Swarm and Evolutionary Computation 50 (2019) 100484

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo

# Leveraging diversity in computer-aided musical orchestration with an artificial immune system for multi-modal optimization

Marcelo Caetano [a,*], Asterios Zacharakis [b], Isabel Barbancho [c], Lorenzo J. Tardón [c]

[a] INESC TEC, Sound and Music Computing Group, Porto, Portugal
[b] School of Music Studies, Aristotle University of Thessaloniki, Greece
[c] ATIC Research Group, ETSI Telecomunicación, University of Málaga, Spain

ARTICLE INFO

*Keywords:*
Musical orchestration
Multi-modal optimization
Artificial immune systems

ABSTRACT

The aim of computer-aided musical orchestration (CAMO) is to find a combination of musical instrument sounds that perceptually approximates a reference sound when played together. The complexity of timbre perception and the combinatorial explosion of all possible musical instrument sound combinations make it very challenging to find even one orchestration for a reference sound. However, finding only one orchestration is seldom enough given the creative nature of the compositional process. Compositional applications of computer-aided musical orchestration can greatly benefit from multiple orchestrations with diversity. In this work, we use an artificial immune system (AIS) called opt-aiNet to search for combinations of musical instrument sounds that minimize the distance to a reference sound encoded in a fitness function. Opt-aiNet was developed to maximize diversity in the solution set of multi-modal optimization problems, which results in multiple alternative orchestrations for the same reference sound that are different among themselves. We compared the diversity and the similarity of the orchestrations proposed by opt-aiNet (CAMO-AIS) against a standard genetic algorithm (CAMO-GA) and Orchids, which is considered the state of the art for CAMO, for 13 reference sounds. In general, CAMO-AIS outperformed CAMO-GA and Orchids for several measures of objective diversity. We performed a listening test to evaluate and compare the perceptual similarity of the orchestrations by CAMO-AIS and Orchids. CAMO-AIS generated orchestrations that were perceived to be as similar to the reference sounds as those returned by Orchids. Therefore, CAMO-AIS has higher diversity of orchestrations than Orchids without loss of perceptual similarity.

## 1. Introduction

Orchestration is understood as "the art of blending instrument timbres together" [1]. Initially, orchestration was simply the assignment of instruments to pre-composed parts of the score, which was dictated largely by the availability of resources, such as what instruments and how many of each are available in the orchestra [2,3]. Later on, composers started regarding orchestration as an integral part of the compositional process whereby the musical ideas themselves are expressed [2,4]. 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 [4,5]. Orchestration remains an empirical activity largely due to the difficulty to formalize the required knowledge [1,2,6]. Diversity has been identified as an important property that can provide the composer with multiple alternatives given the highly subjective nature of musical orchestration combined with the complexity of timbre perception [7].

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) [4,8–13]. CAMO tools automate the search for instrument combinations that perceptually approximate a reference timbre commonly represented by a reference sound [1,6]. The combinations found are generally included in the score and later played by orchestras in live performances. However, most CAMO tools allow the

* Corresponding author.
*E-mail address:* mcaetano@inesctec.pt (M. Caetano).
*URL:* http://www.researcherid.com/rid/D-7821-2015.

https://doi.org/10.1016/j.swevo.2018.12.010
Received 15 March 2018; Received in revised form 12 October 2018; Accepted 27 December 2018
Available online 6 January 2019
2210-6502/© 2019 Elsevier B.V. All rights reserved.

composer to preview the result of the combinations found using musical instrument sounds from pre-recorded databases, which has been deemed an appropriate rendition of the timbre of the instrument combinations [14].

Early CAMO systems adopted a top-down approach [4,8,9] that consists of spectral analysis and subtractive spectral matching. These works usually keep a database of spectral peaks from musical instruments that will be used to match the reference spectrum. The algorithm iteratively subtracts the spectral peaks of the best match from the reference 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 [7].

The concept of timbre lies at the core of musical orchestration [1,2,6,15,16] largely because instrumental combinations can give rise to new timbres if the sounds are perceived as blended [5,17]. Yet, the top-down approach neglects the exploration of timbral combinations by relying on spectral matching, which does not capture the multi-dimensional nature of timbre. Carpentier et al. [10–13,18] adopted a bottom-up approach that relies on timbre similarity and evolutionary computation to search for instrument combinations that approximate the reference. They use a genetic algorithm (GA) to search for instrument combinations that optimize a fitness function that encodes timbre similarity with feature vectors.

The bottom-up approach represents a paradigm shift toward generative CAMO [11,12,19,20], where the timbre of instrument combinations is compared with the timbre of the reference 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 reference. Timbre spaces [5,21–24] yield features that correlate with dimensions of timbre perception. Models of timbral combination [2,13,17] estimate these features for combinations of musical instrument sounds. Timbre similarity can be estimated as distances in timbre spaces [5], which are calculated as weighed distances between feature vectors [12].

CAMO systems that return only one orchestration seldom meet the requirements of the highly subjective and creative nature of music composition [7]. Often, the composer uses CAMO tools to explore the problem space and find instrument combinations that would be missed by the empirical methods found in traditional orchestration manuals [1,6]. The reference sound guides the search toward interesting regions of the search space and the weights fine-tune the relative importance of perceptual dimensions of timbre similarity encoded in the fitness function [1,6]. Diversity of orchestrations is important in CAMO [7] to allow the exploration of different musical ideas. This work focuses on CAMO algorithms that return multiple orchestrations in parallel as the strategy to address the intrinsic need for diversity in CAMO.

There are two current CAMO systems that return multiple orchestrations in parallel, Orchids [12] which uses multi-objective optimization (MOO) and our approach [20] called CAMO-AIS, which uses multi-modal single-objective optimization (SOO). In Orchids, Carpentier et al. [12] use the well-known multi-objective genetic local search (MOGLS) optimization algorithm [25] to tackle diversity by approximating the Pareto frontier. Each point on the theoretical Pareto frontier corresponds to an optimal solution for a specific combination of objectives in the fitness function given by the weight vector. Consequently, Orchids returns multiple orchestrations that approximate the reference sound differently because the weights emphasize timbre dimensions differently. Orchids prioritizes the objective similarity of Pareto optimal orchestrations over the perceptual similarity controlled by the weights.

In this work, we propose to use CAMO-AIS to optimize a single-objective fitness function with a multi-modal artificial immune system (AIS) called opt-aiNet [26]. The single-objective fitness function uses a fixed set of weights to combine the features, restricting the search to orchestrations that have the same relative importance of perceptual dimensions of timbre similarity. The multi-modal ability of opt-aiNet is illustrated in Fig. 1, where the fitness function is represented by the surface and the optima are the peaks. Opt-aiNet is capable of returning multiple solutions (i.e., orchestrations) in parallel, represented by the black dots, that correspond to local optima of the fitness function. These multiple orchestrations approach the reference similarly because they correspond to the same combination of weights, yet they are different among themselves because each is a unique instrument combination. However, the quality of local optima of this single-objective fitness function is always inferior to the global optimum, which would be closest to the reference according to the fitness value. Consequently, CAMO-AIS trades off the objective similarity given by the fitness func-

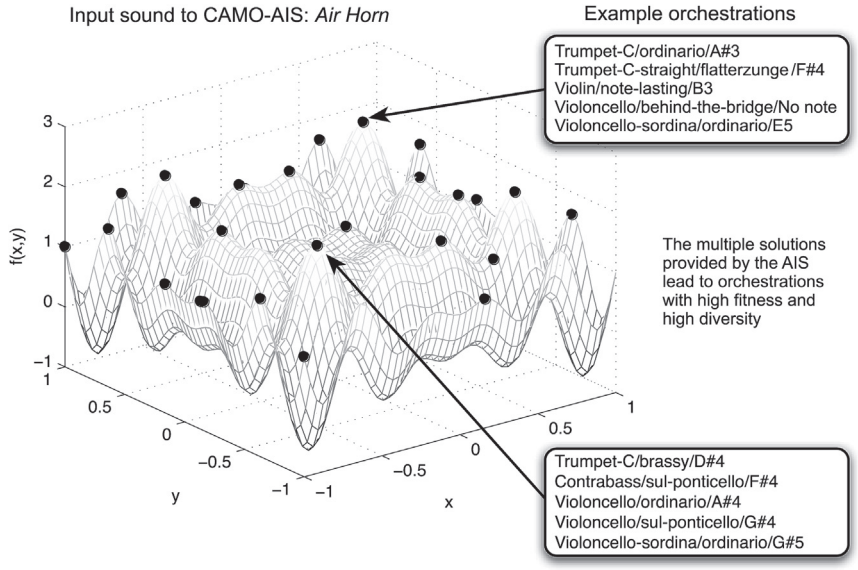



**Fig. 1.** Illustration of multi-modal function optimization in CAMO. The figure shows an objective function with multiple optima. The black dots represent multiple orchestrations returned by CAMO-AIS. Two example orchestrations for the reference sound *air horn* are given following the convention *instrument/playing technique/note*.

tion for perceptual similarity controlled by the weights. The contribution of this work lies in the diversity of orchestrations returned by CAMO-AIS resulting from the multi-modal ability of opt-aiNet.

The remainder of this paper is organized as follows. Section 2 reviews the literature of CAMO. Section 3 discusses theoretical aspects of the diversity strategies used by Orchids and CAMO-AIS. Section 4 presents an overview of our approach to CAMO. Next, Section 5 presents the experiment we performed followed by the evaluation of the results. The evaluation comprises similarity and diversity using objective measures and the subjective ratings from a listening test. Then, Section 6 presents the results, followed by a discussion in Section 7. Finally, Section 8 presents the conclusions and perspectives.

## 2. State of the art of computer-aided musical orchestration

Psenicka [8] 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." SPORCH keeps a database of spectral peaks of musical instrument and uses subtractive spectral matching and least squares to return one orchestration per run. Hummel [9] 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 reference and candidate sounds, disregarding pitch among other perceptual qualities. Rose and Hetrik [4] use singular value decomposition (SVD) to perform spectral decomposition and spectral matching. SVD decomposes the reference 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.

Carpentier et al. [10–13] consider the search for combinations of musical instrument sounds as a constrained combinatorial optimization problem. They formulate CAMO as a binary allocation knapsack problem where the aim is to find a combination of musical instruments that maximizes the timbral similarity with the reference constrained by the capacity of the orchestra (i.e., the database). However, the binary allocation knapsack problem cannot be solved in polynomial time because it was proved to be NP-complete [27]. They explore the vast space of possible instrument combinations with a genetic algorithm (GA) that optimizes a fitness function which encodes timbral similarity between the candidate instrument combinations and the reference sound. They use MOGLS [25] to return multiple instrument combinations in parallel that are nearly Pareto optimal. Later, Esling et al. [19] added the ability to perform dynamic orchestrations by representing the temporal variation of timbral features.

