type: Bowed Strings]: 15 Cents [Signal type: Vocal (both)]: 25 Cents [Signal type: Wind / Organ]: 10 Cents		[Signal type: Instrumental]: 15 Cents [Signal type: Vocal (both)]: 25 Cents	
Onset sensitivity: Noise	[Signal type: Bowed Strings]: 17 % [Signal type: Vocal (Classical)]: 35 % [Signal type: Vocal (Jazz & Pop)]: 30 % [Signal type: Piano / Pl. Strings]: 5 % [Signal type: Piano / Pl. Str. (hist.)]: 8 % [Signal type: Wind / Organ]: 6 % [Signal type: Percussion]: 4 %		[Signal type: Instrumental]: 17 % [Signal type: Vocal (Classical)]: 35 % [Signal type: Vocal (Jazz & Pop)]: 30 %	
Onset sensitivity: Level	[Signal type: Vocal (both)]: 8 dB [Signal type: Piano / Pl. Strings]: 9 dB [Signal type: Piano / Pl. Str. (hist.)]: 10 dB [Signal type: (other)]: 8 dB		[Signal type: Instrumental]: 8 dB [Signal type: Vocal (both)]: 8 dB	
Onset sensitivity: Noise time window	[Signal type: Percussion]: 50 ms [Signal type: Piano / Pl. Str. (hist.)]: 60 ms [Signal type: (other)]: 100 ms		100 ms	
Onset sensitivity: Power rise threshold	[Signal type: Strings]: 6 dB [Signal type: Vocal (Classical)]: 10 dB [Signal type: Vocal (Jazz & Pop)]: 12 dB [Signal type: Piano / Pl. Str. (both)]: 15 dB [Signal type: Wind / Organ]: 12 dB [Signal type: Percussion]: 80 dB		[Signal type: Instrumental]: 6 dB [Signal type: Vocal (Classical)]: 10 dB [Signal type: Vocal (Jazz & Pop)]: 12 dB	
Sustain phase begin threshold	[Clef: Treble]: 50 ms [Clef: Alto]: 60 ms [Clef: Bass]: 70 ms			
Offset sensitivity: Power drop thr.	[Signal type: Piano / Pl. Str. (both)]: 15 dB [Signal type: (other)]: 12 dB			
Offset sensitivity: Spectral drop floor level	[Signal type: Piano / Pl. Str. (both) / Percussion]: -70 dB [Signal type: (other)]: -60 dB			
Offset sensitivity: Spectral drop offset ratio	[Signal type: Strings / Wind / Org.]: 40 % [Signal type: Vocal (Classical)]: 30 % [Signal type: (other)]: 20 %			
[processing:] Audio frames per block	2048			
[processing:] Window increment	256			
Impulse noise ratio: plosive	–	[Sound quality: 1]: 22 % [Sound quality: 2]: 26 % [Sound quality: 3]: 32 % [Sound quality: 4]: 36 % [Sound quality: 5]: 53 % [Sound quality: 6]: 80 %	–	–

