Deliverable D4.11

Evaluation report on the second prototype tool for the automatic semantic description of music pieces

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	31 October 2018 (M33)
Submission date	31 October 2018 (M33)
Report availability	Public (X), Confidential ( )
Deliverable type	Report (X), Demonstrator ( ), Other ( )
Task leader	QMUL
Authors	Johan Pauwels, Anna Xambó
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

Background 4

1 Introduction 5

1.1 Analysis service outline and improvements 5

1.2 Evaluation procedure 6

1.3 Terminology 6

2 System Description 7

2.2 Back-End 7

2.3 Front-End 8

2.4 Visualisations 9

2.4.1 Linear Visualisation 9

2.4.2 Circular Visualisation 9

4 User Study 11

4.1 Study Design 11

4.2 Findings 11

4.3 Lessons Learned 16

5 Conclusion 17

6 References 18

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

Executive Summary

As part of the Audio Commons Ecosystem, a number of tools are provided for the automatic analysis of audio content without the need for human intervention. These tools are designed for extracting i) musical audio properties for music pieces and music samples, and ii) non-musical audio properties for any kind of sounds. A second prototype of these tools has been recently published as a webservice and this deliverable will discuss the evaluation of the tool for the semantic annotation of music pieces.

Concretely, a web app named “Jam with Jamendo” has been created to showcase new possibilities for discovering music that are made possible by providing audio-based descriptors in the Audio Commons Ecosystem. In this app, users are invited to select a number of chords and the system will then return music pieces that contain them so that users can play along with the music and thereby practice the chords.

We completed a study in which seven participants were invited to try out the system and evaluate two forms of chord visualisations. The feedback of this study will be used to inform further iterations of the application.

The chord descriptors (including confidence measure) that are used in the app are available through a public API, but the search procedure was customly developed for this application. In the future, we aim to expose similar search functionality for chords and other descriptors through a public API.

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

Background

WP4 has the aim of developing tools for the semantic annotation on musical pieces. Towards this goal, Deliverable D4.3 described the implementation of several automatic annotation tools. An evaluation of these tools was performed in Deliverable D4.5. One of the points for improvement was that the first prototype consisted of a number of fragmented tools that weren't easy to use. To address this point, the second prototype, described in Deliverable D4.8, unified these tools into a web interface that provides a single consistent (programming) interface that is easier to use for developers in the Audio Commons Ecosystem.

Furthermore, the tools of the prototype were working offline, whereas another motivation to turn this prototype into an online service was to make it easier to integrate the analysis into the rest of the Audio Commons Ecosystem (consisting of content providers and a mediator connecting them together and providing a central point of access). While the integration of this analysis service with the mediator is still ongoing, the availability on the web of the descriptors computed by the service already opens up new possibilities for applications.

This deliverable focuses on evaluating the benefits of online availability of descriptors the semantic descriptors within the ACE. We present a case study in which an audio-based descriptor is used to augment the metadata of a content provider. This allowed us to create a web app that makes a novel way of exploring and retrieving music content possible.

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

1 Introduction

1.1 Analysis service summary and recent improvements

Since it has been reported first in [D4.8](), the analysis service has undergone some small changes. As a reminder, it consists of a single API endpoint, which has now moved to http://audio-analysis.eecs.qmul.ac.uk/function/ac-analysis. In order to make its input format more compatible with the mediator search results of WP2, the file to process is now passed as the single option `id`, in the format `content-provider:provider-id`.

Content of the Europeana archive has now been integrated as well, so possible values for the content-provider are:

- jamendo-tracks
- freesound-sounds
- europeana-res

The requested descriptor can be specified by appending its name to the URL as http://audio-analysis.eecs.qmul.ac.uk/function/ac-analysis/<descriptor>. The essentia music descriptor developed by the MTG has been added to the service, so the current list of supported descriptors includes:

- chords (with confidence measure)
- instruments
- beats-beatroot
- keys
- essentia-music

Finally, a content-type for the response needs to be passed as an HTTP header, which can be one of:

- application/json
- text/plain
- text/rdf
- text/csv

Not all descriptors support all content-types, but specifying an unsupported one will return an error message that lists the supported content-types.

