AcouRe: Streamlining the generation of immersive soundscapes from monophonic sources

Catarina Cruz
University of Porto
ee12172@fe.up.pt

Eduardo Magalhães
INESC TEC & University of Porto
eduardom@fe.up.pt

Gilberto Bernardes
INESC TEC & University of Porto
gba@fe.up.pt

ABSTRACT

The growing number of curated and crowdsourced online sound databases has been transforming several branches of music from production to consumption. In this context, the operational basis of sound design practice has been gradually shifting to incorporate repurposing strategies. One of the most pressing matters in this new context is the retrieval and transformation of audio content to meet the creative endeavours of the sound designer. The latter is the focus of our work, which aims to streamline the sound design process in the context of post-production, particularly when adopting large amounts of poorly annotated data. Different methodologies at the crossroads of machine listening and learning (namely, pattern recognition) are adopted to propose the automatic identification of sound events components from soundscape recordings to be repurposed in the temporal and spatial domains. A prototype interface, named AcouRe, implements the proposed methodologies. A heuristic evaluation to assess the usability of AcouRe was conducted by two expert sound designers.

INTRODUCTION

The need to extend and re-arrange (in time and space) pre-recorded soundscapes so that they match the content of an edited (visual) scene in post-production is a recurring and time-consuming task in sound design. A typical approach resorts to the manual identification of ‘optimal’ loop points in the temporal dimension of an audio source. However, if the loop is perceived as such, i.e., repetitive, the quality and the credibility of the experience can be disrupted. Recent approaches point towards the adoption of stochastic recombination methods, as an expeditious solution to the problem. Representative examples of stochastic methods for soundscape processing are earGram (Bernardes, 2020), Seed (Bernardes, Aly & Davies, 2016), Tapestrea (Misra, Wang & Cook, 2007), Manipulation and Resynthesis with Natural Grains (Hoskinson & Pai, 2001). For a comprehensive literature review of the topic of automatic soundscapes construction please refer to Magalhães (2013).

Furthermore, the current interest in non-linear multimedia narratives in games and extended realities (e.g., virtual and augmented reality), to cite a few, requires solutions for the production of dynamic sonic content that on one hand spatially matches the immersive experience of the visual content (Begault, 2000) and also to the users’ real-time experience. In this context, spatial audio techniques such as Ambisonics (Frank, Zotter & Sontacchi, 2015) and soundfield synthesis (Spors, Teutsch, Kuntz & Rabenstein, 2004) have been adopted by sound designers to enhance the immersive character of the scenes through 3-dimensional auditory experiences. Tools such as Sound Particles, where a soundscape is built upon the behaviour of its smallest elements (Sound Particles, 2020)

Our work stems from the identified problems and aims to propose an interface, AcouRe, for soundscape repurposing by expanding upon the aforementioned open questions in the temporal and spatial domain, beyond the existing linear audio solutions centred around the typical Digital Audio Workstations (DAW), such as Reaper or Ardour (Kronlachner, 2014).

ACOURE

AcouRe processes user-defined soundscapes to automatically segment an audio source into its component sonic events, i.e., smaller identifiable parts. It further labels each event either as *background* or *foreground* based on Schafer's (1994) taxonomy of keynote sound and signal sounds. Following Bernardes (2020), foreground events are modelled and repurposed in time using the pattern matching algorithm Factor Oracle, which enables the automatic recombination of audio source events with enhanced levels of variation, while retaining the structural and sonic (or timbral) attributes of the original source. Foreground events are further clustered into classes according to their level of perceptual similarity, which is shown to the user via a 2-dimensional interface. Each foreground event can then be temporally arranged in a timeline and spatialized by re-encoding and re-rendering its spatial position. Structurally-aware repurposing of existing soundscapes towards flexible source duration, content manipulation, and fluid spatial manifestations of sound objects are revisited in our design. As a contrasting example to Sound Particles, our interface identifies the smallest elements and recombines those into new semantic clusters and adaptive layouts from the spatial and temporal re-configuration. Figure 1 shows the architecture behind the interface developed in this work, how both backend and frontend work together toward our goals.

