Deliverable D4.9

Evaluation report on the tool for manual annotation of musical content

Grant agreement nr	688382
Project full title	Audio Commons: An Ecosystem for Creative Reuse of Audio Content
Project acronym	AudioCommons
Project duration	36 Months (February 2016 - January 2019)
Work package	WP4
Due date	July 31st (M30)
Submission date	August 6th (M31)
Report availability	Public (X), Confidential ( )
Deliverable type	Report (X), Demonstrator ( ), Other ()
Task leader	MTG-UPF
Authors	Xavier Favory, Eduardo Fonseca, Frederic Font
Document status	Draft (), Final (X)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Table of contents

Table of contents 2

Executive Summary 3

1 Motivation 4

2 The AC Refinement Annotator 6

3 Experiment 9

3.1 Methodology 9

3.1.1 Task 9

3.1.2 Context, participants and procedure 9

3.1.3 Survey 10

3.1.4 Interview 10

3.2 Results 10

3.2.1 Survey 10

Usability 10

Engagement 11

3.2.2 Produced labels 12

3.2.3 Interviews and transcriptions 13

4 Discussion 14

4.1 Particularity of the task 14

Sound sources are difficult to recognize 14

Complexity of the categories 14

Highly variable amount of effort produced 14

4.2 Useful features 14

5 Conclusions and Future Work 16

5 References 17

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Executive Summary

This deliverable presents the Audio Commons Refinement Annotator, a tool for the manual refinement of previously existing annotations of audio content. It is a web-based interface that intelligently guides users on the refinement process of annotations from a large variety of sound concepts. This tool is being integrated into Freesound Datasets, a platform for the creation of open audio datasets.

One of the challenges in making use of Creative Commons audio content comes from the fact that it is provided by various sources and authors with different backgrounds and levels of expertise. Therefore, the content is often unstructured and not properly annotated, which hinders its efficient retrieval. Moreover, there is a scarcity of tools and agreed methods to aid users in the task of annotating audio content through established common procedures. Intelligently guiding users on the annotation process would allow a reliable, uniform and complete description of the content which will therefore facilitate its sharing.

In the first section of this document, we motivate the need of novel tools for the manual review and refinement of annotations previously assigned to some audio resource. More specifically, we focus on annotating content with a large set of predefined concepts. The deliverable D4.6 Release of tool for the manual annotation of musical content presents a prototype of a manual annotation tool, which guides the annotation process by providing an iterative approach for defining first the content type and then the relevant music related properties. On the other hand, the deliverable D5.4 Release of tool for the manual annotation of non-musical content describes a tool that allows to carry out the annotation tasks of both *generation* and *refinement*. A preliminar evaluation of this prototype revealed that it was prefered to provide a simpler and more focused tool. Therefore, in this document we present a tool that shares the ideas from D4.6, and further extends it in the context of refining pre-assigned labels. We call this tool the Audio Commons Refinement Annotator, and it is specially designed for the refinement of labels belonging to a large-vocabulary of audio concepts organized in a hierarchy, including musical attributes.

In the second section, we introduce the Audio Commons Refinement Annotator, a web-based tool for the refinement of pre-assigned labels - which guides the user in the process verifying and modifying annotations of audio samples with a wide range of sound categories. We present its evaluation carried out with four users, for which we applied a mixed methods approach combining human-computer interaction (HCI) metrics with behavioral and qualitative data analysis. We propose a topic-oriented discussion about the challenges arisen and possible solutions when annotating audio content in a post-process scenario such as the Freesound Datasets platform. Finally, we end this report with a summary of the work done and sketch the next steps to be carried out for the integration of the tool in a crowd-sourcing scenario.

This deliverable is complemented by Task 5.4 which focuses on evaluating another tool for the manual annotation of audio content.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

1 Motivation

In a previous deliverable we presented a prototype for the manual annotation of musical content (D4.6 Release of tool for the manual annotation of non-musical content). There, we proposed a web-based interface that guides users on the annotation process of music sample and music pieces. The design of the prototype was based on a rather basic hierarchical scheme where first the user identifies the type of musical content she wants to upload, and then she selects musical attributes from a set of properties relevant for the selected content type.

On the other hand, in the deliverable D5.4 Release of tool for the manual annotation of non-musical content, we proposed a tool for the manual annotation of non-musical content, although it was designed to allow the annotation of audio with a large vocabulary of sound concepts, including musical and non-musical content. In particular, this tool was implemented for the use case of Google's AudioSet Ontology - a hierarchically structured collection of 632 categories of everyday sounds [Gemmeke17]. This interface was developed in order to deal with two annotation issues: i) the lack of labels, and ii) the lack of specificity in the labels.