All together, these elements form an API call of the form below:

```
curl -v
'http://audio-analysis.eecs.qmul.ac.uk/function/ac-analysis/<descriptor>?id=<content-provider>:<id>' -H 'Content-Type: <content-type>'
```

For example:

```
curl -v
'http://audio-analysis.eecs.qmul.ac.uk/function/ac-analysis/chords?id=jamendo-tracks:1560969' -H 'Content-Type: application/json'
```

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

1.2 Evaluation procedure

The most straightforward usage of the analysis API is to use it to complement the metadata of the music pieces in the Audio Commons Ecosystem with audio-content based descriptors. For this, a single lookup of the descriptors needs to be made based on the id of the music piece. These descriptors can be calculated on request whenever needed. By precalculating descriptors and storing them in a database, however, new retrieval scenarios become possible based on querying by descriptors.

As a use case of this querying by descriptors, a music discovery web app called "Jam with Jamendo" was developed that aims to suggest new practice material to music learners of all skill levels. It is based on a query-by-chords (QbC) principle. Users select a number of chords they want to practice in a graphical interface and the system returns a list of music pieces that contain those chords. By selecting a music piece, a player is made available that displays the chords synchronised to the audio, such that the users have an indication of what chords to use when playing along with the music. A user study was performed in which this web application was presented to seven participants in order to gain feedback on the prototype. The findings of this study are presented in this report.

The search functionality that accesses the caching database of the analysis service is currently not part of the public API, but was customly developed for this app. The plan is to make this search functionality extensible such that it could include descriptors other than chords and make it publicly available through the API too. This way, third parties can develop similar novel ways of discovering Creative Commons content based on content-based descriptors.

1.3 Terminology

AudioCommons: reference to the EC H2020 funded project AudioCommons, with grant agreement nr 688382.

Audio Commons Initiative: reference to the AudioCommons project core ideas beyond the lifetime and specific scope of the funded project. The term "Audio Commons Initiative" is used to imply i) our will to continue supporting the Audio Commons Ecosystem and its ideas after the lifetime of the funded project, and ii) our will to engage new stakeholders which are not officially part of the project consortium.

Audio Commons: generic reference to the Audio Commons core ideas, without distinguishing between the concept of the initiative and the actual funded project.

Audio Commons Ecosystem (ACE): set of interconnected tools, technologies, content, users and other actors involved in publishing and consuming Audio Commons content.

Audio Commons content (ACC): audio content released under Creative Commons licenses and enhanced with meaningful contextual information (e.g., annotations, license information) that enables its publication in the ACE.

Content creator: individual users, industries or other actors that create audio content and publish in the ACE through content providers.

Content provider: services that expose content created by content creators to the ACE.

Content user: individual users, industries or other actors that use the content exposed by content providers and created by content creators in their creative workflows.

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

Tool developer: individual users, industries or other actors that develop tools for consuming (and also potentially publishing) Audio Commons content.

Embeddable tools: tools for consuming Audio Commons content that can be embedded in existing production workflows of creative industries.

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

2 System Description

This section describes the implementation of the Jam with Jamendo application.