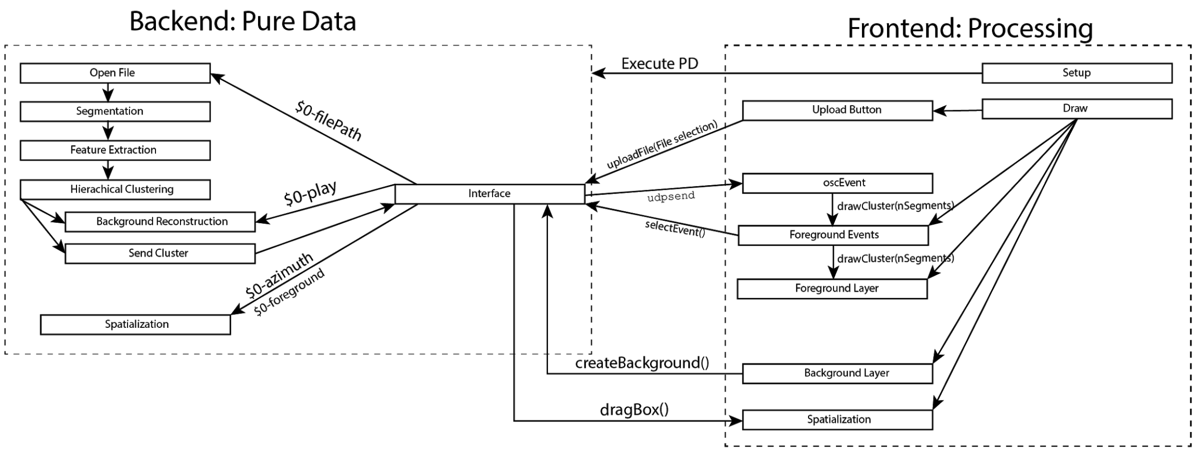

Figure 1: Architecture of AcouRe's interface

The AcouRe interface contains four different sections, Figure 2 shows the design of the interface, allowing the user to manipulate background and foreground elements, in the temporal and spatial domains separately. After the upload of the recorded soundscape, the foreground events will be displayed in the section dedicated to them, providing acoustic information. There are two timelines the user can manipulate: i) the first one, the background layer, gives the freedom to choose the length of the soundscape's background ii) the second, the foreground layer, allows the user to manipulate the temporal position of foreground events. The other two sections are dedicated to the foreground events, one is for the visualization of the cluster and how the events relate to each other acoustically and the second one is for spatial manipulation.

