**Labeled Data Format for Classification Tasks**

For the annotated data of the raw audio files, we adopt the basic format of Raven.

**Format:**

* Begin Time (s): represents the beginning time of a labeled detection. This column is mandatory, and it has float data type.
* End Time (s): represents the ending time of a labeled detection. This column is mandatory, and it has float data type.
* Low Freq (Hz): represents the low frequency of a labeled detection. This column is optional (or with null value), and it has float data type.
* High Freq (Hz): represents the high frequency of a labeled detection. This column is optional (or with null value), and it has float data type.
* Begin File (or Audio Filename): represents the corresponding audio filename. This column is mandatory, and it has string data type.
* Category: This column is mandatory, and it has string data type. This column indicates granularity level of the annotated detections. It can either be at sound source level (such as “animal”, “machine”, “nature”), or at species name level, or sub-species/population level, or sound/call type level, or focal/non-focal level. If the annotation is only for one category (to classify the corresponding presence/absence), then the values in this column are all the same.
  + Animal Taxonomic Classifications: if the labeled sounds are for animal calls, we encourage users to find corresponding Taxonomic Serial Number (TSN) at [Integrated Taxonomic Information System (itis.gov)](https://www.itis.gov/) (ITIS), which provides TSN for different taxonomic Hierarchies (class, family, species, subspecies, etc.) and use it as the value for this column “Category”. For example, the screenshot below shows the search result for “blue whale”, which has TSN “180528”.

Graphical user interface, text, application

Description automatically generated

* If the correct taxonomic classification can not be found in ITIS or the users cannot decide the correct TSN, the users can use customized strings to fill this column. For example, when searching “blue whale”, there are multiple corresponding results, and the users can use “blue whale” if correct TSN cannot be decided. Graphical user interface, text

  Description automatically generated
* If the annotations are at finer granularity (such as call type), users can also use customized strings concatenated by hyphen (‘-‘) indicating the hierarchy of label, for example “Black-capped Chickadee - alarm call”.
* If there are multiple categories’ labels (for example, multiple species calling at same time) that correspond to same beginning/ending timestamps, users can create multiple rows of annotations where each row correspond to one label.
* Label: represents the label of a detection, indicating the presence or absence of detected sound/call from the corresponding category. This column is mandatory, and it has Boolean data type. If the annotation is only for one category, then the labeled data must have annotations for both “1” (i.e., presence) and “0” (i.e., absence). If the annotation is for multiple categories, it is also highly recommended to have annotations for both “1” (i.e., presence) and “0” (i.e., absence) for each category.
* Label confidence: this column is optional, and it has float data type, and it has values between 0.0 and 1.0 that provides confidence level of how much trust we have in the label.

***Example 1: binary labels when annotating one species’ calls from multiple audio files***

* ***If the users can map “Category” to the correct TSN***

A screen shot of a graph

Description automatically generated with low confidence

* ***If the users can not map “Category” to the correct TSN***

Table

Description automatically generated

***Example 2: multi-class categories when annotating from multiple audio files***

Table

Description automatically generated