MLPC 2025 Task 1: Data Annotation

Tara Jadidi, Florian Schmid, Paul Primus

March 2025

Context

Kepler Intelligent Audio Labs (KIAL), a soon-to-be-founded innovative AI startup, aims to collect a general-purpose dataset with strong temporal annotations to train sound event detection systems for their customer (see Figure 1 for a schematic illustration of sound event detection).

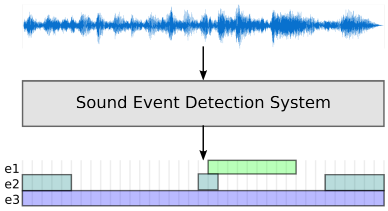


Figure 1: **Sound Event Detection (SED)** systems take an audio recording as input and predict acoustic events of interest ($e1, e2, e3$) *including* their temporal onsets and offsets.

To make this data set reusable for many potential future customers, KIAL decides to annotate sounds with free text annotations instead of relying on a fixed set of categories (see Figure 2). For a detailed description of the project, look at the project description on our Moodle page.

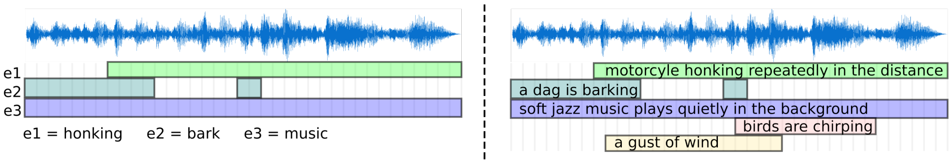

e1
e2
e3
e1 = honking
e2 = bark
e3 = music
motorcyle honking repeatedly in the distance
a dag is barking
soft jazz music plays quietly in the background
birds are chirping
a gust of wind

Figure 2: *Left:* Typically, SED data sets are only annotated for a fixed set of labels ($e1, e2, e3$). *Right:* Instead of relying on a fixed set of labels, KIAL decided to use arbitrary free text descriptions. Text annotations can be mapped to labels of interest ($e1, e2, e3$) by matching synonyms or using word embedding models.

KIAL has divided the project into these five stages:

1. **Data Collection:** Collect a large dataset of audio recordings, containing a large variety of potential target sounds and events. (✓)
2. **Data Annotation:** Once the data is gathered, human annotators carefully annotate sounds in the recordings with textual descriptions and temporal onsets and offsets. (← we are here)
3. **Data Analysis:** The dataset undergoes a thorough examination through exploratory data analysis, allowing for a deeper understanding of its characteristics and potential challenges before further processing.
4. **Model Training & Selection:** KIAL's R&D team needs to demonstrate the usefulness of the new approach to the managers by *training machine learning models* for one or multiple fictitious customers (each customer is represented by a fixed set of labels).

5. **Challenge:** A customer requests a sound event detection system for a fixed set of labels. Provide predictions on a hidden test set to prove to the customer that the developed system has a high detection performance and win an (also fictitious) contract.

KIAL has already collected the dataset by scraping publicly available audio recordings on the web (step 1). Step 2 (the focus of this task) will be a collective effort, joining the forces of all individual course participants. We will assist you with steps 3, 4, and 5 by providing pre-computed audio features, text embeddings, and suggestions on how to convert text annotations to labels.

Task Outline: Data Annotation (10 points)

In this phase, we will annotate a dataset of audio recordings with textual descriptions using Label Studio. Luckily, we are a group of more than 350 students and three instructors, which will allow us to collect a reasonably sized data set for the machine learning experiments in the next phases. To complete this phase, each student is required to *annotate 40 recordings* with textual descriptions and temporal on- and offsets. **Deadline:** Monday, March 24th, 23:59
Successful completion of this task is mandatory to pass the course.

Below is a step-by-step guide to complete Task 1: Data Annotation.

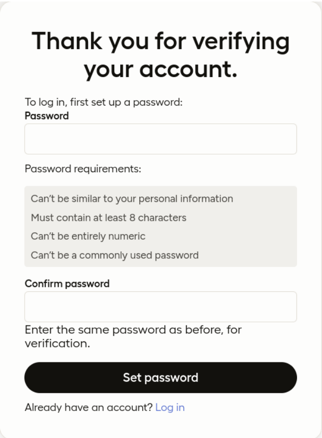
Step 1: Read the Annotation Task Description

Read all instructions in this document carefully before starting the annotation.

Step 2: Get Access to Label Studio & Overview

We will send an announcement via Moodle as soon as the annotation interface is available for you. You should then receive an email titled ”Verify your email” from `hello@humansignal.com` to the email address set in KUSSS (My Setting → E-Mail-Address). Use the link provided in the mail to join the Johannes Kepler University organization on *Human Signal (Label Studio)* by accepting the invitation. Sign up with:

Username: `k<student ID>@students.jku.at`, e.g., `k1234567@students.jku.at`
Password: Choose your own password.