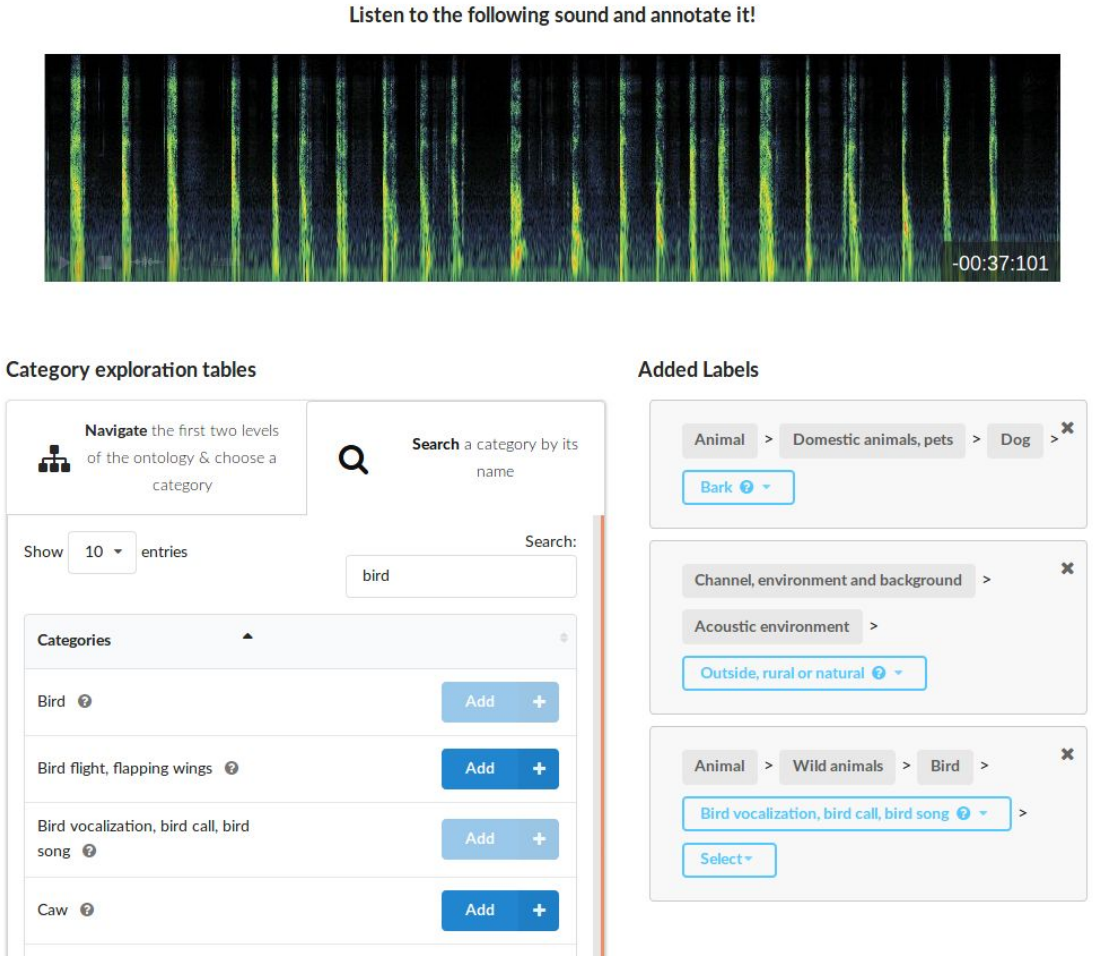

Figure 1. Screenshot of the annotator interface of D5.4 implemented in the Freesound Datasets platform.

As can be seen in Figure 1, the exploration tables (Figure 1 left) attempt to facilitate the *generation* of new labels, while the *refinement* stage (Figure 1 right) allows to further specify the labels. Both of this

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

problems lead to unstructured and not properly annotated Creative Commons licensed audio content and hence they are important issues that hinder its sharing and retrieval.

A preliminary evaluation of this tool revealed its significant complexity, requiring a substantial effort on the user side. Nevertheless, the interface was found to have great potential and usefulness. In order to simplify the tool and lower the barrier for its usage, we decided to split such tool into two independent tools: one focused on the generation of labels and another focused on the refinement of existing labels. The generation tool is described and evaluated in D5.5 (Evaluation report on the tool for manual annotation of non-musical content).

The refinement stage of the interface shown in Figure 1 is based on narrowing down the context of a label by moving down a hierarchy of sound concepts, conceptually much like the prototype of D4.6. Given their shared philosophy, we decided to take the best ideas of both and further develop an improved and more versatile interface for the hierarchical specification of labels. This tool, which is described and evaluated in this document, allows to remove, refine or specify a number of existing labels associated to an audio resource.

Furthermore, by doing this we obtain a tool that is complementary to that of D5.5, i.e. it attempts to solve the refinement issue using a simple approach that complements that of the generation tool. Similarly to D5.5, we are interested in applying this methodology in the context of large-vocabulary sound taxonomies that can be used for both musical and non-musical sounds. In particular, we use the recently released AudioSet Ontology as use case - a hierarchical collection of categories that has been well received by the research community and that includes more than 100 musical-related categories encompassing from musical instruments to music genres or moods. However, the tool presented here could be applied to any music-related taxonomy following a hierarchical graph.

This tool can be useful, for example, for the refinement of huge amounts of Creative Commons audio scattered across the web that can be part of the AudioCommons Ecosystem. Much of this content presents annotations of some sort, either provided by the users that uploaded those audio contents or automatically generated by third party algorithms. However, the fact that this type of audio content is already labeled does not necessarily mean that the labels are the most appropriate. Indeed, it often occurs that these labels lack the sufficient specificity to allow proper reusability of the content. For example, the labels *Organ*, *Piano* or *Harpsichord* can be much more descriptive and useful than a simple *Keyboard*. From the perspective of sound designers and in general creative minds, having labels that tell the differences and nuances among different instruments (for example) can be interesting for certain applications. Moreover, organizing these aspects with a well-established, rich-enough taxonomy, such as that of AudioSet, will promote uniformity and consistency in the labeling, which in turn will ease the retrieval and reusability of the audio content.