"Advanced" parameter name	Onsets / Durations	Articulation	Pitch Vibrato	Portamento
Impulse noise ratio: fricative	–	[Sound quality: 1]: 11 % [Sound quality: 2]: 13 % [Sound quality: 3]: 16 % [Sound quality: 4]: 27 % [Sound quality: 5]: 47 % [Sound quality: 6]: 80 %	–	–
Volume development threshold	–	[Clef: Treble]: 2 dB [Clef: Alto]: 2 dB [Clef: Bass]: 3 dB	–	–
Reverb duration factor	–	[Reverb: Small st.]: 1.0 [Reverb: Large st.]: 1.5 [Reverb: Concert h.]: 2.25 [Reverb: Church]: 3.375	–	–
Overlap compensation factor	–	[Overlap c.: On]: 1.6 [Overlap c.: Off]: 1.0	–	–
Vibrato rate: Minimum	–	–	[Signal: Instr.]: 4.2 [Signal: Vocals]: 4.0	–
Vibrato rate: Maximum	–	–	[Signal: Instr.]: 9.2 [Signal: Vocals]: 10.0	–
Rate threshold: moderate	–	–	[Signal: Instr.]: 6.2 [Signal: Vocal Cl.]: 6.0 [Signal: Vocal Pop]: 5.0	–
Rate threshold: fast	–	–	[Signal: Instr.]: 7.2 [Signal: Vocal Cl.]: 7.0 [Signal: Vocal Pop]: 6.0	–
Vibrato range: Minimum	–	–	[Signal: Instr]: 20 ¢ [Signal: Vocal Cl.]: 30 ¢ [Signal: Vocal Pop]: 20 ¢	–
Vibrato range: Maximum	–	–	[Signal: Instr.]: 200 ¢ [Signal: Vocal Cl.]: 600 ¢ [Signal: Vocal Pop]: 400 ¢	–
Range threshold: medium	–	–	[Signal: Instr.]: 40 ¢ [Signal: Vocal Cl.]: 120 ¢ [Signal: Vocal Pop]: 80 ¢	–
Range threshold: wide	–	–	[Signal: Instr.]: 60 ¢ [Signal: Vocal Cl.]: 220 ¢ [Signal: Vocal Pop]: 160 ¢	–
Section duration threshold	–	–	150 ms	–
Development threshold	–	–	[Signal: Instr.]: 10 ¢ [Signal: Vocal Cl.]: 40 ¢ [Signal: Vocal Pop]: 20 ¢	–
Vibrato shape: Correlation threshold	–	–	[Signal: Instr.]: 0.2 [Signal: Vocal Cl.]: 0.65 [Signal: Vocal Pop]: 0.50	–
Smoothing window length	–	–	70 ms	–
Note segmentation	–	–	"Without glides & segmented"	–
Glide detection: Minimum pitch difference	–	–	–	[Signal: Instr.]: 60 ¢ [Signal: Vocal]: 60 ¢

"Advanced" parameter name	Onsets / Durations	Articulation	Pitch Vibrato	Portamento
Glide detection: Minimum hop difference	–	–	–	3 ¢
Glide detection: Maximum hop difference	–	–	–	[Signal: Instr.]: 50 ¢ [Signal: Vocal]: 70 ¢
Glide detection: Minimum duration	–	–	–	[Signal: Instr.]: 50 ms [Signal: Vocal]: 20 ms
Glide detection threshold: Onset Proximity	–	–	–	[Signal: Instr.]: 1000ms [Signal: Vocal]: 2000ms
Link threshold	–	–	–	[Signal: Instr.]: 70 ¢ [Signal: Vocal]: 100 ¢
Range threshold: medium	–	–	–	250 ¢
Range threshold: large	–	–	–	550 ¢
Duration threshold: moderate	–	–	–	120 ms
Duration threshold: long	–	–	–	210 ms
Dynamics threshold	–	–	–	1 dB
Index scaling factor	–	15.5	0.1471	0.0008

6. "Advanced" Parameters

6.1) Parameters at work in all "advanced" Expressive Means outputs:[16]

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Normalise audio [toggle]	[on]	normalises signal to an amplitude maximum close to 1.0 (= 0 dBFS)	only toggle off if your recording is extracted from an already normalised audio, or if purposefully is a very quiet one
pYin: Yin threshold distribution	Beta (mean 0.15)	settles the threshold for fundamental frequencies (f0) considered as pitch: lower numbers narrow the number of pitch candidates per frame. Pitch estimation is based on analysis of overtones and on a probabilistic course estimation.	use higher numbers (e.g., Beta 0.20) if the plugin doesn't find instances ("NN") where there should some, or in case of significant gaps in the pitch track curve. Use the "single value" options, e.g., in case of continuously large pitch intervals. Higher numbers are more likely to affect voice confusion.

[16] "Portamento" plugin misses the "Sustain phase" and "Offset" parameters. For further details on the pYin parameters, see Matthias Mauch / Simon Dixon (2014), "pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions", in: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014).*

