Crowd-Sourcing the Sounds of Places with a Web-Based Evolutionary Algorithm

Alexander E.I. Brownlee
Computing Science and Mathematics
University of Stirling, UK
sbr@cs.stir.ac.uk

Suk-Jun Kim
School of Language, Literature, Music
and Visual Culture
University of Aberdeen, UK
UK
s.kim@abdn.ac.uk

Szu-Han Wang
Centre for Clinical Brain Sciences
(Section of Synaptic Biology)
University of Edinburgh, UK
s.wang@ed.ac.uk

Stella Chan
School of Health in Social Science
(Section of Clinical Psychology)
University of Edinburgh, UK
stella.chan@ed.ac.uk

Jamie A. Lawson
School of Language, Literature, Music
and Visual Culture
University of Aberdeen, UK
r01jal16@abdn.ac.uk

ABSTRACT

The sounds we associate with particular places are tightly interwoven with our memories and sense of belonging. We describe a platform designed to assist in gathering the sounds a group of people associate with a place. A web-based evolutionary algorithm, with human-in-the-loop fitness evaluations, ranks and recombines sounds to find collections that the group rates as familiar. An experiment covering four geographical locations shows that the process does indeed find sounds deemed familiar by participants.

KEYWORDS

sound, soundscapes, human-in-the-loop, mood

1 INTRODUCTION

The collective memories of a place and its people are tightly interwoven with identity and a sense of belonging. Sound is a fundamental part of memory, and teasing out the aural identity of a place is a powerful way to understand the collective memory of that place [2]. Most soundscape compositions have focussed on the effect of sounds recorded from specific places [7, 8]. This relies on a belief that in people's awareness of places, sounds somehow plays a role, a personal and subjective assumption not yet fully examined.

We demonstrate the Distributed Evolutionary Algorithm for collecting the Sounds of Places (DEASP), a tool for building such soundscapes for a given place. An evolutionary algorithm (EA) runs periodically to assemble short sound samples into groups. A web-based interface allows participants to evaluate the groups; these evaluations determining fitness for the EA. The value of DEASP over a simple voting approach for samples is that the EA's iterative nature allows participants to hear and rank combinations of sounds chosen by others. This means the final choices are arrived at collaboratively. Human-in-the-loop and interactive EAs are commonly used in situations where no clear fitness function exists [1]. Previous approaches are mostly single-user: DEASP captures the preferences of a large group, eliciting a "collective" sound associated with a given place. Full details of the system and experiments can be found in a technical report [3].

2 METHODOLOGY

DEASP was implemented as a web application. The main loop is a conventional EA, with each solution being a list of 10 samples.

The population is initialised with solutions generated at random. Participants evaluate pairs of solutions, choosing the one they find most familiar with respect to the place. When enough choices are collected, they are converted to fitnesses, and a single iteration of the EA runs to generate a new population. The process repeats.

For the evolution, 2 elite solutions are maintained. The remaining 8 solutions are generated via crossover and mutation. To achieve rapid progress, 3-tournament selection is used. Each offspring is subjected to 2-point crossover and mutation whereby one sample in the solution is replaced with another chosen at random from the database (restricted to prevent repeats in a solution).

2.1 Fitness evaluation

Participants compare pairs of solutions. A count is kept of times a solution was voted as more familiar. Once no evaluations had happened for 7 days, the votes are used to compute fitness to drive selection for the EA. Our fitness measure accommodates 3 factors:

(1) solutions chosen more often should have higher fitness
(2) not all solutions were heard the same number of times
(3) of solutions winning the same fraction of votes, we prefer those evaluated most often, giving more confidence they reflect the crowd's preference.

Thus the fitness f of a solution x is given by:

$$f(x) = \begin{cases} v(x)/t(x) + d * t(x)/10000, & \text{if } t(x) > 0 \\ 0.5, & \text{otherwise} \end{cases} \quad (1)$$

where $v(x)$ is the number of times x was voted for, $t(x)$ is the total number of times x was heard by participant, and $d = 1$ if $v(x)/t(x) \geq 0.5$; otherwise $d = -1$. $v(x)/t(x)$ reflects the rate that x was deemed more familiar than another solution. $t(x)/10000$ acts as a tiebreaker: for two solutions with equal $c(x)/t(x)$ ratio, the one evaluated more often will be given either a higher fitness for familiar solutions or a lower fitness for unfamiliar solutions.
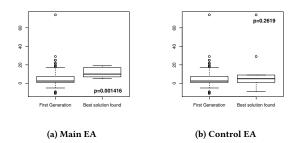
(a) Main EA

(b) Control EA

Figure 1: Independent familiarity test results for Aberdeen city centre. Wilcoxon rank-sum test p-values are inset.

