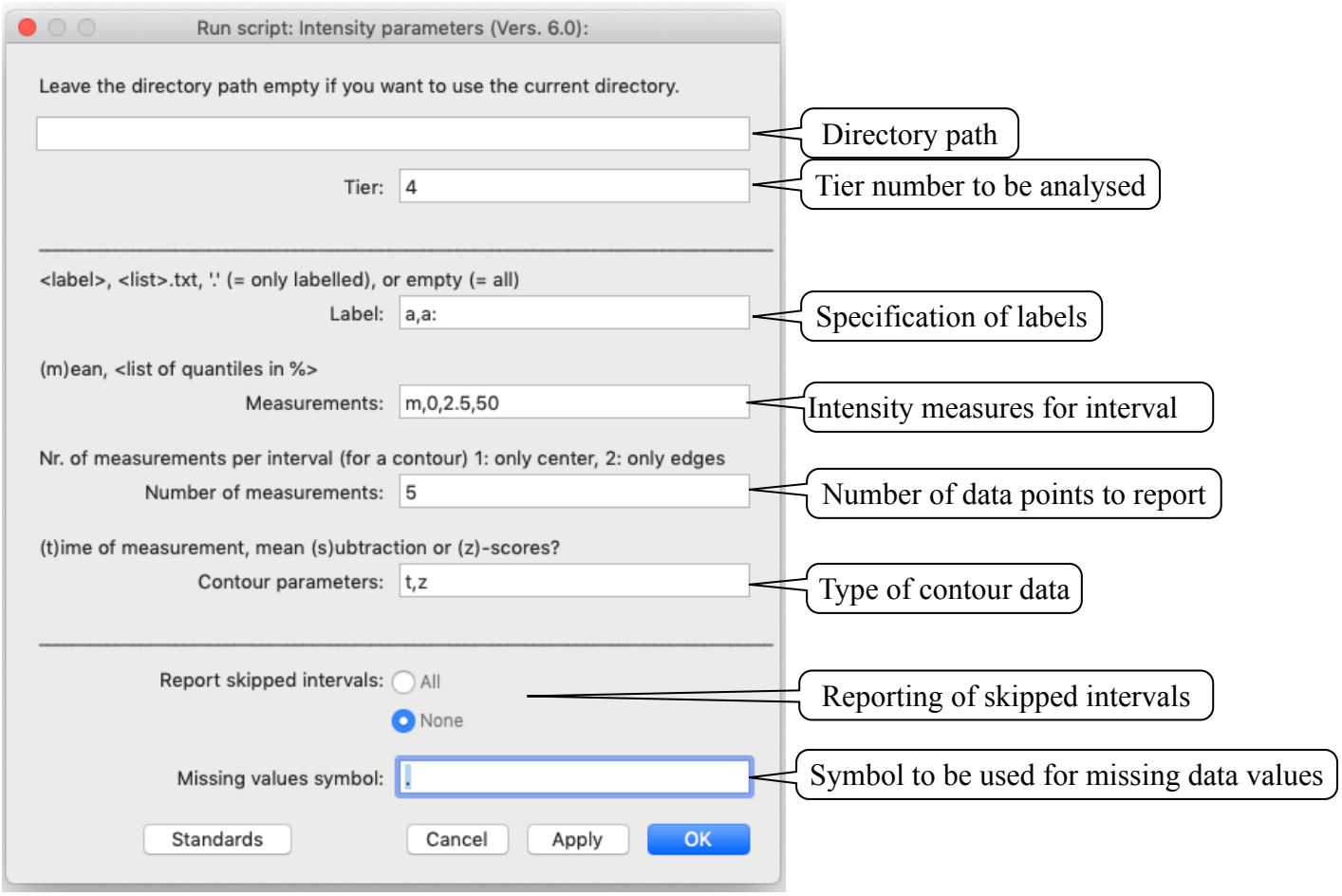
Intensity_6_0_0.praat

Task:
This script opens all sound files in a directory and associated TextGrids (if they exist), computes intensity mean, standard deviation and/or percentiles of whole files or of all intervals, which can be specified in various ways, together with intensity values of several time points (e.g. only at center or edges of an interval/file or to provide data for intensity contours) and writes the results to a text file with the name "intensity_results_<date>_<time>.txt". The starting position and the length of the interval are reported along with the intensity data. Sound files without an associated TextGrid file will be treated as one interval on its own.