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
pYin: Suppress low amplitude pitch estimates	0.1	settles the threshold for low-power ('quieter') frequencies	use higher numbers if the plugin finds too many instances or if the pitch track seemingly arbitrarily jumps around if this could be due to quieter frequencies. Keep in mind that raising may also cut off tone ends of the main voice.
pYin: Use non-standard precise YIN timing (slow)	[off]	aligns pitch analysis to precise time raster. If switched off, pitches will appear a few hops delayed	switching on will deliver exact time-pitch alignments, but will take about five times longer for calculation
Spectral detection range minimum frequency	100 Hz	settles the lower end of frequency range considered for analysis	lower end of your audio's voice register minus some margin (e.g., violin register starts at a pitch of c. 196 Hz)
Spectral detection range maximum frequency	4,000 Hz	settles the higher end of frequency range considered for analysis	upper end of your audio's voice register plus some margin (e.g., violin register roughly ends at a pitch of c. 3,136 Hz)
Minimum onset interval	100 ms	settles the duration of pausing onset detection after an onset has been found	the shortest note durations within your audio: fast passages require short onset intervals, slow passages allow for long intervals (and, therefore, less false-positives)
Moving pitch average window	150 ms	determines length and pitch difference threshold of the floating pitch change detection window. New onsets are detected if the pitch difference of the window's left edge and the window's average first exceed, then fall below "Pitch" threshold again.	the note durations and the amount of pitch vibrato within your audio: short note durations require short windows, slow and wide vibratos long windows (i.e., most violin performances may benefit from short, most vocal performances from long sizes)
Onsets: Use pitch	[on]	activates pitch onset detection	switch off if your audio represents highly impulsive onsets (such as for piano, plugged strings) or if it has no pitch at all (percussion)
Onset sensitivity: Pitch	15 Cents	threshold that first has to be exceeded, then fallen below again to define a new pitch onset	the amount of pitch vibrato: wide vibratos require higher thresholds; however, onset detection is more precise with low values
Onset sensitivity: Noise	17 %	determines the ratio of FFT bins in between detection range that trigger a spectral onset detection by experiencing a level rise above threshold within a certain time (see "Onset sensitivity: Level" and "Noise time window")	the presence of technical artefacts (surface noise etc) within your audio: the more noise, the higher thresholds are needed in order not to prompt false onsets – same may apply to high amounts of pitch vibrato and portamento
Onset sensitivity: Level	8 dB	level rise threshold that has to be exceeded to prompt a "Spectral rise" onset	the level dynamics within your audio: large differences between loud and quiet passages allow for high thresholds, small differences (e.g., in strongly compressed audio) require low thresholds
Onset sensitivity: Noise time window	100 ms	settles the time window for FFT bin level change detections	the more noise, pitch vibrato, and portamento, the shorter the threshold should be
Onset sensitivity: Power rise threshold	6 dB	settles the audio's raw power (overall level) rise threshold that has to be exceeded within "Moving pitch average window" to prompt a "Power rise" onset	the level dynamics within your audio (see above, "Onset Sensitivity: Level")

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Sustain phase begin threshold	60 ms	determines an estimated mean of all IOIs attack and decay phases (i.e., skips their noisy onset parts for level development considerations)	attack and noisiness of the performance (noisy onsets require higher values), register and type of instrument in your audio (i.e., low voices usually reach sustain phases later than high voices; strings usually reach sustain phases later than wind instruments or vocals). Consider an adjustment, e.g., if level development almost always is returned as being ">" (decreasing)
Offset sensitivity: Power drop threshold	12 dB	settles the audio's raw power (overall level) drop threshold that has to be exceeded within "Moving pitch average window" to prompt a "Power drop" offset	the level dynamics within your audio (see above, "Onset Sensitivity: Level")
Offset sensitivity: Spectral drop floor level	-60 dB	settles the level floor threshold for spectral onset and offset detection	the level dynamics within your audio (see above, "Onset Sensitivity: Level")
Offset sensitivity: Spectral drop offset ratio	40 %	defines the ratio of FFT bins at sustain phase begin with levels above the „Spectral drop floor level" that has to fall below this level to define a new offset	choose lower values if your audio represents either singers, percussion, or otherwise highly impulsive instruments (piano, plugged strings)
[processing:] Audio frames per block	2048	affects window sizes for the Fast Fourier Transformation (FFT) = 'resolution' of the plugin's time-frequency discrimination ability	the sampling rate of your audio: 2048 (as a power of 2) is recommended for 44.1 kHz; 1024 may be more suited for a rate of 22.05 kHz. Lower values generally prompt a better resolution in time, higher values a better resolution in frequency domain.
[processing:] Window increment	256	settles the window overlapping property: 256 in context of 2048 audio frames per block equates to 1 - (256 / 2048 = 0.125) = 0.875 = 87.5 % overlap	lower values generally prompt better FFT resolutions in time by keeping frequency resolution, but will take significantly more time for calculation. Default value proved to provide a good balance. We strongly advice to leave the default as it is, for it has been the base for the thresholds given