3 EXPERIMENT AND RESULTS

A run of DEASP with 82 volunteers covered four places: University Campus and City Centre in both Stirling and Aberdeen. 16325 sound samples, 3-5s long, were made by dividing longer recordings. These are an artistic output in themselves, complementing existing soundscapes, but were also the optimisation building blocks, 10 being assembled to form a solution. 8 EAs ran, each with its own population: one EA for those people familiar with each place ("main" EA), and a "control" EA for people unfamiliar with each place. Participants evaluated sounds for both EAs each session.

Each volunteer participated in 1 to 5 sessions, for a total of 194 responses. Each response included mood before and after, a preference of groups of sounds for one familiar and one unfamiliar place, and the independent ratings of the individual samples they had heard. A major issue was rapid drop off in engagement, meaning that the EAs only ran for 1–5 generations. Despite this, some convergence did occur: samples in the 4-5th generation populations were surprisingly consistent. The fittest sounds found by each run are here: https://www.whatisthegrid.co.uk/FinalSounds.html

3.1 Improved familiarity

As participants heard the sounds, they marked parts of the sound as (un)familiar. Ideally *familiar* marks will be more frequent in the final population than the starting population. We took the count of *familiar* marks and subtracted the count of *unfamiliar* marks for each sample. Boxplots of this measure are in Figure 1, for Aberdeen City Centre, where the EA reached 5 generations. For the main EA, there is an uplift in familiarity. A two-tailed Wilcoxon rank-sum test found $p = 0.001416$, showing the difference to be statistically significant. For the control EA there is no significant uplift ($p = 0.2619$): reasonable because the participants had no familiarity with the place. The same trend occurs to varying degrees for the other locations where the EA exceeded 2 generations. Full results can be find in [3]. We conclude that DEASP is indeed able to find a consensus of familiar sounds for a group of people sharing a connection to a place, when participation is high enough.

3.2 Decrease in moods

Participants also indicated their mood on a 1-5 rating of 10 descriptors (the I-PANAS-SF scale [9]) before and after evaluating sounds. Several moods reduced in strength: scores for *determined*, *attentive*

and *active* all decreased by a statistically significant margin during the evaluations. We believe that this is a potentially important factor to consider for any human-in-the-loop evaluations if further evaluations are carried out in quick succession (an additional confounding factor to the known problem of user fatigue [5]).

4 FUTURE WORK

Each place presents a specific acoustic signature (reverberation, diffusion, etc.), and while this work implicitly considers acoustics by using recordings taken from each place, it would be worth teasing this out separately. We also plan to investigate further the connections between participant demographics, mood and each place.

A similar platform could be applied to collaborative design or decision making problems where a group chooses components based on qualitative evaluations of them in combination (e.g., choosing content of a standard toolkit). It would also be interesting in investigate suitable surrogate fitness models [4, 6] to extend the search without increasing the number of human sound evaluations.

REFERENCES

[1] H.J. Aljawawdeh, C.L Simons, and M. Odeh. 2015. Metaheuristic design pattern: Preference. In *Proc. GECCO Companion*. ACM, 1257–1260.
[2] Karin Bijsterveld. 2013. *Soundscapes of the Urban Past : Staged Sound as Mediated Cultural Heritage.* Transcript Verlag.
[3] A.E.I. Brownlee, S-J. Kim, S-H. Wang, S. Chan, and J.A. Lawson. 2019. *Crowd-Sourcing the Sounds of Places with a Web-Based Evolutionary Algorithm (Detailed Implementation & Results).* Technical Report. Stirling. hdl.handle.net/1893/29310.
[4] A. E. I. Brownlee, J.R. Woodward, and J. Swan. 2015. Metaheuristic Design Pattern: Surrogate Fitness Functions. In *Proc. GECCO Comp.* Madrid, Spain, 1261–1264.
[5] Praminda Caleb-Solly and Jim E Smith. 2007. Adaptive surface inspection via interactive evolution. *Image and Vision Computing* 25, 7 (2007), 1058–1072.
[6] Yaochu Jin. 2011. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm Evol. Comput.* 1, 2 (2011), 61–70.
[7] Suk-Jun Kim. 2010. Imaginal Listening: a quaternary framework for listening to electroacoustic music and phenomena of sound-images. *Organised Sound* 15, 1 (2010), 43–53.
[8] David Kolber. 2002. Hildegard Westerkamp's Kits Beach Soundwalk: shifting perspectives in real world music. *Organised Sound* 7, 1 (2002), 41–43.
[9] Edmund R Thompson. 2007. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of cross-cultural psychology* 38, 2 (2007), 227–242.