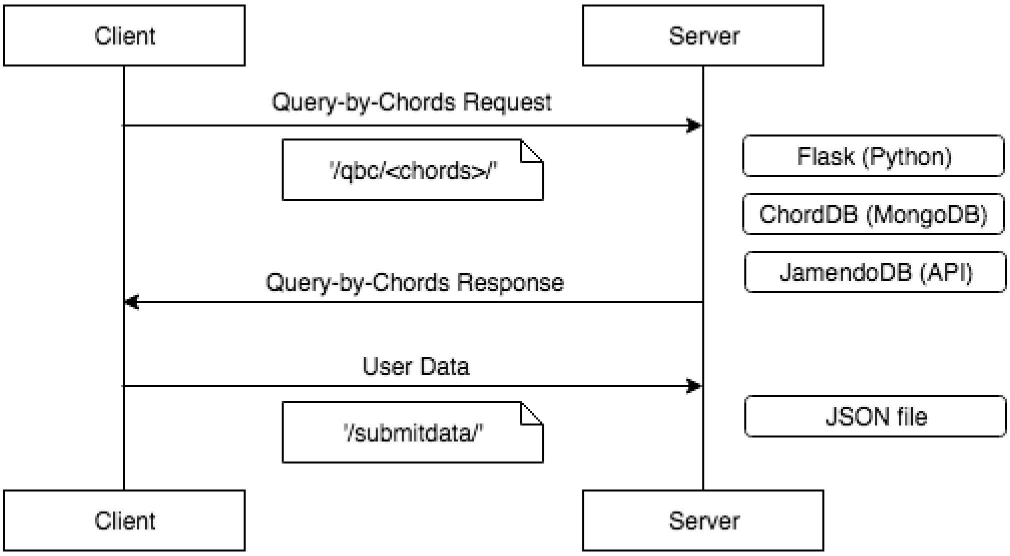


Figure 1: Architecture of the Jam with Jamendo system.

Figure 1 shows the architecture of the Jam with Jamendo application. This is a client-server web application based on a query-by-chords (QbC) approach. The system is designed for supporting music learning with a musical instrument (e.g. guitar, piano, bass guitar, harp). The workflow is based on letting users pick a set of chords that they know and want to practise. Based on this QbC request, a curated list of songs from the Jamendo music collection is provided. Information about the chords within the song is shown in synchrony with the music, so that music learners can follow along with their instrument. Compared to more traditional approaches to music learning, this approach allows the user to discover new bands that are licensed under Creative Commons.

2.2 Back-End

The back-end of the system consists of a MongoDB, a type of document-oriented database based on JSON. It is populated with automatic chord transcriptions of nearly 100 000 songs of the Jamendo Licensing[1] catalogue. Jamendo Licensing is an opt-in service offered by the Jamendo music platform that aims to license out the content of users who signed up to it for commercial purposes. The audio and metadata of these music pieces is accessible through the Jamendo API[2].

This catalogue is lightly curated for recording quality, ensuring that all songs within this are of sufficient technical standards to actively promote, without judging their artistic merits. Bad recordings, copyrighted, and joke submissions are flagged. Of the original 200 000 tracks that are present in the Jamendo Licensing catalogue, about half are retained by this process. The list of files

[1] https://licensing.jamendo.com
[2] https://developer.jamendo.com

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

that are of sufficient standard is not available through the Jamendo API, but was obtained directly from Jamendo. At the time when the dataset underlying Jam with Jamendo was created, 99 960 audio files were accessed through their API. These files were processed by the chord estimation algorithm presented in D4.5. In addition to the estimated chord sequence, the algorithm also calculates a confidence per file. This confidence indicates how sure the algorithm is of its output. This confidence measure was used to sort the list of suggested songs in decreasing order such that song where the chord transcription accuracy is high are ranked first.

A chord vocabulary of 60 chords was selected: five chord types for each of the 12 possible roots. The chord types are major, minor, dominant 7th, major 7th and minor 7th, selected as the most popular chord types in a study of popular music and provide a nearly complete coverage for most songs [Burgoyne, 2011]. Songs containing only the selected chords are then retrieved by the front-end and presented to the user. The decisions to only return songs containing only the chords selected was made to avoid users being presented with unrequested or unknown chords, which they possibly do not know how to play. A consequence is that the more chords are selected, the larger the subset of songs suggested. At the same time, we deem it unsatisfactory if a large number of chords are selected and songs get returned that contain only a small number of distinct chords. Therefore, preference is given to songs that contain as many as possible of the specified chords. In order to deal with the inevitable errors in chord output due to the automatic transcription system, the suggested songs that have the same number of matching query chords are further sorted in decreasing order of confidence in the estimated chord sequence.

2.3 Front-End

The front-end of the app has been built based on a model-view-controller approach. The technologies used include Python's Flask[3] (model), Bootstrap[4] (view) and jQuery[5] and JavaScript (controller). As Figure 1 illustrates, the model (Flask server) manages the QbC requests and communicates with the databases ChordDB (chord sequences and confidence) and JamendoDB (audio and metadata) to retrieve a list of songs that contain the requested chords, as explained in Section 2.2.
The view implements the query and the playback interfaces. In the QbC interface, users can select a set of chords based on their knowledge from a provided set of 60 chords. The playback interface shows a list of selected songs and two chord visualisation modes, a linear view and a circular view. The controller manages the user's interaction with the GUI and secures the asynchronous communication with the Jamendo API to stream the songs. A companion webpage with extra information and a demo video can be found online[6].