6.2) Additional parameters in "Expressive Means (advanced): Articulation" only:

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Impulse noise ratio: plosive	26 %	settles the threshold of FFT bins experiencing a significant level change necessary to be exceeded to return a "plosive" onset indication	the presence of technical artefacts (surface noise etc) within your audio: the more noise, the higher thresholds are needed in order not to prompt exaggerated impulse noise indications
Impulse noise ratio: fricative	13 %	settles the threshold of FFT bins experiencing a significant level change necessary to be exceeded to return a "fricative" onset indication	ibid.
Volume development threshold	2 dB	settles the threshold of minimum dB development during tone's sustain phase necessary to be exceeded to return a "volume development" indication ("increasing" / "decreasing")	the level dynamics within your audio: large differences between loud and quiet passages allow for high thresholds, small differences (e.g., in strongly compressed audio) require low thresholds

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Reverb duration factor	1.5	scales "Impulse noise ratios" by the given factor in order to compensate for reverb / echo (both of which cause more FFT bins to experience a significant change due to overlapping sounds)	the amount of reverb assumed for or measured in your audio. As a rough reference: <150 ms: factor *1 150–600 ms: factor *1.5 600–1,500 ms: factor *2.25 >1,500 ms: factor *3.375
Overlap compensation factor	1.6	scales "Impulse noise ratios" by the given factor in order to compensate for potentially overlapping tones (which cause more FFT bins to experience a significant change)	the predominant tone "Duration" type within your audio: long tones ("lungo", i.e. tone durations of >0.95 IOI duration) benefit from an active overlap compensation (otherwise their onsets may falsely be labelled as of "plosive" or even "affricative")
Index scaling factor	15.5	scales the results from multiplicative cumulation of all Articulation aspects (Noise * Volume development * Duration) to a standardised scale that features a mean of 50 Index points	either if you adjusted the hard-coded factors per Articulation aspect, or in case you prefer another scale range

6.3) Additional parameters in "Expressive Means (advanced): Pitch Vibrato" only:

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Vibrato rate: Minimum	4 Hz	settles lower rate threshold for oscillations to be considered as "vibrato elements"	research that suggests lower or higher vibrato rates in vocal / instrumental performance, or in case of 'wow and flutter' ("Gleichlaufschwankungen") that causes artificial pitch vibratos
Vibrato rate: Maximum	10 Hz	settles upper rate threshold for oscillations to be considered as "vibrato elements"	ibid.
Rate threshold: moderate	6.2 Hz	lower boundary for "vibrato elements" to be classified as of a "moderate" rate / speed (elements below that will be considered as being "slow")	research that suggests other thresholds (= terciles) than suggested
Rate threshold: fast	7.2 Hz	lower boundary for "vibrato elements" to be classified as of "fast" rate / speed (elements below that will be considered as being "moderate")	ibid.
Vibrato range: Minimum	20 Cents	settles lower range threshold for oscillations to be considered as "vibrato elements"	research that suggests lower or higher vibrato ranges in vocal / instrumental performance, or in case of 'wow and flutter' ("Gleichlaufschwankungen") that causes artificial pitch vibratos
Vibrato range: Maximum	200 Cents	settles upper range threshold for oscillations to be considered as "vibrato elements"	ibid.
Range threshold: medium	40 Cents	lower boundary for "vibrato elements" to be classified as of "medium" range (elements below that will be considered as being "narrow")	research that suggests other thresholds (= terciles) than suggested
Range threshold: wide	60 Cents	lower boundary for "vibrato elements" to be classified as of "wide" rate / speed (elements below that will be considered as being of "medium" range)	ibid.