Recently, Antoine et al. [15,28–30] proposed to use supervised classification to generate orchestrations from semantic descriptors of timbre. Currently, i-Berlioz [30] uses support vector machines (SVM) to generate instrumental combinations that would match one of the five supported semantic descriptors of timbre, namely breathiness, brightness, dullness, roughness, and warmth. Antoine et al. have chosen to use semantic descriptors of timbre to enable the composer to focus on a more specific sound quality [30] by *restricting* the number of orchestrations returned by i-Berlioz. Therefore, i-Berlioz is conceived to minimize the diversity of orchestrations returned under the assumption that the five semantic terms unequivocally describe the timbre of the result. However, most research on timbre perception suggests otherwise [5]. In CAMO, this redundancy in the description of timbre translates as multiple instrument combinations approximating the reference timbral description (with varying degrees of similarity). This work focuses on the diversity of orchestrations as aesthetic alternatives for the composer.

In a previous work [20], we adapted an artificial immune system (AIS) called opt-aiNet [26] to return multiple combinations of musical instrument sounds whose timbral features approximate those of a reference sound. The sound database used contained 1439 sounds from the RWC Musical Instrument Sound Database [31,32] selected from 13 instruments played with 3 dynamics. We compared the results with multiple runs of a standard GA (CAMO-GA) using 10 reference sounds.

In this work, we compare the diversity of orchestrations returned by CAMO-AIS against CAMO-GA and Orchids, the state of the art of CAMO using four different musical instrument sound databases (see Section 4.3). We orchestrated 13 reference sounds with CAMO-GA, CAMO-AIS, and Orchids under the same conditions (whenever comparable) and then we compared the diversity of the orchestrations using multiple objective measures. Finally, we performed a listening test to evaluate the perceptual similarity of the orchestrations returned by CAMO-AIS and Orchids.

## 3. Diversity strategies in computer-aided musical orchestration

Diversity is very important in CAMO given the creative nature of the compositional process. The composer is rarely interested in a single combination (i.e., an orchestration) that optimizes some objective measure(s) with a reference sound [7]. Instead, the reference sound is typically used to guide the search towards a region of interest in the vast space of timbral combinations. Very often, the composer will use subjective criteria not encoded in the objective measure(s) guiding the search to choose one or more orchestrations of interest. Therefore, a CAMO algorithm should be capable of returning several orchestrations that are all similar to the reference sound yet dissimilar among themselves, representing different alternative orchestrations for that reference sound. In that case, diversity provides the composer with multiple choices when orchestrating a reference sound, expanding the creative possibilities of CAMO beyond what the composer initially imagined. In this work, we are especially interested in comparing the ability of CAMO-AIS, CAMO-GA, and Orchids to generate diverse orchestrations.

### 3.1. Multi-objective versus multi-modal single-objective optimization

Carpentier et al. [12] propose to use a multi-objective optimization strategy to tackle diversity. The objective similarity with the reference sound is encoded as multiple independent single-objective distance measures $D$ (see Section 4.7 for further details) which are combined with weight vectors $\vec{\alpha}$ as

$$\vec{E}_j = \vec{\alpha_j} D_j, \quad \text{with} \quad \sum_j |\alpha_j| = 1, \tag{1}$$

where $j$ is the index of dimensions of the feature space, $\vec{\alpha_j}$ are the vector components of $\vec{\alpha}$ with magnitude $|\alpha_j|$. Carpentier et al. [12] use MOGLS to find efficient solutions [25] that correspond to different combinations of the weights $\vec{\alpha}$ that maximize diversity along the Pareto front.

We propose to approach the problem as single-objective and use the multi-modal optimization ability of opt-aiNet to find multiple local optima that maximize diversity in the feature space. Therefore, in CAMO-AIS, opt-aiNet minimizes the following distance (fitness) function

$$F = \sum_j |\alpha_j| D_j, \quad \text{with} \quad \sum_j |\alpha_j| = 1. \tag{2}$$

### 3.2. Maintenance of diversity in opt-aiNet

The multi-modal ability of opt-aiNet emerges from the property of maintenance of diversity, which allows opt-aiNet to return multi-

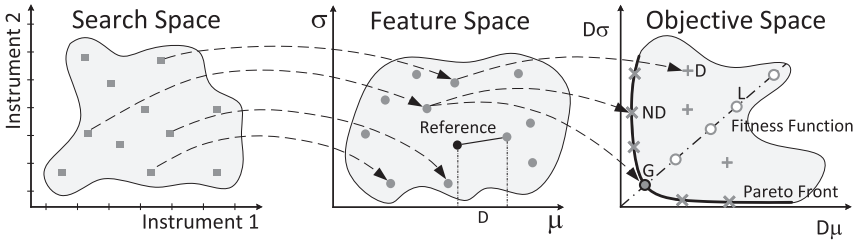



**Fig. 2.** Illustration of the different spaces in CAMO. The left-hand panel shows the search space, the middle panel shows the feature space, and the right-hand panel shows the objective space. Each point in the decision space is an instrument combination (orchestration) that has a corresponding position in the feature space. The reference sound can also be seen in the feature space. The distances $D$ between points in the feature space and the reference sound are calculated in the feature space. Weight vectors $\vec{\alpha}$ map points in the feature space to the objective space.

ple local optima of the fitness function being optimized upon convergence. Fig. 1 shows a multi-modal function with multiple global and local optima represented by the multiple peaks. Standard optimization methods commonly only return one solution (i.e., one black dot) corresponding to one local optimum of the fitness function. The property of maintenance of diversity in opt-aiNet translates as multiple solutions returned in parallel, corresponding to several local optima of the fitness function. Two different measures are involved in the property of maintenance of diversity in opt-aiNet, namely the fitness function and the affinity measure. Fitness is a measure of the quality of a candidate solution and is used to explore promising regions of the search space. Affinity is a measure of distance between the current solutions and is used to eliminate candidate solutions with lower fitness that are close to high fitness solutions.

At each iteration, opt-aiNet uses the immunological principles of clonal expansion, mutation, and suppression to evolve a population of candidate solutions in an immune network. Clonal expansion and mutation expand the size of the pool of candidate solutions during the exploration of regions of the search space associated with high fitness. Then, suppression cuts back down the current population by keeping only the best solutions in regions within a radius $\rho$. Maintenance of diversity is achieved by eliminating the antibodies whose affinity is lower than $\rho$ from the network while keeping the ones with the highest fitness. The result is illustrated in Fig. 1, where only the best individual per peak of the fitness function is returned.

Similarity with the reference is measured with the fitness function of eq. (2), while affinity is measured as Euclidean distances between candidate orchestrations in the feature space (see Section 4.8). Both the fitness function and the affinity measure use the features described in Section 4.4, which, in turn, capture perceptual aspects of the sounds. Consequently, fitness $F$ is inversely proportional to perceptual similarity of orchestrations with the reference sound and affinity is proportional to perceptual dissimilarity between orchestrations. Therefore, in CAMO-AIS, maintenance of diversity translates as orchestrations that are all similar to the reference yet different from one another.

### 3.3. Diversity in orchids and in CAMO-AIS

Fig. 2 illustrates the different diversity strategies between Orchids and CAMO-AIS. Fig. 2 shows the search space (also called decision space in the multi-objective optimization literature), the feature space, and the objective space. Each point in the search space is an orchestration represented as an instrument combination that has a corresponding position in the feature space. The features encode perceptual aspects of the sounds and are used in the single-objective distances $D$ between the orchestrations and the reference sound. Weight vectors $\vec{\alpha}$ map points in the feature space to points in the objective space. The same point in the feature space can be mapped to different points in the objective space by different weight vectors $\vec{\alpha}$. On the other hand, a fixed weight vector $\vec{\alpha}$ always maps points in the feature space to a straight line in the objective space. Thus $\vec{\alpha}$ can be interpreted as specifying the direction by which a solution approaches the theoretical optimum at the origin of the objective space.

The right-hand panel in Fig. 2 shows the Pareto front with non-dominated solutions (ND) illustrated as "X" and dominated solutions (D) illustrated as "+". The locus of the fitness function in the objective space can also be seen as a straight line containing the global optimum (G) illustrated as the filled "O" and the local optima (L) illustrated as the empty "O". Note that dominated solutions D can coincide with local optima L and, in turn, non-dominated solutions ND can coincide with the global optimum G. Thus CAMO-AIS returns solutions L that were discarded by MOGLS because there is a solution G closer to the reference in the same direction in the objective space (i.e., specified by the same $|\alpha_j|$). MOGLS provides a set of efficient solutions that approach the reference sound in different directions. Perceptually, each solution returned by MOGLS would be closer to the reference sound according to different criteria emphasized by the different weight vectors $\vec{\alpha}$. On the other hand, CAMO-AIS returns solutions that always approach the reference in the same direction, emphasizing the same perceptual similarities. The trade-off is that the quality of the solutions decreases when they are local optima L. In this article, we investigate the consequences of these different approaches to CAMO in terms of similarity and diversity. We aim to show that CAMO-AIS returns orchestrations with higher diversity than Orchids without loss of perceptual similarity with the reference.

## 4. Computer-aided musical orchestration with an artificial immune system (CAMO-AIS)

### 4.1. Overview

Fig. 3 shows an overview of CAMO-AIS. The sound database is used to build a feature database, which consists of acoustic features calculated for all sounds prior to the search for orchestrations. The same features are calculated for the reference sound being orchestrated. The combination functions estimate the features of a sound combination from the features of the individual sounds. The evaluation function uses these features to estimate the similarity between combinations of features from sounds in the database and those of the reference sound. The search algorithm opt-aiNet is used to search for combinations that approximate the reference sound, called orchestrations.

