Cloud-Smart Musical Instrument Interactions: Querying a Large Music Collection with a Smart Guitar

LUCA TURCHET*, Department of Information Engineering and Computer Science, University of Trento, Italy
JOHAN PAUWELS, Centre for Digital Music, Queen Mary University of London, United Kingdom
CARLO FISCHIONE, Department of Network and Systems Engineering, KTH Royal Institute of Technology, Sweden
GYÖRGY FAZEKAS, Centre for Digital Music, Queen Mary University of London, United Kingdom

Large online music databases under Creative Commons licenses are rarely recorded by well-known artists, therefore conventional metadata-based search is insufficient in their adaptation to instrument players' needs. The emerging class of smart musical instruments (SMIs) can address this challenge. Thanks to direct internet connectivity and embedded processing, SMIs can send requests to repositories and reproduce the response for improvisation, composition or learning purposes. We present a smart guitar prototype that allows retrieving songs from large online music databases using criteria different from conventional music search, which were derived from interviewing thirty guitar players. We investigate three interaction methods coupled with four search criteria (tempo, chords, key and tuning) exploiting intelligent capabilities in the instrument: i) keywords-based retrieval using an embedded touchscreen; ii) cloud-computing where recorded content is transmitted to a server that extracts relevant audio features; iii) edge-computing where the guitar detects audio features and sends the request directly. Overall, the evaluation of these methods with beginner, intermediate and expert players showed a strong appreciation for the direct connectivity of the instrument with an online database and the approach to the search based on the actual musical content rather than conventional textual criteria, such as song title or artist name.

CCS Concepts: • **Information systems → Music retrieval**; • **Applied computing → Sound and music computing**; • **Human-centered computing →** *Sound-based input / output*; • **Computer systems organization →** *Cloud computing*.

Additional Key Words and Phrases: Internet of Musical Things, Smart Musical Instruments, Music Information Retrieval

ACM Reference Format:
Luca Turchet, Johan Pauwels, Carlo Fischione, and György Fazekas. 2020. Cloud-Smart Musical Instrument Interactions: Querying a Large Music Collection with a Smart Guitar. *ACM Trans. Internet Things* 1, 1, Article 1 (January 2020), 28 pages. https://doi.org/10.1145/3377881

1 INTRODUCTION

The *Internet of Musical Things (IoMusT)* is an emerging research area positioned at the intersection of the fields of Sound and Music Computing and the Internet of Things [37]. The IoMusT refers to the network of *Musical*

*This is the corresponding author

Authors' addresses: Luca Turchet, Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, Trento, 38123, Italy, luca.turchet@unitn.it; Johan Pauwels, Centre for Digital Music, Queen Mary University of London, 327 Mile End Road, London, E1 4NS, United Kingdom, j.pauwels@qmul.ac.uk; Carlo Fischione, Department of Network and Systems Engineering, KTH Royal Institute of Technology, Malvinas Väg 6B, Stockholm, 10044, Sweden, carlofi@kth.se; György Fazekas, Centre for Digital Music, Queen Mary University of London, 327 Mile End Road, London, E1 4NS, United Kingdom, g.fazekas@qmul.ac.uk.

2577-6207/2020/1-ART1 $15.00
https://doi.org/10.1145/3377881

Things. We formally define Musical Things as devices dedicated to production, interaction with, and reception of musical content. Musical Things embed electronics, sensors, data forwarding and processing software and network connectivity enabling the collection and exchange of data for musical purposes.

A prominent class of Musical Things is that of *smart musical instruments (SMIs)* [33]. Such instruments are characterized by built-in sensors, actuators, embedded intelligence, and wireless connectivity to local networks and to the Internet. According to the vision proposed in [33], the embedded intelligence of an SMI is characterized by five core capabilities: 1) knowledge management, i.e., the capability of maintaining knowledge about itself and the environment; 2) reasoning, i.e., the capability of making inferences on the acquired knowledge; 3) learning, i.e., the capability of learning from previous experience; 4) human-SMI interaction, i.e., the capacity to interact with the player in ways that extends the conventional sound production, such as adaptation and proactivity; 5) SMI-Musical Things interaction, i.e., the capability of wirelessly exchanging information with a broader network of Musical Things.

Musical Things may also consist of servers hosting online music databases. Today, various online music databases exist as well as services for them. Some of these are available under Creative Commons licenses (e.g., Freesound, Jamendo) and offer a large amount of musical content that is largely unknown. This spurs the exploration of novel methods of discovery [44]. Despite the possibility to explore and play the songs from such databases through text-based music search engines and playlists, to our knowledge there is still a gap in interaction mechanisms making these large collections of music adapted to the instrument players' needs (e.g., learning or composition). SMIs offer opportunities to bridge this gap. Thanks to their direct connectivity to the Internet and embedded sound delivery system, SMIs can send a request for musical content to remote repositories, receive a response in the form of a music piece, and reproduce it. This allows an SMI user to play along with downloaded musical content. Such a content is emitted by the instrument itself, for instance for improvisation, composition, learning, or rehearsing purposes.

However, to date, the challenge of connecting SMIs to cloud-based services has been scarcely addressed in academic and industrial applications. Such an interconnection has the potential to enable novel kinds of Ubiquitous Music activities. Ubiquitous Music [17] is a branch of the Sound and Music Computing field which develops and analyze musical activities supported by ubiquitous computing concepts and technology [28, 41]. A notable pioneering example is the Sensus Smart Guitar developed by MIND Music Labs, which has been used in conjunction with Spotify [36]. Here, a smartphone application is used by guitar players to stream content to the Sensus guitar via Bluetooth, i.e., songs selected from Spotify. The guitar players could then play along the streamed songs thanks to the instrument capability of reproducing both the performed guitar sounds and the audio downloaded from Spotify. Along similar lines, the Smart Mandolin described in [32] has been interfaced to the Freesound online audio content repository [11]. Using a smartphone app, the mandolin player could select content to be retrieved prior to performing through a set of keywords and structure it for the purpose of creating a desired accompaniment [34].

Various content-based music search engines have been developed by the music information retrieval (MIR) [5] research community (see e.g., [6, 15, 18, 39, 43]). Recently, Xambó et al. presented "Jam with Jamendo", a web-based prototype that allows novice and expert musicians to discover songs in the Jamendo music repository by specifying a set of chords [44]. The Jamendo music collection was indexed by an automatic chord estimation technique [24]. The purpose of such a system was to provide a more pleasurable practice experience by suggesting novel songs to play along with, instead of practicing isolated chords alone, or practice repeatedly with the same song.

Jam with Jamendo was tested by practitioners of various instruments, who interacted with desktop or laptop computers to display and manipulate the app. Such kind of interactions have been recently investigated by Martinez-Avila et al. [20]. The authors conducted an ethnographic study of the preparation activities of working musicians with the aim of understanding the challenges of interacting with supporting resources at the same time as playing. They observed that guitar players' interactions with supporting digital and physical resources is interwoven into their embodied musical practices, which are typically encumbered by having their instrument in hand, and often by playing. Based on those findings, the authors envisioned the augmentation of musical instruments with intelligent components as a possible solution to minimize such encumbered interactions. The features and capabilities of SMIs allow one to limit the amount of devices to be connected and, therefore, have the potential to solve issues with such kinds of interactions.

In this paper, we investigate a system that connects a smart guitar to an online music repository to retrieve music content directly from the instrument (i.e., without the need for any additional external equipment), so the player can play along with the downloaded content. At methodological level, our study was conducted following established techniques from the human-computer interaction field [29], in particular during the phases of requirements discovery, design, and evaluation. Firstly, we gather requirements for such a system by conducting interviews with guitar players in order to understand their needs and expectations. We chose the guitar because it is one of the most widespread musical instruments. Secondly, we present a smart guitar prototype that allows novice and expert musicians to discover songs in the Jamendo music collection by specifying a set of musical search criteria which can be extracted from audio content using Semantic Audio techniques [10]. Such criteria extend the type of the criteria usually adopted by conventional search engines applied to music collections (such as song title or artist name). Specifically, we focus on retrieval by chords, beat per minute (BPM), key, and tuning. Little research has been conducted thus far on the task of querying a repository using such criteria, especially when utilized simultaneously. The repository was indexed along such criteria leading to a database offering a wider range of search possibilities compared to previous efforts.

We investigate three methods to conduct the search process via such criteria by exploiting the intelligent capabilities of the instrument. The first consists of a keywords-based retrieval controlled by a touchscreen embedded in the instrument itself. The second method consists of a cloud-computing application where the content recorded by the guitar is transmitted to a server that automatically extracts from the audio signal the features related to the criteria and then performs the search. This case represents an example of distributed intelligence applications for smart instruments that was envisioned in [33] and thus far not explored. The third method consists of an edge-computing application [30] where the guitar itself automatically detects the features related to the search criteria from audio resulting from the players' interaction with the instrument and sends the request to the repository. We evaluate and compare the three methods in terms of both technical performances and user experience. Finally, we discuss the implications and limitations of our study.

2 REQUIREMENTS GATHERING

A set of individual semi-structured interviews were conducted with the aim of understanding typical requests, processes, needs, and expectations of the guitarists for a proposed system that allows their instrument to retrieve musical content from a cloud-based repository. We collected demographics from participants in order to search for correlations between their characteristics (e.g., years of musical expertise) and their answers.

2.1 Participants

Thirty participants were recruited (6 females, 24 males), aged between 23 and 55 (mean = 33.4, standard deviation = 8.2), and belonging to different nationalities (Italian, Swedish, British, Greek, Indian, Turkish, Australian, USA). The guitar, either classical, acoustic, or electric, was their main musical instrument. On average, they had been playing it for about 19 years. Ten of them considered themselves as beginner guitar players, ten as intermediate, and ten as expert.

2.2 Procedure

The interviews were conducted either face-to-face (a total of 16 participants) or via video-conferences (a total of 14 participants). The interviews lasted 30 minutes on average. The questions and the structure of the interview were chosen after a pilot study conducted with three guitar players, who were not included in the experiment.

First, participants were introduced to the concept of smart instruments and their capabilities. This included the description of the topic at hand, the interconnection of a smart guitar with a large collection of music. They were explained that they could use the system for composition, instrument practice and learning purposes, and that the system could retrieve both known and unknown musical content. Second, they were asked to answer to the following questions:

(1) *Imagine you have a touchscreen embedded in the smart guitar for the retrieval of songs from a database. Which criteria would you use to retrieve the music you want to play with?*
(2) *Imagine that the smart guitar can automatically extract some information from what you played for 30 seconds. Which kind of information would you use to retrieve the music you want to play with?*

Participants were asked to avoid considering the types of requests that can be found in conventional search engines such as title, artist name, or band name. The reason is that we wanted to explore novel ways to retrieve musical content from a repository. During the interviews, to avoid any kind of bias no example of criteria was provided by the experimenter.

2.3 Outcomes of interviews

First, we analyzed the participants' answers by enumerating the identified criteria. Second, we conducted an inductive thematic analysis [3] by generating codes from the open-ended answers to the questions. The codes were further organized into themes that reflected patterns. Table 1 lists the criteria identified by participants for both the first and second questions, along with the number of participants that identified each criterion. It is noticeable that some criteria are common across the two conditions, while others are specific to one or the other.

