A Compositional Tool for Computer-Aided Musical Orchestration

Marcelo Caetano[1]

[1]INESC TEC, Porto, Portugal

E-mail: [1]mcaetano@inesctec.pt

Abstract: **The aim of computer-aided musical orchestration (CAMO) is to find a combination of musical instrument sounds that approximates a target sound or a desired timbral quality. The difficulty arises from the complexity of timbre perception and the combinatorial explosion of all possible instrument mixtures. The estimation of perceptual similarities between sounds requires a model capable of capturing the multidimensional perception of timbre, among other perceptual qualities of sounds. The state of the art uses Genetic algorithms (GAs) to explore the vast space of possible instrument combinations with a fitness function that encodes timbral similarity between the candidate instrument combinations and the target sound. However, GAs tend to lose diversity during the search, resulting in only one solution that commonly corresponds to a local optimum because GAs cannot guarantee to return the global optimum (i.e., the best solution). Diversity results in the ability to provide the composer with multiple choices when orchestrating a sound instead of searching for one solution constrained by choices defined a priori. Therefore, diversity can expand the creative possibilities of CAMO beyond what the composer initially imagined. This work uses an artificial immune system (AIS) called opt-aiNet to search for candidate orchestrations. Inspired by immunological principles, opt-aiNet returns multiple good quality solutions in parallel while preserving diversity. The intrinsic property of maintenance of diversity allows opt-aiNet to return all the optima (global and local) of the fitness function being optimized upon convergence, which translates as orchestrations that are all similar to the target yet different from one another.**

Keywords: Musical orchestration, musical instrument sounds, musical timbre, artificial immune systems

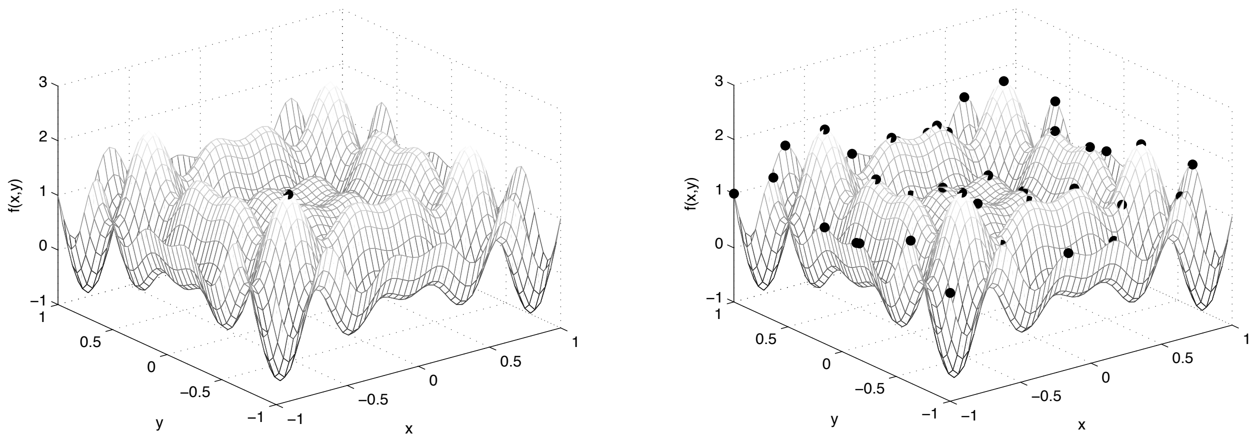
1 INTRODUCTION

Orchestration refers to composing music for an orchestra [Kendall 1993]. Initially, orchestration was simply the assignment of instruments to pre-composed parts of the score, which was dictated largely by availability of resources such as the number and type of instruments available [Handelman 2012, Kendall 1993]. Later on, composers started regarding orchestration as an integral part of the compositional process whereby the musical ideas themselves are expressed [Rose 2009]. Compositional experimentation in orchestration arises from the increasing tendency to specify instrument combinations to achieve desired effects, resulting in the contemporary use of timbral combinations [McAdams 1995, Rose 2009, Abreu 2016]. The development of computational tools that aid the composer in exploring the virtually infinite possibilities resulting from the combinations of musical instruments gave rise to computer-aided musical orchestration (CAMO) [Carpentier 2006,2007,2010a,b, Hummel 2005, Psenicka 2003, Rose 2009]. Most of these tools rely on searching for combinations of musical instrument sounds from pre-recorded datasets to approximate a given target sound. Early works [Hummel 2005, Psenicka 2003, Rose 2009] resorted to spectral analysis followed by subtractive spectral matching.