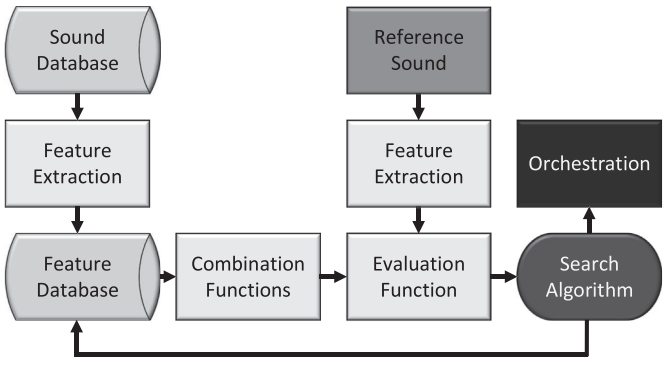
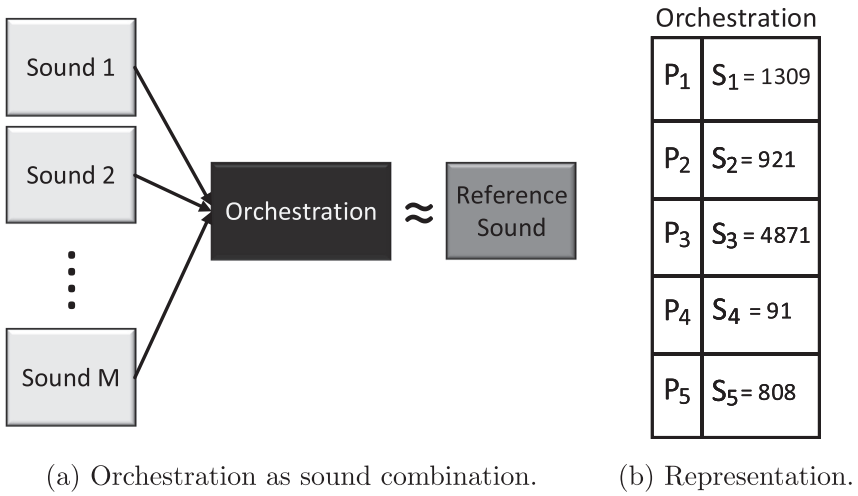



**Fig. 3.** Overview of CAMO. The figure illustrates the different components of the CAMO approach adopted.




(a) Orchestration as sound combination. (b) Representation.

**Fig. 4.** Representation of orchestrations. Part (a) illustrates the orchestration as a combination of sounds that approximates the reference. Part (b) shows the internal representation of each orchestration in CAMO-AIS.

### 4.2. Representation

Fig. 4a illustrates an orchestration as a combination of sounds from the sound database that approximates the reference sound when played together. Fig. 4b shows the representation used by CAMO-AIS, in which an orchestration has $M$ players $p(m)$, and each player is allocated a sound $s(n) \in S$, where $n = [1, \ldots, N]$ is the index in the database $S$, which has $N$ sounds in total. Thus an orchestration is a combination of sounds $c(m,n) = \{s_1(n), \ldots, s_M(n)\}$, $\forall\, s_m(n) \in S$. Fig. 4b shows $c(m,n)$ represented as a list, but the order of players $p(m)$ does not matter for the orchestration. Each sound $s_m(n)$ corresponds to a specific note of a given instrument played with a dynamic level, and $s_m(n) = 0$ indicates that player $p(m)$ was allocated no instrument.

#### 4.2.1. Discrete search space

Originally, opt-aiNet [26] was designed to optimize functions of continuous variables, performing the search in continuous vector spaces. In our work, the search space is discrete because the representation of orchestrations $c(m,n)$ is a vector of discrete indices $n$ of sounds in the database, as shown in Fig. 4b. Most of the operations of the continuous version of opt-aiNet work for discrete vectors as well. The exception is the original mutation operator which used a continuous random variable to add a small perturbation to the vectors being mutated. Thus we adapted the mutation operator for discrete vectors using a probability of mutation to determine if the vector will undergo mutation. The probability of mutation $\chi$ is calculated as

$$\chi = \exp(-\gamma \widehat{F}) \tag{3}$$

where $\gamma$ is a constant and $\widehat{F}$ is the normalized fitness value of the combination vector $c(m,n)$ being mutated. For each index $n$, a uniform random variable $u(0,1)$ will determine if the corresponding sound $s(n)$ is replaced by another sound from $S$. If $u(0,1) < \chi$ then a new $s(n) \in S$ is chosen from another uniform distribution $u(1,N)$, where $n \in \mathbb{N}$ and $n \leq N$. Following our previous work [20], we set $\gamma = 1.2$.

### 4.3. Musical instrument sound databases

This work uses four musical instrument sound databases, namely Studio Online (SOL),[1] Real World Computing (RWC),[2] Philharmonia,[3] and Iowa.[4] All the experiments reported in Sec. 6 used all four sound databases above for all the orchestrations algorithms, except for Orchids, which uses SOL by default and it cannot be replaced.

SOL comes bundled with Orchids with a total of 24,320 sounds from 33 instruments from 5 families, namely brass, keyboards, plucked strings, strings, and woodwinds. SOL has sounds played with over 550 classical, experimental, and extended articulations and playing techniques as well as 6 dynamics, *pianissimo*, *piano*, *mezzo piano*, *mezzo forte*, *forte* and *fortissimo*.

RWC music database contains a total of 37,372 sounds from 36 instruments from 5 families, namely brass, keyboard, popular, strings, and woodwinds. For each instrument, RWC has sounds played with 3 dynamics (*pianissimo*, *mezzo forte*, and *fortissimo*) and different playing techniques. RWC provides up to 3 variations for each instrument, where each variation corresponds to a different instrument manufacturer and a different musician.

Philharmonia has a total of 13,680 sounds from 58 instruments from 5 families, namely brass, percussion, plucked strings, strings, and woodwinds. Philharmonia has sounds played with 3 dynamics (*pianissimo*, *mezzo forte*, and *fortissimo*) and a total of 73 extended articulations and playing techniques.

Iowa has a total of 4,483 sounds from 19 instruments from 3 families, namely brass, strings, and woodwinds. In Iowa, each instrument sound was played *pianissimo*, *mezzo forte*, and *fortissimo* for all the notes comprising the entire instrumental range. For stringed instruments, the total range of sounds was recorded for each string.

### 4.4. Feature extraction

Traditionally, timbre is considered as the set of attributes whereby a listener can judge that two sounds are dissimilar using any criteria other than pitch, loudness, or duration [5]. Therefore, we consider pitch, loudness, and duration separately from timbre dimensions. The features used are fundamental frequency $f_0$ (pitch), frequency $f$ and amplitude $a$ of the contribution spectral peaks $A$, loudness $\lambda$, spectral centroid $\mu$, and spectral spread $\sigma$. The fundamental frequency $f_0$ of all sounds $s(n)$ in the database is estimated with Swipe [33]. The spectral centroid $\mu$ captures brightness while the spectral spread $\sigma$ correlates with the third dimension of MDS timbre spaces [21–24]. All the features are calculated over short-term frames (see window size in Table 1) and averaged

[1] https://www.uvi.net/ircam-solo-instruments.html.
[2] https://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-i.html.
[3] https://www.philharmonia.co.uk/explore/sound_samples.
[4] http://theremin.music.uiowa.edu/MIS.html.

**Table 1**
Parameters of the experiment.

| **Orchestration Algorithms** | | | |
|---|---|---|---|
| Parameter | **CAMO-AIS** | **CAMO-GA** | **Orchids** |
| Maximum Number of Players | 5 | 5 | 5 |
| Number of Iterations | 500 | 500 | 500 |
| Initial size of population | 50 | 50 | 50 |
| Maximum size of population | Auto | 200 | 200 |
| Mating pool size | – | 200 | 200 |
| Number of Clones | 20 | – | – |
| **Sound Analysis** | | | |
| Parameter | **CAMO-AIS** | **CAMO-GA** | **Orchids** |
| Window Type | Hamming | Hamming | Hamming |
| Window Size (ms) | 46.4 | 46.4 | 60 |
| Hop Size (ms) | 23.2 | 23.2 | 10 |
| FFT Size | 4096 | 4096 | 4096 |
| Maximum Number of Partials | 25 | 25 | 25 |

across all frames. Orchids uses the same features calculated similarly.

#### 4.4.1. Contribution spectral peaks

The spectral energy that sound $s(m)$ contributes to an orchestration is determined by the contribution spectral peaks vector $\vec{A}_m(k)$. In what follows, only peaks whose spectral energy (amplitude squared) is at most 35 dB below the maximum level (i.e., 0 dB) are used and all other peaks are discarded. These peaks are stored as a vector with the pairs $\{a(k), f(k)\}$ for each sound $s(m)$, where $k$ is the index of the peak. The contribution spectral peaks $\vec{A}_m(k)$ are the spectral peaks from the *candidate* sound $s(m)$ that are common to the spectral peaks of the *reference* sound $r$. Eq. (4) shows the calculation of $\vec{A}_m(k)$ as

$$\vec{A}_m(k) = \begin{cases} a_s(k) & \text{if} \quad (1+\delta)^{-1} \le f_s(k)/f_r(k) \le 1+\delta \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $a_s(k)$ is the amplitude and $f_s(k)$ is the frequency of the spectral peak of the *candidate* sound, and $f_r(k)$ is the frequency of the *reference* sound.

Fig. 5 illustrates the computation of spectral peak similarity between the *reference* sound and a *candidate* sound. Spectral peaks are represented as spikes with amplitude $a(k)$ at frequency $f(k)$. The frequencies $f_r(k)$ of the peaks of the *reference* sound are used as reference. Whenever the *candidate* sound contains a peak in a region $\delta$ around $f_r(k)$, the amplitude $a(k)$ of the peak at frequency $f_s(k)$ of the *candidate* sound is kept at position $k$ of the contribution spectral peaks vector $\vec{A}_m(k)$. Following our previous work [20], we set $\delta = 0.025$.

#### 4.4.2. Loudness

Loudness $\lambda$ is calculated as

$$\lambda = 20 \log_{10}\left(\sum_k a(k)\right), \tag{5}$$

where $a(k)$ are the amplitudes at frequencies $f(k)$.