[3] http://flask.pocoo.org
[4] https://getbootstrap.com
[5] https://jquery.com
[6] https://www.audiocommons.org/jam

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

2.4 Visualisations

Traditionally, visualisations of single sounds are based on the perceptual domain, such as spectrograms (Isaacson, 2005; Orio, 2006). Here we propose to explore an intuitive, cognitive-perceptual model that allows the user to hear by seeing the chords. We provide two visualisations of the chords in the suggested songs for the user to play along with:

1. Textual information about the current and future chord only.
2. Textual information about the current and future chord within the context of a chromatic circle of chords.

2.4.1 Linear Visualisation

We implemented a linear visualisation inspired by the concept of a moving time axis. This approach is similar to the visualisation used in the Riffstation web application[7], which contains a train of moving chords. Our approach is stripped-down, more space-efficient version that focuses only on showing the present chord in black text and the next chord in grey text, as shown in Figure 2a.

2.4.2 Circular Visualisation

As illustrated in Figure 2b, we developed a circular visualisation inspired by the circular representation of chords. In particular, we implemented a chromatic circle, closely related to the circle of fifths. We used the web-based JavaScript library D3 [Bostock et al, 2011], which is designed to create interactive data visualisations. In particular we used the template for D3 sunburst visualisation (partition layout)[8]. The colours were arbitrarily distributed using an ordinal colour scale under the *schemeCategory20b* option[9].

[7] https://play.riffstation.com
[8] https://bl.ocks.org/denjn5/f059c1f78f9c39d922b1c208815d18af
[9] https://bl.ocks.org/pstuffa/3393ff2711a53975040077b7453781a9