Regarding the first question (query by touchscreen condition), the following criteria were identified:

Music content type. Fourteen participants reported that the system should allow them to select different types of musical content, such as songs with full ensemble versus backing tracks, songs with or without certain musical instruments, or original songs versus cover, or live recordings versus the album version. Specifically, twelve participants reported that they would have used the system to retrieve backing tracks without the guitar part for practicing or recreational purposes. Along these lines, 8 participants reported that they would have used the absence of certain instruments in the ensemble as filters for the music to be retrieved (e.g., find music without vocals), while 9 participants would have used the presence of a certain instruments as selection criteria (e.g., songs with drums, bass, keyboard, and vocals).

Genre/style. Almost all participants (26) explicitly reported that the touchscreen should display keywords to select the genre and subgenre/style of the music to be retrieved. This was among the most popular criteria mentioned by participants on average.

BPM, tempo signature, and rhythm. Twelve participants indicated BPM as criterion (in terms of actual numbers and/or in terms of verbal descriptors such as fast or slow), four the rhythm (intended as rhythmic pattern selected from a list), and 4 the tempo signature (e.g., 4/4, 6/8, etc.).

Difficulty level. Nine participants pointed out that an important criterion, especially in contexts of learning, is the difficulty level of the guitar part in a certain music to be retrieved.

Chords and scales. Eleven participants reported that the system should allow them to select a set of chords and return musical pieces that contain those chords (either randomly positioned or in an exact sequence). For instance, a participants asked for the possibility to search songs allowing her to practice barre chords. Along the same lines, three participants indicated that they would have searched by musical scale (e.g., G pentatonic, D

Table 1. Identified criteria for question one (query by touchscreen) and question two (query by playing) along the number of participants for each criterion.

	Query by touchscreen Number of participants	Query by playing Number of participants
Absence of instruments	8	-
BPM	12	10
Chords	11	20
Difficulty level	9	-
Music content typology	14	-
Genre/style	26	10
Key	-	11
Melody	-	22
Melodic richness	2	-
Mood	4	-
Playing technique	-	12
Popularity	6	-
Presence of instruments	9	-
Rhythm	4	15
Role of the guitar	1	-
Scales	3	8
Tempo signature	4	7
Timbre	-	6
Type of guitar	1	-
Tuning	1	1

minor harmonic, etc.), i.e., the system should return musical pieces that allow for practicing that particular scale. In general, chords and scales criteria were deemed as particularly useful for learning and practicing purposes.

Mood. Four participants indicated the mood of the musical piece (e.g., happy, relaxed, sad) as a desirable search criterion. This criterion was deemed important when the system is used for recreational or composition purposes.

Popularity and endorsements. Six participants reported that popularity and/or endorsement would have been important criteria, especially if given by guitar players and for unknown musical pieces.

Melodic richness. Two participants indicated that they would filter the musical content by its melodic richness, namely search for pieces that contain or lack a dominant melodic component or pieces that feature several kinds of melodies or polyphony.

Type, role and tuning of the guitar. One participant reported that they would have performed the search using the guitar type (i.e., electric, classical, or acoustic) as a criterion. One participant indicated that the role of the guitar (leading or accompaniment) would have been an important criterion to her. Moreover, one participant highlighted that the tuning of the guitar (e.g., in A = 440 Hz or A = 432 Hz) would be an important selection criterion, since the downloaded music would need to match the guitar tuning in order to enable the guitarist to play along with it.

Furthermore, the following themes were identified:

Context of use and usefulness. Ten participants indicated that they would use the system for learning and practicing purposes rather than in live performance. On the contrary, three would have use it in live performances for instance for experimental music concerts or in situations that require to accommodate requests for songs from the audience, such as weddings or during busking. All such participants also commented positively on the usefulness of the system for their envisioned purposes, which could help their current practices: eight participants reported that they normally use streaming services such as YouTube to search for backing tracks to practice with, apps providing MIDI tracks for guitar learning purposes, or Spotify to play along with their favorite artists and bands.

Regarding the second question (related to queries by playing), the following search criteria were identified:

Genre/style, guitar technique, and timbre. Twelve participants reported that the system should be able to detect, from the short recording, the type of guitar technique (e.g., fingerpicking, power chords, arpeggios, fast-alternate picking, tapping, use of the tremolo bar). Along the same lines, six participants mentioned the kind of timbre of the guitar as a criterion for the search (e.g., involving a certain effects processing the guitar sound such as distortion or delays). For 10 participants, this information was deemed as useful to retrieve musical content belonging to a particular genre, subgenre, or style. For instance, a participant suggested that by playing loud, distorted power chords a user could retrieve music belonging to Trash Metal; another participant indicated that using specific melodic patterns such as pentatonic scales and using the bending technique could be a criterion to retrieve music related to blues. Six participants pointed out that the information related to playing technique and/or timbre could be used to retrieve musical pieces involving that particular technique and/or timbre. For instance, a participant commented that by playing with a smart guitar set up with a certain chain of sound effects and amplifier type (e.g., a flanger effect coupled with a Mesa Boogie cabinet) a user would expect to retrieve pieces involving that specific setup.

Chords, key, and scales. Twenty participants indicated that the system could automatically detect sequences of chords played by the user and retrieve musical content involving those chords (exactly those or more), or even their variations (such as harmonic substitutions according to the functional harmonic theory [26]). Eleven participants indicated the key as a criterion, which could be utilized if the recording contained a single chord (i.e., a recording with one chord means that the system has to retrieve musical content having the played chord as the key, while a recording with more than one chord means that the system has to retrieve music containing those chords). In the same vein, 8 participants reported that playing scales could lead to the retrieval of songs containing those scales or chords/keys associated to them.

BPM, tempo signature, and rhythm. Ten participants commented that they would have played the guitar, with strummed chords, stopped strumming, or even by creating percussive sounds with their hands on the guitar, to retrieve music with the BPM associated to their playing. Fifteen participants indicated that the system should be able to automatically detect the rhythmic patterns of strummed chords or stopped strumming (including accents), and consequently retrieve musical pieces that contain those patterns or similar. Seven participants also suggested that from the rhythmic patterns the system could be able to detect the tempo signature and retrieve musical pieces having it.

Melody and tuning. A large amount of participants (namely twenty-two) indicated that they would primarily use the system to make queries to the database by playing melodies. For instance, participants would expect that the system could retrieve backing tracks that fit well with the kind of melody played (e.g., in terms of genre/style, harmonic content, or BPM). Moreover, they would expect that the system could retrieve songs for which they forgot the title (with respect to this, four participants made a specific comparison between the system capabilities and those of services performing automatic recognition of songs such as Shazam). Notably, four participants commented that they would also use the system to check whether melodies that they have composed are too similar to those of existing musical pieces, with the aim of avoiding risks of plagiarism. One participant commented on the importance of retrieving content that follows the tuning of the guitar.

Finally, the following theme was identified:

Appreciation. Five participants very much appreciated the idea of a system allowing them to retrieve musical pieces with criteria related to their musical playing parameters (including the expressive ones), especially if they could complement them with the conventional criteria of search engines. They defined as useful and "cool" a system capable of understanding what the musician plays and how, since it could provide a more natural and immediate way of performing the search, as well as they deemed that the system would have the potential to facilitate their composition practices.

2.4 Discussion

The results of the interviews allowed to understand needs and expectations for a system connecting a smart guitar to an online repository for music retrieval purposes, as well as identify a set of criteria that could act as requirements informing its design and development. It is noteworthy that participants indicated criteria both in terms of editorial/social metadata (such as difficulty level or popularity) and content-related metadata (such as BPM or key). Some criteria were common to the two envisioned systems (such as chords or BPM), others were specific to each system and the kind of query it enable (such as presence/absence of instruments for the query by touchscreen system, and guitar technique for the query by playing system).

Notably, correlations were observed between the level of participants' musical expertise and the type of envisioned criteria. For instance, eight professional and experienced guitar players imagined queries more complex (such as queries by timbre or guitar technique) than those of beginners and less experienced players. The most complex queries, such as the automatic detection of a certain style or the retrieval of musical pieces similar to a given melodic sequence, pose significant research challenges which in some cases represent open problems, such as the definition of a concept of similarity or the detection of features related to expressivity (see e.g., [1, 42]).

3 DESIGN

Starting from the results of the analysis of the interviews, we designed three proof-of-concept systems that attempt to cater to some of the interviewed guitar players' needs and expectations. For this purpose, we considered low-cost hardware and freely available resources, such as access to online music repositories as well as open source music information retrieval methods to be applied to both the guitar sound and large music collections. While guitar players indicated several criteria, for practical reasons related to their possible and available implementation with the available resources, we chose the following ones: chords, key, BPM, and tuning. The three systems to conduct the search process via such criteria are the following:

System 1: Query by touchscreen. The first system consists of a keywords-based retrieval that the user performs via a touchscreen embedded in the guitar. The user is enabled to select among 60 chords (for each of the 12 notes: major, minor, dominant, major seventh, and minor seventh), 24 keys (12 major and 12 minor), a range of BPM between 40 and 240 (in a continuous way), and among 3 tunings (A = 432 Hz, A = 440 Hz, A = 444 Hz). Moreover, the user at the moment of selecting the BPM can activate a metronome delivering the selected BPM (the metronome sounds are delivered by the smart guitar itself).

System 2: Query by playing via cloud computing. In the second system the user utilizes the touchscreen embedded in the guitar to start and stop a recording. The recording is then transmitted to a server that automatically extracts features from it related to the criteria, then performs the search. The user is allowed to perform queries by either playing chords or a melody, and is allowed to repeat a recording before submitting the query (in case the user is not satisfied with what he/she played).

System 3: Query by playing via embedded computing. Akin to System 2, in the third system the touchscreen embedded in the guitar is used to start and stop a recording, which has to contain either chords or a melody.

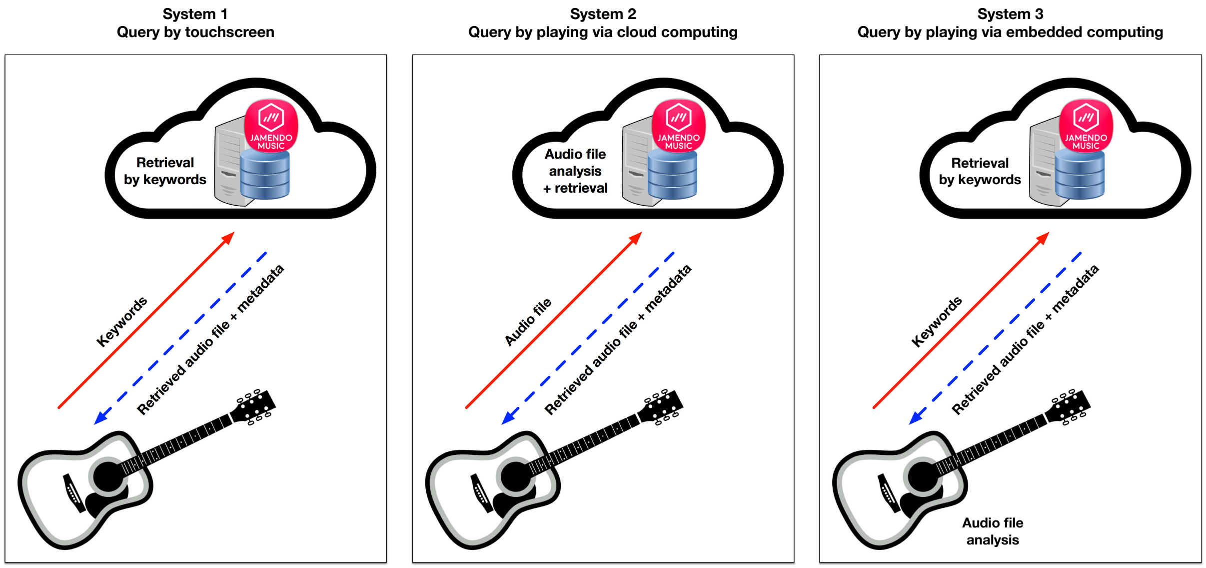


Fig. 1. A schematic diagram of the three systems.