There are more parameters that control the behaviour of the script than displayed in the form window (the size of the form window fits into a screen with 1024+ points vertical resolution). These parameters can be set below the form statement in the script (see the **Programming**) section.

Parameters:



Please read the **Programming** section for more internal parameters that can easily be changed.

Directory path:
The script handles all sound files in a directory specified in this field. If this field is left empty, the script will handle all sound files in the directory where the script was started (i.e., the script is placed in the same directory as the sound and TextGrid files).

Tier number to be analysed:
The number of the interval tier to be analysed.If this is set to "0" the whole file will be taken as one interval.

Specification of labels:
The intervals that should be analysed can be specified in several ways (in case no TextGrid file is found, this field is ignored):

<label>:
Giving a label (e.g. a:) or a list of labels separated by commas or spaces (e.g. i:,I u:,U) will only report intervals that have this label. This function is case sensitive.

<list>.txt:
Giving a text file (e.g. label_list.txt) will report all intervals that are listed on a line-by-line basis in a raw text file (not a Word or Pages file). Note that the extension .txt must be given in this field. Example of such a text file:

```
a
a:
ae
```

'.':
Using a dot (.) will report values for every labelled interval.

empty:
Leaving this field empty will report values for all labelled and unlabelled intervals.

Intensity measures for interval:
This is a string, separated by commas, that define what will be reported.

'm' (Report means and standard deviation of an interval):
Setting this switch will report the mean and standard deviation of each interval selected by **Specification of labels**. If no TextGrid file is found, the whole sound file will be used.
(In **Programming → unit$** PRAAT's energy, sone or dB can be selected; default is dB.)

<numerical values> (Specification of quantiles to be reported):
Quantiles (actually percentiles) of an interval will be reported by specifying values separated by commas. Values must be given as percentages between 0 and 100. The program will always make the reporting symmetrical to the lower and upper quantile range. For example, a specification like. 0, 2.5, 10, 50 will report the Minimum (0%), 2.5%, 10%, Median(50%), 90%, 97.5%, and Maximum (100%) quantiles. If no TextGrid file is found, the whole sound file will be used.

Number of data points to report:
The number of data points along an interval/file. If '0' is given, only the means and/or quantiles are reported. '1' will measure the intensity at the center of the interval/file, '2' will report data at the left and right edges, '3' at edges and center. Higher numbers will report data at more points. Note that this is the number of data points, not intervals: '10' will report data at 0%, 11.11%, 22.22%, … 77.78, 88.89%, 100% of the length of the interval/file. The intensity data will always be reported for the number given in this field, even when the **Type of contour data** field below is left empty.

Type of contour data:
The data can be reported with the actual (absolute) time of the measurement, as intensity values, values with subtracted mean of the interval/file or as z-scores. (Cf. <https://sites.google.com/site/tonemodelling/anaposts/z-transformdoesnotworkforpitchcontoursflat> why z-scores are not always appropriate; the argument given there for pitch holds for intensity as well.)

t:
The (absolute) time of a measurement is reported.

s:
The mean of an interval/file is subtracted from the intensity measure to normalise the data to the deviation from a mean during an interval in dB:

$$s\ [\text{dB}] = \textit{Intensity}\ \text{at time point [dB]} - \textit{Intensity mean}\ \text{of segment [dB]}$$

z:
The mean of an interval/file is subtracted from the intensity measure and divided by its standard deviation to normalise the data as z-scores:

$$z = \frac{\textit{Intensity}\ \text{at time point [dB]} - \textit{Intensity mean}\ \text{of segment [dB]}}{\textit{Intensity stdev}\ \text{of segment [dB]}}.$$

Reporting of skipped intervals:
Handling of intervals which are excluded by the **Specification of labels:**

All:
All intervals excluded from the computations are reported with File, Label, Start(s) and Duration(ms) only (see **Result file** below), all other values are set to the missing value symbol (see below). This function can be helpful to see the context of a particular interval analysed.

None:
No intervals that are excluded are reported.

Symbol for missing values:
When PRAAT cannot compute a value it uses internally the string "- undefined -". This script replaces this string by the string given in this field. For a subsequent analysis of the data with JMP, the dot indicate missing values, for R it would be NA.