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

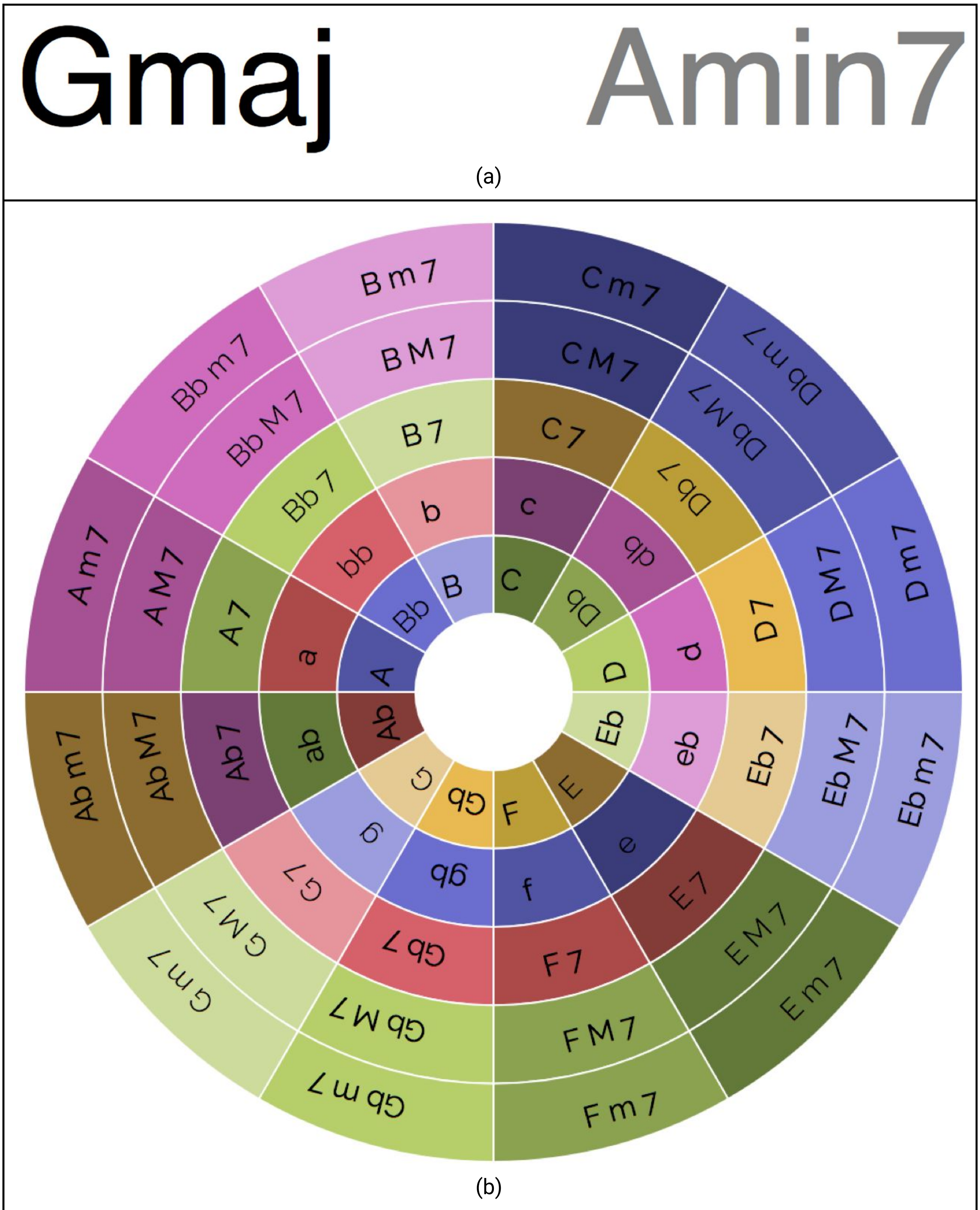


Figure 2: Screenshot of the linear (a) and circular (b) visualisation

4 User Study

4.1 Study Design

An experiment was designed so that we could observe either in person or remotely the participants interacting with the system while practising with their musical instrument.
With each of the participants, we had a conversation about their opinions on the visualisation modes while using the application inspired by thinking-aloud techniques [Rogers, 2011]. We captured their thoughts in real-time by taking notes as a form of verbal protocol.

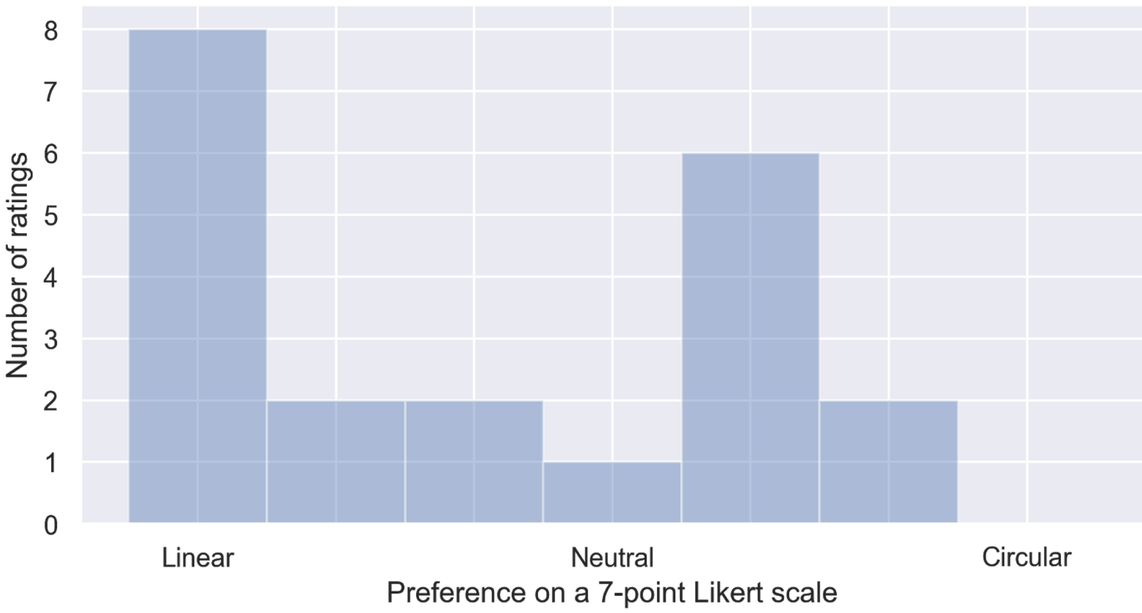
The participants were asked to bring their musical instrument or use a provided electric guitar and: (1) choose a set of chords that they know from the web application; (2) explore the suggested three songs containing the chords by using the two visualisation modes and playing along; (3) rate their preferred visualisation mode from a 7-point Likert scale ranging from "strongly prefer linear" (StP1), "prefer linear" (P1), "slightly prefer linear" (SlP1), "neutral" (N), "slightly prefer circular" (SlP2), "prefer circular" (P2) and "strongly prefer circular" (StP2). They were asked to repeat this task three times in independent trials. The task was expected to be completed in 45--60 minutes. A final post-questionnaire gathered demographic information.