#### 4.4.3. Spectral centroid

The spectral centroid $\mu$ is calculated as

$$\mu = \sum_k f(k) \frac{|a(k)|^2}{\sum_k |a(k)|^2}. \tag{6}$$

#### 4.4.4. Spectral spread

The spectral spread $\sigma$ is calculated as

$$\sigma = \sum_k (f(k) - \mu)^2 \frac{|a(k)|^2}{\sum_k |a(k)|^2}. \tag{7}$$

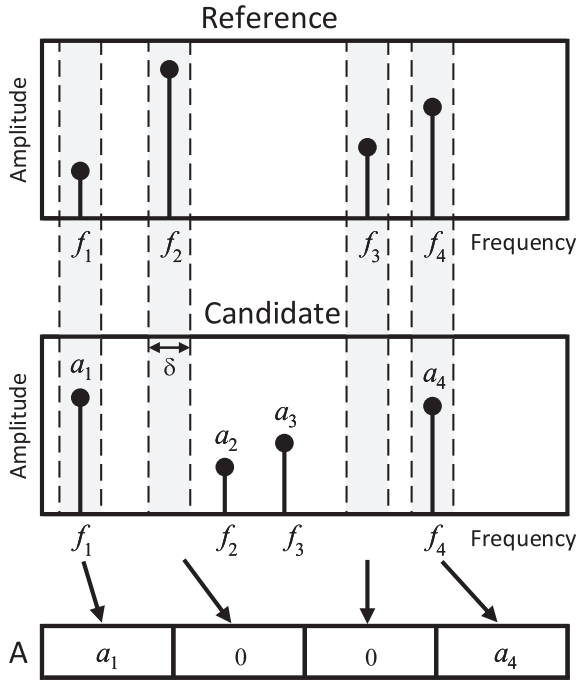



**Fig. 5.** Contribution spectral peaks $\vec{A}_m(k)$. The figure shows the representation of the contribution spectral peaks of a candidate sound.

### 4.5. Pre-processing

Prior to the search for orchestrations of a given reference sound $r$, the entire sound database $S$ is reduced to a subset $S^r$ of sounds that will be effectively used to orchestrate $r$. All the sounds whose contribution spectral peaks vector $\vec{A}_m(k)$ is all zeros are eliminated because these do not contribute spectral energy to the orchestration. Similarly, all the sounds whose $f_0$ is lower than $f_0^r$ are eliminated because these add spectral energy outside of the region of interest and have a negative impact on the final result. Partials with frequencies higher than all frequencies in $r$ are disconsidered because these are in the high-frequency range and typically have negligible spectral energy.

### 4.6. Combination functions

The sounds $s(m,n)$ in an orchestration $c(m,n)$ should approximate the reference $r$ when played together. Therefore, the combination functions estimate the values of the spectral features of $c(m,n)$ from the

features of the isolated sounds $s(m,n)$ normalized by the RMS energy $e(m,n)$ [13]. The combination functions for the spectral centroid $\mu$, spectral spread $\sigma$, and loudness $\lambda$ are given respectively by

$$\mu_c = \frac{\sum_m^M e(m)\,\mu(m)}{\sum_m^M e(m)}, \tag{8}$$

$$\sigma_c = \sqrt{\frac{\sum_m^M e(m)\left(\sigma^2(m)+\mu^2(m)\right)}{\sum_m^M e(m)} - \mu_c^2}, \tag{9}$$

$$\lambda_c = 20\log_{10}\left(\sum_m^M \frac{1}{K}\sum_k^K a(m,k)\right). \tag{10}$$

The estimation of the contribution spectral peaks of the combination $\vec{A}_c$ uses the contribution vectors $\vec{A}_s$ of the sounds $s(m,n)$ in $c(m,n)$ as

$$\vec{A}_r = \left\{\max_{k\in K}\left[\vec{A}(m,1)\right], \max_{k\in K}\left[\vec{A}(m,2)\right], \ldots, \max_{k\in K}\left[\vec{A}(m,N)\right]\right\}. \tag{11}$$

Orchids uses the same combination functions [12,13].

### 4.7. Distance functions

Each distance $D_j$ in eq. (2) measures the difference between the features from the reference sound $r$ and the candidate orchestration $c_q(m,n)$, where $q$ is the index of the orchestration among all the candidates for $r$, as follows

$$D_\mu = \frac{|\mu(c_q) - \mu(r)|}{\mu(r)}, \tag{12}$$

$$D_\sigma = \frac{|\sigma(c_q) - \sigma(r)|}{\sigma(r)}, \tag{13}$$

$$D_\lambda = \frac{|\lambda(c_q) - \lambda(r)|}{\lambda(r)}. \tag{14}$$

The distance between the contribution vector of the reference sound $\vec{A}_r$ and the contribution vector of the orchestration $\vec{A}_c$ is calculated as

$$D_{\vec{A}} = 1 - \cos\left(\vec{A}_r, \vec{A}_c\right). \tag{15}$$

Orchids uses the same distance functions [12,13].

Ultimately, the weights in CAMO-AIS are an aesthetic choice by the composer to determine the *perceptual* direction from which *all* the orchestrations should approximate the reference. In Orchids, each solution corresponds to a different set of weights, so the composer implicitly chooses a different perceptual direction by selecting an orchestration among the pool of solutions returned. In CAMO-AIS, the weights allow the composer to interactively explore the vast space of compositional possibilities and still have multiple orchestrations to choose from. The weights used in this work are $\alpha_{\vec{A}} = 0.6$, $\alpha_\lambda = 0.2$, $\alpha_\mu = 0.1$, and $\alpha_\sigma = 0.1$. These weights were validated in informal experiments with composers.

### 4.8. Affinity measure for suppression

Suppression discards candidate orchestrations that have affinity below a given threshold $\rho$. The affinity between two candidate orchestrations $c_q$ and $c_u$ is the Euclidean distance

$$\omega(q,u) = \sqrt{\sum_{j=1}^{J=4}\left(c_q(j) - c_u(j)\right)^2}, \tag{16}$$

where $c(1) = f_0$, $c(2) = \lambda$, $c(3) = \mu$, and $c(4) = \sigma$ are the dimensions of the reduced feature space where suppression operates. Following our previous work [20], the suppression threshold used is $\rho = 0.01$.

## 5. Evaluation

The quality of an orchestration depends on how similar it is to the reference sound [7]. Ideally, all orchestrations found should be as similar to the reference sound as possible. However, diversity is also important. Multiple solutions should be different from one another to represent alternatives, giving the composer options to choose from. Therefore, we evaluate the similarity and the diversity of the orchestrations generated by CAMO-AIS and compare with CAMO-GA and Orchids. We use objective and perceptual measures in the evaluation. Both CAMO-GA and CAMO-AIS use the same representation and optimize the fitness function $F$ of eq. (2) with the same weight values and equivalent parameters, whereas Orchids uses MOGLS to optimize eq. (1). Loss of diversity in standard GAs commonly results in many individuals converging to the same local optimum, whereas opt-aiNet returns multiple local optima. First we will compare the diversity of the orchestrations with several different objective measures. The aim is twofold, to compare the multi-modal ability of opt-aiNet (CAMO-AIS) against the standard GA (CAMO-GA) when optimizing $F$ and to compare the approaches behind Orchids (MOO) and CAMO-AIS (multi-modal SOO). Finally, we will compare the perceptual similarity of the orchestrations by CAMO-AIS and Orchids.

The experiment consisted in orchestrating $R = 13$ reference sounds with CAMO-GA, CAMO-AIS, and Orchids using the subset of the sound databases $\hat{S}$ described in Section 5.2. All reference sounds were orchestrated using the *static* mode in Orchids, which corresponds to the description found in Ref. [12]. We selected $Q = 8$ orchestrations for each reference generated by all methods, resulting in a total of 24 orchestrations per reference. CAMO-AIS returns the orchestrations ordered by fitness value, from lowest $F$ to highest (or from closest to the reference to the farthest according to $F$) so we simply use the first $Q = 8$. In Orchids, however, there is no natural ordering of the solutions because all the solutions returned correspond to the best solution found for a particular weight vector $\vec{\alpha}$. The orchestrations proposed by Orchids are ordered according to their position index in the population, so we simply selected the first $Q = 8$. CAMO-GA uses a standard GA with uniform crossover with 0.7 probability, uniform mutation with 0.2 probability, roulette wheel selection, and elitism (top 5% individuals). Similarly to CAMO-AIS, CAMO-GA returns orchestrations ordered by fitness value, so we retrieve the first $Q = 8$ orchestrations. Section 5.1 explains the reference sounds, section 5.2 details the subset $\hat{S}$ of the sound databases used in the experiment, and section 5.3 lists the parameters of the experiment.

### 5.1. Static and nearly harmonic reference sounds

All methods are used to generate *static* orchestrations with *nearly harmonic* musical instrument sounds. The term *static* orchestrations emphasizes that the features are averaged across the duration of the sounds so the feature vectors do not contain information about the temporal variation of these features during the course of the sounds. *Nearly harmonic* means that the spectrum of the musical instruments used contains partials with frequencies nearly harmonically related. Therefore, we used both *static* and *nearly harmonic* as criteria to select appropriate reference sounds. We chose sounds that present relatively little temporal variation and some degree of harmonicity, such as sirens and notes from instruments not found in the orchestra (e.g., synthesizers).

It is also important to choose reference sounds that are distributed relatively evenly in the feature space so these do not concentrate around one particular region and pose different challenges to orchestrate. Fig. 6 illustrates the distribution of the reference sounds in the reduced feature space (see Section 4.8). To visualize the relative distribution of the reference sounds, Fig. 6 was obtained with classic MDS [34] analysis of the feature vectors with dimensions $f_0$, $\mu$, $\sigma$, and $\lambda$ calculated from the reference sounds.

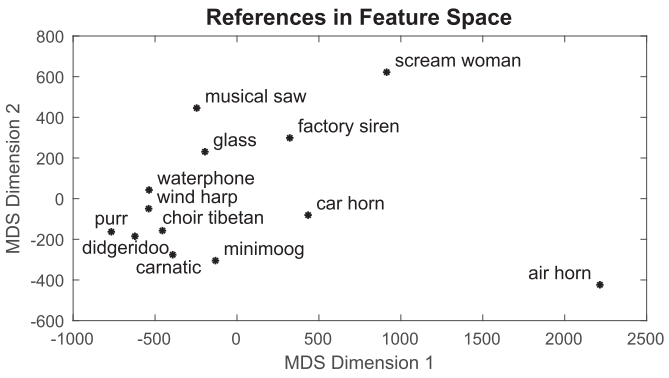



**Fig. 6.** Reference sounds in MDS representation of feature space.

### 5.2. Static and nearly harmonic musical instrument sounds

For all musical instrument sound databases used (see Section 4.3), we selected a subset $\widehat{S}$ played with non-time-varying articulations such as *ordinario* and *non-vibrato* as the most appropriate to orchestrate *static* reference sounds. Appendix A contains tables that show the size of the subspace $\widehat{S}^r$ for each reference, the total number of possible combinations in $\widehat{S}^r$ with $M$ players, and the total number of orchestrations returned by CAMO-AIS per reference sound, which is the number of local optima found. SOL appears in Table A.3, RWC appears in Table A.4, Philharmonia appears in Table A.5, and Iowa appears in Table A.6.

