

Demo of Voice-based Interface for Soundscape Composition: Composing Soundscapes by Vocally Querying Online Sounds Repositories

Alex Zanetti, Luca Turchet

University of Trento

Trento, Italy

luca.turchet@unitn.it

Abstract—This demo 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.

I. 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 and continued by focusing mostly on musical applications, with pioneering works of Barry Truax [1]. “Soundscape composition” refers to a sound-based art form that concerns the creation of sonic environments [2]. This art form has grown from acoustic ecology [3] and soundscape studies [4].

Composed soundscapes are used widely in various contexts, including movies, music performances, artistic installations, and virtual environments. 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 (<http://www.freesound.org>), a collaborative repository of audio samples developed at and maintained by the Music Technology Group of Universitat Pompeu Fabra [5], [6]. 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 [7].

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 [8]. Freesound enables designers to create third-party applications exploiting its audio content in live applications by granting access to the database through a REST API [9].

Various systems for soundscape composition have been developed, including real-time, interactive, non interactive, automatic, and even tangible. 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 [10]. 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 demo 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 [11]. 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.

Fig. 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,

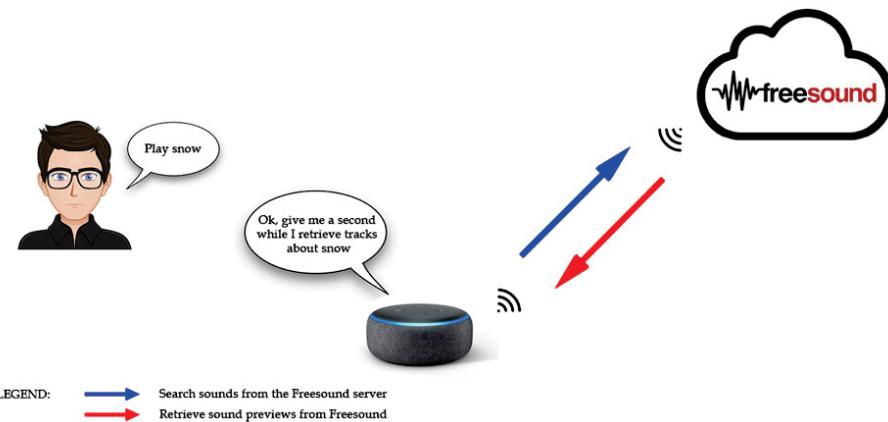


Fig. 1. Schematic representation of the implemented Internet of Audio Things ecosystem

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 (<https://github.com/mtg/freesound-python>).

REFERENCES

- [1] B. Truax, "Electroacoustic music and the soundscape: the inner and outer world," *Companion to contemporary musical thought*, vol. 1, pp. 374–398, 1992.
- [2] ——, "Soundscape composition as global music: electroacoustic music as soundscape," *Organised Sound*, vol. 13, no. 2, pp. 103–109, 2008.
- [3] K. Wrightson, "An introduction to acoustic ecology," *Soundscape: The journal of acoustic ecology*, vol. 1, no. 1, pp. 10–13, 2000.
- [4] H. Westerkamp, "Linking soundscape composition and acoustic ecology," *Organised Sound*, vol. 7, no. 1, pp. 51–56, 2002.
- [5] F. Font, G. Roma, and X. Serra, "Freesound technical demo," in *Proceedings of the ACM international conference on Multimedia*. ACM, 2013, pp. 411–412.
- [6] E. Fonseca, J. Pons Puig, X. Favery, F. Font Corbera, D. Bogdanov, A. Ferraro, S. Oramas, A. Porter, and X. Serra, "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, 2017, pp. 486–493.
- [7] F. Font, T. Brookes, G. Fazekas, M. Guerber, A. La Burthe, D. Plans, M. Plumbley, M. Shaashua, W. Wang, and X. Serra, "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, 2016.
- [8] X. Favery, E. Fonseca, F. Font, and X. Serra, "Facilitating the manual annotation of sounds when using large taxonomies," in *Proceedings of the 23rd Conference of Open Innovations Association FRUCT*. IEEE, 2018, pp. 60–64.
- [9] V. Akkermans, F. Font, J. Funollet, B. de Jong, G. Roma, S. Togias, and X. Serra, "Freesound 2: An improved platform for sharing audio clips," in *Proceedings of the International Society for Music Information Retrieval Conference*, 2011.
- [10] L. Turchet and M. Barthet, "Jamming with a smart mandolin and Freesound-based accompaniment," in *IEEE Conference of Open Innovations Association (FRUCT)*. IEEE, 2018, pp. 375–381.
- [11] L. Turchet, G. Fazekas, M. Lagrange, H. Shokri Ghadikolaei, and C. Fischione, "The internet of audio things: state-of-the-art, vision, and challenges," *IEEE Internet of Things Journal*, 2020 (accepted).