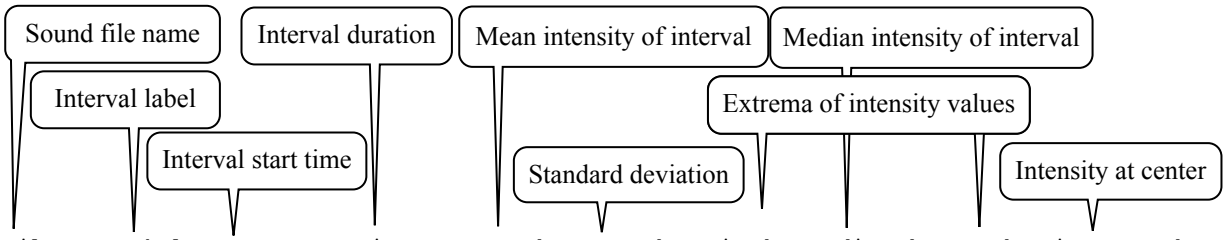
Result file:
The script generates a raw text file with tab-delimited data and a header line. The file name is of the form “intensity_results_<date>_<time>.txt” with <date> of the form ‘yymmdd’ (i.e. 2-digit year, 2-digit month, 2-digit day) and <time> of the form ‘hhmmss’ (i.e. 2-digit hour in 24 hour format, 2-digit minutes, 2-digit seconds).

For example, a file with the name “intensity_results_201112_153701” was created on the 12th of November 2020 at 3 pm, 37 minutes and 1 second. The parameters controlling the computation are listed at the end of every result file.

The result file for parameter settings

Label: n,t
Measurement: m,0,50
Number of measurements: 1
Contour parameters:
Report skipped intervals: None
Missing values symbol: .

to report mean, minimum, median, maximum and values at the the center of and [n] or [t] intervals will look like:



File	Label	Start(s)	Duration(ms)	Mean(dB)	StDev(dB)	Min(dB)	Median(dB)	Max(dB)	i_center(dB)
g071a000	n	1.0455	82.5	71.40	1.30	68.53	71.94	72.70	71.93
g071a000	t	1.1281	14.5	69.38	0.67	68.57	69.40	70.55	69.28
g071a000	n	1.3933	78.6	71.94	2.02	65.75	72.69	73.32	73.14
• • •									
g071a000	t	8.2948	70.4	18.34	21.71	-27.22	20.15	46.16	21.62
g071a000	n	8.8649	77.6	67.25	0.11	66.97	67.27	67.37	67.35

```
Script: Intensity_6_0_0.praat
Analysis started: 12-Nov-20 15:37:01
Tier: 4
Labels: n,t
Step rate: 0.005 s
Low F0: 50 Hz
Computation units: dB
Minimal length: 0 ms
```

In this next example, an output is displayed for parameter settings

Label: list.txt
Measurement: m,2.5
Number of measurements: 4
Contour parameters: z
Report skipped intervals: All
Missing values symbol: .

to report mean, 2.5% percentile and z-scores at 4 positions for each of the labels listed in the file *list.txt.* (The selection of z-scores automatically forces also the reporting of the intensity values at the 4 positions.)

The file *label.txt* contained the 3 lines:

```
n
m
N
```

Note that the 2.5% quantile is the dB-value below which 2.5% of all data in the interval falls (97.5% boundary for all data above 97.5%), whereas e.g. t_33.33% is a time point at 33.33% of the total duration of the interval (and i_33.33% is the dB-value at that time and z_33.33% is the z-score at that time).

