Voice-based interface for accessible soundscape composition: composing soundscapes by vocally querying online sounds repositories

Luca Turchet
Department of Information Engineering
and Computer Science
University of Trento
luca.turchet@unitn.it

Alex Zanetti
Department of Information Engineering
and Computer Science
University of Trento
alex.zanetti@unitn.it

Figure 1: Schematic representation of the implemented Internet of Audio Things ecosystem.

ABSTRACT

This paper presents an Internet of Audio Things ecosystem devised to support soundscape composition via vocal interactions. The ecosystem involves a commercial voice-based interface and the cloud-based repository of audio content Freesound.org. The user-system interactions are exclusively based on vocal input/outputs, and differ from the conventional methods for retrieval and sound editing which involve a browser and programs running on a desktop PC. The developed ecosystem targets sound designers interested in soundscape composition and in particular the visually-impaired ones, with the aim of making the soundscape composition practice more accessible. We report the results of a user study conducted with twelve participants. Overall, results show that the interface was found usable and was deemed easy to use and to learn. Participants reported to have enjoyed using the system and generally felt that it was effective in supporting their creativity during the process of composing a soundscape.

AM'20, September 15–17, 2020, Graz, Austria

ACM ISBN 978-1-4503-7563-4/20/09…$15.00
https://doi.org/10.1145/3411109.3411113

CCS CONCEPTS

• **Human-centered computing** → **Sound-based input / output**; • **Software and its engineering** → *Software libraries and repositories*; • **Applied computing** → Sound and music computing.

KEYWORDS

Internet of Audio Things, voice assistant, conversational AI, online sound repository, Freesound

ACM Reference Format:
Luca Turchet and Alex Zanetti. 2020. Voice-based interface for accessible soundscape composition: composing soundscapes by vocally querying online sounds repositories. In *Proceedings of the 15th International Audio Mostly Conference (AM'20), September 15–17, 2020, Graz, Austria.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3411109.3411113

1 INTRODUCTION

The term "soundscape" refers to all the sounds that can be heard in a specific location. This sonic environment is the aural counterpart of the term landscape referred to visually-related items in an environment. Research on real soundscapes started with R. Murray Schafer, among others, in late sixties [27] and continued by focusing mostly on musical applications, with pioneering works of Barry Truax [32, 33]. "Soundscape composition" refers to a sound-based art form that concerns the creation of sonic environments [10, 34, 46]. This art form has grown from acoustic ecology [47] and soundscape studies [46].

Composed soundscapes are used widely in various contexts, including movies [11, 25], music performances [17, 34], artistic installations [7, 24], and virtual environments [12, 40, 41]. To date, soundscape composition is facilitated by the availability of high-quality commercially available sound effects libraries, conceived especially for creating environmental sounds in movies. In recent years, large repositories of sounds are becoming available online. One of the most popular and freely available online repositories is Freesound[1], a collaborative repository of audio samples developed at and maintained by the Music Technology Group of Universitat Pompeu Fabra [14, 16]. The Freesound database provides a collection of several hundreds of thousands of crowd-sourced non musical and musical sounds licensed under Creative Commons, and is part of the Audio Commons Initiative [15].

The Audio Commons Initiative is a recent endeavor aiming to bridge the gap between audio content producers, providers and consumers through a web-based ecosystem. The approach combines techniques from music information retrieval (to extract creative metadata to automatically annotate audio content) and the semantic web (to structure knowledge and enable intelligent searches). Content aggregators part of the Audio Commons ecosystem, such as Freesound, provide access to audio data through user-facing and application programming interfaces (APIs). In Freesound, the available metadata information about the sounds depends on what has been provided by authors during uploads including tags, descriptions or file names [13]. Freesound enables designers to create third-party applications exploiting its audio content in live applications by granting access to the database trough a REST API [1].

Various systems for soundscape composition have been developed, including real-time [43], interactive [44], non interactive [28, 31], automatic [42], and even tangible [21]. On the other hand, recently researchers are exploring initiatives to combine embedded systems for Internet of Things with Audio Commons ecosystems in order to create new forms of artistic interaction with audio content [29, 36]. However, to the best of authors' knowledge, a tool for soundscape composition based on vocal interactions and leveraging Audio Commons ecosystems has not been devised yet. Such a system may be proven particularly useful for visually-impaired sound designers and those without the use of their hands.