Finally, it is worth mentioning that the tool presented here will be deployed in the Freesound Datasets platform,[1] a platform developed by our group for the collaborative creation of open audio datasets labeled by humans. In this context, it will be instrumental as a post-annotation step in which users of the platform collaboratively contribute to the annotation of content and ground truth generation.

[1] https://datasets.freesound.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

2 The AC Refinement Annotator

The development of the first manual annotation tool of musical content presented in D4.6 showed the potential of giving a way to iteratively define music properties. More specifically, the Audio Commons Music Annotator prototype allows, for a given audio sample, to first choose a content type and then annotate relevant properties associated with the content type previously selected. Moreover, the informal evaluation of the prototype presented in D5.4 revealed that this idea of iteratively defining sound related properties was appreciated by the users, and seemed to ease the manual annotation process when relying on a large set of acoustic categories. This approach allows to narrow down the context of a label, by taking advantage of the hierarchical relationships provided by a taxonomy such as AudioSet. In this deliverable, we present the evaluation of the AC Refinement Annotator, which takes inspiration from the two mentioned prototypes.

Annotation of audio content can be useful for several use cases, e.g., when a provider publishes content in the Audio Common Ecosystem, or in a post-processing stage, where users collaboratively annotate the content, for instance in the Freesound Datasets platform. The tool allows to focus on a single sound resource at a time, which is accessible from a player displaying the spectrogram of the sound to facilitate the localization and understanding of sound events in the clip (Figure 2). Some labels are previously proposed. These labels are automatically recommended based on a tag-matching method which takes advantage of Freesound uploaders' tags[2]. For the proposed labels, the annotator can examine its location in the hierarchy. He also can examine siblings and children of the proposed categories. Figure 3 shows how the children categories of the proposed label "Guitar" are displayed in a dropdown, which allows to modify the label and define it more precisely. For all the labels, some popup show description and examples when available (Figure 4). Moreover, it is possible to duplicate a label using the copy icon at the top right corner. This allows for instance to specify a label by adding two of his children categories.

When reviewing the proposed labels, the user is also asked to verify the *presenceness* in the audio clip. The user must choose one of the following response type:

Response type	Meaning
Present and predominant	he type of sound described is clearly present and predominant. This means there are no other types of sound, with the exception of low/mild background noise.
Present but not predominant	The type of sound described is present, but the audio clip also contains other salient types of sounds and/or strong background noise.
Not Present	The type of sound described is not present in the audio clip.
Unsure	I am not sure whether the type of sound described is present or not.

Table 1. Response types for the validation of a proposed or refined category.

A typical workflow would consist in:

1. Listen to the sound sample (Figure 2, top)
2. Inspect the proposed labels (Figure 2)
3. Refine the proposed labels by inspecting the related siblings and children (Figure 3 & 4)
4. Validate the presence of the proposed or refined category

[2] A list of tags were manually assigned to all the AudioSet Ontology categories, which allowed to perform tag-based queries to the Freesound database. More information can be found in [Fonseca17].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