### 5.3. Parameters of the experiment

Table 1 lists the parameters of the experiment for the orchestration methods CAMO-AIS, CAMO-GA, and Orchids as well as for the sound analysis [35]. We used the presets in Orchids and in our previous work [26, 20] for CAMO-AIS. Whenever possible, we use the same parameter value for CAMO-AIS, CAMO-GA, and Orchids, such as the number of players $M$, the maximum number of partials $K$, and the maximum number of iterations because these parameters can potentially impact the result. $M$ and $K$ directly affect the spectrum of the combination because fewer partials result in lower similarity, while fewer players would have a detrimental effect as well. The number of iterations must be large enough to guarantee convergence, otherwise the similarity of the result might also be affected.

### 5.4. Evaluation of diversity

Given a reference sound and an orchestration method, the evaluation of perceptual diversity requires determining if each orchestration is perceptually different from the others. Each set of 8 orchestrations requires 28 pairwise comparisons. A total of 13 reference sounds and 3 methods would require 1092 pairwise evaluations. So we opted for an objective evaluation of diversity instead.

We can evaluate objective diversity in 2 spaces shown in Fig. 2, the search space and the feature space. In the search space, diversity translates as unique combinations. In the feature space, the positions of sounds reflect perceptual relations among them. So diversity in the feature space can be associated with diversity along the perceptual dimensions associated with the features used. We use the distribution of the orchestrations in the feature space to estimate the diversity of the orchestration set. We propose to use the variance of the positions in the feature space as measure of objective diversity.

Orchids is the only CAMO algorithm that operates in the objective space. The calculation of the objective measure of diversity in the objective space requires the final weights $\alpha$ and distances $D$. However, Orchids does not allow access to either $\alpha$ or $D$ to calculate the diversity in the objective space. Therefore, it is not possible to calculate diversity in the objective space for any of the CAMO algorithms tested.

#### 5.4.1. Objective diversity in the search space

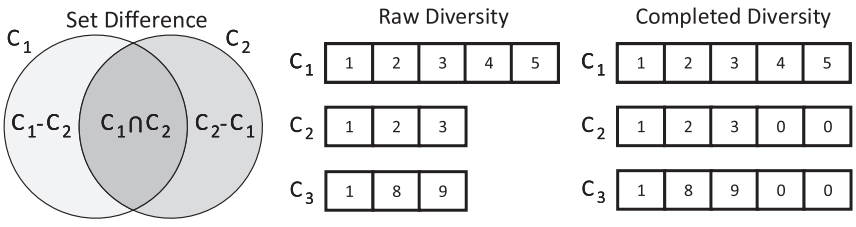
The orchestrations are represented in the search space as illustrated in Fig. 4b. In general, we expect different combinations of sounds to result in different orchestrations. However, we need to differentiate between a pair of orchestrations with $M-1$ identical sounds and 1 different sound and another pair of orchestrations where all $M = 5$ sounds are unique. So we propose to measure the difference $\epsilon$ in the combinations by simply counting the number of different sounds in each pair of orchestrations and dividing by the maximum number of players, written formally as

$$\epsilon(c_1, c_2) = \frac{1}{M}\text{card}(c_1 - c_2), \tag{17}$$

where $c_1$ and $c_2$ are combinations, $c_1 - c_2$ is the set difference between $c_1$ and $c_2$, "card" is the set *cardinality* operator, and $M$ is the maximum number of players. The cardinality of a set $\text{card}(c)$ is the number of elements in $c$ and $c_1 - c_2 = \{s \mid \in c_1 \text{ and } s \notin c_2\}$ is the elements $s$ in $c_1$ that are not in $c_2$.

Fig. 7a illustrates the difference between two sets $c_1$ and $c_2$ with a Venn diagram. Fig. 7a shows that the difference of two sets is a disjoint set because $c_1 = (c_1 - c_2) \cup (c_1 \cap c_2)$ and $c_2 = (c_2 - c_1) \cup (c_1 \cap c_2)$, so $(c_1 - c_2) \cap (c_2 - c_1) = \emptyset$.

Consequently, the difference operator is not commutative because $c_1 - c_2 \neq c_2 - c_1$. Therefore, $\epsilon(c_1, c_2) = \epsilon(c_2, c_1)$ only when $\text{card}(c_1) = \text{card}(c_2)$ because only then the number of remaining elements is the same. In other words, $\epsilon(c_1, c_2) = \epsilon(c_2, c_1)$ only when the orchestrations $c_1$ and $c_2$ being compared have the same number of players $M$. However, the orchestrations typically have between $M = 1$ and $M = 5$ players. Fig. 7b illustrates the measure of diversity for orchestrations $c_1$, $c_2$, and $c_3$ with different numbers of players. Note that $\epsilon(c_1, c_2) = 2/5$ but $\epsilon(c_2, c_1) = 0$. Also from Fig. 7b, $\epsilon(c_1, c_3) = 4/5$ but $\epsilon(c_3, c_1) = (c_1, c_2) = 2/5$. This is called *raw diversity*, as opposed to the *completed diversity* shown in Fig. 7c, which replaces missing players with 0 standing for *no instrument* allocated to player $m$. Fig. 7c shows that




(a) Set difference. (b) Raw diversity. (c) Completed diversity.

**Fig. 7.** Diversity in the search space. (a) Illustrates the difference between two sets $c_1$ and $c_2$. (b) Illustrates the raw diversity measure. (c) Illustrates the completed diversity measure.

now $\epsilon(c_1, c_2) = \epsilon(c_2, c_1) = 2/5$ and $\epsilon(c_1, c_3) = \epsilon(c_3, c_1) = 4/5$.

The final measure of diversity for each orchestration $c_q$ is

$$\overline{\epsilon}(c_q) = \frac{1}{Q-1} \sum_{u=1}^{Q} \epsilon(c_q, c_u), \tag{18}$$

which is simply the mean value of eq. (17) between $c_q$ and every other orchestration $Q - 1$. Finally, for each reference sound, we have

$$\widehat{\epsilon}(r) = \frac{1}{Q} \sum_{q=1}^{Q} \overline{\epsilon}(c_q), \tag{19}$$

which is the mean of the individual diversity for each orchestration $c_q$.

Different sound (or instrument) combinations do not necessarily correspond to perceptually different orchestrations. Oftentimes, sounds considered different are, in fact, a different playing style or articulation of the same note for the same instrument. The features used capture dimensions of sound perception such that diversity in the feature space should be a better indicator of perceptual diversity.

#### 5.4.2. Objective diversity in the feature space

The diversity of the orchestrations in feature space is proportional to the variance of their distribution in feature space. Thus we measure diversity in the feature space with an objective measure that captures the variance of the orchestrations in the same reduced feature space as the affinity is calculated. We propose to use principal component analysis (PCA) to indirectly estimate the variance of the orchestrations. The measure uses how much of the variance is captured by the first principal component, as shown in Fig. 8. Fig. 8a illustrates the case of maximum variance where the first principal component explains approximately 50% of the variance whereas Fig. 8b illustrates the case of minimum variance where the first principal component explains 100% of the variance. The measure of diversity $\varepsilon$ is given by

$$\varepsilon = \frac{J(1-E)}{J-1}, \text{ with } E \in \left[\frac{1}{J}, 1\right] \text{ and } \varepsilon \in [0, 1], \tag{20}$$

where $E$ is the percentage of variation explained by the first principal component, and $J$ is the number of dimensions of the reduced feature space. The dimensionality $J$ of the original space imposes a theoretical limit to the minimum of $E$ given by $1/J$ as specified in eq. (20). For the example in Fig. 8a, the maximum variance explained by the first principal component is 50% because $J = 2$. Note that $\varepsilon = 0$ when $E = 1$ and $\varepsilon = 1$ when $E = 1/J$. In other words, minimum diversity corresponds to PCA explaining 100% of the variation and maximum diversity corresponds to PCA explaining $\frac{100}{J}$ % of variation.

### 5.5. Evaluation of similarity

The fitness values from CAMO-GA and CAMO-AIS can be used as the objective measure of similarity. However, we cannot compare with Orchids because the application does not give access to either $\alpha$ or $D$ of the orchestrations returned. Therefore we performed a listening test to evaluate the perceptual similarity of the orchestrations compared to the reference sound for CAMO-AIS and Orchids. CAMO-GA was not included in the listening test because it uses the same fitness function as CAMO-AIS.

#### 5.5.1. Objective similarity

Objective similarity is measured with the fitness value from eq. (2). Fitness represents the distance to the reference sound $r$ such that smaller values of $F$ correspond to orchestration that are closer to the reference.

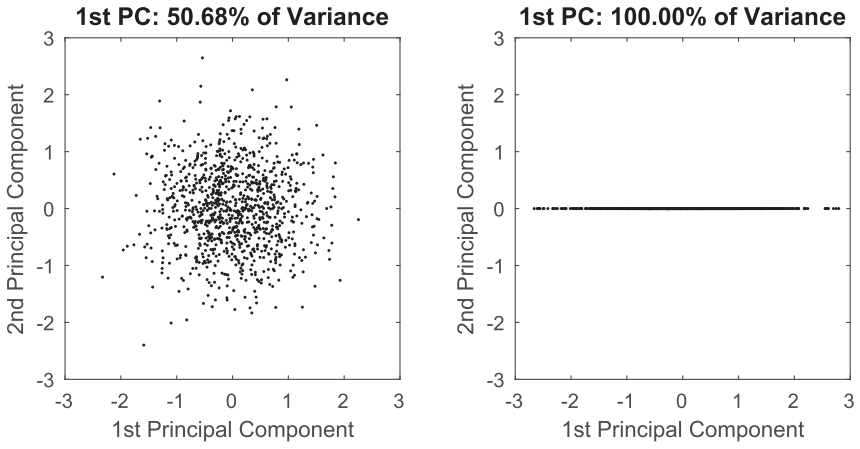
#### 5.5.2. Perceptual similarity

We designed and conducted an online listening test to evaluate the perceptual similarity of orchestrations to a number of preselected reference sounds. In total, 47 listeners (mean age: 33.7, age range: 19–60) participated in the listening test, all of which reported practicing a musical instrument, professional experience of audio processing, familiarity with a listening test procedure and listening to the stimuli with the use of high quality headphones. All participants provided informed consent, were free to withdraw at any point and were naive about the purpose of the test. The listening test can be found at http://camo.inesctec.pt.