In this paper we explore the use of a speech-based system able to interface with Audio Commons ecosystems for the retrieval of online audio content and its repurposing into soundscape composition practice. We present a prototype involving a commonly available vocal interface used to query content from Freesound and utilize it to generate a soundscape in real-time. This application is positioned within the context of the emerging Internet of Audio Things (IoAuT) field, an extension of the Internet of Things paradigm to the audio domain [37]. The developed IoAuT ecosystem was devised to support soundscape composition by leveraging interactions only based on audio input/outputs, differently from the conventional methods for retrieval and sound editing which involves a browser and programs running on a desktop PC. This study targets sound designers interested in soundscape composition and in particular those with visual and hand impairments.

[1]http://www.freesound.org

The remainder of this paper is organized as follows. Section 2 presents an overview of related works. Section 3 describes the developed IoAuT ecosystem, while Section 4 presents a user study that assessed it. Finally, Section 5 provides summarizing conclusions.

2 RELATED WORKS

In this section we review key works on technologies related to the proposed architecture.

2.1 Voice-based interfaces for accessible interactions

The emergence and widespread availability of speech recognition and synthesis systems embedded in mobile and in-home digital assistants (e.g., Google Home, Apple's Siri, Amazon Echo), as well as mobile screen readers and chatbots, are fostering novel interactive applications to support communication, collaboration, and information seeking. This increasing availability is also providing new opportunities for broad, accessible interaction by voice. This is due to the fact that voice-based interfaces do not require visual and motor skills needed for text input through a keyboard, which lowers the barriers of entry and use for older adults and people with disabilities.

A number of studies have investigated the use of voice-based interactions for the control of interactive applications [22], in particular for accessibility purposes [5]. Examples within this domain include interfaces substituting input devices (such as the mouse [19]), targeting various kinds of applications (e.g., web navigation [9] or computer games [20]) and various categories of users (including the visually-impaired [3], elderly population [23], people with motor impairments [26] or cognitive disabilities [18]). Overall, these studies show the effectiveness of voice-based interfaces for replacing other kinds of interaction modalities.

2.2 The Audio Commons initiative and its artistic use

The Audio Commons Initiative [15] provides an ecosystem through which sound designers and musicians can access audio content with various tools, including interfaces based on web browsers (e.g., Freesound [14], Jamendo [2]), audio plugins, or live coding tools [48]. This web-based approach that provides access to distributed audio content in a user-friendly way, aims to bridge the gap between audio content producers, providers and consumers. This is achieved in a different way from methods based on traditional digital audio workstations and digital musical interfaces, which were conceived to operate with local audio content (for example personal recordings gathered by the musician).

Various systems have recently leveraged the creative opportunities offered by the Audio Commons ecosystem thanks to its audio content search engine informed by semantic metadata and audio content-based features. This search engine enables quick access to hundreds of thousands of sounds from various online content providers according to requirements matching the sound designers' or composers' needs. Playsound [30] is a web-based tool designed for beginners or advanced musicians willing to explore music composition based on semantic ideation and spectrogram sound inspection. In a different vein, other ecosystems based on

the Internet of Audio Things [37] and Internet of Musical Things [38] paradigms interface Audio Commons repositories with devices based on embedded systems such as wearables [29] or smart musical instruments [35]. The study reported in [29] proposed a sonic wearable interface letting users trigger and transform sounds downloaded from Freesound through body-based gestural interactions tracked by e-textile sensors. Along the same lines, the ecosystems reported in [36] and [39] used sounds retrieved from Freesound and Jamendo onto smart instruments [35] for music learning, improvisation, composition, and participatory performance purposes.

3 OVERVIEW OF THE ECOSYSTEM

The implemented IoAuT ecosystem aimed to enable the creative use of content retrieved from Freesound via a voice-based interface. Figure 1 shows a schematic representation of its main components, user-system vocal interactions, and data flow. The voice assistant utilized was Alexa of the Amazon Echo device, which was connected to the Internet via a Wi-Fi router. The system was implemented using the Software Development Kits provided by Amazon for the development of programs for the Alexa vocal assistant. The program leveraged the Freesound API using the Python client released on the Freesound developer Github page [2].