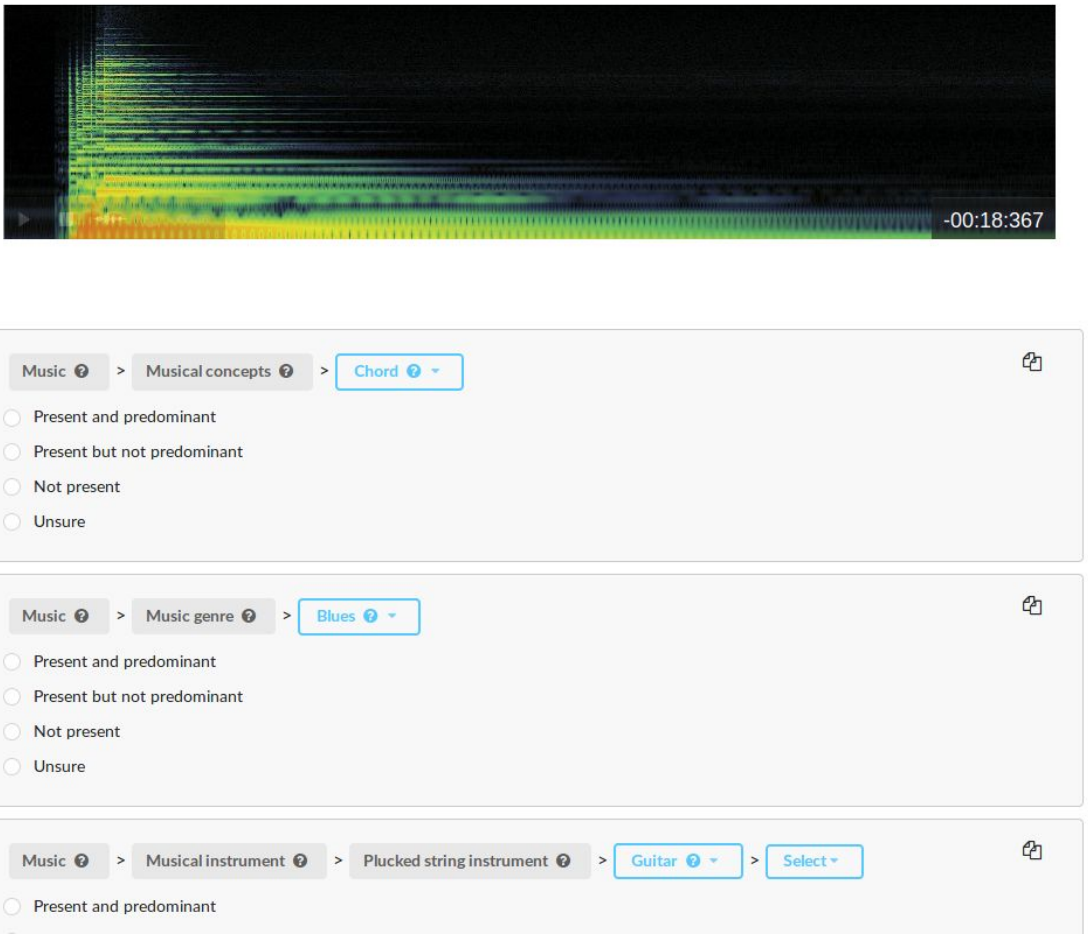

Figure 2. Screenshot of the Audio Commons Refinement Annotator displaying a sound sample and its three suggested label *paths*.

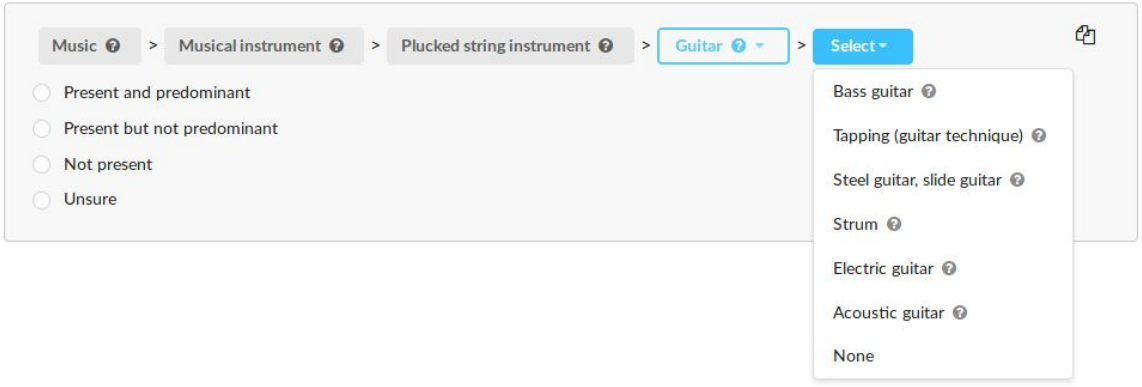

Figure 3. Screenshot of the Audio Commons Refinement Annotator showing a dropdown displaying the children categories of “Guitar”.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382

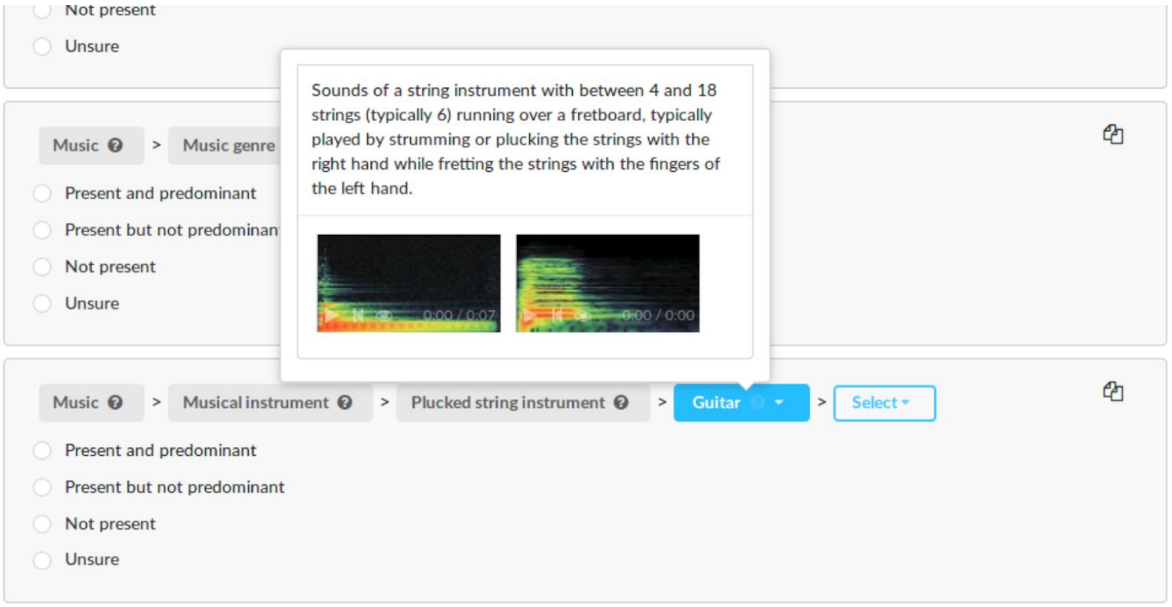

Figure 4. Screenshot of the Audio Commons Refinement Annotator showing the description and examples of the “Guitar” category in a popup.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382