The recorded audio, however, is analyzed on the embedded system itself and the result of the computation is then transmitted in terms of keywords to the server, which performs the search using them.

Furthermore, we adopted an additional design choice which did not emerge from the interviews: the guitar's automatic configuration based on the metadata associated to the retrieved music content. Specifically, for all the three systems the BPM of the retrieved music was utilized to automatically configure the delay time parameter of a delay with feedback effect (so to guarantee that the delay repetitions would fit well with the downloaded music). On the one hand, this choice was taken to explore the concept of a system capable of intelligently setting up parameters of the sound engine and of the touchscreen according to parameters of the retrieved musical content. On the other hand, this allowed us to investigate the users' experience in interacting with such a system and gathering feedback useful for future development iterations. Moreover, the music title and the actual parameters of the retrieved music were displayed on the touchscreen, to provide the user with a visual feedback about the outcomes of the retrieval process. Figure 1 illustrates the three designed systems, their components, and data flow.

4 IMPLEMENTATION

The apparatus comprised of the following components:

Jamendo music repository. Jamendo[1] hosts digital music content released predominantly under Creative Commons licenses. It provides access to music tracks from independent artists that are free for personal use. Part of the Jamendo catalogue is offered for commercial licensing, e.g., for in-store radios or as background music for videos or games. About 100K tracks are actively promoted by Jamendo for this purpose (which implies that they conform to higher recording standards) and these tracks were indexed along the four aforementioned criteria

[1]https://www.jamendo.com/

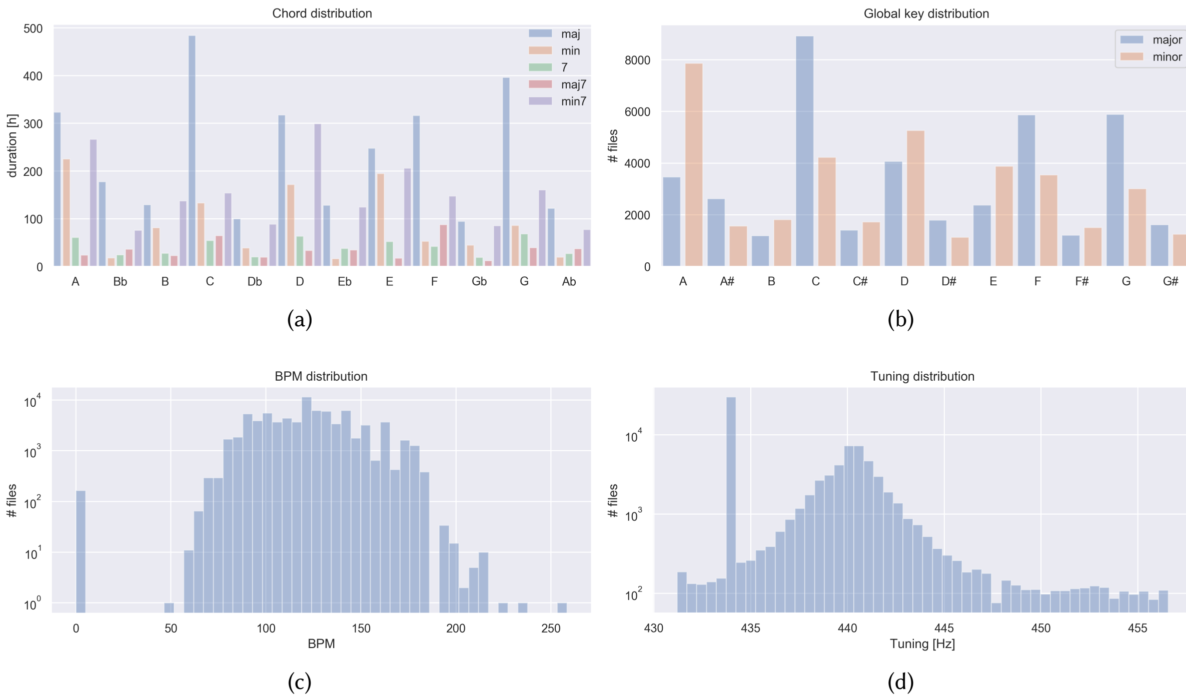


Fig. 2. Distribution of the musical content available on Jamendo, according to the four criteria.

to create a searchable database described below. Figure 2 illustrates the distribution of the analyzed subset of musical content available on Jamendo for the four search criteria at the time of writing.

Smart guitar. The smart guitar prototype (see Fig. 3) comprised a conventional classical guitar enhanced with: a contact microphone; a small loudspeaker; a touchscreen (created by attaching a smartphone to the top part of the guitar and developing apps for it); wireless connectivity, via Wi-Fi dongle and a Wi-Fi router (both featuring the IEEE 802.11ac standard over the 5GHz band); the router was extended with a 4G dongle to enable Internet connectivity (however, during the experiments Internet was provided by connecting the router to other routers allowing for greater speed and wider transmission band); the Bela-mini low-latency audio processing board [21]; an embedded battery, powering both the computing board and the router.

The smart guitar audio engine was coded in the Pure Data real-time audio processing environment and comprised: a component processing the strings sounds with a delay with feedback effect in chain with a reverb effect; a player of the musical content retrieved from Jamendo; a component to record the guitar sounds.

The smartphone and the computing board exchanged messages over Wi-Fi leveraging the Open Sound Control (OSC) protocol over the User Datagram Protocol. The components of a Wi-Fi system were optimized for live performance following the recommendations reported in [22] to reduce latency and increase throughput. The router was configured in access point mode, security was disabled, and only the IEEE 802.11ac standard was supported. Three apps for the smartphone were developed to perform the three kinds of music retrieval (see Figure 4). As a result of the retrieval, the sound engine and the touchscreen were configured automatically as described in section 3.

Fig. 3. The developed smart guitar prototype.

For the first system, the smartphone sends the keywords that are selected in the app to the computing board, while for system 2 and 3 the smartphone controls the recording process on the board. In all three cases, a Python script processes the inputs and constructs a call to the search API running in the cloud. For system 1, the keywords are passed directly to the search service. System 2 uploads the recorded guitar sound to the search service. System 3 first extracts keywords for the four criteria through a local analysis of the recorded sound and then sends these keywords to the service. For system 2 and 3 the user is allowed to perform queries by playing either a sequence of chords or a melody.

Search service. The search API is a web service that is constructed according to the Function-as-a-Service (FaaS) paradigm. It consists of a persistent MongoDB database and short-lived worker processes that are created each time an API call is received (potentially in parallel). The database contains the automatic analysis of the 100K Jamendo tracks according the four proposed musical criteria. Chord indexing was performed by the same algorithm that was used for Jam with Jamendo [25, 44], the remaining criteria (key, BPM, and tuning) were calculated by the open-source Essentia extractor [2]. The service is hosted in-house and is built using OpenFaaS[2] and Docker Swarm technologies.

The worker processes search the database according to the parameters passed through the API call. These parameters can either be specified directly as keywords (system 1 and 3) or indirectly as a raw audio file (system 2). In the latter case, the audio file is analyzed with the same exact tool-chain that was used to index the Jamendo audio, to avoid a potential source of retrieval errors caused by inconsistencies between the two instances of audio content analyses. For system 3, the audio analysis already happens on the Bela board, once again using the same tool-chain for the same reason, such that the actual API call is identical to the one of the first system.

For each set of search parameters, a number of tracks in the Jamendo catalogue that correspond optimally to the desired criteria is returned. For the BPM, this means that tracks whose BPM falls within 5% of the requested BPM are withheld, with preference given to the ones that are closest to the requested value. Regarding tuning, all tracks whose detected tuning falls within 0.4% of the requested tuning is withheld, again with a preference for the smallest deviation. For the key search, this means that tracks with exactly the same key as the requested one are selected. Finally, a track matches with respect to chords if it is comprised of only the requested chords or a subset thereof. Systems 2 and 3 are queried by audio recordings for which the users can indicate whether chords

[2]https://www.openfaas.com/

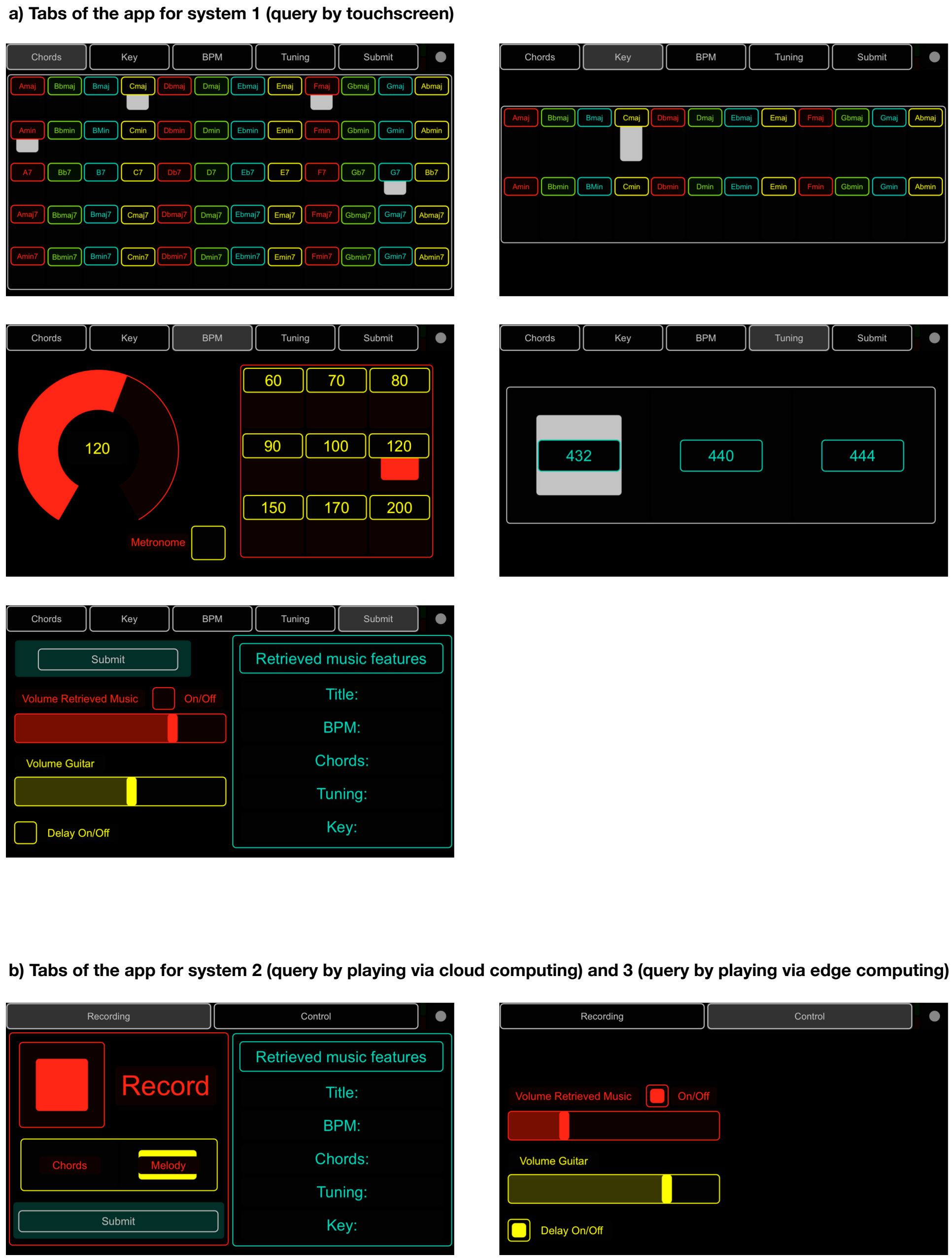


Fig. 4. Layouts of the apps for the three systems. a) Screenshot of the tabs of the app for system 1 (query by touchscreen), displaying the user selections of Cmaj, Amin, Fmaj and G7 in tabs Chords, Cmaj in tab key, 120 beats-per-minute in tab BPM, 432Hz in tab Tuning, and the volume level for the retrieved song and guitar in tab Submit; b) Screenshot of the tabs of the app for both system 2 (query by playing via cloud computing) and 3 (query by playing via edge computing), displaying the user selection of the recording of a melody in tab Recording and the adjustment of the volume level for the retrieved song and guitar in tab Control, as well as the activation of the delay effect.

or a melody is recorded. When chords are recorded, or in the case of system 1, the intersection of the four search criteria is returned to the smart guitar. When a melody is recorded, no chord search is performed since trying to recognize chords in a melody line does not make much sense. The final result then consists of the intersection of the three remaining criteria. In all cases, no tracks are returned to the smart guitar when the intersection is empty.