Figure 2 illustrates the designed interactions between the user and Alexa to achieve the task of composing a soundscape. At the outset after the Alexa wake up word for starting the program (*"Alexa, Open Freesound"*), the user is provided with a welcome sentence which also acts as a helper (*"Welcome, you can say "Play", "Get my kept tracks", "Delete kept tracks" or "Help". During the reproduction you can say "Alexa, pause" to keep current track, "Alexa, previous" to play the previous track or "Alexa, next" to play the next track"*). However, this long sentence is not reproduced each time the program starts, but only if the program has not been used in the previous 5 minutes. This in order to speed up the interaction (the user can directly pronounce the desired command).

When the user pronounces *"Play [name of the sound to be retrieved]"*, the program will retrieve the tracks corresponding to that sound according to some criteria definable by the user. By default, the retrieval is based on the highest ranking attributed by Freesound search engine. Nevertheless, the system also allows to retrieve sounds by ascending order of duration (with *"Play [sound name] by duration"*), date of creation from the most to the last recent one (with *"Play [sound name] by creation"*), or by ascending order of ratings of Freesound users (with *"Play [sound name] by rating"*). As soon as the sounds are retrieved, their preview is reproduced in sequential order, preceded by a word stating the number of the track (e.g., *"First"*) so the user can then save the wanted track by recalling the number associated to it. If a user wants to save the track currently being reproduced s/he can just issue the keyword *"Alexa, pause"*.

During the reproduction of the tracks preview, the user is empowered to stop the reproduction and go directly to the next or to the previous file. Moreover, at any time the user can vocally adjust the volume of the device, even during the reproduction of the retrieved audio content. If before or after the reproduction of a track the user pronounces *"Get my kept tracks"*, the system produces a mix with the tracks previously saved. As soon as the mix is created, it is reproduced and the corresponding file is saved on the cloud. When the user pronounces *"Delete kept tracks"* all the tracks saved for the mixes are deleted and removed from the cloud. A useful design choice was that of storing the downloaded file on the cloud, in the space allocated by Amazon. This allows to avoid to re-download a same content in the case a user decides to search it more than once, or to use it for more than one mix.

In the case the retrieval time for a sound exceeds the 5 seconds, the user is notified that the retrieval process is taking long and s/he needs to wait (e.g., for the sound name "snow", *"Ok, give me a second when I retrieve the tracks about snow"*). The user can always interrupt an issued command by pronouncing *"Cancel"*, and if s/he pronounces *"Quit"* the program is quitted. If the search does not produce any result, the user is notified accordingly. Finally, issuing the command *"Help"* triggers a set of sentences briefly describing all the possible commands.

The development of the system could not parallel all the intended designs due to a series of technical limitations imposed by the development tools available on the Amazon Alexa SDK. Firstly, the maximum number of sounds retrieved needed to be set to just three. This was due to the limited computational resources of Amazon Echo that does not allow to download and reproduce simultaneously more tracks. Due to some technical constraints of the structure of Amazon Alexa, all retrieved tracks to be mixed (up to three) could only be reproduced simultaneously from their beginning. Another technical limitation is due to the fact that Amazon Alexa does not allow to issue user-defined commands after the audio reproduction of the retrieved content: Alexa terminates the session as soon as the track starts to be played, allowing exclusively the use of keywords related to the playback (i.e., *"pause", "stop", "resume", "previous", "next"*). This is the reason why we used *"Alexa, pause"* instead of a more reasonable command *"Alexa, keep"* to keep a track for the subsequent mix.

4 EVALUATION

The user study aimed at preliminary assessing the usability of the system and participants' experience in interacting with it. A total of 12 participants took part to the evaluation (8 males, 4 females, aged between 18 and 25, mean age = 22.4). Participants did not report any auditory, visual or motor impairment. All of them had a musical background and they reported to have an average experience with musical software for editing (e.g., digital audio workstations or other sound design and sound editing tools) equal to 4±1.65 assessed on a 7-point Likert scale. The experiments were conducted in part in a laboratory of University of Trento and in part at the home of participants. Participants took on average one hour to complete the experiment. The procedure, approved by the local ethics committee, was in accordance with the ethical standards of the 1964 Declaration of Helsinki.