3 Experiment

The main characteristic of collaborative multimedia collection is that the content is provided by people from different backgrounds and expertises. This tends to produce content that is not uniformly annotated, and does not allow its efficient retrieval. Moreover, crowdsourcing emerged as a powerful tool for making the process of annotating large sound collection scalable. In these contexts, there is a need for proposing new manual interfaces to properly annotate audio content, with labels that are comparable and of same nature. In this experiment, we advance our user-driver design process of proposing new annotation tools for annotating audio content from a large variety of types. We take advantage of the AudioSet Ontology which provides a hierarchical taxonomy of very broad acoustic categories. We use the Audio Commons Refinement Annotator as a technology probe to observe its use in a real context, to evaluate its functionalities and to inspire new ideas [Hutchinson03]. One of our goals is to propose a method that will guide people in providing annotations that are as consistent as possible.

3.1 Methodology

3.1.1 Task

We selected some sounds from the Freesound Datasets platform featuring one or more of the following aspects: (i) containing multiple sources, (ii) presenting background noise or (iii) hard to recognize. This process resulted in a list of 15 sounds whose labels must be refined by the users with the help of the Audio Commons Generation Annotator. At the end of the task, they were provided a questionnaire, followed by a semi-structured interview.

3.1.2 Context, participants and procedure

We gathered four participants with different level of expertise. We will use A, B, C and D letters for referring to them in this document. B is quite familiar with Freesound content and the challenges around its accurate annotation. A and D have a bit of experience in using other annotation tools for the annotation of audio content. C is rather not very familiar with this sort of work. All the participants were non native english speakers, but declared to have an excellent level.

Some guidelines were shown to them, together with verbal explanations given by the examinator. First, the context of the study was explained: evaluate novel interfaces for the manual annotation of audio content with large vocabulary of sound concepts. They were instructed, for every audio clip, (i) to evaluate the label according to the four response types; (ii) if the evaluation result is one of the “present” responses, try to specify the label as much as possible by going deeper down in the hierarchy; (iii) also, it is possible to duplicate a label i order to add other related labels that are in a close hierarchical level. The taxonomy was presented as a hierarchical structure containing over 600 audio categories. Upper levels in the hierarchy contain broader sound concepts while lower levels are formed by more specific categories. These categories mostly include concepts related to: (i) Sound events, source or production mechanisms (e.g., ‘Bark’, ‘Tearing’, …); (ii) Categories describing aspects of sound (e.g., ‘Reverberation’, ‘Boing’, ...).

While users were performing the task, they were asked to think out loud, and share their comments or doubts. The examiner was present during all the experiment, making sure that no major issue was preventing participants from successfully performing the task, to support them in case of doubts, and to transcribe relevant participants actions and comments.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382

3.1.3 Survey

The survey was divided in two parts. It first included usability related questions (SUS usability scale) [Brooke96], and then overall feedback on engagement and learning. The SUS questionnaire investigates dimensions related to interest, complexity, ease of use and simplicity/difficulty, integration and consistency. The overall feedback asses the English language level of the participant, and the levels of engagement, learning, novelty and quality of category retrieval.

3.1.4 Interview

We conducted semi-structured interviews with the four participants after they completed the task and the survey. We used open-ended questions and specific questions related to observed behaviours during the performance of the task. We used thematic analysis to identify emerging themes from participants' answers. These are the questions that were asked to participants, from the which some interesting discussions emerged:

- How intuitive was the interface?
- What about the different features?
    - Dropdown
    - Popup
    - Duplicate
- How difficult was the task?
- Did you have any doubt? How did you react? Could you solve it?
- Would you find it useful to see the sounds metadata from Freesound?
- Would it help to have consecutive sounds with similar labels (one topic)? Like choosing a family?

3.2 Results

3.2.1 Survey

Usability

No significant differences were found that correlate with the level of expertise of the four participants. Figure 5 illustrates the results obtained to the usability questionnaires (SUS). Participants strongly agreed that the interface was easy to use. They also tended to agree to items related to confidence in using the system and the quick learning of its use. The participants found the system slightly unnecessarily complex. We believe this is due to the taxonomy complexity, which makes it sometimes hard to retrieve some categories, as can be observed in Figure 7.

As a result, the Audio Commons Refinement Annotator obtained on overall a relatively high usability score according to the SUS metric (M=77.5, SD=8.90). Figure 6 shows the individual usability score for the four participants.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