The returned tracks are played back through the sound engine on the smart guitar by downloading them first using the Jamendo API. The web service returns up to 10 tracks if sufficient tracks are available that correspond to the requested search criteria. They are ordered in descending order according to the fitness of their match. By default the first track is used, the remainder are fallback options. Due to the dynamic nature of the Jamendo catalogue and API, it is possible that after the one-time indexing process completed, the audio for a specific result became temporarily or permanently unavailable for playback through Jamendo's API. In such cases, the results are tried in order, until either an available track is found or the end of the result list is reached. In the latter case, no result is played back or displayed on the touchscreen, similarly to the case of an empty result set.

5 EVALUATION

Firstly, we conducted a technical evaluation of the three systems and compared their performances. Secondly, we conducted a user study to assess the usefulness of the systems and better understand the experiences of guitar players interacting with them.

5.1 Technical Evaluation

Facilitated access to content is an important objective of our work. The time spent on retrieving relevant content can be considered an important criterion for measuring how well a system helps this activity. Therefore, we considered the total time taken to retrieve the desired musical content as a metric to technically evaluate the three systems.

total retrieval time = embedded analysis time + cloud service query time + audio and metadata retrieval time

Where

- *embedded analysis time* is the time taken by System 3 to analyze the recorded audio on the computing board according to the four defined musical characteristics. For Systems 1 and 2 it is zero by definition.
- *cloud service query time* is the time it takes to query the cloud search service as measured on the computing board. It includes the transmission time of the query (keywords for System 1 and 3, an audio file for System 2), the processing time on the server and the transmission time of the response. Since we aimed to measure elapsed time from the perspective of the SMI with an embedded device, we did not directly compare cloud vs. embedded analysis times.
- *audio and metadata retrieval* is the time it takes to query the Jamendo API for the audio and metadata of the result returned by the search service, as measured on the computing board. This time is common to all three systems.

Table 2 shows the results of the tests. The results are the median values of 20 repetitions, expressed in seconds. All three timings are a function of the query input: longer audio recordings take longer to analyze and to send to the server. The database query depends on the exact search parameters and the music retrieval time correlates with the length of the result audio file. Therefore, it is more interesting to compare the timings relative to each other, while the absolute timings are only indicative of the order of magnitude of the entire process. The input to the three systems has been aligned such that they all return the same result, which was also still available through the Jamendo API, in order to make the comparison fair. More precisely, for the assessment of the performances

Table 2. Technical performances of the three systems according to the timing metrics.

	System 1 (touchscreen)	**System 2 (cloud computing)**	**System 3 (embedded computing)**
embedded analysis time	0 s	0 s	41.676 s
cloud service query time	1.091 s	4.237 s	0.985 s
audio and metadata retrieval time	1.823 s	1.808 s	1.812 s
total retrieval time	2.914 s	6.045 s	44.473 s

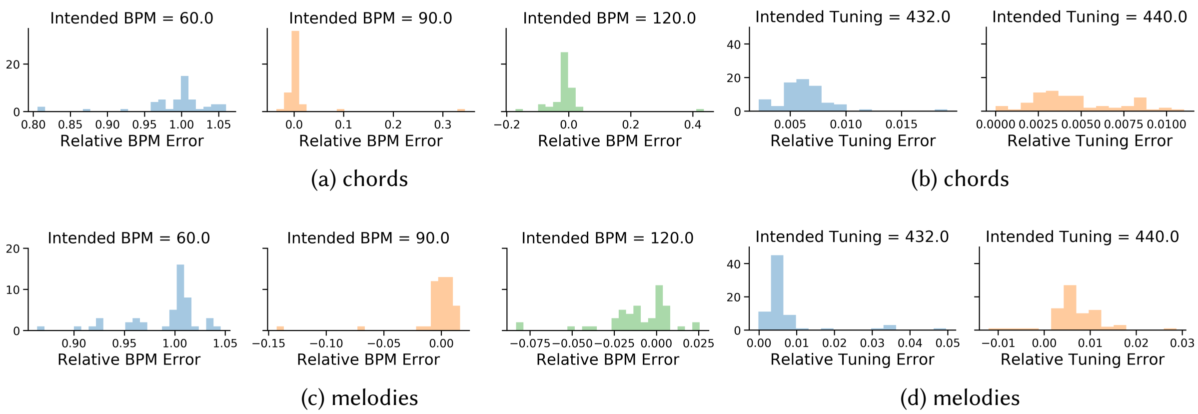
of the queries involving the act of playing as an input (for System 2 and 3) we used a 20 sec audio file in wav format. This duration was selected on the basis of the average time taken by participants of the experiments reported in Section 5.2 to make their recordings when using System 2.

We notice that System 1 is the fastest, which was expected because it uses keywords directly instead of requiring the conversion from an audio query to keywords. The time it takes to analyze the audio query on the server, observed as the difference in query time between System 2 and System 1, is an order of magnitude smaller than the time it takes to analyze the audio on the current embedded board. This result can be explained by the significantly higher computing power of the cloud server. The consequence is that the analysis time is the dominant factor in the total retrieval time of System 3, whereas in the other two systems the audio retrieval is also a significant contributor. Finally, the fact that the actual query time of System 1 is virtually identical to the one of System 3 suggests that the two processes are nearly identical from that point on.

Secondly, all three systems rely on the same automatic indexing procedure of the Jamendo audio. Due to the imperfections of state-of-the-art audio content analysis methods, the metadata used for indexing is inherently noisy. A good illustration of this can be seen in Figure 2c, which shows that an obviously wrong BPM of zero has been found in case of a non-negligible number of tracks. These errors will affect the three systems equally however. No ground truth is available for the entire database. It would be prohibitively time-consuming to manually annotate all 100K tracks according to the four criteria, so no rigorous testing can be undertaken. Informal testing shows that most of the errors are musically related, so they won't affect the user experience significantly from the technical perspective. Also the cases where the automatic analysis breaks down completely, such as when zero BPM is detected, are not problematic, since those files will never be returned by the search process. In any case, unsatisfactory search results – due to bad indexing and other reasons – will be accounted for using negative points in the user experience evaluation.

The difference between System 1 and Systems 2 and 3 is that for the latter two, automatic content analysis is also used during the querying. This increases the chance that analysis errors will have a negative influence on the query results and consequently the overall user experience. In order to quantify the errors on the query side, a collection of representative audio queries was created and this was used to evaluate the audio analysis algorithms.

A total of 228 recordings were made by a professional guitar player. These involved 114 chord sequences and 114 melodies. Specifically, for each of the 24 keys (12 major and 12 minor), we recorded chords sequences and melodies with 3 BPM (i.e., 60, 90, 120) and 2 tunings (A = 432 Hz and A = 440 Hz, these were chosen on the basis of the tuning distributions reported in Figure 2d). The features of the recordings were devised with the aim of testing the accuracy of the second and third systems according to the four search criteria. Each of the chord sequences consisted of 6 chords, where the first and second chords were identical (which leads to a total of 5 different chords in each sequence). Based on conventional tonal harmony theory [26], the sequences for the 12



(a) chords (b) chords (c) melodies (d) melodies

Fig. 5. Distribution of the relative errors for tempo and tuning of the recorded chord and melody queries.

major keys followed the pattern I - IV - V - VI - V7 - I, while the sequences for the 12 minor keys followed the pattern I - IV - VII - III - V7 - I. Two strums were performed for each chord in the sequence with the exception of the last for which three strums were performed. Each of the melodies consisted of an octave followed by a descending and an ascending arpeggio. The notes were separated by identical time intervals. All recordings were performed using a metronome to keep the tempo constant.

These recordings were then analyzed by our toolchain. Since they were created according to specific values for each of the four criteria, a formal evaluation of the audio analysis can be made. For the numerical values (tempo and tuning), the relative error is plotted in Figure 5. A relative error (defined as the difference between detected and intended value, normalized by the intended value) of zero means that the numerical value was recognized correctly, whereas positive and negative errors indicate the percentage of over- or under-estimation of the value respectively. Especially relevant for BPM is that double and half tempos, common errors with BPM detection [12], show up as 1 and -0.5.

From the BPM errors displayed for chords and melodies in Figures 5a and 5c, we can see that for target BPMs of 90 and 120, all detected BPMs only deviate a few percents, but that recordings with a BPM of 60 are almost consistently recognized as their double value (120 BPM). The reason is likely due to a strong prior towards higher BPMs in the tempo-tracking algorithm, which can produce values between 40 an 208 BPM [8]. If a 120 BPM song is returned when one of 60 BPM has been requested, it is still possible to play along at 60 BPM, but the character of the song is likely to be significantly different.

The musical key of the recordings is a single label per recording, so we can easily compare detected keys with intended keys and categorize errors into a number of musically related keys [23]. Apart from correctly detected keys, we distinguish adjacent keys (correctly detected scale, but tonic a fifth down or up removed from the correct tonic), relative keys, parallel keys (correct tonic, but other scale), chromatic keys (correct scale, but tonic a semitone up or down). The last category typically is a consequence of wrongly detected tuning. All non-musically related errors are grouped as "other keys". In Table 3a, the key evaluation for the chord and melody queries is listed together with the baseline performance of a random classifier. We can see that nearly all detected keys are musically related to the intended key, and that it is unlikely that such a error distribution happens by chance. In more than 70% of cases, the key recognition is perfect.

Finally, for the chord-based search to work, it is important that the set of chords that is played is correctly recognized. We know the five different chords that are being played in each recording, so standard information

Table 3. Key and chord evaluation results for the recorded chord and melody queries.