[2] https://github.com/mtg/freesound-python

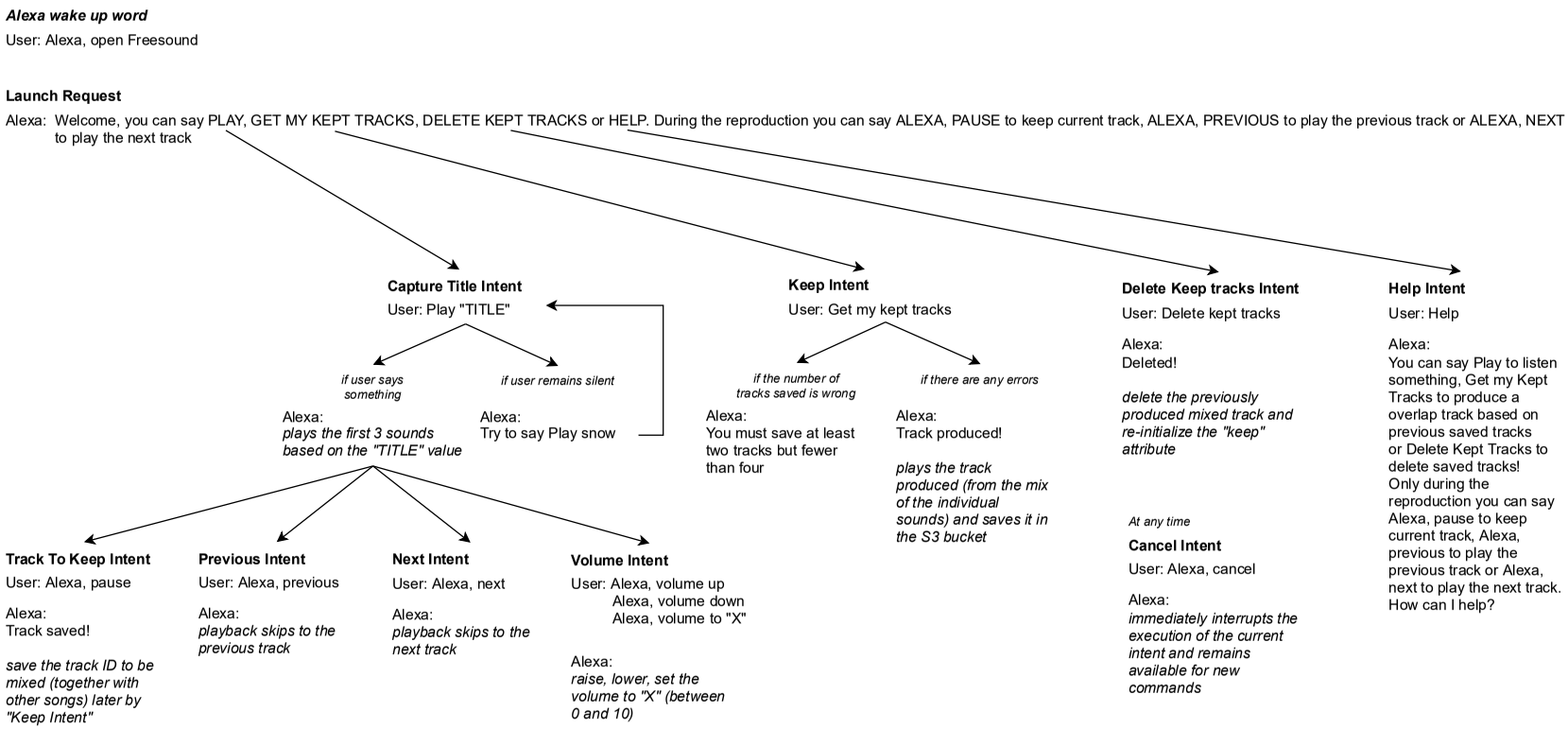

Figure 2: Vocal commands and Alexa responses for the soundscape composition task.

The evaluation procedure consisted of the following steps. Firstly, participants were debriefed about the experiment and were asked to interact with the web browser interface of Freesound to retrieve and listen to sounds of their choice. Secondly, they underwent a familiarization phase where they tried the system. The experimenter practically demonstrated to each participant the available vocal commands, which were also listed on an instruction sheet. Together with the experimenter, participants accomplished the task of composing two different soundscapes: a sea shore and and snowy environment. Specifically, as the system allows up to three simultaneous tracks, they were asked to choose three sonic events within these environments: subjects were asked the following question: "Imagine that you are right now along a sea shore: which sounds do you think you would hear?"

Thirdly, participants started the main experiment. This consisted of creating three different soundscapes. Participants were allowed to freely select the environments and the three sonic events within them. After having composed each of the three soundscapes, they were asked to fill an ad-hoc questionnaire. This was partly inspired by the System Usability Scale questionnaire [6] and the questionnaire to calculate the creativity support index [8]. The questionnaire was devised to assess the usability of the system, investigate the degree of creativity fostered by the system, and understand its hedonic qualities [45].

