Audiogram database requirements description and implementation guide

Version 2019-11-21

Abstract

This document specifies the entities, their properties and relations to be represented in the audiogram database. For each element its significance for representing audiophysiological measurements is indicated and examples are given. This is accompanied by notes on how the element is represented in the described version of the database. The information in this document was compiled by Denise Jäckel and Christian Bölling on the basis of domain experts' feedback and the audiogram data collection represented in MS Excel spreadsheets. The database schema was developed by Christian Bölling. The logical data model – they database schema – is specified in a MySQL workbench file, accompanying this document.

Status of this document

May be superseded

This document describes the audiogram database requirements and contains implementation notes for the initial database schema as represented in the accompanying MySQL workbench file as of 2019-11-21. Other documents may supersede this document. It is expected that the database structure and possibly also the requirements will evolve further.

Background information

The information compiled here is based on audiophysiological measurement data that was manually collected and consolidated from scientific publications. The purpose of the database is to make this measurement data and the relevant experimental metadata available in a structured form that is useful for review and combination of these findings. The original data, as it is reported in the publications, was curated to ensure that data collated as a single audiogram contains only data points from measurements taken in the course of a single experimental investigation. Therefore, a single audiogram, in the context of the audiogram database, comprises data obtained from a single experimental investigation using a single set of methodical parameters for conducting the experiments.

Description of specific entities

Audiogram Number: *Is a successive number in the Excel table, which is assigned consecutively for each audiogram.*

- **Audiogram ID**
  - Example: 1
  - Implementation: represented in audiogram_experiment.id and used as primary key.

Taxonomic information: *Information on the animal species is recorded in a tree structure. The recommended workflow is to* ***record the latin name*** *(binomial*

name) of the species, and then to use the Open Tree of Life API to get the remaining taxonomical lineage.

Each node in the tree stores the following fields:

- **Open Taxonomy Tree ID**: an identifier pointing to an entry in the Open Tree of Life reference taxonomy v3.0.
    - Example: 124215
    - Implementation: represented in taxon.ott_id and used as primary key.
- **Unique name**: the Latin name of the animal. Generally, this would be a species or subspecies name, but higher ranks are also possible.
    - Example: *Orcinus orca*
    - Implementation: represented in taxon.unique_name
- **Rank**: the taxonomical rank of this taxon.
    - Example: ‘species’
    - Implementation: represented in taxon.rank, should be one of 'subspecies', 'species', 'genus', 'family', 'order', 'class', 'phylum'
- **Parent**: the OTT ID of the parent taxon of this taxon, as recorded in the lineage obtained from the Open Tree of Life reference Taxonomy.
    - Example: 124219, which is the OTT ID of the genus *Orca*.
    - Implementation: represented in taxon.parent
- **Left** and **right** ordering keys: trees and hierarchies in relational databases are often implemented as nested sets. This is a model that allows traversing the tree without the need for recursive queries, which are not supported natively by some relational database engines. The left and right keys can be used to determine the (down- or upwards) hierarchy membership for each node, thus making taxonomy queries significantly simpler and faster.
    - Example: the genus *Spheniscus* would be represented in the taxonomy tree by a node with left key 100 and right key 109, meaning that all species in this genus have a left key greater than 100 and a right key less than 109.
    - Implementation: taxon.lft and taxon.rgt

Taxonomic information – higher ranks: *as explained above, the taxonomic lineage of each species is recorded in a tree structure. Nodes for higher ranks are identical to species nodes, except for the ‘rank’ value.*

Vernacular names:
Additionally to the taxonomical information, each **species** node stores the vernacular name of the animal. No vernacular names are stored for higher ranks. Vernacular names of subspecies are stored if applicable, otherwise the species name is used.

The recommended workflow is to fill-in the vernacular names during data import, by collecting them from the Wikidata API.

-

Common name: *the usual name is entered here, both in English and in German*
    - **German name** (Example: Schwertwal)

        - implementation: represented in taxon.vernacular_name_german
    - **English name** (Example: orca)
        - implementation: represented in taxon.vernacular_name_english

Literature information: *is divided into the long and short form and the DOI. Note that more than one publication can be related to an audiogram. In particular, data that was already published can be published again in conjunction with data from other experiments in a follow-up publication. In such cases, data from different experimental investigations is always allocated to different audiograms.*

The recommended workflow is to pull the citation information from the DOI-foundation's API by resolving the DOI during data import. As not all publications have a DOI, especially older ones, the "source long" and "source short" fields can be filled manually as a fallback.