4.2 Findings

Seven participants assess the web app, four female, three male, with an age range of 26--43 years old (*M*=33.86, *SD*=7.06). In the context of usability, seven users has been found to be a suitable number for a small-scale user study [Nielsen, 1993]. We present our findings as a preliminary study that is helpful to identify potential challenges of the prototype and design the next iteration. From here on we will use the nomenclature of P1--P7. We thus gathered 21 independent trials. Three of the participants did the study remotely while the other four did the study co-located. The participants' musical skill level was intermediate (3 participants) and advanced (4 participants). Their experience with music technology varied from less than 2 years (1 participant), 4--6 years of experience (1 participant), to more than 10 years of experience (5 participants). The instruments used for the study included electric guitar (3 participants), piano (2 participants), Roli Seabord (1 participant) and a synth (1 participant).

Figure 3 shows the aggregation of the 7-point Likert scale regarding the preferred visualisation mode. The linear visualisation was the most popular: on 12 occasions the linear visualisation was preferred, 8 of which were a strong preference. In contrast, there were 8 occurrences in which the circular visualisation was preferred, 6 of those being a slight preference.
The number of chords selected by the participants ranged from 1 to 24, with an average of 6.81 and standard deviation of 5.74.

Figure 3: Bar plot of the ratings of the two visualisation modes (*N=21*)



The queries are visualised in Figure 4 as the proportion of total queries that contain each chord.
We can see that there was a strong preference for major and minor chords, with *Cmaj*, *Gmaj* and *Amin* as the most popular ones.
In general, the most popular roots were natural notes.
This preference largely reflects the chords that are taught first in music curricula, and therefore should not come as a surprise.