Specifically, the questionnaire comprised the following questions to be evaluated on a 7-point Likert scale (1 corresponds to *strongly disagree* and 7 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 results I got out of the system.*
- *[Quick learning.] I would imagine that most people would learn to use this system very quickly.*
- *[Exploration.] It was easy for me to explore many different ideas using this system.*
- *[Expressiveness.] The system allowed me to be very expressive.*
- *[Creativity.] I was able to be very creative while composing a soundscape with this system.*
- *[Immersion.] I became so absorbed in the soundscape composition activity that I forgot about the system or tool that I was using.*
- *[Results Worth Effort.] What I was able to produce was worth the effort I had to exert to produce it.*

At the end of the experiment, 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?*

4.1 Results

Figure 3 illustrates the results of the questionnaire items. As it is possible to notice from the figure, with exception of items Complexity and Immersion the evaluations were all above neutrality. Overall, the system was found usable and was deemed not difficult to use and to learn. The highest average response was found for item Enjoyment, which indicates that participants experienced positively the interactions afforded by the system. Moreover, generally participants felt that the system was able to support their creativity during the process of composing a soundscape.

Participants' answers to the open-ended questions were analyzed using an inductive thematic analysis [4]. The analysis was conducted by generating codes, which were further organized into themes that reflected patterns, as described below.

Speed and easiness of creation. Nine subjects reported that they liked very much the speed with which they were able to create the soundscapes. They also stated to have found the interface easy to use. In particular, a feature that was particularly appreciated was that of being able to quickly listen to the previews of the sounds, which was deemed as very useful. They stated that the vocal interaction to retrieve and listen to the sounds snippets appeared to them faster than the usual interaction with the browser. They also highlighted the fact that the vocal interaction could allow to save time (e.g., *"It is the fastest method to search and listen to the previews of the sounds"*).

Concept and novelty. Five participants stated to have strongly appreciated the idea behind the system, i.e., the direct connectivity of the vocal assistant with an online database and the approach to the search based exclusively on vocal interactions (e.g., *"I liked the most the fact that a function usually accomplished via a desktop computer can be done entirely with a vocal assistant"*. The concept of the system was considered innovative, especially because it enables sound designers to retrieve audio content in a more immediate way rather than using a textual search in a browser form. Moreover, six participants commented to have enjoyed interacting with the system (e.g., *"I find this system very useful an I enjoyed using it"*).

Sound availability and expressiveness. Nine participants expressed strong satisfaction for being capable of retrieving any type of sounds in an immediate way (e.g., *"I enjoyed being able to get immediately whatever type of sound comes to my mind and use it for composing a soundscape straight after"*). In particular, four of them also commented positively on the expressive power of the system, that allows to compose any kind of soundscape (e.g., *"Having a database so big I can produce whatever soundscape"*). These features were deemed by three participants to be effective in stimulating creativity (e.g., *"This huge sound availability facilitates and fosters the creation process"*). However, three participants also suggested that the system could be improved by allowing the retrieval of audio content from other sources than Freesound, including other online repositories as well as the sounds available on their own computer.

Ubiquitous use. Seven participants reported that for them the added value of the developed system lies in its ubiquitous nature, which avoids the need of multiple devices such as PCs and loudspeakers (e.g., *"I can bring Echo where I want without problems and I can use it at any moment to compose a soundscape when I feel like to do so"*). Moreover, two of them reported that the ubiquitous, standalone, and "always on" nature of the system has the potential to improve the workflow of idea generation for soundscape composition.

Inclusiveness. Six participants commented that the system is inclusive as it allows anyone to compose in an easy manner a soundscape, including those who do not have a background in audio