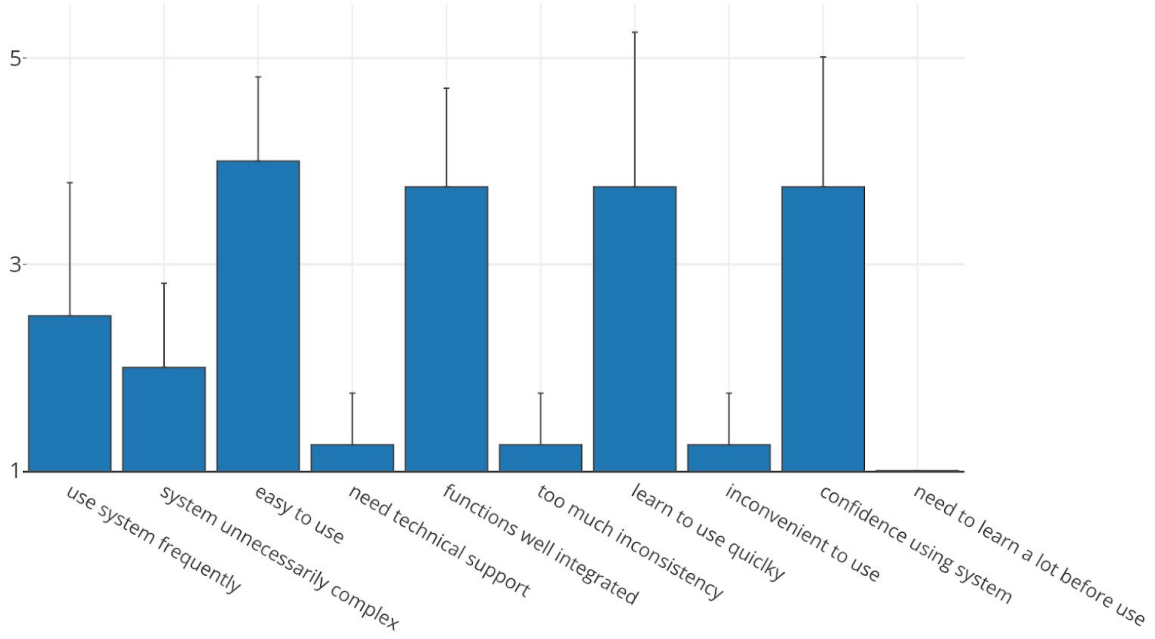

Figure 5. Mean and standard error for the usability questionnaire items (SUS). Value of 1 corresponds to strongly disagree, 3 to neutral, and 5 to strongly agree.

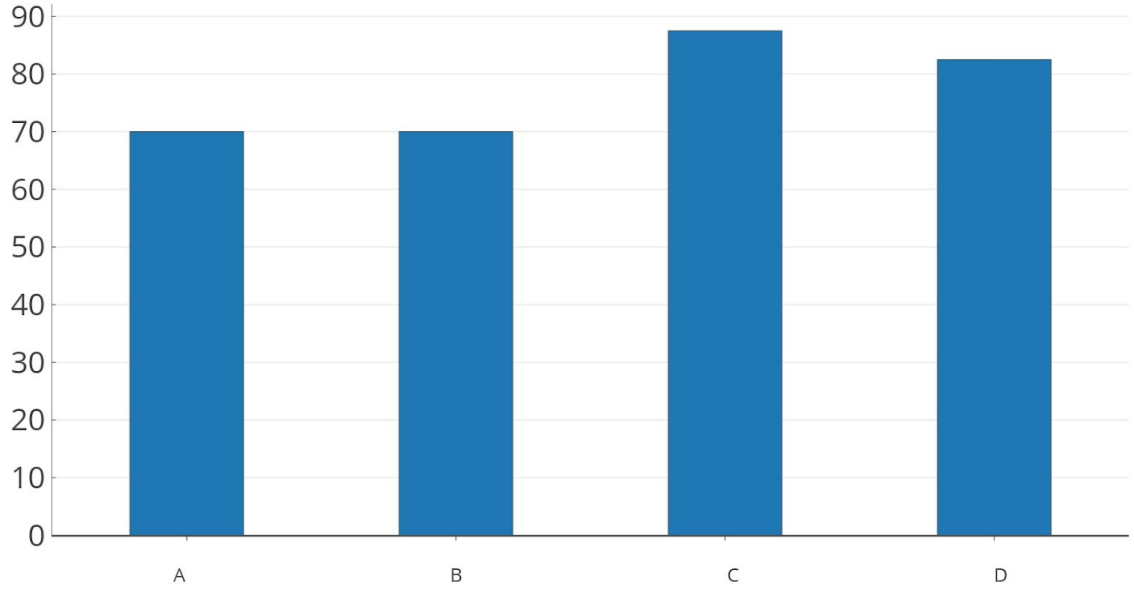

Figure 6. Individual overall usability scores for the four participants

Engagement

Further analysis was conducted on the learning and satisfaction related to category retrieval. Figure 7 shows the mean and standard error of the different questions. Participants did not totally agree on the amount of knowledge they learnt by performing the task with the proposed interface. The system

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

appeared to be quite novel to all of them, whereas the category retrieval performance was not very high.