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Section duration threshold	200 ms	compensates for potential pitch track overlaps by interpolating "vibrato elements'" characteristics towards the associated tones' onsets and offsets	the tone durations within your audio: longer durations may benefit from a higher, shorter durations from a lower threshold in order for the analysis to be more accurate (however, the default of 200 ms will work for most instances as it corresponds to the duration of one vibrato cycle within a rate of 5 Hz)
Development threshold	10 Cents	settles the minimum range development over consecutive "vibrato elements" necessary for them to be considered as of "increasing" or "decreasing" quality	research that suggests lower or higher vibrato ranges in vocal / instrumental performance, or in case of 'wow and flutter' ("Gleichlaufschwankungen") that causes artificial pitch vibratos
Vibrato shape: Correlation threshold	0.2	determines the correlation of oscillations with a cosine curve in order to be considered as "vibrato elements"	the irregularity and flexibility of vibrato instances within your audio. Performances that feature highly varying vibrato will profit from lower thresholds.
Index scaling factor	0.1471	scales the results from multiplicative cumulation of all Vibrato aspects (Duration * Rate * Max. Range * Development) to a standardised scale that features a mean of 50 Index points	either if you adjusted the thresholds or the hard-coded factors per Vibrato aspect, or in case you prefer another scale range
Smoothing window length	70 ms	settles the window for pitch track smoothing that precedes vibrato rate (speed) analysis (cf. Hähnel in note 3, p. 272)	the mean rate (speed) of vibrato instances within your audio: very fast rates may benefit from shorter smoothing windows
Note segmentation	Without Glides and Segmented	Removes "vibrato elements" that cross IOIs, or may be part of a glide, or both. Options: (1) None (2) Segmented (3) Without Glides (4) Without Glides and Segmented	the performance tradition represented in your recording: e.g., most singers tend to employ vibrato and portamento together, which is why the "Segmented" option (only) would be more suited. However, "Segmented" and "Without Glides" analyses may miss significant parts of onset-crossing vibratos, while "None" may lead to considerable distortion of vibrato rates (speeds) and, in particular, ranges

6.4) Additional parameters in "Expressive Means (advanced): Portamento" only:

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Glide detection: Minimum pitch difference	60 Cents	sets the minimum range in Cents for a glide to qualify as portamento	(1) the amount of reverberation in your audio: much reverb may cause simple pitch changes to be identified as glides due to the overlap of pitches (2) the type of instrument and performance tradition represented in your recording: instrumentalists may by nature employ shorter glides (due to technical limitations) than singers
Glide detection: Minimum hop difference	10 Cents	sets the minimum pitch increase / decrease between consecutive FFT hops during the "Minimum duration" of a glide to qualify as portamento	serves to rule out minimal as well as solely technically caused glides: change according to the quality of your pitch track data

Parameter name	**default value** in "advanced" mode	**…what it does:**	**…change it according to:**
Glide detection: Maximum hop difference	50 Cents	sets the maximum pitch increase / decrease between consecutive FFT hops during the "Minimum duration" of a glide to qualify as portamento	serves to disqualify simple pitch steps as portamentos: reduce according to the amount of reverberation in your audio, but keep in mind that this may rule out very fast portamentos as well
Glide detection: Minimum duration	70 ms	sets the minimum duration of a glide movement in order to qualify as portamento	(1) the pitch vibrato rates (speeds): slow rates may cause vibratos to be identified as portamentos (2) overlapping pitches that may cause simple pitch changes to be considered as glides In both cases, increase the threshold to solve the problem, but keep in mind that this may rule out very fast portamentos as well.
Glide detection threshold: Onset Proximity	350 ms	only glides that are within threshold duration to an onset will qualify as portamento	serves to rule out IOI-internal glides (as portamentos are by definition connections of two or more notes): change according to the predominating duration of the tones – slow passages may benefit from longer proximity values
Link threshold	70 Cents	sets the minimum pitch gap between glide and its starting or target notes, respectively, in order to classify it accordingly (gaps below the threshold suggest direct connections to the respective notes)	reverberation may require a lower value, fragmentary or otherwise low-quality pitch track data may require a higher value
Range threshold: medium	250 Cents	lower boundary for glides to be classified as of "medium" range (glides below that will be considered as being "small")	research that suggests other thresholds (= terciles) than suggested
Range threshold: large	550 Cents	lower boundary for glides to be classified as of "large" range (glides below that will be considered as "medium")	ibid.
Duration threshold: moderate	120 ms	lower boundary for glides to be classified as of "moderate" duration (glides below that will be considered as "short")	ibid.
Duration threshold: long	210 ms	lower boundary for glides to be classified as of "long" duration (glides below that will be considered as "moderate")	ibid.
Dynamics threshold	1 dB	settles the dB threshold of minimum deviation from glide's respective starting and target tone levels during the glide to qualify as "louder" or "quieter"	level dynamics within your audio: large differences between loud and quiet passages allow for high thresholds, small differences (e.g., in strongly compressed audio) require low thresholds
Index scaling factor	0.0008	scales the results from multiplicative cumulation of all Portamento aspects (Direction * Link * Range * Duration * Dynamics) to a standardised scale that features a mean of 50 Index points	either if you adjusted the thresholds or the hard-coded factors per Portamento aspect, or in case you prefer another scale range

