Proceedings of the 2nd AES Workshop on Intelligent Music Production, London, UK, 13 September 2016

AN EVOLUTIONARY COMPUTATION APPROACH TO INTELLIGENT MUSIC PRODUCTION INFORMED BY EXPERIMENTALLY GATHERED DOMAIN KNOWLEDGE

Alex Wilson and Bruno Fazenda

Acoustics Research Centre
University of Salford
{A.Wilson1@edu.,B.M.Fazenda@}salford.ac.uk

ABSTRACT

A novel methodology for intelligent music production has been developed using evolutionary computation. Mixes are generated by exploration of a "mix-space", which consists of a series of inter-channel volume ratios, allowing efficient generation of random mixes. An interactive genetic algorithm was used, allowing the user to rate mixes and guide the system towards their ideal mix. Currently, fitness evaluation is subjective but can be aided by specific domain knowledge obtained from a large-scale study of real mixes.

1. BACKGROUND

Intelligent music production (IMP) has been an active research topic for over a decade. One aim is the development of systems which perform common tasks: level-balancing, equalisation, panning, dynamic range compression and application of artificial reverberation. Many previous IMP systems developed were modelled as expert systems wherein a music production task is solved by optimisation, and domain knowledge, obtained by examining industry "best-practice" methods, is used to determine the optimisation target [1]. Drawbacks to this method include the fallibility of this type of domain knowledge and the fundamental assumption that there is a global optimum, i.e. one mix which all users would agree is best. Subjective evaluation suggested that existing systems struggled to compete with human-made mixes [2], perhaps due to a lack of what we would perceive as creativity. Additionally it has been suggested that mix engineers prefer their own mix to those of their peers [2]. Consequently, IMP tools would benefit from increased interactivity and subjectivity, to determine user-specific "personal" global optima in the solution space, instead of a single "universal" global optimum.

2. CONCEPT

We propose to use interactive evolutionary computation (IEC) to solve this problem, being well-suited to aesthetic design problems which are non-linear and non-deterministic [3]. The flowchart in Fig. 1 demonstrates the method, with an interactive genetic algorithm (IGA). The solution space we explored is a "mix-space" which theoretically represents all the mixes that it is possible to create with a finite set of tools [4]. For level-balancing, the gains $\mathbf{g}$ of all n tracks are selected from a unit hypersphere in $\mathbb{R}^n$. This hypersurface has $n-1$ dimensions, representing a series of inter-channel volume ratios, Φ (see Fig. 2). This method has the advantage that all random mixes generated are unique and have equal loudness (after normalising the loudness of tracks beforehand). The fitness function for optimisation is subjective, allowing mixes to be generated based on any perceptual description, such as "warmth", "punch" or "clarity" or simply "preference".

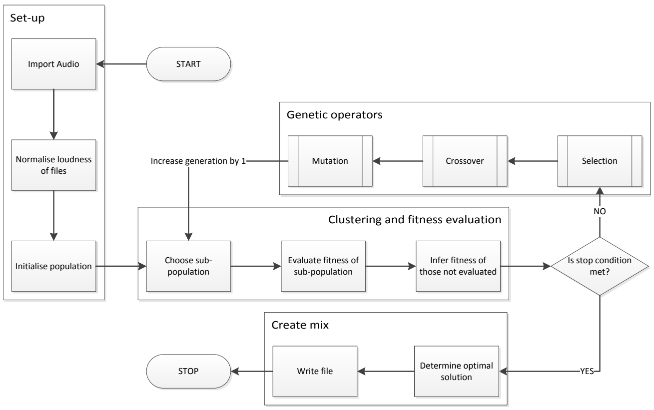

Figure 1: Flowchart of intelligent mixer using IGA

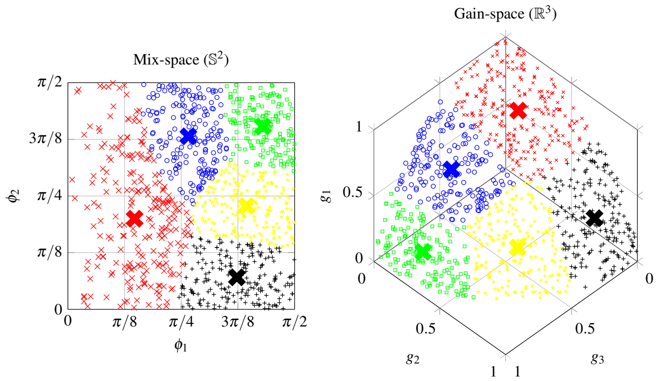

Figure 2: k-means with cosine distance metric (spherical k-means), clustered in gain-space, for a simple 3-track mixing task. The population size is 1000 (deliberately large, for visualisation purposes) and the number of clusters is 5.