(a) key evaluation

	chords	melodies	random
correct keys	70.14%	71.53%	4.17%
adjacent keys	9.72%	6.94%	8.33%
relative keys	6.94%	2.08%	4.17%
parallel keys	9.03%	9.03%	4.17%
chromatic keys	0.69%	3.47%	8.33%
other keys	3.47%	6.94%	70.83%

(b) chord evaluation

	chords
precision	88.29% ± 12.75%
recall	83.06% ± 14.20%
f-measure	85.36% ± 12.91%
cardinality	4.71 ± 0.48

retrieval measures can be used to evaluate the retrieval performance. The figures in Table 3b show that on average, slightly fewer chords are recognized than there actually are in the recordings, but with good precision and a fairly balanced recall. In addition, many of the wrongly recognized chords are harmonically related. A full breakdown of the errors is out of the scope of this paper however.

5.2 User experience evaluation

An evaluation with users was conducted to investigate the experience of guitar players according to the two types of interactions involved, i.e., using the touchscreen or playing the intended query on the guitar. From the interaction standpoint, System 2 and System 3 are identical, while the time to retrieve the desired musical content is always greater for System 3 (see Table 2). Therefore, only System 1 and System 2 were utilized in the user evaluation. During the requirements gathering phase (see Section 2), some differences in the participants emerged that were related to their proficiency with the guitar (i.e., experts envisioned queries more complex, such as those involving a search by timbre or guitar technique). As a consequence, experiment aimed at searching for correlations between the degree of participants' expertise and their evaluation scores.

5.3 Participants

Thirty participants were recruited (1 female, 29 males) aged between 19 and 55 (mean = 33.5, standard deviation = 10.1), and belonging to different nationalities (Italian, Spanish, and Swiss). They all played a variety of instruments, but the guitar, acoustic or electric, was their main musical instrument. On average, they had been playing the guitar for about 15 years. Ten of them considered themselves beginner guitar players, ten were intermediate and ten expert. None of the participants took part in the interviews described in Section 2. All of them reported to play along with songs streamed from services such as YouTube, Spotify, Soundcloud. Specifically, 13 of them use this method frequently, nine sometimes and eight rarely.

5.4 Procedure

The experiments were conducted in part in a laboratory of University of Trento, in part in schools of music and in part at the home of participants.

Each participant was invited to try the two systems. Participants were divided in two groups. To eliminate familiarity biases, half of the group used System 1 first, while the other half used System 2. Participants were asked to try *search via touchscreen* twice, and *search via playing* four times (two chord-based queries and two melody-based). For each downloaded piece, they were also asked to play along with it. After having tried a system, participants were asked to fill an ad-hoc questionnaire. This was partly inspired by the System Usability Scale questionnaire [4] and the questionnaire to calculate the creativity support index [7]. The questionnaire was devised to assess the usability of the systems, investigate whether a system was more suitable for composition, practice, or casual enjoyment purposes, understand the hedonic qualities of the systems [40], and to assess users' preferences for one system over the other.

Specifically, the questionnaire comprised the following questions to be evaluated on a visual analogue scale (VAS). In this scale, zero corresponds to *strongly disagree* and ten stands for *strongly agree*:

- *[Frequency.] I think that I would use this system frequently.*
- *[Complexity.] I found the system complex to use.*
- *[Enjoyment.] I enjoyed using this system.*
- *[Satisfaction.] I was satisfied with the query results I got out of the system.*
- *[Quick learning.] I would imagine that most people would learn to use this system very quickly.*
- *[Composition.] The system is helpful for exploring ideas for composition.*
- *[Practice.] The system is helpful for practicing.*
- *[Fun.] The system is helpful for having fun while playing.*

In addition, participants were asked to answer the following open ended questions:

- *What did you like the most in the system?*
- *What did you like the least in the system?*
- *How would you improve the system?*
- *What is the added value of the system?*
- *Do you think that your way of searching the musical content would improve compared to the use of streaming services and involving the computer?*

At the end of the experiment, participants were asked to answer the following questions:

- *Which of the two ways of search do you prefer? Why?*
- *Which of the two systems has more potential? Why?*

Participants were allowed to select among three possibilities: first system, second system, and no preference. Finally, participants were also allowed to leave an open comment about their experience. On average, participants took one hour to complete the experiment.

5.5 Results

5.5.1 VAS-based questions. Figure 6 shows the results of the VAS-based questions related to the user experience. Responses were not normally distributed, therefore all analyses are henceforth reported using non-parametric statistical tests. Firstly, the questionnaire data were analyzed using the Wilcoxon Signed-Rank Test to assess differences between responses to the two systems for each category of participants' expertise level. The analysis showed that beginners and intermediates deemed that they would use the touchscreen-based system significantly less frequently than the playing-based system (respectively, z = -1.84, $p < 0.05$ and z = -1.97, $p < 0.05$) and that the touchscreen-based system was significantly more complex to use than the playing-based system (respectively, z = -2.25, $p < 0.05$ and z = -3.09, $p < 0.01$); beginners also were found to enjoy the playing-based system significantly more than the touchscreen-based system (z = -1.99, $p < 0.05$). All other comparisons were non significant. Secondly, a Kruskal-Wallis test was carried out to compare the responses of the three groups of participants and for each system. Significant differences between groups were found for frequency of use ($\chi^2(2) = 6.53$, $p < 0.05$), complexity ($\chi^2(2) = 8.17$, $p < 0.05$), and satisfaction ($\chi^2(2) = 5.48$, $p < 0.05$), only for the touchscreen-based system. Wilcoxon Signed-Rank pairwise tests were then carried out for the three pairs of groups (adjusted using the Benjamini & Hochberg correction). Results showed that beginners would use the touchscreen-based system significantly less

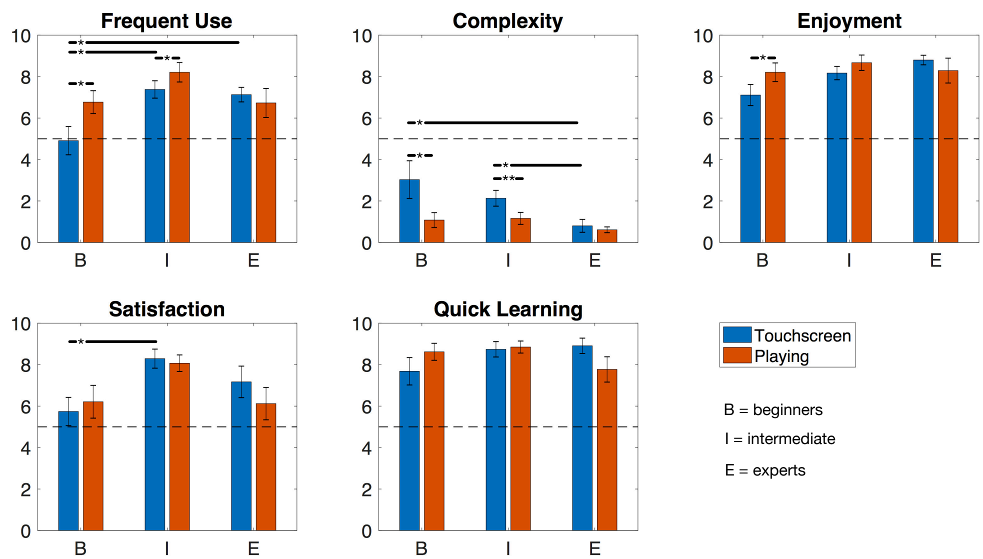


Fig. 6. Graphical representation of the mean and the standard error for participants' answers to the VAS-based questionnaire items related to the user experience with the two systems, in relation to the participants' expertise level. Legend: * represents $p < 0.05$ and ** $p < 0.01$.

frequently than intermediates and experts (both $p < 0.05$); experts find the touchscreen-based system significantly less complex than beginners and intermediates (both $p < 0.05$); intermediates were significantly more satisfied with the results returned by the touchscreen-based system compared to beginners ($p < 0.05$).

Figure 7 shows the results of the VAS-based questions related to the envisioned purpose of the two systems. These questionnaire data were first analyzed with a Wilcoxon Signed-Rank Test to assess differences between responses related to the two systems for each category of participants' expertise level. Results showed that intermediates considered that they would use the touchscreen-based system significantly less frequently than the playing-based system (z = -2.04, $p < 0.05$ respectively) and conversely, experts considered that they would use the touchscreen-based system significantly more frequently than the playing-based system (respectively, z = -1.77, $p < 0.05$). All other comparisons were non significant. Secondly, a Kruskal-Wallis test was carried out to compare the responses of the three groups of participants for each system. Significant differences between groups were found for item practice ($\chi^2(2) = 7.37$, $p < 0.05$) only for the touchscreen-based system. The Wilcoxon Signed-Rank pairwise test (adjusted using the Benjamini & Hochberg correction) showed that experts considered that the touchscreen-based system is helpful for their learning practices significantly more than intermediates ($p < 0.05$). Thirdly, a Kruskal-Wallis test was carried out to compare the responses of each group of participants for each of the three purposes (composition, practice, recreational/fun) for both systems. Significant differences between the three purposes were found for beginners and for the playing-based system only ($\chi^2(2) = 7.8$, $p < 0.05$). Wilcoxon Signed-Rank pairwise tests showed that beginners would use the playing-based system significantly less for practicing purposes compared to both composition and fun purposes (both $p < 0.05$).

5.5.2 Open ended questions. The open ended questions were analyzed with an inductive thematic analysis [3]. The following themes were identified as common to both systems:

Concept and novelty. By far the most common comment (twenty-seven participants) was suggesting a strong appreciation for the idea behind the system, i.e., the direct connectivity of the instrument with an online database

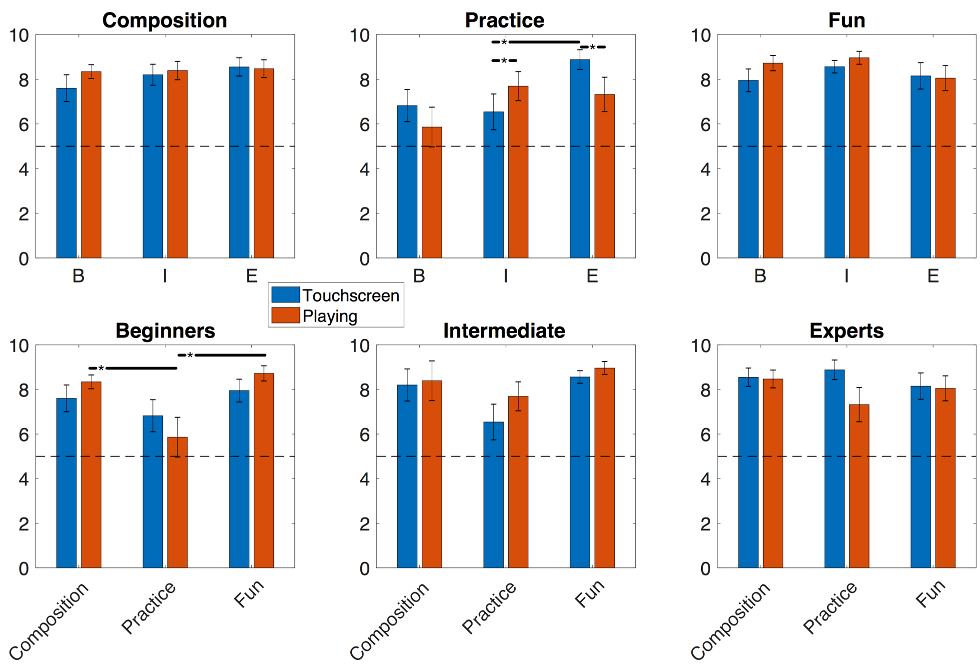


Fig. 7. Two complementary graphical representations of the mean and the standard error for participants' answers to the VAS-based questionnaire items related to the envisioned purpose of the two systems, in relation to the participants' expertise level. Legend: * represents $p < 0.05$.

and the approach to the search based on the actual musical content rather than conventional textual criteria, such as song title or artist name (e.g., *"What I liked the most is the possibility of finding some songs on the basis of the chords and tempo or speed over which I want to play"*). The concept of the system was considered innovative, especially because it enables musicians to retrieve music that corresponds more directly to their actual needs and desires (by playing what they have on their mind). Eight participants considered this aspect stimulating for their musical creativity and considered that the system facilitates the learning process (e.g., *"This system can improve the way of learning while practicing, as well as having fun with what you want to play without being constrained by songs that may have characteristics not matching what you really desire"*).