Thank you for verifying your account.

To log in, first set up a password:

Password

Password requirements:

- Can't be similar to your personal information
- Must contain at least 8 characters
- Can't be entirely numeric
- Can't be a commonly used password

Confirm password

Enter the same password as before, for verification.

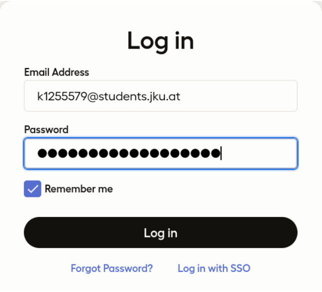
Already have an account? Log in

Figure 3: Sign up for a Human Signal account with your student email.

Important: *Use your generic JKU student email* instead of your actual email to sign up and log in. Everything sent to `k...@students.jku.at` will be forwarded to the email address set in the KUSSS.

Once the registration is complete, you should be able to log into your Human Signal account here:

https://app.heartex.com/user/login/



Log in

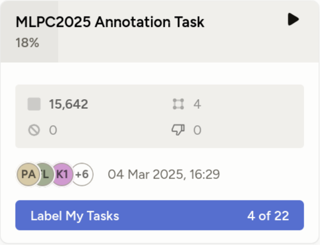
Email Address: k1255579@students.jku.at

Password: ●●●●●●●●●●●●●●●●●●

☑ Remember me

Figure 4: Log into your Human Signal account with your student email.

The landing page displays a box titled *MLPC2025 Annotation Task* (Fig. 5), which gives a short overview of your progress.



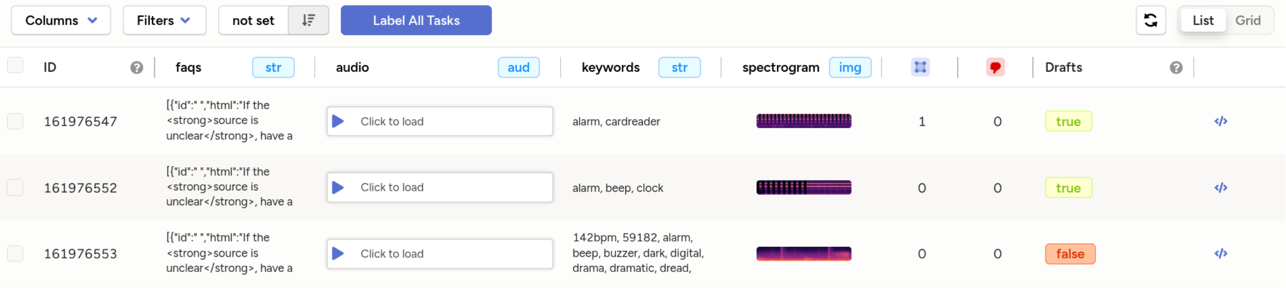
MLPC2025 Annotation Task
18%

15,642 4
0 0

PA L K1 +6 04 Mar 2025, 16:29

Figure 5: Click on `Label all Tasks` to start the annotation interface. Use the play button to open the *Data Manager*.

Click the play button in the top right corner to open the *Data Manager*. The *Data Manager* (Fig. 6) gives a more detailed overview of all pending and completed annotation tasks.



□	ID	faqs str	audio aud	keywords str	spectrogram img			Drafts	
□	161976547	[{"id":" ","html":"If the <strong>source is unclear</strong>, have a	▶ Click to load	alarm, cardreader		1	0	true	</>
□	161976552	[{"id":" ","html":"If the <strong>source is unclear</strong>, have a	▶ Click to load	alarm, beep, clock		0	0	true	</>
□	161976553	[{"id":" ","html":"If the <strong>source is unclear</strong>, have a	▶ Click to load	142bpm, 59182, alarm, beep, buzzer, dark, digital, drama, dramatic, dread,		0	0	false	</>