- **DOI** (Example: 10.1121/1.427121)
    - implementation: represented as publication.doi
- **Source long** (Example: Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., Henry, K. R. (1999) Killer whale (Orcinus orca) hearing: Auditory brainstem response and behavioral audiograms. Acoustical Society of America. 106 (2): 1134-1141)
    - implementation: represented as publication.citation_long
- **Source short** (Example: Szymanski et al., 1999)
    - implementation: represented as publication.citation_short
-
    -
-

Individual: *Information about the test animal is collected here in a coherent manner. If the data can be clearly assigned to particular individuals, the following information is recorded separately. However, if no clear separation of the data can be established, the data shall be collected for all animals together and separated by a semicolon. If the information is identical for both animals, they need not be separated with a semicolon. It is then sufficient to record this information only once.*

- **Number of experimental animals:** *The exact number of animals that have been tested and included as data in this audiogram is given here*
    - Example: 2
    - Implementation: The number of experimental animals involved is represented implicitly in the number of records from table individual_animal associated to the corresponding record in the table audiogram_experiment by a record in table test_animal. This design enables representation of the involvement of one or more usually anonymous individual animals in an audiogram experiment and at the same time enables representation of repeated involvements of the same individual animal in different experiments. Records in individual_animal are identified with artificial primary keys. If the animal can actually be identified individually (by a unique name or ID given to it by its keeping authority), this

representation enables linking the same animal to different experiments. Individual animals that are not identified by a unique name in the information sources may actually be represented by more than one entry in table individual_animal. This is, with respect to the intended data usage, irrelevant. Unless the animal is uniquely identified by an individual name, it can simply not be decided if two records in the table individual_animal represent the same or different individuals.

- **Measurements:** *Here it is recorded how often the animal was tested.*
  - Example for 2 individuals: 2
  - implementation: represented as audiogram_experiment.number_of_measurements

- **Sex:** *the sex of the animal is indicated as "female" or "male*
  - Male / Female (Example for 2 individuals: Female; Female)
  - implementation: represented as individual_animal.sex

- **Name:** *If the name of the test animal is mentioned, it is entered here in order to be able to check whether the same animal appeared in several publications.*
  - Example for 2 individuals: Yaka; Vigga
  - implementation: represented as individual_animal.individual_name

- **ID:** *Own ID generated here during source evaluation. This ID is assigned across audiograms if the same animal is found to have been sampled in different audiograms.*
  - Example for 2 individuals: 1; 2
  - implementation: represented as individual_animal.id, used as primary key and referenced in table test_animal
  -
- **Life stage:** *Information about the life status and the age of the animal is recorded.*
  - Juvenile / Sub-adult / Adult (Example for 2 individuals: adult; adult)
    - implementation: represented as test_animal.life_stage
  - Age in months (Example for 2 individuals: 192 to 216; 312 to 336)
    - represented as test_animal.age_min_in_month and test_animal.age_max_in_month

- **Status of liberty:** Division into "wild" and "captive". Here freedom means that the animal was not brought into a scientific facility but tested on the beach and then released again. Animals living in captivity remain captive for a certain period of time.
  - Wild/ Captive / Stranded
    - implementation: represented as test_animal.liberty_status
    -
- **Duration of captivity (in months):** *Here you can see how long the animal has been in captivity.*
  - *o* Example for 2 individuals: 312; 192
  - implementation: test_animal.captivity_duration_in_month

-
  -
  -

Experiment information: *contains information about where and how the experiment was conducted*

- Locality: *The exact position of the experiment is recorded. In most cases, this is a scientific facility.*
  - **Name of the facility (**Example: Marine World)
    - implemented as facility.name
  - **Latitude** (Decimal degrees, example: 38.13777778)
    - implemented as audiogram_experiment.latitude_in_decimal_degree
  - **Longitude** (Decimal degrees, example: 122.232)
    - implemented as audiogram_experiment.longitude_in_decimal_degree

- **Position of the animal:** *The exact position of the animal is recorded here and assigned to four predefined categories.*
  - totally underwater / head just below water surface / head half out of water / outside of the water
  - implemented as audiogram_experiment.position_of_animal

- **Distance to sound source (in m):** *Here you can see how far away the sound source was from the animal.*
  - Example:
  - implemented as audiogram_experiment.distance_to_sound_source_in_meter

- **Test environment:** *Helpful information about the test environment can be entered in the form of texts.*
  - Example: The test pool, filled with sea water, was about 4 m deep and 15 m in diameter.
  - implemented as audiogram_experiment.test_environment_description

- **Medium: Whether the audiogram was obtained in air or underwater.**
  - Example: "water"
  - implemented as audiogram_experiment.medium

- **Method:**
  - Behavioral: go-no go / pressing a paddle / pushing something
  - Electrophysiological: Auditory evoked potentials (AEP): auditory brain-stem responses (ABR)
  - implementation: The methods are represented in a separate table method and referenced in audiogram_experiment.measurement_method_id – see the implementation note below

- **Position of the 1st/2nd/3rd electrode:**
  In the case of electrophysiological subjects, the exact positions of the electrodes are indicated. The positions of the individual electrodes are recorded separately.