Improved workflow through ubiquitous standalone system. Five participants commented positively on the standalone nature of both systems, praising their compactness and integration within the instrument, without the need of external devices such as PCs or smartphones. They all mentioned the benefits resulting from the ease of setup compared to conventional settings involving the instrument and the PC. They also appreciated the amplification system embedded into the guitar and the resulting possibility to listen to the music directly from the instrument. Two of them also indicated that the system enhances portability and could be particularly useful for those intending to use the system while traveling.

Along the same lines, twenty-four and twenty-seven participants reported that the touchscreen-based system and the playing-based system would drastically improve their workflow, in terms of comfort and time taken, compared to conventional setting involving the instrument, the computer, and external loudspeakers (e.g., *"With this guitar it is all easier and faster: typically I would need to choose a song, download it on the laptop, understand which key and chords it has, play it via the speakers, and then handle the guitar effects, like setting the delay* *parameters"*). Three participants particularly highlighted how the smart guitar would facilitate actions that typically require some efforts, such as using the keyboard to write the title of the query song while holding the guitar (e.g., *"The fact that it is all integrated into the instrument is important since it avoids laborious movements towards/from the computer while having the guitar in your hands"*). It is interesting to notice that two participants commented that while using YouTube or Spotify during their musical practice, they are displeased by the presence of advertisements, which delays the act of playing and breaks the flow of a playful moment. A smart guitar relying on Creative Commons content would avoid such issue.

Surprise and exploration. One of the most appreciative comments common to both system (reported by seven participants) was related to the possibility of discovering musical content that was unknown (e.g., *"I appreciate the most that this system allows me to discover something unknown, which differs from standard searches with title of the song or name of the artist"; "I enjoyed the randomness of the found piece, as it may be useful to listen and know new sonorities"*). For five participants, it is precisely this element of surprise that fosters the exploration of new content available on the database. This also serves as an encouragement to use the systems and, especially for beginners, to go beyond the own comfort zones (e.g., *"This system allows me to explore unknown content so I can widen my knowledge of musical pieces and try to play on new styles"*). Nevertheless, two participants also suggested that the system should also enable the retrieval of known songs.

Automatic configuration. Five participants (all of whom were either expert or intermediate guitar players) particularly appreciated the automatic configuration feature of the delay time parameter in synchronous relation with the BPM of the downloaded piece. They found this functionality very useful as it would allow them to save a considerable amount of time to find the right value for that parameter using conventional interfaces for the delay effect. Notably, one guitar player also suggested that similarly, other sound effects could be automatically configured to suit well with the retrieved musical content.

Beside such positive aspects, participants also commented on the following issues affecting both systems and also provided suggestions for improvements:

Display of more retrieved results. Eleven guitar players reported that they would prefer that the interface could display more than one result, so that they could choose from a list and select the songs that they like the most. Two participants also suggested to allow the musician to create playlists.

Extension of the search criteria. Several guitar players felt important to complement the available search criteria with others in order to refine the search and make it less unpredictable (if so desired). These include the genre (ten participants), the presence or absence of instruments (six participants), as well as the song title (three participants) and the name of the artist or the band (three participants). Two participants also commented that both systems could allow the search by specifying chords other than those available, such as the suspended or the diminished ones. On the other hand, five participants suggested that the system should retrieve songs having only the chords specified (rather than songs including also extra chords), and two also requested that the desired chords should appear in the song in the order they provided. These aspects were considered particularly important during the use of the systems for learning purposes.

Synchronous chords display. Three participants suggested that both systems could be complemented with an external device, such as a smartphone or a tablet, displaying in real-time the chords occurrence when the retrieved song was reproduced. This suggestion was conceived especially for supporting the learning practices of beginners.

The following theme specific to the touchscreen-based system was identified:

GUI interaction. The possibility to interact with a graphical user interface (GUI) to select certain musical parameters was very much appreciated by 20 participants. This was considered especially helpful for learning harmony theory as well as for practicing chords on different tempi (e.g., *"The added value of this system is that it allows to challenge yourself by selecting particular chords for various BPMs"*). On the other hand, the search by the key and the chords related to it was considered as a limit to those with little theoretical background. In the same vein, four participants recommended to include in the GUI the possibility to recall some presets of chords sequences, especially for training purposes (e.g., *"For beginners it would be nice to have some recommendations for the harmonic possibilities, such as once the key is set, the GUI displays only the chords fitting with that key"*). In a different vein, two subjects suggested to include a functionality in the GUI allowing one to accomplish the search by specifying fewer criteria than those available.

Regarding the playing-based retrieval system, the following themes were identified:

Dissatisfaction with some of the results. Seven participants reported that the system returned musical content that was not satisfactory enough for their purposes and did not correspond to their expectations (e.g., *"I would like that the search could return results more similar to what I played"; "The result is not too close to my execution"*). On the one hand, this was attributed to the actual content of the database, which was requested to be wider. On the other hand, the dissatisfaction was attributed by participants to the limits in the recognition capabilities of the system, which should not only extract the four addressed features, but also other parameters like rhythm, style, timbre, or the mood (e.g., *"I would improve the accuracy of the recognition, detecting also ghost notes and chords that are "dirty", which are more linked to my style"*). Moreover, three of such participants reported that they would have expected a song with a melody similar enough to the one they played.

Latency. A crucial issue was that in some cases the time to retrieve the music piece was long enough to break the flow of the experience. This largely depended on the duration of the recording made, the speed of the available Internet connection, and in part also to the type of query made, which could have forced the system to make multiple trials before returning a piece. Seven participants explicitly mentioned this in their comments, comparing such a latency with the one more reasonable of common streaming services. Nevertheless, typically such participants provided long recordings (about 35 seconds) and with several different chords or with various melody patterns.

From the questions at the end of the experiment, a clear preference emerged for the second system compared to the first one: twenty-two participants (nine beginners, eight intermediates, five experts) reported a preference for retrieving musical content via the act of playing, rather than interacting with a touchscreen to select keywords. These preferences of participants' can be explained by motivation identified in the following themes:

More immediate, user-friendly and adaptable system. Seven participants addressed their preference for the playing-based system because of its immediacy and directness with the act of playing compared to the interaction using the touchscreen (e.g., *"The playing-based system is more immediate to use, you don't have to select the parameters but you simply have to play. This is easier compared to selecting the criteria with the touchscreen"; "I like more the workflow of the second system: with the touchscreen-based system you have to select manually any single option, instead the second the system automatically recognizes your intended choices"*). In particular, six participants commented that the playing-based system was more immediate and intuitive to use due to the fact that no music theory knowledge is required (e.g., *"The touchscreen-based system requires a theoretical background, whereas with the playing-based system you just play what you know. The touchscreen-based system can be useful when you are proficient and you want to practice a certain set of chords."; "I prefer the playing-based system because it allows to express more naturally your ideas even without having solid musical background"*). Along the same lines, five participants mentioned the preference for the second system, because of its ability to adapt to the musician (e.g., *"The second system adapts to you, instead with the touchscreen-based system you have certain options to select from"; "I prefer the method involving playing because it adapts much more to the actual abilities of the musician"*).

More comfortable and faster retrieval. Four participants attributed their preference for the second system to the type of interaction afforded, which was deemed more comfortable, with the benefit of a faster search process (e.g., *"I prefer it because it allows me to detach my hands from the instrument much less time compared to the touchscreen system"; "The difference is that with the touchscreen-based system I need to think and choose what to* *play using the GUI, while with the automatic recognition system the search process is much easier and faster, as I just record and play"*).

More engaging and more creative interactions. Five participants attributed their preference for the second system to its greater level of induced or required engagement, which also fosters the use of the system and generates more curiosity towards its results (e.g., *"It is a more recreational and funny experience"; "Playing requires a more active engagement of the musician, and this encourages more the utilization of the system"; "It spurs more my curiosity for the surprise of discovering what the system has retrieved"; "There is a greater relation between the musicians and their expectations"*). Two participants highlighted that playing is a more creative act compared to clicking some widgets on a touchscreen (e.g., *"I preferred it for the simplicity of the search and the possibility to express your own compositional ideas"*).

More innovative approach and more capabilities. Six participants reported to have assigned their preference for the second system because they appreciated most the innovative approach based on the automatic recognition of what is being played (e.g., *"I find revolutionary the possibility to shape my backing tracks and my musical culture on the basis of what I play. This is in contrast with what happens normally, where one builds his own way of playing on songs and melodies that are already known"*). In addition, four participants stressed the fact that the possibility of querying the database by means of a melody was their main factor for the preference towards the second system, compared to the touchscreen-based system that is devoid of it.

Five participants (one intermediate, four experts) did not express a preference for any of the two systems. The thematic analysis of such answers lead to the following theme:

Different purposes for different categories of users. The five participants attributed their lack of preference to the fact that the systems can target different purposes catering the needs of musicians with different expertise level, and therefore can be seen as complementary (e.g., *"I think that the two systems have two different aims: by touchscreen for practicing purposes, by playing for fun purposes"; "The playing method is more intuitive and immediate to use for beginners. The touchscreen method is more suitable for guitar players that know exactly what to search, with greater theoretical and practical background"*). Three participants also commented that the two systems are complementary.

Regarding the preference about the potential of the two systems, twenty-one participants (eight beginners, six intermediates, seven experts) indicated that the playing-based system has more potential compared to the touchscreen-based system. The thematic analysis identified the following theme motivating such a choice:

Results closer to the search intentions. Five participants motivated their preferences for the playing-based system with the fact that future developments could lead to searches much closer to the actual desire of the musician (e.g., *"It has the potential to make the search much closer to what you really want"*). Specifically, three participants mentioned expressivity as an aspect crucial in their preference (e.g., *"The second system has a greater potential because can capture the expressivity of the musician, thing that is not possible with the first method"; "With the playing-based system there is the possibility for searches more linked to the expressivity of the musician"*).

More users. Three participants considered that the playing-based system had more potential because it can be used by a wider basis of guitar players, including beginners that do not possess a solid theoretical background. This was corroborated by the envisioned capability of the system to adapt to the expertise level of the guitar player (e.g., *"Musicians can search songs that are suitable to their level of expertise: typically one has to do this himself, but with this system it could be automatic"*).

Five participants (one beginner, one intermediate, three experts) did not express a preference for a system regarding its potential, whereas four participants (one beginner, three intermediates, eight experts) considered that the touchscreen-based system had more potential compared to the playing-based method. The latter was motivated by the following theme:

Greater range of applicability. Two participants indicated that the touchscreen-based method may have more potential because it can be adapted to various instruments (e.g., *"it can be used for any instrument, without* *being related to the specific acoustic signal of the guitar"*). Nevertheless, contrarily to what those participants assumed, the developed playing-based systems were instrument-agnostic.