3. METHOD

EC typically requires a large population of candidate solutions. To increase the population size beyond that which a user could realistically evaluate, before becoming fatigued, the fitness of a rated sub-population is extrapolated to nearby solutions [5]. Figure 2 shows the population clustered into c clusters. The mixes closest to each cluster centroid are chosen for audition and user-evaluation.

To aid this extrapolation we introduce findings from a recent large-scale study of music mixes which revealed tolerance ranges for low-level audio features [6]. This can be used to augment the fitness of the population alongside the subjective ratings provided to a subset of the population, effectively adding a penalty to mixes which are unlikely to be created by a real engineer, while still giving the user the authority to override these heuristics.

While clustering is performed in the gain-space, genetic operations take place in the mix-space. Currently, the system uses roulette selection and uniform crossover with mutation. These operations could also be performed in the gain-space if solved on the sphere.

Typically, in EC, the optimal solution is considered to be the solution with the highest fitness. However, many problems that can be addressed by IEC are perceptual and as such do not require *exact* solutions but rather seek to identify an area of the solution space in which many fit solutions exist which are perceptually similar [3]. In a music mixing problem there is a limit to the precision required when determining gain values, as small adjustments in the gain of individual tracks will not be perceived.

Determining the region of optimal solutions employed kernel density estimation (KDE). Figure 3 shows the univariate KDE result, with the values of ϕ having evolved towards specific modal values. These values are converted back to gain-space in order to construct the final mix.

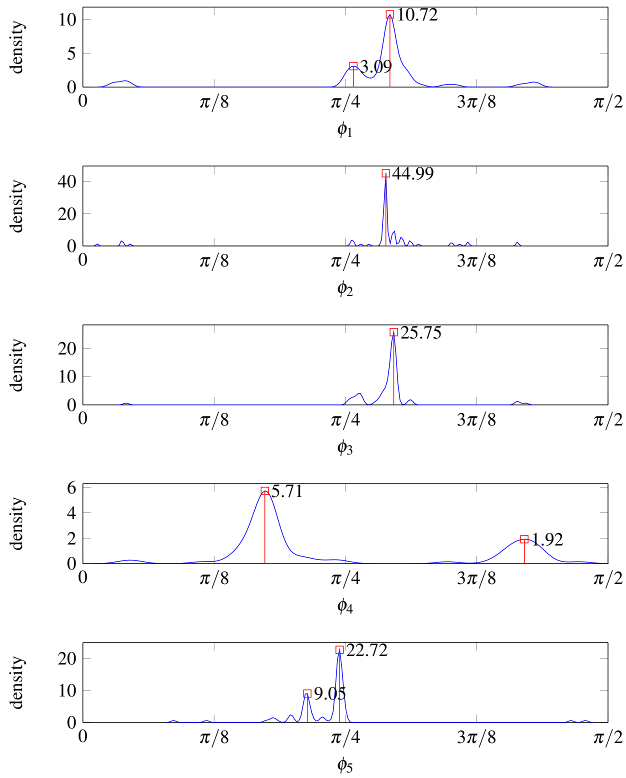

Figure 3: Kernel density estimation, showing modes in mix-space for a 6-track mixing session. The position of each mode is highlighted along with the density value. These values of $\phi_{1...n-1}$ are transformed to $g_{1...n}$ to create the final mix. Note, that in this example, multiple optimal mixes are possible, due to the multi-modal nature of ϕ_4.

4. CONCLUSIONS

Early results indicate that the system can produce a variety of mixes, suited to varying personal taste. As this system makes minimal assumptions as to what makes a good mix, or possibly no assumptions, it learns from the expertise of the user, rather than the traditional approach, which assumes the novice user learns from the expert system. We believe this approach can be used to further expand the study of IMP, to deliver personalised object-based audio to consumers and to increase the understanding how music is mixed.

5. REFERENCES

[1] P. Pestana and J. D. Reiss, "Intelligent audio production strategies informed by best practices," in *AES 53rd International Conference: Semantic Audio*, (London, UK), 2014.

[2] B. De Man, M. Boerum, B. Leonard, R. King, G. Massenburg, and J. D. Reiss, "Perceptual evaluation of music mixing practices," in *Audio Engineering Society Convention 138*, (Warsaw, Poland), 2015.

[3] H. Takagi, "Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation," *Proceedings of the IEEE*, vol. 89, no. 9, pp. 1275–1296, 2001.

[4] A. Wilson and B. M. Fazenda, "Navigating the mix-space: theoretical and practical level-balancing technique in multitrack music mixtures," in *Proceedings of the Sound and Music Computing Conference*, (Maynooth, Ireland), 2015.

[5] J.-Y. Lee and S.-B. Cho, "Sparse fitness evaluation for reducing user burden in interactive genetic algorithm," in *Fuzzy Systems Conference Proceedings*, (Seoul, Korea), 1999.

[6] A. Wilson and B. M. Fazenda, "Variation in multitrack mixes: Analysis of low-level audio signal features," *J. Audio Eng. Soc*, vol. 64, no. 7/8, pp. 466–473, 2016.