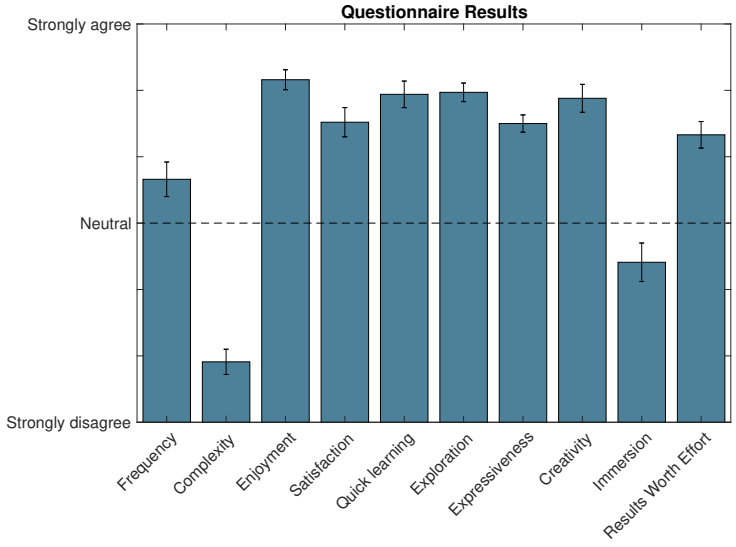

Figure 3: Mean and standard error of the questionnaire items (evaluated on a 7-point Likert scale).

editing (e.g., *"Thanks to this system no musical skills are required to compose a soundscape nor knowledge of audio editing is necessary"*). However, it is also worth noticing that three participants stated that the system is not advanced enough to accomplish more complex functions, for which a computer with an audio editing software is still needed.

System limits. Three participants felt limited in the interaction possibilities afforded by the system, requesting more features. Firstly, they reported that the main issue at interactive level was that the system needed to be restarted after the reproduction of the list of sounds, which hampered a bit the workflow. Secondly, the possibility of mixing maximum three tracks was deemed a limit to the creative potential of a sound designer (e.g., *"I would need to mix more than three tracks to express my ideas well"*). Another limit was the system inability of starting, during the mixing process, the reproduction of a track at any moment decided by the user (while the system only allowed to reproduce all mixed tracks simultaneously starting from the beginning). Furthermore, four participants also felt limited in the sonic control of the tracks, requesting features such as the adjustment of the volume of each track, the application of filters to each track or to their mix, or the trimming of the tracks. One user also suggested to add a command to send via e-mail the composed audio file to her own e-mail address once or to host it on an online repository.

Irrelevance. Four participants reported that the retrieved sounds did not fully correspond to the keyword they had said, which highlights the importance to have better tags in Freesound (e.g., *"The sounds not always correspond to what I said. More accuracy is needed"*).

5 DISCUSSION AND CONCLUSION

This paper presented a novel voice-based system capable of supporting sound designers in the creative practice of composing a soundscape. Overall, the user study revealed that participants learned to use the system given a short training period, and that they were able to achieve good results. The evaluation however highlighted also some weaknesses of the system, which however were mostly due to the SDK technical limitations encountered during the implementation.

We believe that the preliminary results presented in this paper are useful to provide directions to designers of voice-based interactive systems focusing soundscape composition and related sound design practices. However, it is worth noticing that our study has some limitations. Firstly, a small sample size was involved. We plan to continue our evaluation of the system by recruiting more participants, in particular those with various visual and hand impairments, to try out our system. Notably, whereas the presented study has implications for visually-impaired sound designers, the evaluation was conducted involving participants with no impairments. In future work we plan to assess the developed system with visually-impaired individuals. Moreover, in future work we plan to increase the number and the complexity of the interactions afforded by the system. We are also looking into studying the learning curve of the interface through a longitudinal study to determine the level of training necessary for people to achieve sufficient proficiency. Such a longitudinal study will also help reveal any potential issues with fatigue of the vocal cord after pro-longed use. Through our evaluations so far we have not encountered any major complaints of vocal fatigue from our participants.

Another limitation of our study is represented by the fact that the number and complexity of the user interactions afforded by the system were constrained by the possibilities offered by the API of Amazon. It is possible to envision several avenues to extend the development of the system and improve the quality of the interactions available to the users. Some of them where requested by the participants of the experiment, although were already devised by the authors (e.g., the possibility of mixing a number of tracks higher than three and the possibility to have mixes with sounds not beginning simultaneously). One can also make the interactions more complex, but in that case it is necessary to also

consider the users' cognitive load such as the ability to remember several different commands. However, most of these avenues are currently constrained by the technical limits of the devices running the personal assistants (e.g., little computational resources).

Potentially, the proposed approach could find application in the musical domain, where online repositories of tracks of individual musical instruments/singers could be exploited for composition purposes. Along the same lines, interactions mediated by vocal input/output could be used for retrieving musical pieces from music repositories (e.g., the Jamendo repository), by means of vocal queries based on the musical features (e.g., chords, beat per minute, key) rather than using conventional search criteria based on textual inputs on browsers (in a similar fashion of what reported in [39] for the case of smart musical instruments). These and other avenues are currently unexplored and call for more research on vocal interactions between users and online sound repositories.