(The lines of the result file are broken up in this listing into 3 parts, to fit onto the page width.)

```
File     Label  Start(s) Duration(ms) Mean(dB)     StDev(dB)    2.5%(dB)     97.5%(dB)  •••
• • •
g071a000 @      1.3639   29.3         .            .            .            .          •••
g071a000 n      1.3933   78.6         71.94        2.02         66.42        73.31      •••
g071a000 f      1.4718   115.3        .            .            .            .          •••
g071a000 a      1.5871   53.9         .            .            .            .          •••
g071a000 N      1.6410   65.7         71.92        0.40         70.91        72.34      •••
g071a000        1.7067   0.06         .            .            .            .          •••
• • •

(••• continuation of 1 header and 6 data lines •••)
• • •
••• t_0%(s)t_33.33%(s)  t_66.67%(s)  t_100.00%(s) i_0%(dB) i_33.33%(dB) i_66.67%(dB) i_100.00%(dB)  •••
••• .      .            .            .            .       .            .            .              •••
••• 1.3933 1.4194       1.4456       1.4718       72.18   73.30        72.28        66.40          •••
••• .      .            .            .            .       .            .            .              •••
••• .      .            .            .            .       .            .            .              •••
••• 1.6410 1.6629       1.6848       1.7067       72.34   72.17        71.86        70.93          •••
••• .      .            .            .            .       .            .            .              •••
• • •

(••• further continuation of  1 header and 6 data lines)
• • •
••• z_0%(z)  z_33.33%(z)  z_66.67%(z)  z_100.00%(z)
••• .        .            .            .
••• 0.12     0.68         0.17         -2.74
••• .        .            .            .
••• .        .            .            .
••• 1.04     0.62         -0.13        -2.45
••• .        .            .            .
• • •

Script: Intensity_6_0_0.praat
Analysis started: 12-Nov-20 15:54:30
Tier: 4
Labels: list.txt
Step rate: 0.005 s
Low F0: 50 Hz
Computation units: dB
Minimal length: 0 ms
```

Information for programming:
Some parameters can be set underneath the 'form' section in the script. These are:

unit
Unit for mean of interval computing.
unit = 1 energy
unit = 2 some
unit = 3 dB (default)

Position reporting in steps (1, 2, 3,…) or percentage (0%, 25%…) of interval
position_in_percentage = 0 Positions and data are reported as step number within an interval (e.g. t_1(s), t_2(s), t_3(s)…)
position_in_percentage = 1 Positions and data are reported as percentage of interval length (e.g. t_0.00%(s), t_25.00%(s), t_50%(s)…)

Intensity computing parameters:
step_rate = 0.005
low_F0 = 50

Minimal length (in milliseconds) of an interval to be considered
This parameter can be set to exclude too short intervals. Setting it to '0' excludes none.
minimal_length_ms = 0

Maximal number of intervals for reporting the contour
This parameter should only prevent the computation of too many time points per interval (e.g. due to accidental wrong input).
max_number_of_measurements = 50

Directories:
The script uses internally separate strings for sound, TextGrid, result and support directories. Users who use separate directories for these can adjust these names in the script.

Path_name:
By default, only the TextGrid or sound file name is listed. If the full path should be reported, path_name should be set to 1.
path_name = 0 No report of full path name
path_name = 1 Full path name is reported

ext$ (sound file names):
The default extension for sound files is ".wav". This parameter can be changed in the script.

sep$ (separator symbol):
In the (columns of) data in the result file are separated by this symbol. The default is tab$ (tabulator) but users might use e.g. the comma for an csv file.

user_feedback:
The script reports which file is being handled and the percentage of all files in a directory that have been handled. By setting this switch, any output (other than error and warning messages) will be suppressed. This will decrease processing time, but there is no feedback other than an increasing size of the result file.
user_feedback = 0 Gives user feedback
user_feedback = 1 No user feedback

np_string$ (noprogress string):
PRAAT itself reports its activity when computing intensity. This outputs can take substantial processing time. (This string is actually positioned where in PRAAT the noprogress is written.)
np_string$ = "" (= empty string) normal PRAAT feedback
np_string$ = "noprogress" No PRAAT feedback

dummy_data_header (dummy data header line):
Statistic programs like JMP decide the type of data for each column on basis of the first data line. To force correct data-type assignment (due to missing data in the first data row) a dummy data line of text, 0 and 0.0 can be generated to force correct data-type assignment.
dummy_data_header = 0 No dummy data line
dummy_data_header = 1 Dummy data line with "Dummy" for strings and "0.0" for numerals

duration_in_ms (duration reporting):
duration_in_ms = 1 Duration is reported in milliseconds
duration_in_ms = 0 Duration is reported in seconds

Current version and date:
6.0.0, 12-nov-2020

Known problems:
None

Planned extension:
Handling of point tiers
Allow spaces in interval labels

Contact:
reetz.phonetics@gmail.com