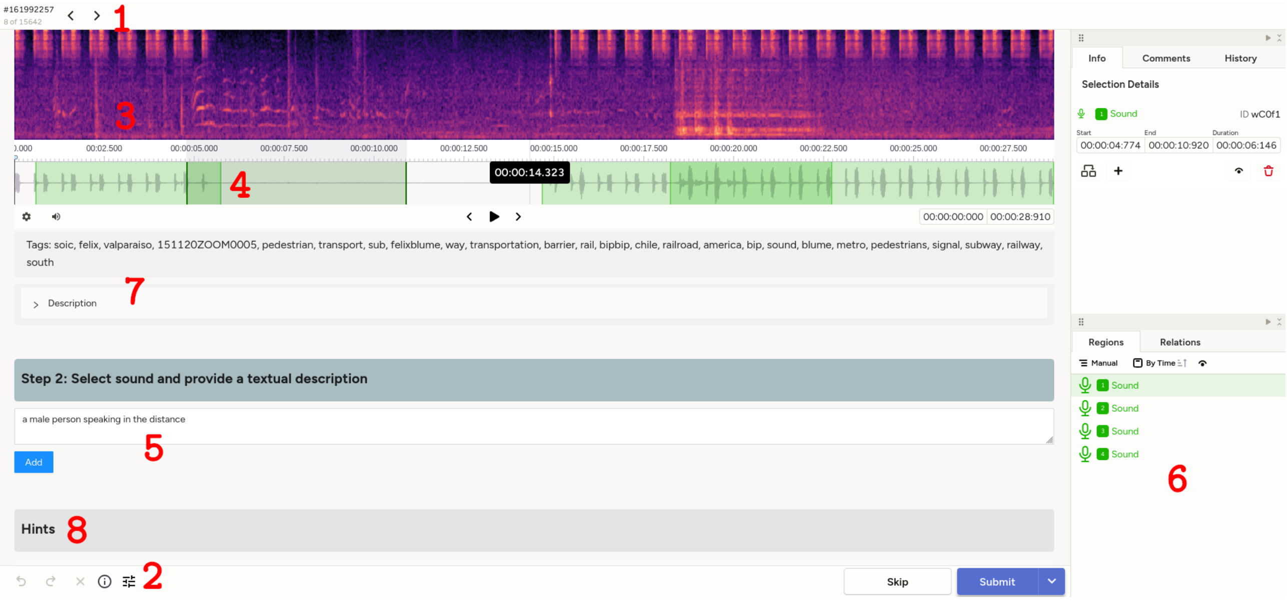
Figure 6: The *Data Manager* gives an overview of all tasks assigned to you. You can use it to revisit incomplete and submitted annotation tasks.

Click on the `Label all Tasks` button to start the annotation interface. A pop-up with a summary of the labeling instructions will appear.

Step 3: Get Familiar with the Annotation Interface

Fig. 7 highlights the most important sections of the labeling interface. The following paragraphs explain each component in greater detail.

1. **Navigation:** Use the left and right buttons to navigate through your queue of annotation tasks. *The order of the annotation tasks in the queue might change after submitting a task or exiting the current annotation session.* You're progress will not be lost; you can always find the incompletely annotated task in the *Data Manager*.



Tags: soic, felix, valparaiso, 151120ZOOM0005, pedestrian, transport, sub, felixblume, way, transportation, barrier, rail, bipbip, chile, railroad, america, bip, sound, blume, metro, pedestrians, signal, subway, railway, south

Step 2: Select sound and provide a textual description

a male person speaking in the distance

Figure 7: The annotation interface.

2. **Lower Menu Bar:** *Left Side*: Allows you to undo or redo interaction with the interface, reset the interface, view the labeling instructions, and change default interface behavior. *Right side*: Use the `Submit` button once you are done with annotating the audio recording. The `Skip` button *must only* be used for offensive or disturbing sounds (more info in the disclaimer below).

3. **Spectrogram Plot:** The spectrogram plot gives an audio-visual representation of the sound. The plot is roughly aligned with the waveform and might be helpful in finding onsets and offsets.

4. **Waveform Plot:** Annotate regions here by clicking on the onset position of a sound and dragging your mouse to the offset position. To annotate an overlapping region, click above the green area. You can modify the event boundaries afterward with your mouse. Select a region in the regions menu (see point 6) or by directly clicking on it (not possible for overlapping sounds). *Most useful hotkeys:*

   - Use `Esc` to unselect a region.
   - Use `Backspace` to delete a selected region.
   - Use `Space` or `Ctr+P` to play or pause the sound.
   - Use `Ctrl+C` and `Ctrl+V` to copy and paste a selected region.
   - Use `Ctrl+Z` and `Ctrl+Shift+Z` to undo and redo interactions.

   Find more hotkeys in the lower menu bar → settings.

5. **Text Annotation Area:** The text field only appears if a region is selected (click on the region or select the corresponding list item in the region menu). Type your textual description into this text field and click `Add` once you're done. *The text field must not be empty.* You can modify or delete the annotation later by selecting the region and clicking on the respective button. *Each region must have exactly one textual annotation.*

6. **Region Menu:** The region menu gives an overview of all the marked regions. You can use it to select regions (e.g., if they are overlapping) and to review and edit textual annotations.

7. **Metadata:** The metadata section contains useful hints about the content of the audio recording. The tags are always displayed. To show the textual description, click on the `Description` dropdown. However, be cautious: *Metadata may be incomplete or incorrect.*

8. **Help:** Shows a compilation of hints and frequently asked questions from our initial trials.