5.6 Discussion and summary of the evaluation

The quantitative results reported in Figure 6 about participants' experience revealed differences between the three groups of guitar players as well as between the two systems. Such quantitative results were corroborated by the qualitative results of the thematic analyses of the open ended questions.

The touchscreen-based system was considered more complex to use by beginners and intermediate guitar players compared to the playing-based approach. This was corroborated by results of the envisioned use of the systems, where beginners and intermediates reported that would use the playing-based system more frequently than the touchscreen-based system. Accordingly, results show that beginners enjoyed significantly more the playing-based approach rather than the one relying on the touchscreen. Notably, these differences between the two systems were not found for expert guitar players. Along the same lines, the touchscreen-based approach was considered more complex to use by beginners and intermediates compared to experts. Results also show that beginners would use the touchscreen-based system less frequently than intermediates and experts, and that their level of satisfaction for the results returned by the system was significantly smaller than those of intermediates.

These differences in complexity, envisioned frequency of use and enjoyment between the two systems and groups were motivated in part by the need of a theoretical musical background to use the touchscreen-based system and in part by the more difficult, less intuitive, and less immediate interaction afforded by it. These aspects were more relevant for beginners than intermediates, whereas, reasonably, the experience of expert guitar players did not significantly differ along the investigated dimensions.

The differences expressed by participants regarding their experience partly reflected their envisioned purpose of the two systems (see Figure 7). Intermediates appear to prefer the playing-based approach for practicing purposes, whereas the contrary happens for experts. These results can be explained by the fact that intermediates find the retrieval of the desired musical content easier using the act of playing, and letting the system automatically recognize the criteria of their search. For beginners this does not happen, likely because playing is still a difficult activity for them, which could lead to imprecise recordings. As a consequence, this may lead to the retrieval of undesired content, which would be detrimental during the process of learning. Under other conditions, the touchscreen-based system was not preferred by beginners for practicing purposes, likely because of the required theoretical knowledge. These results for beginners are corroborated by the fact that they considered the playing-based system more helpful for composition and recreational purposes rather than for practice purposes. Such differences were not present for the touchscreen-based system.

Conversely, the experts judged the touchscreen-based system as more helpful for the process of learning compared to the playing-based method. This in part was due to the fact that they envisioned the issues for beginners mentioned above and in part because the touchscreen-based system spurs more reflection on the query to make, especially accounting for the player's theoretical knowledge and the need for practicing a musical content with precise musical parameters. This is also supported by the fact that experts judged the touchscreen-based approach as significantly more suitable for practicing compared to intermediates, for whom the immediacy of the interaction enabled by the playing-based method is a very important factor. Notably, no significant differences were found between the three purposes for either intermediates or experts, leading to the conclusion that they would use both systems for composition, practicing or having fun.

6 GENERAL DISCUSSION

The interviews conducted with the first batch of thirty guitar players allowed one to establish an understanding of the needs and expectations of guitar players from systems interconnecting an instrument to an online repository for retrieving musical content. The developed prototypes focused only on a small subset of the criteria identified by the interviewed participants, the ones that were more easily implementable with the semantic audio tools available at hand. The technical and experiential evaluations allowed us to validate our concept as well as assess the strengths and limitations of the developed prototypes. The first noticeable element emerging from both the interviews and the hands-on experiments was that the proposed search methods, based on musical parameters, were very appreciated by the guitar players. Such an appreciation was attributed not only to the novelty of the concept, but also for the fact that the developed systems can address actual needs of musicians which are not satisfied by current search methods based on song title or artist's name.

The developed systems implement the proposed concept in three different ways at the technical level and in two modalities at the user interaction level. The first system accomplished the query by providing the server with parameters set by the user via a touchscreen, the second system leveraged the cloud computing paradigm where the features are extracted by the server from the content recorded by the guitar player. Whereas the third system applied the edge-computing paradigm by performing the analysis of the recording on the computing unit embedded in the instrument. Notably, the methods presented in this study are applicable to smart instruments other than the smart guitar. Indeed, the three systems followed a musical instrument-agnostic approach.

Nevertheless, the playing-based approach of the second and third systems could be tailored to the peculiarities of the guitar, by applying music information retrieval techniques specific to the guitar's acoustic signal. This would lead to the extension of the parameters with which performing the search (e.g., the style) and the improvement of the detection accuracy, which are aspects that would allow to cater to the needs expressed by participants during the experiments. However, accomplishing such a vision entails the progression of the state of the art in semantic audio and music information retrieval techniques that need to be robust enough to capture and represent nuances of the execution, which represents an open challenge [42].

Notably, during both the interviews and the experiments some guitar players envisioned the timbre as a search criterion, especially considering the sonorities resulting from the applications of various sound effects (e.g., distortion). The retrieval process could exploit the association between the instrument's timbral features with certain musical genres. This possibility could be achieved by a smart instrument without relying on any music information retrieval technique, but leveraging the intelligent capability of the instrument of representing itself [33], which includes the knowledge of which software plugins are utilized in the sound effect chain [38]. Along the same lines, a particular feature of the developed prototypes was the possibility of automatically configuring parameters of the instrument's sound engine on the basis of the retrieved content. Such feature was considered useful especially by guitar players more proficient in sound processing technologies, as it avoids having to deal with continuous adjustments of effects parameters that normally have to be modified for each downloaded music piece.

The time taken by the retrieval process is a crucial aspect for the overall user experience, as emerged from the comments of participants of the experiments. A complicating factor is represented by the computational load that may be rather high for certain tasks to be performed on an embedded system in the third approach. This may lead to long waiting times before obtaining the query result. Our timing benchmarks confirmed this to be the case. Nevertheless, besides the need for more powerful embedded systems or more efficient cloud-based algorithms for audio analysis and database queries, it is also necessary to improve the latency and variability of the network, which are factors contributing to the effective latency in the interaction. The envisioned Tactile Internet [19] is expected to drastically improve network latency issues and enable more time-efficient Internet of Musical Things applications such as the ones discussed here [37].

In theory, Systems 2 and 3 are more susceptible to returning less optimal suggestions, because they use audio analysis both for querying and for indexing the music collection, whereas the querying in System 1 is done through direct specification of the musical characteristics. In the former systems, errors in both analysis stages therefore have the potential to accumulate. Nevertheless, the results of the audio analysis on a representative set of example queries (see Figure 5 and Table 3) shows that any errors being made are usually musically related. Consequently, and because music recommendation is inherently vague, it's not surprising that none of the participants noticed the potentially reduced relevance of the returned audio in practice.

In comparison with earlier examples of query-by-example systems [13], we argue that our Systems 2 and 3 are novel in the sense that they use well-defined high-level musical characteristics to retrieve similar audio, instead of less interpretable similarity measures. It allows to search for audio that is similar in one or more specific characteristics relevant for the use-case, while consciously opting to ignore other characteristics, as opposed to trying to define a general measure of similarity.

Currently, the wireless connection of smart instruments to the cloud is still partially hindered by the lack of real-time standard solutions. For smart instruments, the typical latency that the wireless communication protocols must offer has to be below 20 ms with round-trip links (from the player to the cloud, and back) [37]. To achieve such latency, there are two networking aspects to be considered: wireless access (connection of the smart instrument to an access point or base station) and cloud networking. Concerning wireless access, we have two options: cellular protocols such as the 4-th generation (4G) or the upcoming 5-th generation (5G), or wireless personal/local area networks (WPAN/WLAN) standards (IEEE 802.11-like standards). The state-of-the-art cellular wireless protocols (4G) will not be able to offer 20 ms round-trip latency [16]. Perhaps the upcoming 5G generation of wireless cellular protocols will be able to meet such latency requirements, but only for very short packets and in relatively small geographical areas. Concerning WPAN/WLAN standards, the use of millimeter waves will arguably permit to achieve low latency communications [31], but at the moment we do not have a definite solution. Regarding the cloud networking aspects, i.e., the communication between the wireless access point/base station to the data repository, this is an area that needs substantial technological developments, because low latency networking has not been the primary concern of such technology up until now.

The smart musical instrument will be equipped with embedded hardware and software platforms having relatively small computational resources. Performing data analysis on those platforms may be challenging, especially if we consider that machine learning methods may need large training data sets and state-of-the-art trained models such as neural networks may have a large number of trained parameters that would have to be stored on the device. Here, the alternative would be to perform data analysis computations on the smart instrument devices by algorithms of low computational and communication complexity. This is a nascent topic within the machine learning community in the field of "Machine Learning over Networks". The major issue to perform real-time data analysis on the cloud is that machine learning over communication networks faces the fundamental bandwidth limitations and latency of the communications, as we have mentioned in the previous paragraph. The emerging technology of extremely low latency communications will rely on short packets that carry few bits [16]. The smart musical instruments generating data, as many other IoT devices in other domains, may not have enough communication bandwidth to transmit data where it has to be analyzed in real-time, or simply not enough computational power to perform local analysis [9, 14]. Finally, a further problem is that privacy and security in these systems could be threatened. Therefore, finding efficient ways to perform machine learning in these scenarios is still largely an open question.

One of the strengths of the developed prototypes is the novel kinds of interaction with online music repositories they afford. This effectively allows guitar players to overcome typical issues occurring while conducting the search of the musical content they want to play along with, such as the difficulty and the reduction of the freedom of movements due to the use of the computer while holding the guitar, or the time and effort needed to setup [20, 44] a system for a given scenario. The drastic improvement of the workflow reported by the vast majority of participants compared to their usual searching practices is due to the standalone and direct connectivity properties of the smart guitar, which integrates in a unique, portable solution all the technologies needed to perform the search, without relying on external equipment. This has also the benefit to afford ubiquitous musical activities [17, 35] as highlighted by some participants. Such an appreciation is in line with the results of other studies assessing musicians' interactions with smart musical instruments (e.g., [27]). Overall, the developed prototypes, in particular the one implementing the playing-based approach, represent an attempt to respond to the challenge highlighted by Martinez-Avila et al. in [20]: *"We believe that the most promising opportunities for technical intervention in terms of augmenting the guitar lie in improving the relationship between the musician with guitar in hand and the various resources that support them through the rehearsal process [...] The third possible solution is to make the system "smart", i.e., for the system to work out what the musician wants to do in terms of supporting resources and do it automatically for them".*

The two interaction modalities investigated (using the touchscreen vs. playing) were tested under the hypothesis that querying via the act of playing would lead to a more natural and immediate way of performing the search compared to querying with a textual research. This hypothesis, which arose from participants' comments emerged during the requirements gathering phase, found to be confirmed by the evaluations and in the responses reported by several participants testing the prototypes. In general, the results of the user experience evaluations of the prototypes were in line with the expectations expressed by previous participants who were interviewed to gather the requirements for the design of the systems. Compared to the touchscreen-based system, the playing-based system was considered to implement a more innovative approach and offer more capabilities than the first one, as well as to enable more creative and more engaging interactions. Importantly, the easiness of use was one of the main discriminating factors between the interactions afforded by the two systems: the use of the touchscreen-based system assume more advanced knowledge of music theory, which represents an obstacle for novices, especially for practicing purposes. Conversely, the touchscreen-based system was deemed by expert players more helpful for the process of learning compared to the playing-based method. The results of both interviews and experiments also indicated that participants would need to complement the playing-based method with some criteria that can't be specified with the sole act of playing, such as querying by the type of ensemble (e.g., full ensemble or backing track without the guitar). Taken together these results highlight the complementarity of the two approaches and suggest their combination.