  - near the ear / left pan bone / right pan bone / near the blowhole / near the dorsal fin/ on the back/ near the beginning of the dorsal ridge/ posterior to the dorsal ridge (Example: near blowhole (1st electrode), near dorsal fin (2nd electrode), NA (3rd electrode))

- implemented as audiogram_experiment.position_first_electrode and similar columns for the 2nd and 3rd electrode position

- **Year of experiment:** *The exact year or timespan of the experiment are recorded here.*
  - Example for 2 individuals: 1995 to 1996
  - implemented as audiogram_experiment.year_of_experiment_start and audiogram_experiment.year_of_experiment_end, respectively

-
  -
  -

- **Calibration:** *Information about the calibration is recorded here with regard to the Sound Pressure Level (SPL) reference value*
  - Example: between 6–10
  - implemented as audiogram_experiment.calibration , stored in a text field

- **Threshold determination info (in %): *Indicates the percentage at which the threshold was set.***
  - Example**:** 50
  -
  - implemented as audiogram_experiment.threshold_determination_method

- **Duration of test tone (in ms):** *The length of the test tone is specified here.*
  - Example: 1 (at 1 and 2kHz); 0.5 (at all other frequencies)
  - implemented as audiogram_data_point.testtone_duration_in_millisecond
- **Form of the tone:** *The type of sound is selected from the following defined specifications*
  - cosine-gated tone bursts / sinusoidal amplitude modulated (SAM) tone (modulated narrow band noise; modulated rectangular click, pure tone)/ sinusoidal frequency modulated (FM) tone (linear upward frequency modulated sweep/ linear downward frequency modulated sweep/ sinusoidal frequency modulation)
    The methods are represented in a separate table method and referenced in audiogram_experiment.testtone_form_method_id – see the implementation note below.
- presentation of the test tone: *Indication of the manner in which the test tone was presented*
  - **Staircase procedure** (yes / no)
    - implemented as audiogram_experiment.testtone_presentation_staircase
  - **Method of Constants** (yes/no)
    - implemented as audiogram_experiment.testtone_presentation_method_constants
  - **Form of the sound** (click / pipe trains / prolonged / SAM (sinusoidal amplitude modulation) (Example: click)
    - implemented as audiogram_experiment.testtone_presentation_sound_form

- **sedated**: *Indication whether the test animal has been sedated*
  - Yes / no (Example: no)
    - implemented as audiogram_experiment.sedated

- **Sedation Details:** *More detailed information about sedation will be collected here (e.g. medicament for sedation).*
  - Example: /
  - implemented as audiogram_experiment.sedation_details

Audiogram data

- **Frequency (in kHz):** *Each examined frequency is entered here*
  - Example: 1
  - implemented as audiogram_data_point.testtone_frequency_in_khz

- **Sound Pressure Level (SPL) (with reference level according to next field; in dB):** *The corresponding SPL value determined can be found here*
  - Example: 105
  - implemented as audiogram.data_point_sound_pressure_level_in_decibel

- **SPL reference:** *Due to the specified number of possible reference values, a corresponding selection function is performed here.*
  - re 1 µPa
  - re 1 µbar
  - re 1mPa
  - re 20 µPa
  - re 0.0002 dyne cm-²
  - re 1 dyne cm-²
  - re 2x10-4 µbar
  - Implementation: The different references are represented in table sound_pressure_level_reference and referenced in audiogram_data_point.sound_pressure_level_reference_id.

- **SPL reference value**
  - root mean squared (RMS) / Peak-to-Peak (PP)
  - implemented as audiogram_data_point.sound_pressure_level_reference_method

Additional implementation notes

Representation of sound pressure level references

The various sound pressure level references are represented in a separate table `sound_pressure_level_reference` and referenced in the `audiogram_data_point` table as `sound_pressure_level_reference_id`.

In the `sound_pressure_level_reference` table the significance of each sound pressure level reference can be indicated as well as the conversion factors to modern air- and water-borne sound pressure reference values. The proposed contents of this table based on this specification are represented as an MS Excel spreadsheet in the accompanying audiogrambase_import_spl.xlsx file.

Representation of measurement and test tone form methods

These methods, referenced as “Method” and “Form of the tone” above are represented in a separate table method to enable representation of the relation between main methods and their specialized variants as methods are expected to be represented and queried at all levels of granularity. The proposed contents of this table based on this specification are represented as an MS Excel spreadsheet in the accompanying audiogrambase_import_methods.xlsx file.