

Scoring in *Silbido*

Please read the ReadMeFirst file before attempting to use this software.

Introduction

The *silbido* tonal contour detector contains software to automate comparisons between sets of detections. One set of detections is considered to be the reference, or ground-truth detections, and the other is the set to be compared. Ground-truth detections are typically created by an analyst either manually using an annotation tool or in a semi-automated manner with a detection algorithm followed by manual cleanup of the detections, inserting missed detections and removing/adjusting faulty ones¹.

The scoring software expects to have the original audio data and the ground-truth detections in the same directory (folder). It is frequently convenient to organize data into subdirectories and the scoring software will by default examine all subdirectories as well. The detections to be compared can either be in the same directory or directories or in a separate directory that mirrors the structure of the audio data and ground truth (Fig. 1)

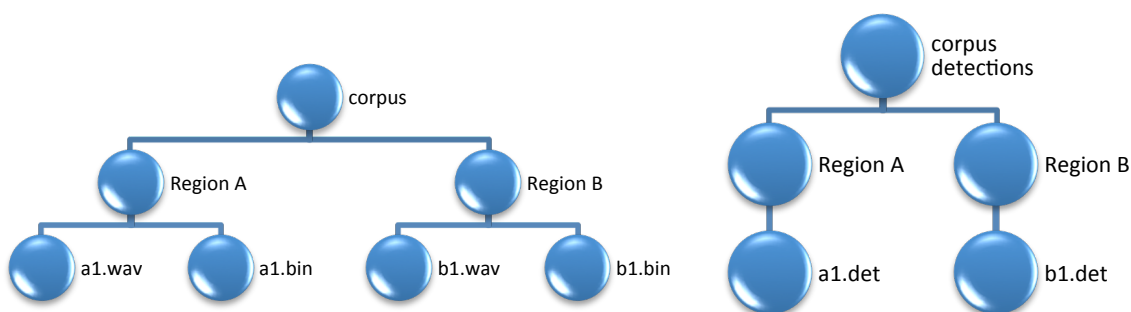


Figure 1 – Directory “corpus” contains audio data (.wav) and ground-truthed detections (.bin). Directory “corpus detections” contains corresponding detections that are to be scored. It is possible to place the detections in the same directory as the audio and ground-truth data, but this is not recommended as a number of new files will be generated. In addition, using multiple directories permits scoring of detections generated from different parameter sets or algorithms by placing each trial in its own directory.

¹ *Silbido* has the dtPlotUIGroundTruth (older and better tested) and dtTonalAnnotate (richer features, but not well tested) annotation tools that can be used in either of the described manners.

Scoring

For each detection file, the scoring software will find the corresponding audio and ground-truth detection. The ground-truth detections are read in and examined to see if they meet the user settable selection criteria which are dependent upon a minimum duration and a percentage of the tonal above a signal to noise (SNR) threshold. SNR is calculated by the scoring system, and currently uses *Silbido's* default settings. This is likely to change in future versions, and can be overridden by modifying the scoring script. We elected not to store the SNR with the ground-truth detections as the noise estimation method can significantly affect the estimated SNR.

Each ground truth tonal is examined. Regardless of whether or not it meets the selection criteria, any overlapping detections are examined to see if they are close enough to be considered a match. After all ground-truth tonals have been processed, any remaining detections are considered false positives *with respect to the ground truth*. It is important to remember that the ground-truth tonals are highly unlikely to be completely error free. Statistics are gathered on the detection group and several tonal contour files are written into the same directory from which the detection file originated. For file *f* the following files will be created:

f.d- The set of false positives, the detections that did not match any of the ground-truth tonals regardless of whether or not the ground-truth tonal met selection criteria.

f_s.d+ The set of correct detections for tonals that met the selection criteria.

f_s.gt+ The set of ground-truth tonals that met selection criteria and were detected. Note that there is no minimal length requirement for the detection, but the quality of the match is captured in other statistics that are gathered.

f_s.gt- The set of ground-truth tonals that met selection criteria and did not have a matching detection.

f_a.d+, f_a.gt+, f_a.gt- Similar tonal sets to f_s.d+, f_s.gt+, f_s.gt- except that all ground-truth tonals are considered rather than only those that meet the selection criteria.

It is necessary to examine all ground-truth tonals including those that do not meet the criteria to ensure that a correct detection of a whistle which did not meet the selection criteria is not added to the set of false positives. As the ground-truth tonals are processed, statistics are gathered that permit the computation of precision and recall rates.

In addition to the correct, missed, and false positive detections, information is gathered on how well each correct tonal matched the ground-truth tonals. Statistics are gathered on the percentage of the ground truth tonal that was matched (coverage), the average absolute deviation in frequency between the ground-truth tonal and matching portions, and the degree to which a ground-truth tonal was fragmented (multiple matching detections covering different portions of the tonal).

Using the software

See the instructions for setting up *silbido*. Once *silbido* has been configured, change the working directory to the directory containing the detections. Remember that you must start triton before using *silbido* components. The `scoreall` function will compute scoring statistics. It supports keyword/value pairs of arguments that are all optional. The following arguments are supported:

'DetExt', Ext - Filename extension for detection files. Default '.det'

'Corpus', Dir - Directory containing audio and ground truth .bin files from which detections were derived.
Default '.' (current directory)

'Detections', Dir - Directory where detections are located. Default '.' which is an abbreviation for the current working directory.

'ResultName', FileBasename - Use FileBasename for .mat/.txt. Default 'score'

'GroundTruthCriteria', [dB, RatioAbove_SNR, MinLen_s] Criteria for determining whether or not to expect each ground truth tonal to be detected. When tonals are detected but we did not expect them to be, they are not counted towards the recall, but they will not be used to penalize the precision. Default: [10, .2, .150]

Example:

```
results = scoreall('Corpus', 'c:/Users/corpora/DCLMMPA2011', 'ResultName', 'myresults');
```

This would compare all detections ending with .det in the current directory to audio data (.wav) and ground-truth (.bin) files in directory c:/Users/corpora/DCLMMPA2011. A results structure array is contained with one item per file. Sample entry, e.g. results(1):

```
falsePos: '\AXW.d-'
    all: [1x1 struct]
    snr: [1x1 struct]
    file: 'C:\Users\corpora\DCLMMPA2011-eval-anon\AXW.wav'
falsePosN: 37
```

The fields contain information about the generated tonal files as well as statistics that are gathered for the precision and recall and tonal quality rates. The tonal files can be viewed with any of *silbido's* plotting or annotation tools. As an example the false positives could be seen using:

```
dtTonalAnnotate(results(1).file, 'TonalsLoad', results(1).falsePos)
```

A copy of the results structure will be saved in myresults.mat (score.mat if the ResultName argument had not been used) and a log of the detection processing will be stored in myresults.txt.

Detection statistics can be gathered using the `dtAnalyzeResults` function:

`dtAnalyzeResults(results)`

which will generate a per file and summary statistics for all files contained in the results structure. These can be saved to a file by providing an optional output file name to `dtAnalyzeResults`.