Each participant evaluated the similarity of 7 reference sounds selected at random from the total pool of 13 references. See section 5.1 for a description of the reference sounds. Each page of the test presented one reference on top followed by the 16 orchestrations, 8 from CAMO-AIS and 8 from Orchids. The presentation order of both the 7 reference sounds and the 16 orchestrations in each page of the test was randomized (uniform distribution). In total, each reference was evaluated by at least 19 participants. Participants rated similarity between sounds using sliders with endpoints labeled *very dissimilar* and *very similar* respectively that corresponded to a hidden scale ranging between 0 and a 100.

## 6. Results

Section 6.1 presents the results for diversity, with diversity in the search space in Section 6.1.1 and diversity in the feature space in Section 6.1.2. Then, Section 6.2 presents the results for similarity, with




(a) Maximum variance. (b) Minimum variance.

**Fig. 8.** Variance of distributions. (a) Illustrates maximum variance. (b) Illustrates minimum variance.

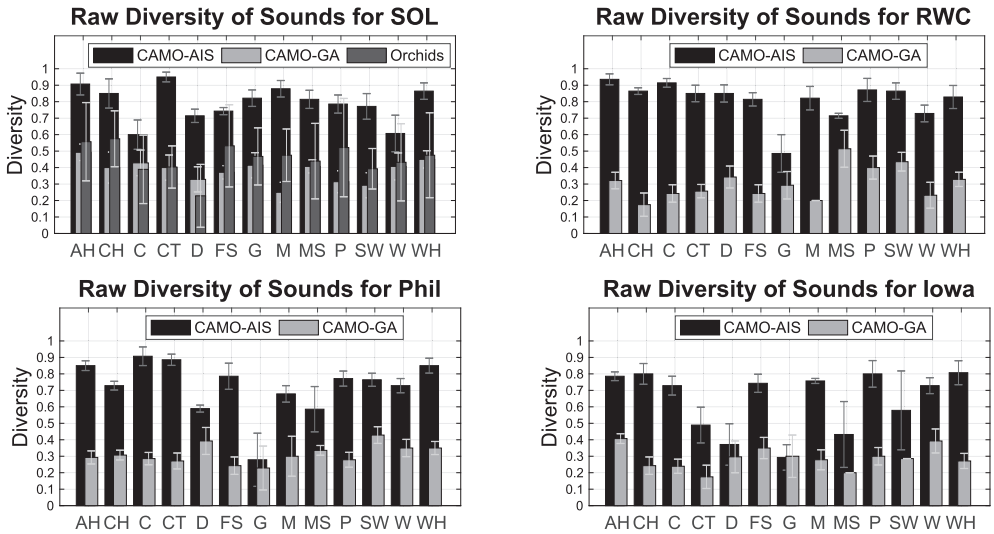



**Fig. 9.** Raw diversity of sounds for all $Q = 8$ orchestrations found for each reference sound using the following sound databases: SOL, RWC, Philharmonia, and Iowa. The labels stand for the following reference sounds: AH (air horn), CH (car horn), C (carnatic), CT (choir tibetan), D (didgeridoo), FS (factory siren), G (glass), M (minimoog), MS (musical saw), P (purr), SW (scream woman), W (waterphone), WH (windharp).

objective similarity in Section 6.2.1 and perceptual similarity in Section 6.2.2. All the data resulting from the experiment is available at https://doi.org/10.5281/zenodo.2533264.

### 6.1. Diversity

#### 6.1.1. Diversity in the search space

Diversity in the search space is estimated with eq. (19) for different sounds and different instruments. Fig. 9 compares the raw diversity of sounds and Fig. 10 compares the completed diversity for the orchestrations returned by CAMO-AIS, CAMO-GA, and Orchids. The bars represent the average values of $\hat{\epsilon}(r)$ and the whiskers are the standard deviation of $\hat{\epsilon}(r)$ around the mean. The mean and standard deviation are summary statistics used as a visual aid to simplify the comparison. Similarly, Fig. 11 compares the raw diversity of musical instruments and Fig. 12 compares the completed diversity of musical instruments.

Figs. 9 and 10 show that CAMO-AIS presented higher diversity of sounds than CAMO-GA or Orchids for all reference sounds using all databases. With SOL, CAMO-GA and Orchids present a similar diversity.

Figs. 11 and 12 show that CAMO-AIS presented higher diversity of instruments than CAMO-GA or Orchids for all references with RWC, Philharmonia, and Iowa. With SOL, Orchids has a higher average diversity for 5 of the 13 references with overlapping standard deviations.

There is no notable difference between *raw* and *completed* diversity for CAMO-AIS or CAMO-GA, revealing that most orchestrations returned sounds allocated to $M = 5$ players. However, for Orchids, the completed diversity tightens the standard deviation around the mean, which is also raised in some cases. For example, AH, D, FS, and P in Figs. 9 and 10. The same trend appears in Figs. 11 and 12. A higher completed diversity for Orchids is indication that several orchestrations found had fewer sounds (or instruments) than the maximum of $M = 5$.

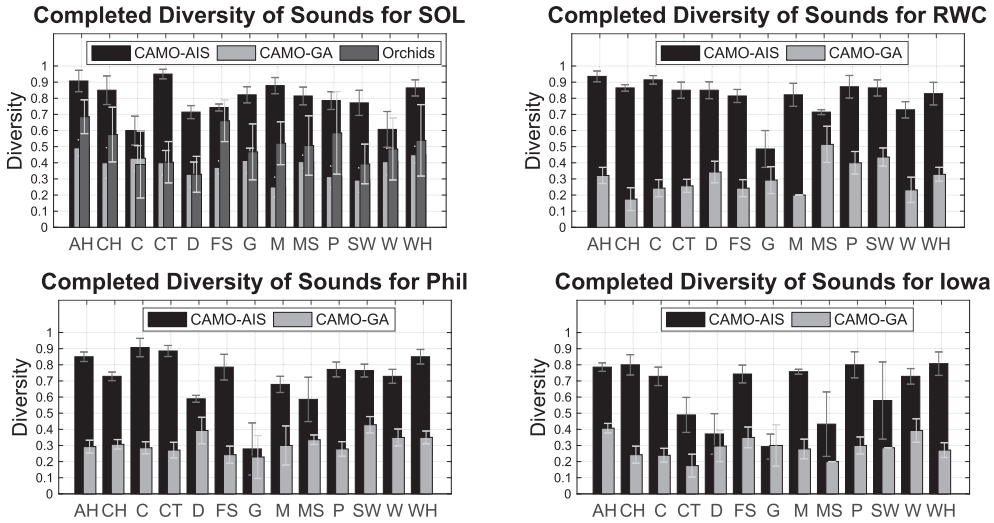



**Fig. 10.** Completed diversity of sounds for all $Q = 8$ orchestrations found for each reference sound using the following sound databases: SOL, RWC, Philharmonia, and Iowa. The labels stand for the following reference sounds: AH (air horn), CH (car horn), C (carnatic), CT (choir tibetan), D (didgeridoo), FS (factory siren), G (glass), M (minimoog), MS (musical saw), P (purr), SW (scream woman), W (waterphone), WH (windharp).

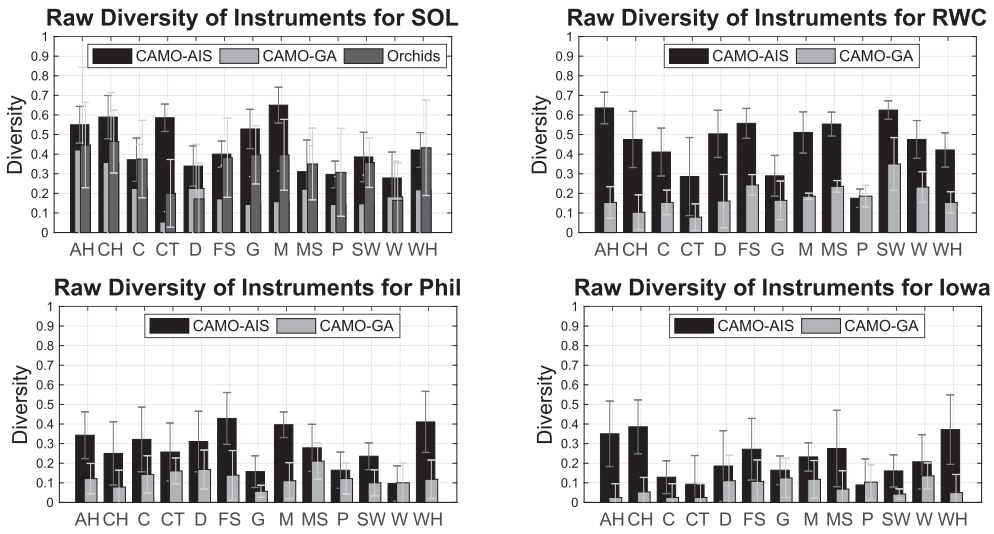



**Fig. 11.** Raw diversity of instruments for all $Q = 8$ orchestrations found for each reference sound using the following sound databases: SOL, RWC, Philharmonia, and Iowa. The labels stand for the following reference sounds: AH (air horn), CH (car horn), C (carnatic), CT (choir tibetan), D (didgeridoo), FS (factory siren), G (glass), M (minimoog), MS (musical saw), P (purr), SW (scream woman), W (waterphone), WH (windharp).

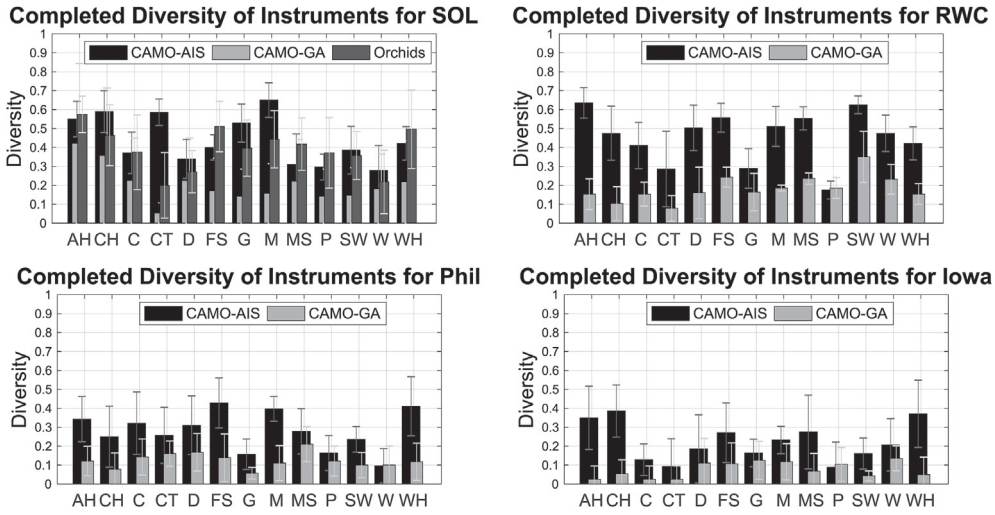



**Fig. 12.** Completed diversity of instruments for all $Q = 8$ orchestrations found for each reference sound using the following sound databases: SOL, RWC, Philharmonia, and Iowa. The labels stand for the following reference sounds: AH (air horn), CH (car horn), C (carnatic), CT (choir tibetan), D (didgeridoo), FS (factory siren), G (glass), M (minimoog), MS (musical saw), P (purr), SW (scream woman), W (waterphone), WH (windharp).