REFERENCES

[1] V. Akkermans, F. Font, J. Funollet, B. de Jong, G. Roma, S. Togias, and X. Serra. 2011. Freesound 2: An improved platform for sharing audio clips. In *Proceedings of the International Society for Music Information Retrieval Conference*.

[2] S. Bazen, L. Bouvard, and J.B. Zimmermann. 2015. Musicians and the Creative Commons: A survey of artists on Jamendo. *Information Economics and Policy* 32 (2015), 65–76.

[3] J.P. Bigham, T. Lau, and J. Nichols. 2009. Trailblazer: enabling blind users to blaze trails through the web. In *Proceedings of the 14th international conference on Intelligent user interfaces*. 177–186.

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

[5] R.N. Brewer, L. Findlater, J. Kaye, W. Lasecki, C. Munteanu, and A. Weber. 2018. Accessible Voice Interfaces. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 441–446.

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

[7] O. Chapman. 2009. The Icebreaker: Soundscape works as everyday sound art. *Organised Sound* 14, 1 (2009), 83–88.

[8] 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.

[9] K. Christian, B Kules, B. Shneiderman, and A. Youssef. 2000. A comparison of voice controlled and mouse controlled web browsing. In *Proceedings of the fourth international ACM conference on Assistive technologies*. 72–79.

[10] J. L. Drever. 2002. Soundscape composition: the convergence of ethnography and acousmatic music. *Organised Sound* 7, 1 (2002), 21–27.

[11] J. d'Escriván. 2009. Sound art (?) on/in film. *Organised Sound* 14, 1 (2009), 65–73.

[12] G. Eckel. 2001. Immersive audio-augmented environments: the LISTEN project. *Proceedings Fifth International Conference on Information Visualisation* 128 (2001), 571–573.

[13] X. Favory, E. Fonseca, F. Font, and X. Serra. 2018. Facilitating the manual annotation of sounds when using large taxonomies. In *Proceedings of the 23rd Conference of Open Innovations Association FRUCT*. IEEE, 60–64.

[14] E. Fonseca, J. Pons Puig, X. Favory, F. Font Corbera, D. Bogdanov, A. Ferraro, S. Oramas, A. Porter, and X. Serra. 2017. Freesound datasets: a platform for the creation of open audio datasets. In *Proceedings of the International Society for Music Information Retrieval Conference*. International Society for Music Information Retrieval, 486–493.

[15] F. Font, T. Brookes, G. Fazekas, M. Guerber, A. La Burthe, D. Plans, M.D. Plumbley, M. Shaashua, W. Wang, and X. Serra. 2016. Audio Commons: bringing Creative Commons audio content to the creative industries. In *Audio Engineering Society Conference: 61st International Conference: Audio for Games*. Audio Engineering Society.

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

[17] J. Freeman, C. Disalvo, M. Nitsche, and S. Garrett. 2011. Soundscape composition and field recording as a platform for collaborative creativity. *Organised Sound* 16, 3 (2011), 272–281.

[18] C. Granata, M. Chetouani, A. Tapus, P. Bidaud, and V. Dupourqué. 2010. Voice and graphical-based interfaces for interaction with a robot dedicated to elderly and people with cognitive disorders. In *19th International Symposium in Robot and Human Interactive Communication*. IEEE, 785–790.

[19] S. Harada, J.A. Landay, J. Malkin, X. Li, and J.A. Bilmes. 2006. The vocal joystick: evaluation of voice-based cursor control techniques. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*. 197–204.

[20] S. Harada, J.O. Wobbrock, and J.A. Landay. 2011. Voice games: investigation into the use of non-speech voice input for making computer games more accessible. In *IFIP Conference on Human-Computer Interaction*. Springer, 11–29.

[21] C.C. Huang, Y.J. Lin, X. Zeng, M. Newman, and S. O'Modhrain. 2015. Olegoru: A Soundscape Composition Tool to Enhance Imaginative Storytelling with Tangible Objects. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. 709–714.

[22] T. Igarashi and J.F. Hughes. 2001. Voice as sound: using non-verbal voice input for interactive control. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*. 155–156.