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

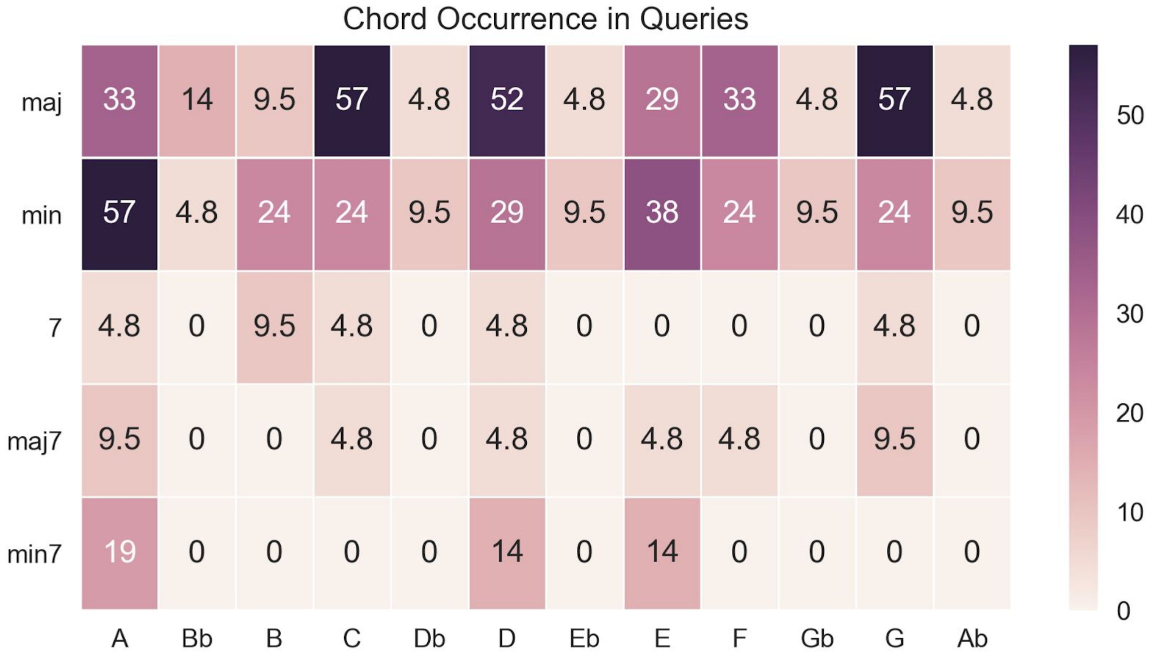


Figure 4: Proportion of the queries (*N=21*) that contain each chord

Tables 1 and 2 summarise the findings from the thinking-aloud conversations with the participants. From Table 1 we observe that even though there was a stronger preference for the linear visualisation, there were arguments in favour of both. The linear visualisation was associated with simplicity and the circular visualisation tended to be associated with more complexity. The use of colours seemed to be more appealing than using black-and-white text; however, they should have a clear meaning. For example, the relationships between chords and chord types, or the specified vs. unspecified chords, could be chromatically reflected.

Table 2 gives an overview of the main aspects that could be improved from the web application: chord recognition, QbC requests, context of the song and practice hints. Overall, the participants found the application to look promising, but a number of them suggested it should blend better into the music learning practice and the real-time activity of playing a song.

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

Type	Benefits	Challenges
Circular	• The fixed position of the chords (2 participants): *no need to read it, less effort* (P2) and *you can see all the chords* (P6). • The relationships between chords (2 participants): it promotes seeing the *harmony, patterns, (...) and finding relationships* (P4). • The colours (3 participants): *it is more pleasing aesthetically* (P7). • Intuitive (1 participant): *the colour can be more intuitive for playing the chord by colour and position'* (P6) and *it provides more information and can be remembered in a more intuitive way* (P6). • Usage (2 participants): *if you can follow the song easily because you know the chords, then you can spend some time looking at the circle and finding relationships* (P4) and *it's fun now that I am more confident to play with the chords.* (P4).	• Missing a global structure (3 participants): *I know what the next chord is, but I don't know when it is going to change* (P6). • Unclear relationships (1 participant): *Make the circle turn such that the tonic is in a fixed position based on key recognition* (P4). • Readability (4 participants): *it is hard to read, there is too much information* (P1) and *unsettling colours* (P3). • Complexity (4 participants): *it is difficult to see what is happening* (P3) and *my knowledge of music theory does not give me a clear understanding of the colour wheel* (P7). • Usage (1 participant): *if the chords change fast they are hard to follow and need to pay more attention than in the timeline* (P4) and *I would not use it for a concert* (P4).
Linear	• Simplicity (2 participants): *I don't feel confident with the chords, so it is easier to see the current and next chord (no distraction)* (P4). • Usage (2 participants): *I find this very useful as an improviser* (P3) and *this works well if you need to pay a lot of attention, for example, if the chords change fast* (P4). • Clarity (2 participants): *it is faster and clearer with the letters* (P7).	• Missing a global structure (3 participants): *I know what the next chord is, but I don't know when it is going to change* (P6) and *for songs that don't change chords every 4 beats it is harder, there are a lot of transitions that kill you* (P4). • Lack of relationships (1 participant): *seeing the connections would be nice (...) for solos see substitutions (that's a problem that I have when learning), for example when doing arrangements'*(P3). • Lack of colours (1 participant): *there could be more colours* (P1).