Step 3.5: Get Familiar with the Annotation Process

Try to identify and annotate sounds in a handful of recordings to get a feeling for the task. textitOnly submit the task if you have annotated all sounds properly!

For each sound in a recording, do the temporal and textual annotation:

1) **Temporal Annotation:** Annotate the sound's onset and offset.

   Potential Problems / Difficulties:

   - **Separate Regions for Separate Sounds:** Annotate each sound as a new region. The regions might overlap, as shown in Figure 7.
   - **Gaps between repeated events:** Separate two events of the same kind if there is a noticeable gap or pause between them. If the pauses are shorter than 1 second, you may merge them into a single annotation.
     Example: Separate two dog barks if there is a noticeable pause in between them. Do not try to separate the individual words in a sentence or the individual beats of an alarm tone.
   - **Continuous sounds:** If a sound is present throughout the entire recording, set the onset at the beginning and the offset at the end (luckily, this happens quite often).
     Example: If a recording contains a large crowd of people talking throughout the entire recording, then the annotated region should span the entire recording.
   - **Echo:** If the recording also captures the echo of a sound event, include the echo in the temporal and textual annotation.

   A rule of thumb: Everything that is not covered by a region in the annotation should be silent. Conversely, silent parts should not be covered by a region.

2) **Textual Annotation:** Provide a textual description for each region.

   Each region must be annotated with a sentence describing the sound in a detailed manner. The description should be as specific as possible and contain the following information:

   - **Source and Action (mandatory):** What is the source of the sound?
     Example: "An airplane flying overhead", "A person walking on gravel"
   - **Descriptor:** How does it sound?
     Example: "loud," "metallic," "rhythmic," "muffled"
   - **Temporal:** Does the sound change over time?
     Example: "fading," "repeating," "gradually increasing"
   - **Context:** Where is it happening, or what environment does it suggest?
     Example: "indoors," "echoing," "distant"

   Example of a detailed textual annotation: "A telephone rings loudly in a steady pattern inside a room."

   Potential Problems / Difficulties:

   - **Unknown Sound Source:** If you do not recognize a sound or if the sound source is unclear, have a look at the metadata tags and the textual description; they typically contain valuable information.
     However, be cautious: Metadata may be incomplete or incorrect. If you still cannot identify the source of a sound, use a more general term or terms for source-ambiguous sounds.
     Example: "A metallic thumping sound, repeating rhythmically."
   - **Independent Descriptions:** Descriptions **must not** depend on other descriptions, i.e., each description should make sense by itself.
     Incorrect: "The same dog as before is barking repeatedly."
     Correct: "A dog barking repeatedly."
   - **One Description → One Sounds:** Annotate each sound individually - do not describe multiple sounds in a single annotation (e.g., by connecting separate sound descriptions with words like while, after, or before). Instead, annotate a separate region for each sound.
     Incorrect: "A dog is barking and people are talking."
     Correct: "A dog barking loudly" and "Several people talking indoors" with their respective onsets and offsets.

Once you have completed an annotation, click the `Submit` button.

The more effort you put into these annotations, the better the quality of the data set for the later phases of the project.

Step 4: Annotate the first 40 Recordings in your Queue

Annotate the first 40 recordings in your queue. Based on our experiments, we estimate that the annotation process can take up to four hours if you do it thoroughly (i.e., on average 6 minutes per annotation task; the difficulty between audio recordings varies greatly). We highly recommend splitting the overall workload across two or three days to avoid fatigue.

Annotation Quality & Grading

You will receive full points if you stick to the annotation guidelines on Moodle and annotate in a best-effort manner. However, if your annotations are clearly not best-effort or deviate from the instructions, your previous attempt becomes null and void, and we will ask you to *repeat the full annotation task.*

Disclaimer

All recordings were sourced from the Freesound platform, which is a valuable source of free sound samples (e.g., for the DCASE community[1]). Please note that it was not possible for us to check all recordings for problematic content. Recordings or their metadata might include disturbing content, including but not limited to violence and death. We apologize for any discomfort this may cause. If you find any of the content disturbing, please inform us (email *and* in the comments) and skip the corresponding annotation task. We will remove the corresponding annotation task. Please provide a short explanation in the skip text field. A skipped sound does not count towards the total number of annotated sounds (i.e., you need to annotate the first $40+$ the number of skipped sounds in your task queue).

Help & Contact

If you experience issues with the annotation interface, please post a message in the Moodle Forum.

If you are unable to complete this task due to disabilities or other limitations, including but not limited to hearing, vision, or motor impairments, let us know. We might ask you to file an official request to drop this task through the *Institute für Integriertes Studieren* at JKU.

Acknowledgment

We'd like to thank HumanSignal for providing us with the enterprise version of Label Studio for free.

[1] DCASE website