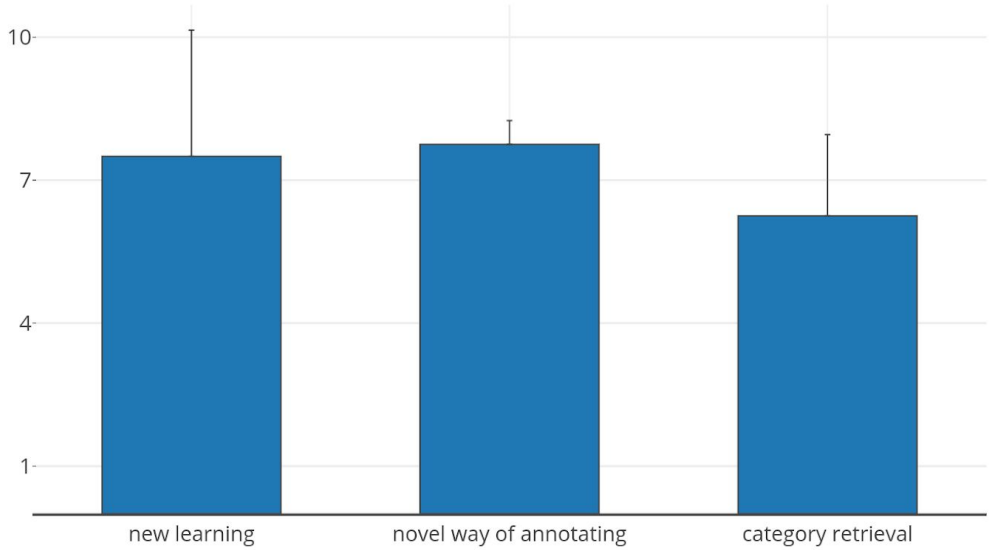

Figure 7. Mean and standard error for the overall feedback questionnaire.

3.2.2 Produced labels

As one of the main ideas is to investigate how different annotators from different levels of expertise annotate sounds, we propose next some statistics that allows us to quantify it. Table 2 shows the total number of labels produced by the participants, and the amount of time they spent annotating. We call *path* to the entire list of labels from the same hierarchical taxonomy tree path (which correspond to each of the label paths that can be seen in Figure 2). We observe that all the annotators produced a similar number of labels for each sound.

Annotator	Total number of labels from paths verified as *Present* (15 sounds)	Total number of paths verified as *Present* (15 sounds)	Time spent annotating
A	119	34	1 hour 20 minutes
B	93	26	55 minutes
C	98	29	35 minutes
D	88	23	40 minutes

Table 2. Number of total labels produced for the selected list of 15 sounds

However, when comparing the provided labels, only few are common among participants. Table 3 shows the average number of provided labels that are common between pairs of annotators, e.g. common to users A and B, to users B and C, and the rest of pairs among the four users A, B, C and D. This suggests that several annotators are needed when annotating content with such a large taxonomy of concepts.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Average of pair annotator common labels	Total number of different label produced
59.2	129

Table 3. Average pair annotator commons labels produced and the total of different labels produced

3.2.3 Interviews and transcriptions

Next we provide selected feedback gathered from the interviews and transcriptions of the participants. Some of them can be somewhat opposite, presumably resulting from the different levels of users' expertise.

- It is not possible to know in advance what children categories are included under a category. Participants would like to see all the ontology. Alternatively, it would be nice to see the children when exploring a label, a visual tree for instance.
- It is difficult because of taxonomy complexity. You need to understand the taxonomy, sometimes its complexity does not help.
- There are some other sounds present, but I cannot add them since they are not proposed.
- Doing only precision of labels would be easier and faster. Removing siblings would minimize some confusion when exploring the ontology.
- It is easy to get familiar with a category by exploring with the dropdown and duplicating labels and adding them.
- I do not remember which was the original proposed label after having made changes.
- Since there is no sound example in this category, I won't explore it.
- This "Wind" category seemed not appropriate, but after having checked its children, I saw a category that was very much appropriate.
- People use spectrograms to localize specific sounds and re-listen to them carefully.
- The task is a little bit deconstructed. The goal should be more explicit.
- Technically, a "Vehicle" is a "Mechanism", but it does not seem to be adequate to add both.
- When I was not confident, I did not try too hard, voted Unsure and went to the next label, not to lose time and effort. I dropped it for things that seemed to demand high competences. I spent more efforts in more basic and mechanical categories.
- It's difficult because I cannot identify well the speaker in the sound, and I see that I should specify more.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

4 Discussion

4.1 Particularity of the task

Sound sources are difficult to recognize

In the context of post-process annotation of audio content, the annotator is typically not the publisher of the content. Hence the annotator usually does not know how the recording conditions were, or what sources were captured. Listening to the sound does not necessarily lead to the identification of the source. However, by proposing some labels, people were guided towards the identification and specification of the sources. When they were not able to identify more precisely the source, they were stopping at a certain level of the taxonomy.

Complexity of the categories

The taxonomy used in the task is presented as a hierarchical structure containing over 600 audio categories. Upper levels in the hierarchy contain broader sound concepts while lower levels are formed by more specific categories. The nature of the categories included varies to a high extent, as detailed in D5.4. In this task, people were guided by being able to iteratively define the labels. However, sometimes this was a source of confusion. For some categories, it is hard to guess in advance what more specific children categories they include. For instance, “Natural Sounds > Wind” includes “Wind noise (microphone)”, which may not be expected since the sound originates from the microphone system (and hence can be understood as a non natural sound).

Highly variable amount of effort produced

It was observed that the different users spend a highly variable amount of time performing the task (from 35 minutes, up to 1 hours and 20 minutes, see Table 2). We believe there are several reasons for this: (i) the task instructions indicate that it is possible to duplicate labels in order to generate additional, more exhaustive annotations; (ii) users were instructed to check the siblings of the proposed category, which can lead to a significant amount of time for category inspection. The combination of these two issues made some users duplicate quite a few labels, and navigate through their siblings. It seems reasonable to think that the task is somewhat complex as, in a way, it is mixing two different annotation exercises to some extent. That is, the task moves from a purely *refinement* scenario towards another one that includes shades of *exploration* and/or *generation*.