Table 1: Benefits and Challenges Between Linear and Circular Visualisations

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

Themes	Design Issues	Design Suggestions
Transcription	Chords are delayed and sometimes the chord transcription is not accurate (4 participants): *chord delays confuses me especially when they change fast* (P2) and *are the chords correct? they don't change aligned to the music (there's a little bit of a lag)* (P4).	Show the information to allow for anticipation (1 participant) *Need to be able to anticipate* (P1) and combine automatic chord recognition with other MIR techniques, e.g. *combine symbolic recognition to improve the chord recognition with the confidence estimation* (P1).
QbC	The QbC interface could align better with music education (3 participants): *it's better now that I reduced the number of chords and I could focus more on playing* (P4); *querying by tonality and adding diminished chords would be useful* (P5) and *the chord structure is not the way I use for classical piano study, but scales with their accidentals. It's then difficult to understand and reproduce the songs following the chords chart.'* (P7).	Reconsider the organization the QbC grid (1 participant): *the chords could be organized by notes e.g. C major, C minor, C 7, and so on, for both expert musicians who are familiar with chords, music solfège, harmony, but also for the novice musicians* (P5).
Song	If the song is unknown by the practitioner, it is difficult to identify when the next chord comes and therefore to practise (4 participants): *Need to know about the musical structure e.g. chorus, verse*(P1); *it would be nice to have a sense of duration beside the chord, how long it is going to last (to allow for anticipation strategies when performing)* (P3) and *when rehearsing with the system you are never told what octave are you in* (P3).	Show more contextual information about change and temporal durations (4 participants): *include beat timing* (P1) and *it would be nice to see when the change is coming* (P3).
Hints	It is unclear how many chords are suitable to include in the QbC request (the 7 participants asked how many chords they should query): *now I've selected a more complex song because it could be more interesting* (P6) and *it's not so useful when there is only one chord* (P6).	Give some (automatic) hints on the types of chords to choose or play (3 participants): *exploring chords that you know with one or two new chords could be interesting*(P1) and *with this kind of music (...) it makes more sense to improvise* (P3).

Table 2: Design Issues and Design Suggestions for Improvement

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

4.3 Lessons Learned

From this experience, we have identified four main areas that we would like to improve in the next design iteration of *Jam with Jamendo*:

- **Improvement of the user experience**: address the delay between current and next chord and provide more contextual information of the song (e.g.\ representation of metrical structure) to facilitate sight-reading.
- **Keep two visualisations**: it seems useful to provide two modes of visualisation so that users can pick according to their musical knowledge, personal taste and needs. The use of colours seems to be useful, yet a colour palette needs be explored more thoroughly for both visualisations.
- **Improvement of the QbC interface**: it might be interesting to suggest chords that go well with the user's selection based on occurrence in the dataset and also suggest alternative chords informed by music theory.
- **Expanding the features of the web application based on real use-cases**: from this user study a number of usages emerged, such as music improvisation or teaching, which can inform the new forthcoming features of the application in line with co-design practices.

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

5 Conclusion

In this deliverable, we presented Jam with Jamendo, a web application designed for learners of musical instruments that supports query-by-chord requests from the music platform Jamendo. It serves as a showcase of the new retrieval scenarios that are made possible by exposing audio-based descriptors in the Audio Commons Ecosystem. We conducted a thinking-aloud user study with seven participants to evaluate the system and compare two visualisation modes: a linear mode and a circular mode. We identified the benefits and challenges of each visualisation and a set of design issues and potential solutions for the next iteration of the web application. In summary, using two visualisation modes is promising, but they need to better support the task of real-time performance within an educational context.

As future work on this application, we plan to improve the existing visualisations and graphical user interface, add more contextual information of the metric structure, and also improve the query resolution mechanism. So far, we have relied on the number and type of chords to implicitly infer the difficulty of music pieces, but it would be useful to present the user with the option to exclude or order pieces according to difficulty. After all, the same chords can be used in a wide range of complex rhythmic patterns. Analysing the rate of chord changes would be a first step towards a solution.

In the wider scope of the Audio Commons project, we will need to integrate the audio-based content description with the mediator developed in WP2. This version of Jam with Jamendo was built with a private version of the analysis service, but the same chord date should be exposed through the public mediator such that third parties can use it to build similar applications.

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

6 References

M. Bostock, V. Ogievetsky, and J. Heer. D3 Data-Driven Documents. IEEE Transactions on Visualization and Computer Graphics, 17(12):2301–2309, 2011.

J. A. Burgoyne, J. Wild, and I. Fujinaga. An Expert Ground-Truth Set for Audio Chord Recognition and Music Analysis. In Proceedings of the 12th ISMIR Conference, pages 633–638, 2011.

E. Isaacson. What You See Is What You Get: On Visualizing Music. In Proceedings of the 6th ISMIR Conference, pages 389–395, 2005.

J. Nielsen and T. K. Landauer. A Mathematical Model of the Finding of Usability Problems. In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, pages 206–213, 1993.

N. Orio. Music Retrieval: A Tutorial and Review. Foundations and Trends® in Information Retrieval, 1(1):1–90, 2006.

Y. Rogers, H. Sharp, and J. Preece. Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, 2011.

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