2 STATE OF THE ART

Psenicka (2003) describes SPORCH (SPectral ORCHestration) as "a program designed to analyze a recorded sound and output a list of instruments, pitches, and dynamic levels that, when played together, create a sonority whose timbre and quality approximate that of the analyzed sound." The method keeps a database of spectral peaks estimated from either the steady state or the attack (for nonpercussive and percussive sounds, respectively) of musical instrument sounds organized according to pitch, dynamic level, and playing technique such as staccato and vibrato. The algorithm iteratively subtracts the spectral peaks of the best match from the target spectrum aiming to minimize the residual spectral energy in the least squares sense. The iterative procedure requires little computational power, but the greedy algorithm restricts the exploration of the solution space, often resulting in suboptimal solutions because it only fits the best match per iteration. Hummel (2005) approximates the spectral envelope of phonemes as a combination of the spectral envelopes of musical instrument sounds. The method also uses a greedy iterative spectral subtraction procedure. The spectral peaks are not considered when computing the similarity between target and candidate sounds, disregarding pitch among other perceptual qualities. Rose and Hetrik (2009) use singular value decomposition (SVD) to perform spectral decomposition and spectral matching using a database of averaged DFTs of musical instrument sounds containing different pitches, dynamic levels, and playing techniques. SVD decomposes the target spectrum as a weighted sum of the instruments present in the database, where the weights reflect the match. Besides the drawbacks from the previous approaches, SVD can be computationally intensive even for relatively small databases. Additionally, SVD sometimes returns combinations that are unplayable such as multiple simultaneous notes on the same violin, requiring an additional procedure to specify constraints on the database that reflect the physical constraints of musical instruments and of the orchestra.

Corresponding author: Marcelo Caetano, INESC TEC, Rua Dr. Roberto Frias, 378, 4200-465, Porto, Portugal, mcaetano@inesctec.pt



a) Standard Genetic Algorithm (GA) **b) Artificial Immune System (AIS)**

Fig.1: Comparison between mono-modal (GA) and multi-modal (AIS) optimization. The AIS used is opt-aiNet.

Spectral matching [Hummel 2005, Psenicka 2003, Rose 2009] provides inherently limited orchestrations because the goal of CAMO is to find an instrument combination that best approximates the target timbre rather than the target spectrum [Tardieu 2007]. The concept of timbre lies at the core of musical orchestration. Yet, timbre perception is still only partially understood [McAdams 1995, Caclin 2005, McAdams 2009]. The term timbre encompasses auditory attributes, perceptual and musical issues, covering perceptual parameters not accounted for by pitch, loudness, spatial position, duration, among others [Krumhansl 1989, McAdams 1995]. Nowadays, timbre is regarded as both a multi-dimensional set of sensory attributes that quantitatively characterize the ways in which sounds are perceived to differ and the primary vehicle for sound source recognition and identification [McAdams 1995]. McAdams and Bruno (1995) wrote that "instrumental combinations can give rise to new timbres if the sounds are perceived as blended, and timbre can play a role in creating and releasing musical tension."

To overcome the drawbacks of spectral matching, Carpentier and collaborators [Carpentier 2006,2007,2010a,b, Tardieu 2007] search for a combination of musical instrument sounds whose timbral features best match those of the target sound. This approach requires a model of timbre perception to describe the timbre of isolated sounds, a method to estimate the timbral result of an instrument combination, and a measure of timbre similarity to compare the combinations and the target. Multidimensional scaling (MDS) of perceptual dissimilarity ratings [McAdams 1995, Caclin 2005] provides a set of auditory correlates of timbre perception that are widely used to model timbre perception of isolated musical instrument sounds. MDS spaces are obtained by equating distance measures to timbral (dis)similarity ratings.