4.2 Useful features

The proposed labels originate from an algorithm based on tag-based queries performed on the Freesound database. The quality of this method varies a lot across the categories drawn from the AudioSet Ontology. This is partly due to the quality of the original tags provided by the sounds’ uploaders, who sometimes add terms that are not accurate. Moreover, they occasionally fail to represent some aspects of the sounds that could be covered with our set of predefined categories. Another source of poor quality comes from the complexity of the natural languages. Polysemy and synonymy can affect our system’s precision and recall performances respectively. Moreover, people tend to use different words which have contrasting meaning for the same thing. When relying on a large set of concepts, there are slight variations of meaning between the categories, that people do not originally consider.

However, the main issue that our tool tries to address is the lack of specificity in the previously existing labels. The hierarchy structuring the audio related concepts assumes that some categories convey more information that others. Therefore, it is important to use labels as specific as possible in order to accurately describe the audio content. Some participants complained that all the hierarchy could not be seen at once. However, we believe that providing a way to iteratively specify the labels is

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382

indeed helpful, since it eases and speeds up the generation of accurate labels by focusing on the most relevant semantic audio context.

Nonetheless, the AC Refinement Annotator allows to do more than just precising labels. It also makes possible to explore sibling categories that sometimes correspond to slightly different concepts. This would allow to correct the automatically generated labels that could result from imprecise user generated tags. Yet, this feature led people to produce variable results in terms of labels produced. It also yields people to get lost in the set of concepts, making them wasting a lot of time. Some of the participants put a lot of effort in checking the siblings, making the task very slow and not actually scalable.

The AudioSet Ontology provides a set of audio related categories together with their descriptions. We augmented it with some sound examples from the Freesound collection. These descriptions and examples are shown in popup elements. Interesting behaviors were observed related to the navigation though the different levels of specificity in the hierarchy. For instance, a participant was sometimes not inspecting, or hesitating to check the children of a category. This occured due to several reasons: (i) since no sound examples were available in the present category, he assumed this would also be the case in deeper hierarchy levels. Hence he decided not to explore this branch due to lack of confidence with it; (ii) He also assumed that since the original category was not appropriate, none of the children would be either (where in fact, one of them was). To mitigate this problem and ease quick inspection of the categories, the popups could include the list of the children categories. In this way, users would not be forced to select the category in order to see what it contains.

The *duplicate label* feature allows annotators to produce more exhaustive and complementary annotations. For example, they can select sibling categories that share a common parent (e.g., “Meow” and “Purr”, children of “Cat”). On the other hand, this feature also makes that users can spend a fair amount of time duplicating the proposed label and exploring the related siblings and children. Currently, this feature duplicates the entire hierarchy path of the labels, from the broadest level in the hierarchy. However, since the main goal is to further specify the proposed label, it would be optimal to duplicate only the essential part of the branch. That is, starting from the proposed label covering the deeper levels of the hierarchy (instead of duplicating the entire path). By doing this, redundant information would not be shown, thereby making the interface clearer and simpler.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382

5 Conclusions and Future Work

In this deliverable we presented the Audio Commons Refinement Annotator, which is used for refining previously existing annotations of audio samples with labels from the AudioSet hierarchical Ontology. We evaluated it using a mixed methods approach combining HCI metrics with behavioral and qualitative data analysis. The results show that the four participants of the study found the system easy to use. The tool seemed to facilitate browsing and using large taxonomies of concepts for annotating audio content. Proposing some automatically generated annotations, together with a fixed vocabulary for the annotation of the content allows to gather more consistent annotations across different annotators. However, the complexity of the task suggests that it could be simplified. Since we evaluated it in a post-process scenario, we see advantages in including this as a new annotation tool in Freesound Datasets, our platform for the collaborative creation of open audio datasets. This new tool can be combined with the validation task already present in the platform, and the tool presented and evaluated in D5.5.

Future work includes improvements on the design, such as making clearer which were the labels originally proposed, or adding some information about children categories in the popups. Moreover, some simplification of the task might be needed. Currently, the tool allows users to produce very different amount of effort, and for instance can lead people to waste a considerable amount of time trying to explore the taxonomy of concepts. For example, the task could focus only on the *precision* of labels, which would consist only in providing more precise annotations for labels corresponding to concepts present in the audio sample. Finally, since this tool will be integrated in the Freesound Datasets platform, quality control mechanisms must be designed to make it suitable for crowd-sourcing.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

5 References

[Fonseca17] Fonseca, Eduardo et al. (2017). “Freesound Datasets: A platform for the creation of open audio datasets”. In: Proceedings of the International Society for Music Information Retrieval Conference.

[Gemmeke17] Gemmeke, Jort F et al. (2017). “Audio Set: An ontology and human-labeled dataset for audio events”. In: Proceedings of the Acoustics, Speech and Signal Processing International Conference.

[Brooke96] Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.

[Hutchinson03] Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., ... & Roussel, N. (2003, April). Technology probes: inspiring design for and with families. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 17-24). ACM.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382