A study limitation might be that the participants' background could have affected the results of the evaluations reported in Sections 2.4 and 5.6. Factors may include gender, age, cultural background, musical background (e.g., musical listening preferences, genre of music played). A deeper analysis of the results of the first experiment (Section 2.3) revealed that no differences could be attributed to these factors. Similarly, the comparison of participants' evaluations in the second experiment (Section 5.5), according to the same factors, did not reveal important differences (although the authors acknowledge a clear gender unbalance). Overall, results indicate that the main factor that caused differences in participants' evaluations can be attributed to the level of guitar playing expertise.

7 CONCLUSIONS

This paper investigated novel methods to retrieve musical content from an online repository of Creative Commons audio, which is an alternative to the conventional process of text-based searches available in common streaming services (e.g., YouTube, Spotify). The key novelties were a search based on parameters of the actual musical content and the direct interaction of a musical instrument with the online music database. Three prototypes were developed which allow novice, intermediate, and expert musicians to explore music pieces from Jamendo's collection, based on four musical parameters: key, chords, BPM, and tuning. While the proposed prototype systems represent a musical instrument-agnostic approach, we focussed on the guitar. Overall, the proposed approaches were welcomed by guitar players who appreciated the possibilities offered by the systems and positively assessed its usefulness within their practices.

In general, the query by playing was the approach preferred by most participants and was judged to be the one with the highest potential to reach a wide basis of guitar players as well as to better cater for their needs. However, it is also the approach more difficult to fully implement, as it assumes the existence of automatic techniques capable of understanding the intention of the musician with a high degree of accuracy. These represent a technical challenge especially in the context of real-time computations on embedded systems.

In future work, we plan to investigate queries based on an extended set of search criteria (e.g., mood), as well as to integrate query-by-playing with textual searches, accounting for criteria not expressible by playing alone (e.g., presence or absence of instruments in the ensemble of the music piece to be retrieved). Moreover, we plan to explore other opportunities for the automatic configuration of the smart guitar's sound engine on the basis of the retrieved content. This represents one of the strengths of the developed prototypes. Furthermore, we plan to apply the developed methods to other smart instruments and involving other online music repositories.

To date, cloud-SMIs interactions represent an unexplored line of research within the emerging field of the Internet of Musical Things. The authors hope that this work could inspire other practitioners to investigate future applications of SMIs interacting with online music collections.

ACKNOWLEDGMENTS

The authors wish to thank all the guitar players that took part to the experiments. This work has been partly funded by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L019981/1 and by the European Union's Horizon 2020 research and innovation programme Audio Commons under grant agreement N° 688382.

REFERENCES

[1] Emmanouil Benetos, Simon Dixon, Zhiyao Duan, and Sebastian Ewert. 2019. Automatic Music Transcription: An Overview. *IEEE Signal Processing Magazine* 36, 1 (2019), 20–30.

[2] D. Bogdanov, N. Wack, E. Gómez Gutiérrez, S. Gulati, P. Herrera Boyer, O. Mayor, G. Roma Trepat, J. Salamon, J.R. Zapata González, and X. Serra. 2013. Essentia: An audio analysis library for music information retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference*. 493–498.

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[4] John Brooke et al. 1996. SUS–A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[5] J.A. Burgoyne, I. Fujinaga, and J.S Downie. 2016. Music Information Retrieval. *A New Companion to Digital Humanities* (2016), 213–228.

[6] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. 2008. Content-based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE 96* 96, 4 (2008), 668–696.

[7] E. Cherry and C. Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction* 21, 4 (2014), 21.

[8] Norberto Degara, Enrique Argones Rua, Antonio Pena, Soledad Torres-Guijarro, Matthew E. P. Davies, and Mark D. Plumbley. 2012. Reliability-informed beat tracking of musical signals. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1 (2012), 290–301.

[9] R. Du, P. Santi, M. Xiao, A. V. Vasilakos, and C. Fischione. 2018. The sensable city: A survey on the deployment and management for smart city monitoring. *IEEE Communications Surveys Tutorials* (2018). https://doi.org/10.1109/COMST.2018.2881008 Accepted. To Appear.

[10] G. Fazekas and T. Wilmering. 2012. Semantic Web and Semantic Audio Technologies. *Tutorial presented at the 132nd Convention of the Audio Engineering Society, Budapest, Hungary* (2012).

[11] F. Font, G. Roma, and X. Serra. 2013. Freesound technical demo. In *Proceedings of the ACM international conference on Multimedia*. ACM, 411–412.

[12] Masataka Goto and Yoichi Muraoka. 1997. Issues in Evaluating Beat Tracking Systems. In *Working Notes of the IJCAI-97 Workshop on Issues in AI and Music – Evaluation and Assessment*. 9–16.

[13] Marko Helén and Tuomas Virtanen. 2010. Audio query by example using similarity measures between probability density functions of features. *EURASIP Journal on Audio, Speech, and Music Processing* 2010 (2010), 1.

[14] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, and T. Melodia. 2019. Machine Learning for Wireless Communications in the Internet of Things: A Comprehensive Survey. *CoRR* abs/1901.07947 (2019). arXiv:1901.07947 http://arxiv.org/abs/1901.07947

[15] J-S. R. Jang, H-R. Lee, and J-C. Chen. 2001. Super MBox: An Efficient/Effective Content-Based Music Retrieval System. In *Proceedings of the Ninth ACM International Conference on Multimedia*. 636–637.

[16] X. Jiang, H. Shokri-Ghadikolaei, G. Fodor, E. Modiano, Z. Pang, M. Zorzi, and C. Fischione. 2018. Low-latency Networking: Where Latency Lurks and How to Tame It. *Proceedings of the IEEE (Accepted for Publication, To Appear)* (2018).

[17] D. Keller, V. Lazzarini, and M.S. Pimenta. 2014. *Ubiquitous music*. Springer.

[18] P. Knees, T. Pohle, M. Schedl, and G. Widmer. 2007. A Music Search Engine Built upon Audio-based and Web-based Similarity Measures. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 447–454.

[19] M. Maier, M. Chowdhury, B.P. Rimal, and D.P. Van. 2016. The tactile internet: vision, recent progress, and open challenges. *IEEE Communications Magazine* 54, 5 (2016), 138–145.

[20] J. Martinez-Avila, C. Greenhalgh, A. Hazzard, S. Benford, and A. Chamberlain. 2019. Encumbered Interaction: a Study of Musicians Preparing to Perform. In *Proceedings of the Conference on Human Factors in Computing Systems*. ACM.

[21] A. McPherson and V. Zappi. 2015. An Environment for Submillisecond-Latency Audio and Sensor Processing on BeagleBone Black. In *Audio Engineering Society Convention 138*. Audio Engineering Society. http://www.aes.org/e-lib/browse.cfm?elib=17755

[22] T. Mitchell, S. Madgwick, S. Rankine, G.S. Hilton, A. Freed, and A.R Nix. 2014. Making the Most of Wi-Fi: Optimisations for Robust Wireless Live Music Performance. In *Proceedings of the Conference on New Interfaces for Musical Expression*. 251–256.

[23] Johan Pauwels. 2016. *Exploiting prior knowledge during automatic key and chord estimation from musical audio*. Ph.D. Dissertation. Ghent University.

[24] J. Pauwels, G. Fazekas, and M. B. Sandler. 2018. Recommending Songs to Music Learners Based on Chord Content. In *Proceedings of the 2018 Joint Workshop on Machine Learning for Music*.

[25] J. Pauwels, K. O'Hanlon, G. Fazekas, and M. B. Sandler. 2017. Confidence measures and their applications in music labelling systems based on hidden Markov models. In *Proceedings of the 18th Conference of the International Society for Music Information Retrieval (ISMIR)*. 279–285. https://ismir2017.smcnus.org/wp-content/uploads/2017/10/195_Paper.pdf

[26] W. Piston. 1948. *Harmony*. WW Norton.

[27] C. Rossitto, A. Rostami, J. Tholander, D. McMillan, L. Barkhuus, C. Fischione, and L. Turchet. 2018. Musicians' Initial Encounters with a Smart Guitar. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*. ACM, New York, NY, USA, 13–24. https://doi.org/10.1145/3240167.3240223

[28] M. Satyanarayanan. 2001. Pervasive computing: Vision and challenges. *IEEE Personal communications* 8, 4 (2001), 10–17.

[29] H. Sharp, Y. Rogers, and J. Preece. 2019. *Interaction design: beyond human-computer interaction*. John Wiley & Sons.

[30] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu. 2016. Edge computing: Vision and challenges. *IEEE Internet of Things Journal* 3, 5 (2016), 637–646.

[31] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi. 2015. Millimeter Wave Cellular Networks: A MAC Layer Perspective. *IEEE Transactions on Communications* 63, 10 (Oct 2015), 3437–3458. https://doi.org/10.1109/TCOMM.2015.2456093

[32] L. Turchet. 2018. Smart Mandolin: autobiographical design, implementation, use cases, and lessons learned. In *Proceedings of Audio Mostly Conference*. 13:1–13:7. https://doi.org/10.1145/3243274.3243280

[33] L. Turchet. 2019. Smart Musical Instruments: vision, design principles, and future directions. *IEEE Access* 7 (2019), 8944–8963. https://doi.org/10.1109/ACCESS.2018.2876891

[34] L. Turchet and M. Barthet. 2018. Jamming with a smart mandolin and Freesound-based accompaniment. In *IEEE Conference of Open Innovations Association (FRUCT)*. IEEE, 375–381. https://doi.org/10.23919/FRUCT.2018.8588110

[35] L. Turchet and M. Barthet. 2019. An ubiquitous smart guitar system for collaborative musical practice. *Journal of New Music Research* 48, 4 (2019), 352–365. https://doi.org/10.1080/09298215.2019.1637439

[36] L. Turchet, M. Benincaso, and C. Fischione. 2017. Examples of use cases with Smart Instruments. In *Proceedings of Audio Mostly Conference*. 47:1–47:5. https://doi.org/10.1145/3123514.3123553

[37] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet. 2018. Internet of Musical Things: Vision and Challenges. *IEEE Access* 6 (2018), 61994–62017.

[38] L. Turchet and P. Kudumakis. 2019. Requirements for a File Format for Smart Musical Instruments. In *Proceedings of the Workshop on Multilayer Music Representation and Processing*. 5–9. https://doi.org/10.1109/MMRP.2019.8665380

[39] G. Tzanetakis and P. Cook. 2000. 3D Graphics Tools for Sound Collections. In *Proceedings of the Conference on Digital Audio Effects*.

[40] I. Wechsung and A. B. Naumann. 2008. Evaluation methods for multimodal systems: A comparison of standardized usability questionnaires. In *International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*. Springer, 276–284.

[41] M. Weiser. 1991. The Computer for the 21st Century. *Scientific american* 265, 3 (1991), 94–105.

[42] G. Widmer. 2017. Getting closer to the essence of music: The Con Espressione manifesto. *ACM Transactions on Intelligent Systems and Technology* 8, 2 (2017), 19.

[43] E. Wold, T. Blum, D. Keislar, and J. Wheaten. 1996. Content-based Classification, Search, and Retrieval of Audio. *IEEE Multimedia* 3, 3 (1996), 27–36.

[44] A. Xambó, J. Pauwels, G. Roma, M. Barthet, and G. Fazekas. 2018. Jam with Jamendo: Querying a Large Music Collection by Chords from a Learner's Perspective. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*. ACM, 30:1–30:7. https://doi.org/10.1145/3243274.3243291

Received XXX XXX; revised XXX XXX; accepted XXX XXX