[23] S. Kopp, M. Brandt, H. Buschmeier, K. Cyra, F. Freigang, N. Krämer, F. Kummert, C. Opfermann, K. Pitsch, L. Schillingmann, et al. 2018. Conversational assistants for elderly users–the importance of socially cooperative dialogue. In *Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications co-located with the Federated AI Meeting*, Vol. 2338.

[24] M. Koutsomichalis. 2013. On soundscapes, phonography and environmental sound art. *Journal of sonic studies* 4, 1 (2013).

[25] M. Leonard and R. Strachan. 2014. More Than Background: Ambience and Sound-Design in Contemporary Art Documentary Film. In *Music and Sound in Documentary Film*. Routledge, 180–193.

[26] A. Pradhan, K Mehta, and L. Findlater. 2018. "Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–13.

[27] R. M. Schafer. 1977. *The Tuning of the World*. Random House Inc.

[28] M. Schirosa, J. Janer, S. Kersten, and G. Roma. 2010. A system for soundscape generation, composition and streaming. In *XVII CIM-Colloquium of Musical Informatics*.

[29] S. Skach, A. Xambó, L. Turchet, A. Stolfi, R. Stewart, and M. Barthet. 2018. Embodied Interactions with E-Textiles and the Internet of Sounds for Performing Arts. In *Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 80–87. https://doi.org/10.1145/3173225.3173272

[30] A. Stolfi, M. Ceriani, L. Turchet, and M. Barthet. 2018. Playsound.space: Inclusive Free Music Improvisations Using Audio Commons. In *Proceedings of the Conference on New Interfaces for Musical Expression*. 228–233.

[31] M. Thorogood and P. Pasquier. 2013. Computationally Created Soundscapes with Audio Metaphor.. In *International Conference on Computational Creativity*. 1–7.

[32] B. Truax. 1992. Electroacoustic music and the soundscape: the inner and outer world. *Companion to contemporary musical thought* 1 (1992), 374–398.

[33] B. Truax. 1996. Soundscape, Acoustic Communication and Environmental Sound Composition. *Contemporary Music Review* 15, 1-2 (1996), 49–65.

[34] B. Truax. 2008. Soundscape composition as global music: electroacoustic music as soundscape. *Organised Sound* 13, 2 (2008), 103–109.

[35] 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

[36] 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

[37] L. Turchet, G. Fazekas, M. Lagrange, H. Shokri Ghadikolaei, and C. Fischione. 2020 (In press). The Internet of Audio Things: state-of-the-art, vision, and challenges. *IEEE Internet of Things Journal* (2020 (In press)).

[38] 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. https://doi.org/10.1109/ACCESS.2018.2872625

[39] L. Turchet, J. Pauwels, C. Fischione, and G. Fazekas. 2020. Cloud-Smart Musical Instrument Interactions: Querying a Large Music Collection with a Smart Guitar. *ACM Transactions on the Internet of Things* 1, 3 (2020), 1–29. https://doi.org/10.1145/3377881

[40] L. Turchet and S. Serafin. 2013. Investigating the amplitude of interactive footstep sounds and soundscape reproduction. *Applied Acoustics* 74, 4 (2013), 566–574.

[41] P. Turner, I. McGregor, S. Turner, and F. Carroll. 2003. Evaluating soundscapes as a means of creating a sense of place. In *Proceedings of International Conference on Auditory Display*. 148–151.

[42] A. Valle, P. Armao, M. Casu, and M. Koutsomichalis. 2014. SoDA: A Sound Design Accelerator for the automatic generation of soundscapes from an ontologically annotated sound library.. In *International Computer Music Conference*.

[43] A. Valle, V. Lombardo, and M. Schirosa. 2009. Simulating the soundscape through an analysis/resynthesis methodology. In *Auditory Display*. Springer, 330–357.

[44] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. 2009. A 3-D immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech, and Language Processing* 18, 6 (2009), 1550–1561.

[45] 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.
[46] H. Westerkamp. 2002. Linking soundscape composition and acoustic ecology. *Organised Sound* 7, 1 (2002), 51–56.
[47] K. Wrightson. 2000. An introduction to acoustic ecology. *Soundscape: The journal of acoustic ecology* 1, 1 (2000), 10–13.
[48] A. Xambó, G. Roma, A. Lerch, M. Barthet, and G. Fazekas. 2018. Live Repurposing of Sounds: MIR Explorations with Personal and Crowdsourced Databases. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*