#### 6.1.2. *Diversity in the feature space*

Diversity in the feature space is measured with $\varepsilon$ from eq. (20). The property of maintenance of diversity of opt-aiNet means that CAMO-AIS returns multiple solutions corresponding to different local optima. The suppression operation in CAMO-AIS eliminates solutions that are close together in the feature space (see Section 3.1). CAMO-GA optimizes the same fitness function as CAMO-AIS with a standard GA. Most individuals tend to the same local optimum when the standard GA converges, resulting in individuals that are closer together. Finally, the MOGLS algorithm behind Orchids operates in the objective space, approximating the Pareto frontier.

Fig. 13 shows a comparison of $\varepsilon$ for the $Q = 8$ orchestrations generated by CAMO-AIS, CAMO-GA, and Orchids for all reference sounds using the sound databases SOL, RWC, Philharmonia, and Iowa. For SOL, CAMO-AIS resulted in higher $\varepsilon$ than Orchids for most references, except W (*waterphone*). However, CAMO-GA resulted in higher $\varepsilon$ than CAMO-AIS for about half the references with SOL. Diversity in the feature space was not consistent across databases for RWC, Philharmonia, or Iowa. CAMO-AIS has higher $\varepsilon$ than CAMO-GA for most references using Philharmonia, whereas CAMO-GA has higher $\varepsilon$ than CAMO-AIS for most references for Iowa. Neither achieved higher $\varepsilon$ for most references using RWC.

### 6.2. *Similarity*

#### 6.2.1. *Objective similarity*

The fitness value $F$ can be used as measure of objective similarity for both CAMO-GA and CAMO-AIS. $F$ measures the distance between each orchestration $c_q$ (where $q$ is the index of the orchestration) and the reference sound $r$ that $c_q$ approximates. Therefore, $F$ reflects the proximity of $c_q$ to $r$ such that a lower $F$ tells the composer that $c_1$ is closer to $r$ than $c_2$, for example. The calculation of the measure of objective similarity for Orchids requires the final weights $\alpha$ and the distances $D$. However, Orchids does not allow access to either $\alpha$ or $D$. Therefore, we have performed a listening test to compare the perceptual similarity between CAMO-AIS and Orchids. Section 6.2.2 presents the results of the listening test.

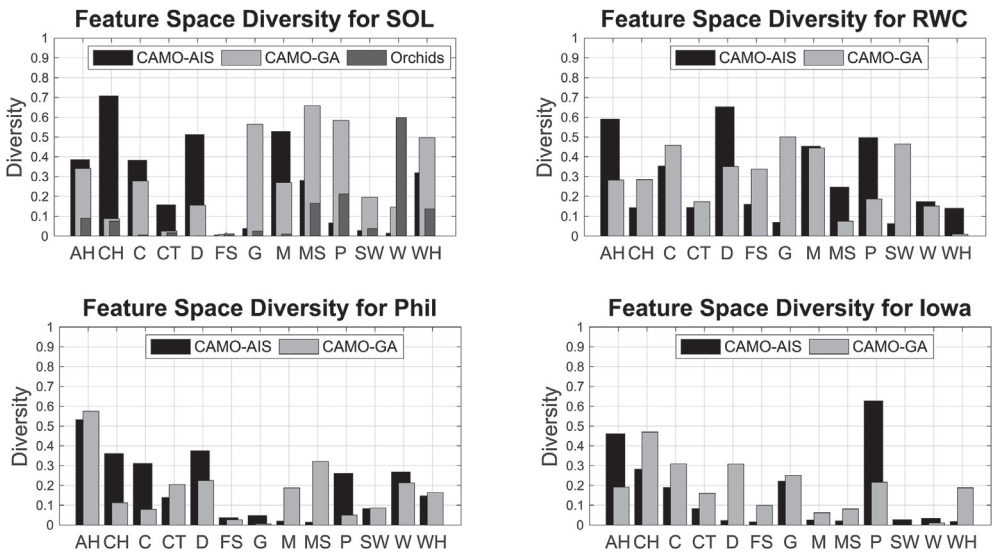



**Fig. 13.** Diversity in feature space for all $Q = 8$ orchestrations found for each reference sound using the following sound databases: SOL, RWC, Philharmonia, and Iowa. The labels stand for the following reference sounds: AH (air horn), CH (car horn), C (carnatic), CT (choir tibetan), D (didgeridoo), FS (factory siren), G (glass), M (minimoog), MS (musical saw), P (purr), SW (scream woman), W (waterphone), WH (windharp).

Fig. 14 shows the fitness values for all $Q = 8$ orchestrations returned by CAMO-AIS (top) and CAMO-GA (bottom) for each reference sound using the four musical instrument sound databases, namely SOL, RWC, Philharmonia, and Iowa. These results show that the fitness values for CAMO-GA and CAMO-AIS have a very similar pattern for each sound database because both optimize the same fitness function *F*. Comparison of the fitness of CAMO-AIS and CAMO-GA in each figure reveals similar relative values for the same reference sounds independently of the CAMO algorithm. For example, the fitness of P (*purr*) for CAMO-AIS and CAMO-GA are both around 0.3 using SOL, 0.2 using RWC and Philharmonia, and 0.5 using Iowa. However, comparison of the fitness of one CAMO algorithm for the same reference sound across databases reveals variation. For example, the fitness of P (*purr*) for CAMO-AIS varies between 0.2 and 0.5 depending on the database used. Taken together, these results show that the fitness values depend on the database used but not on the CAMO algorithm.

*6.2.2. Perceptual similarity*

CAMO-AIS resulted in more diversity than CAMO-GA and Orchids in general. So we investigated whether the increased diversity of the orchestrations affected the perceptual similarity with the reference. The listening test compared the perceptual similarity of the orchestrations from CAMO-AIS and Orchids. CAMO-GA was not included in the comparison because the listening test would become prohibitively long to perform, leading the participants to fatigue.

We averaged the similarity ratings of the participants between each orchestration and the corresponding reference prior to analysis. Then we calculated the value of Cronbach's alpha to test the internal consistency of the ratings among participants (i.e., to test whether there is agreement among participants about the similarities). The value of Cronbach's alpha obtained for most references was higher than 0.8, indicating good agreement. The only exceptions were for references *carnatic*, *factory siren*, and *scream woman*, for which agreement was weak.

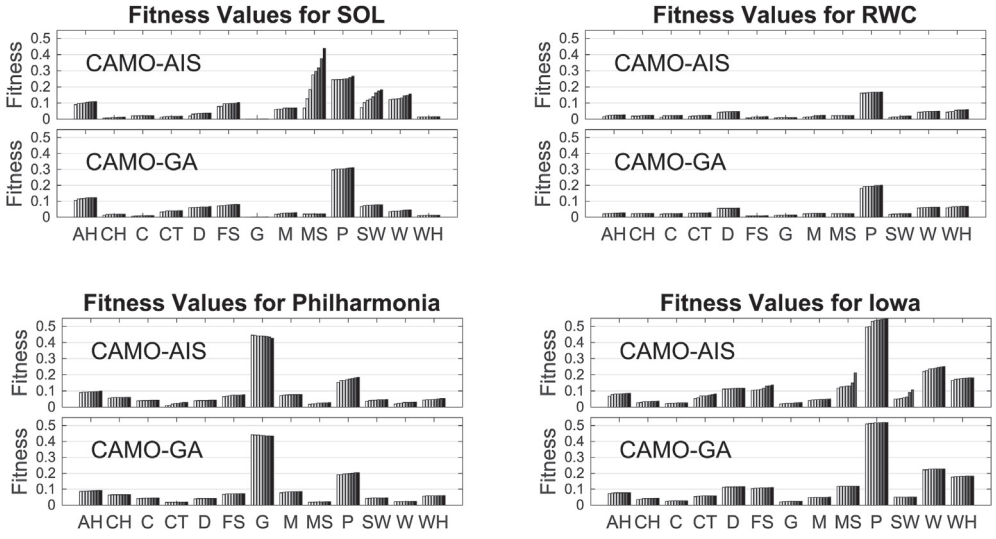



**Fig. 14.** Fitness values of CAMO-AIS (top panel) and CAMO-GA (bottom panel) for all $Q = 8$ orchestrations found for each reference sound using the following sound databases: SOL, RWC, Philharmonia, and Iowa. The labels stand for the following reference sounds: AH (air horn), CH (car horn), C (carnatic), CT (choir tibetan), D (didgeridoo), FS (factory siren), G (glass), M (minimoog), MS (musical saw), P (purr), SW (scream woman), W (waterphone), WH (windharp).

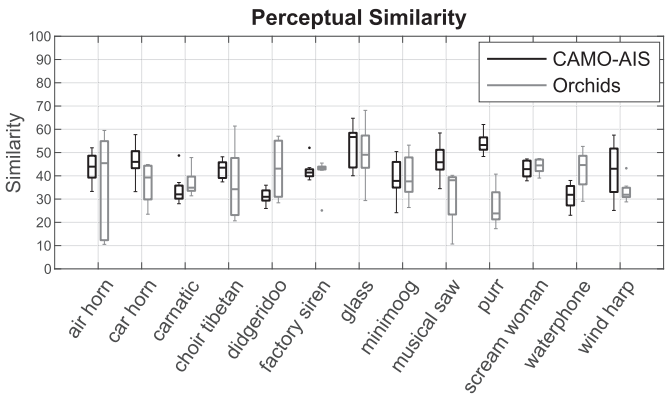



**Fig. 15.** Perceptual similarity of the Orchestrations for each of the 13 reference sounds for both CAMO-AIS and Orchids.