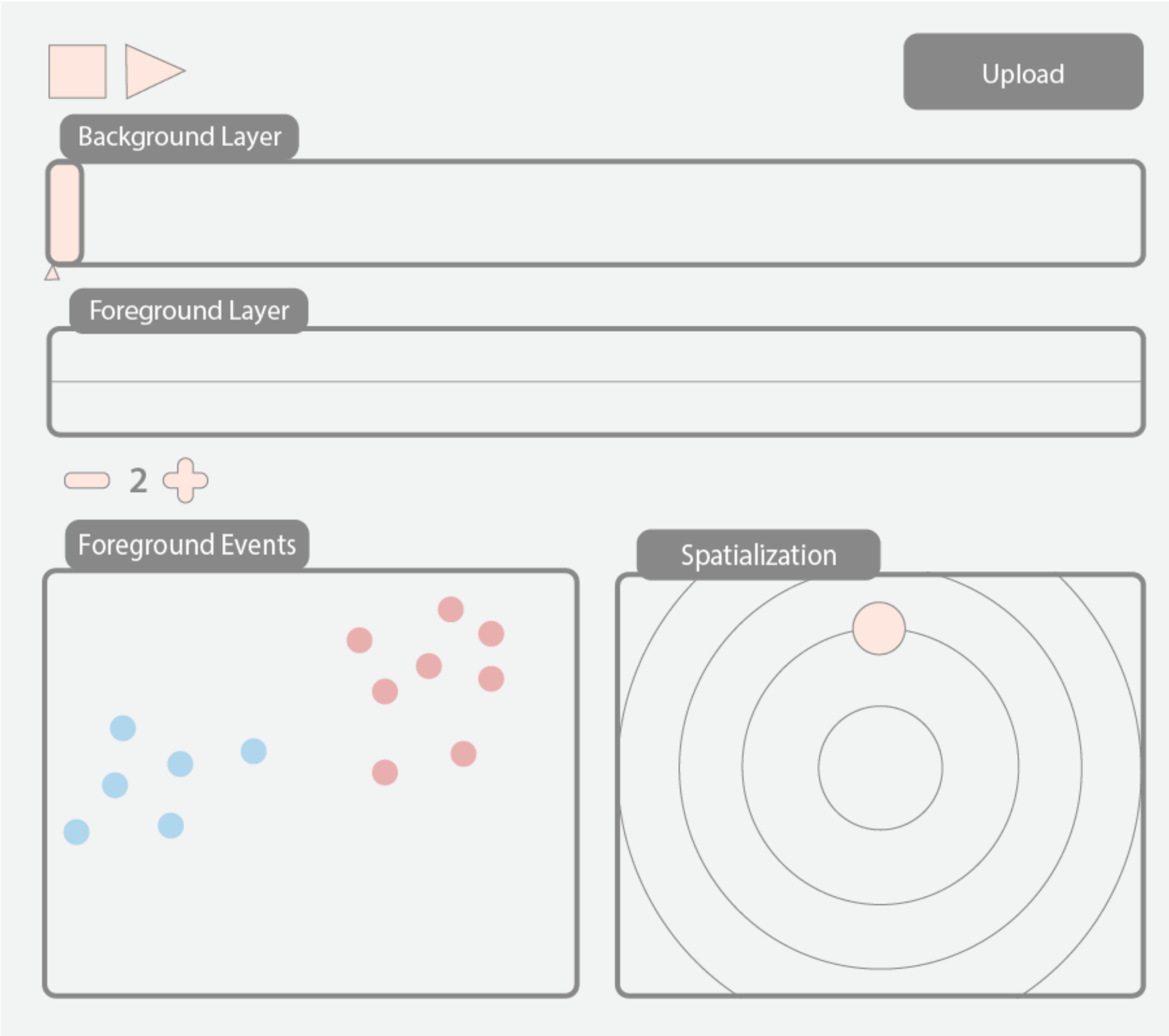

Figure 2: Layout of AcouRe's interface

EVALUATION AND RESULTS

To evaluate AcouRe, we conducted a heuristic evaluation (Nielsen & Molich, 1990) that aims to assess the interface in terms of usability by sound designers. To this end, we selected two participants with knowledge on the use of DAWs as well as some experience with sound design and familiarity with the terms foreground and background. The participants were given a script with a list of tasks to complete with the AcouRe interface. They were asked to take note of problematic behaviour and usability issues on ten heuristics and rank them by severity on a 5-point Likert scale.

Overall, the experiment participants reported that the ten heuristics under evaluation were met in AcouRe. However, some issues with different degrees of severity were identified along with the information on how to tackle these problems in the future. Of note, 1) the need to improve the foreground events visualization as it was reported to be very cluttered, thus making both the temporal and spatial manipulation of the events very cumbersome. 2) Some interface buttons were identified as lacking clear functionality. 3) The temporal alignment of the background and foreground layers was not evident for the evaluators. 4) The spatialization methodology implemented shows some erratic behaviour.

CONCLUSIONS

In sum, the interface of AcouRe is still at a prototype state in relation to its potential. However, the results of this work are very promising and have shown many leads for further development. We believe the interface makes the temporal and spatial manipulation not only easier but more dynamic. The flexible processing and repurposing of blind-source soundscapes are promoted. Though the process is quite dynamic the results are linear, as of the moment of writing.

REFERENCES

Begault, D. R. (2000). 3-D Sound for Virtual Reality and Multimedia. California, United States of America. Academic Press

Bernardes, G., Aly, L., & Davies, M. E. P. (2016). Seed: Resynthesizing environmental sounds from examples. In Proceedings of the Sound and Music Computing Conference (pp. 55-62).

Bernardesm G. (2020). Interfacing sounds: Hierarchical audio content morphologies for creative re-purposing in earGram 2.0. Proceedings of the International Conference on New Interfaces for Musical Expression.

Frank, M., Zotter, F., & Sontacchi, A. (2015). Producing 3D audio in ambisonics. In 57th AES Convention (pp. 1–8).

Hoskinson, R., & Pai, D. (2001). Manipulation and resynthesis with natural grains. In Proceedings of the International Computer Music Conference (pp. 338–341).

Kronlachner, M. (2014). Plug-in suite for mastering the production and playback in surround sound and ambisonics. In 136th AES Convention, pp. 3–7.

Magalhães, E. (2013). Exploração de ambientes sonoros imersivos no contexto multimédia: aplicações na mistura e desenho de som. Master thesis, University of Porto.

Misra, A. Wang, G. & Cook, P. (2007). Musical tapestry: Re-composing natural sounds. Journal of New Music Research, *36*(4), 241-250.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 249-256).

Schafer, R. M. (1994). The Soundscape: The tuning of the World. Destiny Books

Sound Particles. (2020). Sound Particles [Software]. Retrieved from https://soundparticles.com/

Spors, S., Teutsch, H., Kuntz, A. & Rabenstein, R. (2004). Sound Field Synthesis. In: Audio Signal Processing for Next-Generat