7. Hints & tricks: Using supplemental outputs for parameter and threshold refinements

"Expressive Means (advanced)" provides a number of outputs intended as tools to further refine the parameters, classification boundaries, and thresholds at work. Consider for instance Figures 6.1–6.3: the outputs "Pitch Onset Detection Function", "Vibrato-Only Pitch Track", and "Portamento Significant Points" helped with parameter optimisation for the respective recordings.

! Advice: Since the plugins' calculations may take significant amounts of time (proportional to the audio's length and sample rate), we recommend to either

(1) use a short sample of the recording (e.g., 20 seconds) only to determine optimal parameter configurations before prompting the analysis of the whole recording, or
(2) select a portion of the audio only within SV by use of the "Select" tool (the 'cursor', see p. 2), then prompting the plugin and ticking the „Restrict to selection extents“ box.

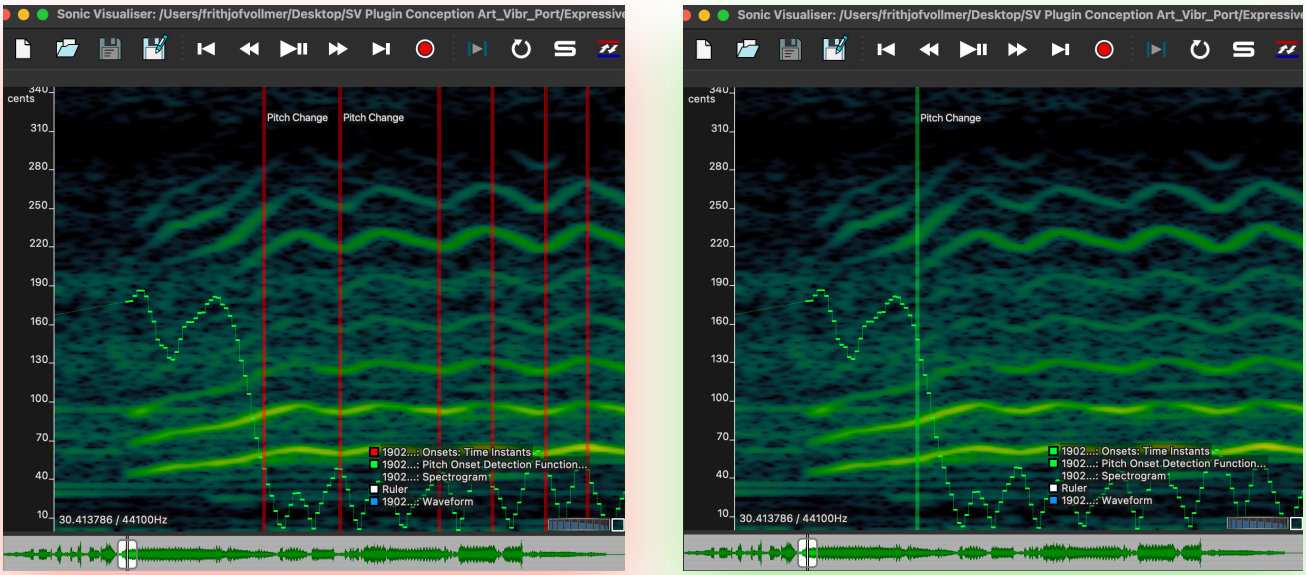

Figure 6.1: Using "Pitch Onset Detection Function" to refine the "Onset Sensitivity: Pitch" parameter (before–after)
On the left: Multiple faulty pitch onsets have been found (red lines). To identify new pitches, the function (green curve) by definition first has to exceed the Cent threshold given by the "Onset Sensitivity: Pitch" parameter for the duration of the "Minimum onset interval", then fall below this threshold again. It becomes clear that the default at 15 Cents is a too strict setting for vocal recordings due to typically wide vibrato ranges. Resetting it to 100 Cents solves the problem (on the right).