In metric MDS spaces, the distance measure directly allows timbral comparison. Models of timbral combination [Carpentier 2010b, Kendall 1993] estimate these features for combinations of musical instrument sounds. Carpentier and collaborators [Carpentier 2006,2007,2010a,b, Tardieu 2007] consider the search for combinations of musical instrument sounds as a constrained combinatorial optimization problem [Carpentier 2010a]. They formulate CAMO as a variation of the knapsack problem where the aim is to find a combination of musical instruments that maximizes the timbral similarity with the target constrained by the capacity of the orchestra (i.e., the database). The binary allocation knapsack problem can be shown to be NP-complete so it cannot be solved in polynomial time. They explore the vast space of possible instrument combinations with a genetic algorithm (GA) that optimizes a fitness function that encodes timbral similarity between the candidate instrument combinations and the target sound.

GAs are metaheuristics inspired by the Darwinian principle of survival of the fittest. The GA maintains a list of individuals that represent the possible combinations of instruments. These individuals evolve towards optimal solutions by means of the biologically inspired operators of crossover, mutation, and selection. Crossover and mutation are responsible for introducing variations in the current population and promoting the exploration and exploitation of the search space. Selection guarantees that the fittest individuals are passed to the next generation gradually converging to optimal regions of the search space. The major drawback of this approach arises from the loss of diversity inherent in the evolutionary search performed with GAs. In practice, the loss of diversity results in only one solution that commonly corresponds to a local optimum because GAs cannot guarantee to return the global optimum (i.e., the best solution). Moreover, running the GA multiple times with the same parameters commonly results in different solutions. Carpentier et al. (2010a) use a combination of local search and constraint strategies in Orchidée[1] to circumvent the issues resulting from loss of diversity. Orchidée has become the de facto state-of-the-art orchestration tool used by composers, composition teachers and students.

[1] http://www.gregoirecarpentier.net/research/orchidee.php

3 PROPOSED APPROACH

3.1 Preliminary Work

My preliminary work [Abreu 2016] used an artificial immune system (AIS) called opt-aiNet [de Castro 2002] to search for multiple combinations of musical instrument sounds whose timbral features match those of the target sound. Inspired by immunological principles, opt-aiNet returns multiple good quality solutions in parallel while preserving diversity. The intrinsic property of maintenance of diversity allows opt-aiNet to return all the optima (global and local) of the fitness function being optimized upon convergence, as shown in Fig.1 b). Fig.1 illustrates a two-dimensional fitness function to be optimized by finding the peaks (i.e., points where the function has maximum amplitude). Fig.1 a) illustrates the mono-modal optimization property of GAs, which typically converge to a unique solution represented as a black dot on a peak. In turn, Fig. 1 b) illustrates the multi-modal ability of the AIS opt-aiNet, capable of returning all optima of the function within the region of interest.

The application of opt-aiNet in CAMO gave rise to Immune Orchestra[2] [Abreu 2016]. The preliminary results suggest that the property of maintenance of diversity translates as orchestrations that are all similar to the target yet different from one another. Thus Immune Orchestra can provide the composer with multiple choices when orchestrating a sound instead of searching for one solution constrained by choices defined a priori [Carpentier 2010a]. Consequently, Immune Orchestra has the potential to expand the creative possibilities of CAMO beyond what the composer initially imagined.

Acknowledgments

The work presented in this paper was partially supported by FourEyes, a Research Line within project "TEC4Growth – Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020" financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[Abreu 2016] J. Abreu, M. Caetano, R. Penha. Computer-Aided Musical Orchestration Using an Artificial Immune System. In: C. Johnson, V. Ciecielski, J. Correia, P. Machado (Eds.) Evolutionary and Biologically Inspired Music, Sound, Art and Design. Springer International Publishing, vol. 9596, pp. 1-16 (2016)

[Caclin 2005] A. Caclin, S. McAdams, B. Smith, S. Winsberg. Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J. Acoust. Soc. Am. 118(1), pp. 471–482 (2005)

[Caetano 2005] M. Caetano, J. Manzolli, F. J. Von Zuben. Application of an Artificial Immune System in a Compositional Timbre Design Technique. Artificial Immune Systems, Lecture Notes in Computer Science, Springer-Verlag, vol. 3627, pp. 389-403 (2005)

[Caetano 2010] M. Caetano, J. J. Burred, X. Rodet. Automatic Segmentation of the Temporal Evolution of Isolated Acoustic Musical Instrument Sounds Using Spectro-Temporal Cues. Proc. Int. Conf. on Digital Audio Effects (2010)