Figure 6.2: Using "Vibrato-Only Pitch Track" to refine the "Note segmentation" parameter

On the left, an implausible vibrato has been detected. Using the "Vibrato-Only Pitch Track" (red curve) reveals that the pitch of the preceding tone has been involved into vibrato range calculation. Changing the "Note segmentation" from "Without glides" to "Without glides and segmented" solves the problem. Other settings may profit from this output as well; e.g., if too few (or, on the contrary, too many) Vibrati have been found: lowering the "Vibrato shape: Correlation threshold" parameter value leads to detection of more potential vibratos, raising to the opposite.

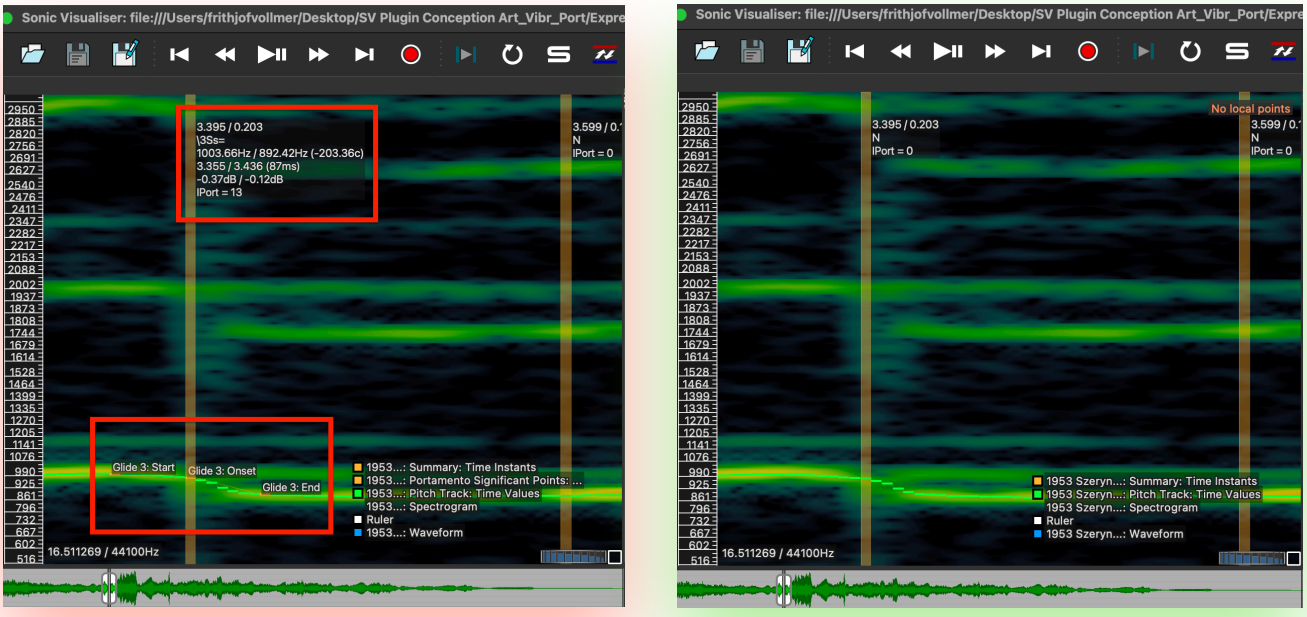

Figure 6.3: Using the "Portamento Significant Points" to locate erroneous glides

On the left, an overlapping of pitch track data led to a false-positive glide detection. "Portamento Significant Points" (orange annotations) helped with locating the ostensible glide, returning its start and end point. Resetting the "Glide Detection: Minimum duration" from 70 to 85 milliseconds (ms) solved the problem (right). Another way would have been to reset "Glide Detection: Minimum hop difference" from 10 to 30 Cents, but that might have negative impacts on long portamentos that naturally proceed slower.