[Carpentier 2006] G. Carpentier, D. Tardieu, G. Assayag, X. Rodet, E. Saint-James. Imitative and generative orchestrations using pre-analysed sound databases. Proceedings of the Sound and Music Computing Conference, pp. 115–122 (2006)

[Carpentier 2007] G. Carpentier, D.Tardieu, G. Assayag, X. Rodet, E. Saint-James. An evolutionary approach to computer-aided orchestration. In: Giacobini, M. (ed.) EvoWorkshops 2007. LNCS, vol. 4448, pp. 488–497 (2007)

[Carpentier 2010a] G. Carpentier, G. Assayag, E. Saint-James. Solving the musical orchestration problem using multiobjective constrained optimization with a genetic local search approach. J. Heuristics 16(5), 681–714 (2010)

[Carpentier 2010b] G. Carpentier, D. Tardieu, J. Harvey, G. Assayag, E. Saint-James. Predicting timbre features of instrument sound combinations: application to automatic orchestration. J. New Music Res. 39(1), 47–61 (2010)

[de Castro 2002] L. N. de Castro, J. Timmis. An artificial immune network for multimodal function optimization. In: Proceedings of the 2002 Congress on Evolutionary Computation, vol. 1, pp. 699–704 (2002)

[Esling 2010] P. Esling, G. Carpentier, C. Agon. Dynamic Musical Orchestration Using Genetic Algorithms and a Spectro-Temporal Description of Musical Instruments. Applications of Evolutionary Computation, LNCS, vol. 6025 (2010).

[Esling 2012] P. Esling, C. Agon. Time Series Data Mining. ACM Computing Surveys, Vol. 45, Issue 1, article 12 (2012)

[Esling 2013] P. Esling, C. Agon. Multiobjective Time Series Matching for Audio Classification and Retrieval. IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 10, pp. 2057-2072 (2013)

[Goto 2002] M. Goto, H. Hashiguchi, T. Nishimura, R. Oka. RWC music database: popular, classical and Jazz music databases. In: Proceedings of the International Society for Music Information Retrieval Conference, vol. 2, pp. 287–288 (2002)

[Handelman 2012] E. Handelman, A. Sigler, D. Donna. Automatic orchestration for automatic composition. In: 1st International Workshop on Musical Metacreation (MUME 2012), pp. 43–48 (2012)

[Hummel 2005] T. Hummel. Simulation of human voice timbre by orchestration of acoustic music instruments. In: Proceedings of the International Computer Music Conference (ICMC), pp. 185 (2005)

[Kendall 1993] R. A. Kendall, E. C. Carterette. Identification and blend of timbres as a basis for orchestration. Contemp. Music Rev. 9(1–2), pp. 51–67 (1993)

[Lartillot 2007] O. Lartillot, P. Toiviainen. A Matlab toolbox for musical feature extraction from audio. In: International Conference on Digital Audio Effects, pp. 237–244 (2007)

[McAdams 2009] S. McAdams, B. L. Giordano. The perception of musical timbre. In: Hallam, S., Cross, I., Thaut, M. (eds.) The Oxford Handbook of Music Psychology, pp. 72–80 (2009)

[McAdams 1995] S. McAdams, S. Winsberg, S. Donnadieu, G. DeSoete, J. Krimphoff. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychol. Res. 58(3), pp. 177–192 (1995)

[Navarro 2015] M. Navarro, M. Caetano, G. Bernardes, L. N. de Castro, J. M. Corchado. Automatic Generation of Chord Progressions with an Artificial Immune System. Evolutionary and Biologically Inspired Music, Sound, Art, and Design, Lecture Notes in Computer Science, Springer-Verlag, vol. 9027, pp. 175-186 (2015)

[Psenicka 2003] D. Psenicka. SPORCH: an algorithm for orchestration based on spectral analyses of recorded sounds. In: Proceedings of International Computer Music Conference (ICMC), pp. 184 (2003)

[Rose 2009] F. Rose, J. E. Hetrik. Enhancing orchestration technique via spectrally based linear algebra methods. Comput. Music J. 33(1), pp. 32–41 (2009)

[Tardieu 2007] D. Tardieu, X. Rodet. An instrument timbre model for computer aided orchestration. In: 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 347–350 (2007)

[2] https://paginas.fe.up.pt/~ee10146/#thesis-results