In other instances, you may like to refine the "Type" classifications on base of a larger set of recordings. To do so, use the "Mean" outputs to determine appropriate thresholds, e.g., for "wide", "medium", or "narrow" Pitch Vibrato ranges (cf. Fig. 4.2):[17]

(1) Add *Expressive Means (advanced)* → *Pitch Vibrato: Mean Maximum Range*
(2) At the left edge of the pane, you'll find a mean $range_{max}$ value of all detected "vibrato elements" now (see Figure 4.6). Export this value, e.g., to a spreadsheet. Proceed to the next recordings.
(3) If you gathered all mean values, calculate the terciles of the whole set.[18]
(4) In the "advanced" plugin modes, set these values as new thresholds (the first tercile for "Range: medium", the second for "Range: wide").
(5) At the plugins' Github page (https://github.com/cannam/expressive-means), you may suggest your values as new hard-coded defaults for the 'semantic' outputs (so you don't need to reset them every time when restarting *Sonic Visualiser*). To do so, open a new "Issue" and describe your case – we will manage the coding and updating.

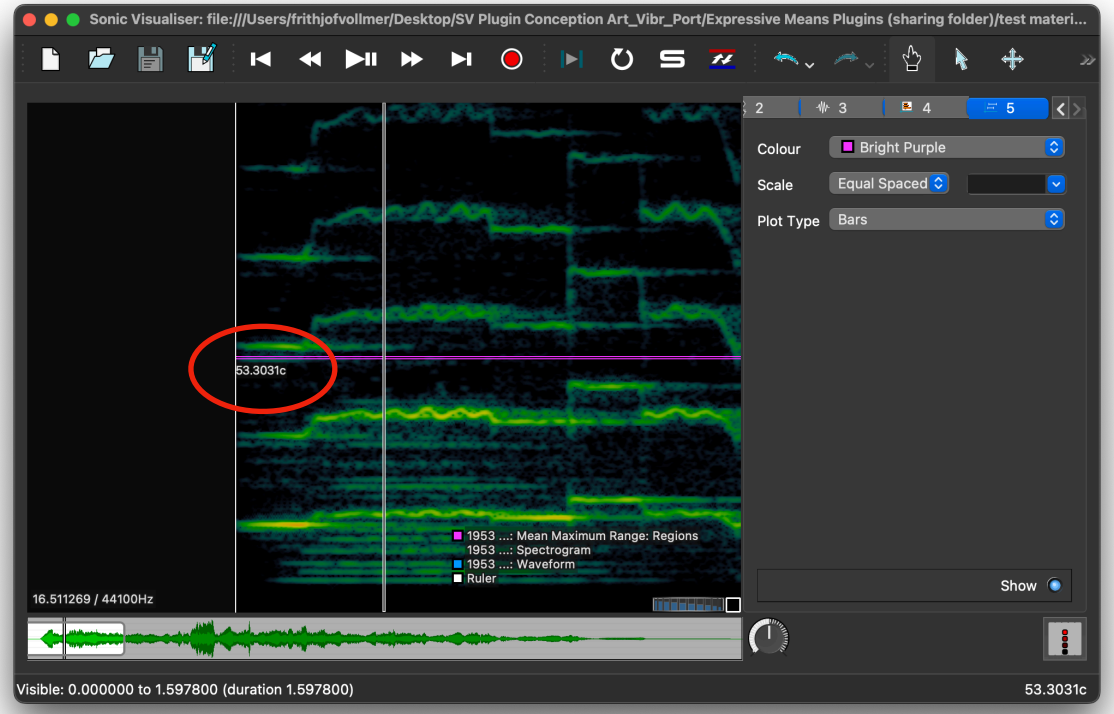

Figure 6.4: Using "Pitch Vibrato: Mean Maximum Range" for threshold determination
value is given in a regions layer (bright purple) at the left edge of the pane

[17] See note 14 on the current 'semantic' thresholds for vocals and for instruments *other* than bowed ones.

[18] Convenient English *Excel* formula: "=PERCENTILE([array]; 1/3)", German *Excel*: "=QUANTIL.INKL([Bereich]; 1/3)"; or "2/3", resp.