Fig. 15 presents a comparison of the boxplots of the mean dissimilarities for the orchestrations of the two methods for each of the 13 reference sounds. Each box is bounded by the 0.25 percentile at the bottom and 0.75 percentile at the top. The median is indicated by a horizontal line and the whiskers show the entire range of ratings, except for the points considered outliers represented by dots. We applied the Shapiro-Wilk normality test to the mean dissimilarity ratings of all orchestrations for both methods to test if their distributions can be approximated by a normal (Gaussian) distribution. Several orchestrations from both methods failed to pass the test (at significance level, $p = 0.05$). In addition, a Levene's test between all 13 pairs of orchestrations showed that pairs 1, 4, 5 and 13 did not have equal variance (at $p = 0.05$ level). Therefore, we employed the non-parametric Wilcoxon signed-rank test, which does not assume data normality, to examine whether differences between the medians of the distributions presented in Fig. 15 are statistically significant.

Table 2 presents the results of the Wilcoxon signed-rank test which indicated that a statistically significant difference between median dissimilarities appears in 3 out of the 13 reference sounds, highlighted in bold in Table 2. CAMO-AIS presented a higher overall similarity (i.e., higher median) compared to Orchids for *musical saw* and *purr*, whereas Orchids presented a higher overall similarity for *waterphone*. All the other reference sounds did not feature a significant difference in overall similarity between CAMO-AIS and Orchids, indicating that both methods generate orchestrations that were considered as perceptually similar for these references. Therefore, the results of the listening test show that CAMO-AIS returns orchestrations that are perceived as similar to the reference as those by Orchids. Taken together, the results of the diversity and similarity analyses support the conclusion that CAMO-AIS has higher diversity of orchestrations than Orchids without loss of perceptual similarity.

**Table 2**
Results of the Wilcoxon signed-rank test of the mean dissimilarities between CAMO-AIS and Orchids for the 13 reference sounds. Bold p values ($\leq 0.05$) indicate a statistically significant difference between the mean ranks of the two groups. The last column shows the actual difference between medians where a positive number indicates CAMO-AIS > Orchids and vice versa.

| Reference | T | z | p value | Median difference |
|---|---|---|---|---|
| air horn | 22 | 0.6 | 0.57 | −1.5 |
| car horn | 27 | 1.3 | 0.21 | 6.7 |
| carnatic | 10 | −1.1 | 0.26 | −2.9 |
| choir tibetan | 26 | 1.1 | 0.26 | 9.3 |
| didgeridoo | 8.5 | −1.3 | 0.18 | −12.1 |
| factory siren | 16 | −0.3 | 0.78 | −2.0 |
| glass | 24 | 0.8 | 0.40 | 7.6 |
| minimoog | 20 | 0.3 | 0.78 | 0.2 |
| musical saw | 36 | 2.5 | **0.01** | 7.7 |
| purr | 36 | 2.5 | **0.01** | 29.4 |
| scream woman | 12 | −0.8 | 0.40 | −1.6 |
| waterphone | 1 | −2.4 | **0.02** | −12.8 |
| wind harp | 30 | 1.7 | 0.09 | 11.2 |

## 7. Discussion

While our analysis demonstrates that CAMO-AIS is capable of generating multiple orchestrations that are similar to the reference sound with diversity, it is also important to consider the inherent temporal aspect of sound perception. We used the average of the feature values across the duration of the sounds, not taking the temporal variations of these features into consideration. Therefore, this work was restricted to *static* orchestrations, which are not suitable for reference sounds that present temporal variation. Reference sounds with a high degree of temporal variation require a fitness function that encodes temporal variations of the features. Additionally, reference sounds with temporal variation are expected to pose a greater challenge to orchestrate using *static* notes of musical instrument sounds. However, most musical instruments from an orchestra can be played with temporal variations, such as *glissando* or *vibrato*. Thus it seems natural to use reference sounds that vary in time and to make future effort to orchestrate references with inherent temporal variation.

In particular, the attack time is not included among the features used to match the references. However, the attack time is the most salient feature in dissimilarity studies and should be considered when searching for orchestrations, especially when orchestrating percussive reference sounds such as a *gong*. Naturally, orchestrating percussive references with percussive musical instrument sounds such as piano notes of plucked violin strings should give better results and should definitely be pursued in the future.

The perceptual similarity between the orchestrations and the references depends not only on the features used but also on the different weights given to each feature. This is probably the major difference between CAMO-AIS and Orchids in this work. CAMO-AIS uses a single-objective approach with fixed weights whereas Orchids uses a multi-objective approach where each solution corresponds to a different set of weights. Combinations with a good match of partials with the reference are perceived as having similar pitch, especially when the $f_0$ is close. Thus the fitness function in this work emphasizes the match of the partials $\vec{A}$ more than the other features used. However, CAMO-AIS generated orchestrations with larger dissimilarities in pitch than those generated by Orchids in general. Therefore, a different similarity measure for $\vec{A}$ might improve the perceptual match for pitch.

The experiments were performed with a fixed set of weights for CAMO-AIS (and CAMO-GA), so naturally the results and conclusions are restricted to the experimental conditions of this work. Other combinations of weights should be tested and compared to extrapolate these results. However, the experimental procedure adopted in this work must be adapted to test other combinations of weights because of the costly listening test. The same applies for the parameters of the algorithms. We used default parameter values for both opt-aiNet and Orchids under the assumption that they are appropriate for the problem at hand. Once again, a parameter tuning experiment would require a prohibitively costly listening test to fine-tune the parameters to obtain maximal perceptual similarity.

In music, timbre is traditionally associated with the musical instrument producing the sound. Combinations of different instruments typically result in complex timbral blends. In general, Orchids resulted in orchestrations closer in pitch than CAMO-AIS but with less diversity measured both in the search space (i.e., combinations) and in the feature space. Most orchestrations returned by Orchids had fewer players than $M = 5$ and repeated sounds. In some cases, Orchids returned groups of orchestrations with the same instruments except for one. Most

orchestrations for CAMO-AIS allocated instruments to all $M = 5$ players, resulting in more complex instrumental combinations, which, in turn, render more complex timbral blends. Therefore, the perceptual diversity for CAMO should rely heavily on timbre.

The measure of objective diversity in the search space developed for this work is based on fundamental concepts of set theory. One of the consequences is the different values of *raw diversity* and *completed diversity* used in the evaluation. *Raw* diversity uses the combinations returned by CAMO-AIS and Orchids directly, while *completed* diversity corresponds to filling with 0 the blank spaces left whenever no instrument is allocated to a player. We decided to include both the *raw diversity* and the *completed diversity* in the evaluation because these provide different perspectives for comparison.

Another consequence of the set-theory-based measure of diversity is the dependence of diversity values on the total number of elements available rather than their proportions. Use of this measure results in lower diversity for instruments than sounds, ignoring the intrinsic connection between the two since each instrument is capable of producing several sounds. A lower diversity value for musical instruments than for musical instrument sounds simply reflects this property of the measure. Thus the only reliable comparison is always between methods with everything else fixed. For example, *raw* diversity of instruments for CAMO and Orchids is comparable.

## 8. Conclusions and future perspectives

Computer-Aided Musical Orchestration (CAMO) methods can help composers find combinations of musical instrument sounds that approximate a reference sound when played together. Composers usually have subjective criteria other than only similarity with the reference when searching for an orchestration [7]. Therefore, CAMO methods that return multiple orchestrations in parallel provide alternatives for the composer. Diversity of orchestrations is very important to provide aesthetic options. In this work, we proposed CAMO-AIS, which uses a multi-modal artificial immune system (AIS) called opt-aiNet to search for orchestrations. The characteristic of maintenance of diversity of opt-aiNet resulted in multiple orchestrations that are considered similar to the reference but that are different among themselves.

We generated 8 orchestrations for 13 reference sounds with CAMO-AIS, with CAMO-GA using a standard genetic algorithm (GA), and with the state-of-the-art system Orchids and compared the results in terms of diversity and similarity. We used several measures of diversity in the search space (of the combinations) and in the feature space to evaluate diversity and we conducted a listening test to evaluate the perceptual similarity of the orchestrations. CAMO-AIS resulted in higher diversity in the search space than both CAMO-GA and Orchids for most references. CAMO-AIS was more diverse than Orchids in the search space for most reference sounds tested. The results of the listening test did not show a statistically significant difference in similarity between CAMO-AIS and Orchids for most references used. Overall, CAMO-AIS generated orchestrations that were considered just as perceptually similar to the references used as those generated by Orchids but with higher diversity. Thus, CAMO-AIS provides more options for the composer without loss of perceptual similarity. So the diversity from CAMO-AIS does not sacrifice the similarity with the reference sound.

This work uses the original opt-aiNet in a proof-of-concept implementation of CAMO-AIS. A natural next step toward future developments of CAMO-AIS is the application of improved versions of opt-aiNet [36–38] published recently. These publications feature algorithmic improvements over the original opt-aiNet targeting specific optimization domains. Continuous optimization [36] would require adaptation to the discrete and combinatorial nature of the representation adopted in CAMO, thus it seems more appropriate to directly use the combinatorial optimization version [38]. However, these algorithms [36,37] were developed for single-objective optimization (SOO) problems. An interesting alternative would be to approach CAMO as a multi-objective optimization (MOO) and use the appropriate AIS [38], following the approach used by Orchids. Alternatively, other optimization algorithms that return multiple local optima such as brain storm optimization [39] or the quantum-inspired immune clonal algorithm [40] can also result in orchestrations with diversity. Finally, multimodal deep learning [41] has the potential to tackle the multidimensional nature of timbre in CAMO. However, the challenge of attaining diversity of orchestrations would be added to the well known difficulty of interpretability of deep learning.

The application of more sophisticated measures of objective diversity in future work has potential to improve further CAMO-AIS. For example, the measure of diversity in the search space developed for this work uses concepts from set theory rather than statistics, as is the tradition in population biology. A natural source of inspiration is swarm and evolutionary computation [42–44], especially multi-objective optimization [45–47], where the concept of diversity has been extensively studied. An appropriate measure of diversity could be used in the affinity calculation to maximize diversity of the combinations directly.

Finally, perceptual diversity lies at the core of CAMO. The perceptual evaluation of diversity still poses a challenge due to the combinatorial nature of the task. In addition, the assumption that orchestrations that are more similar to the reference sound are aesthetically more appropriate could also be investigated.

## Acknowledgments

This work has been partially funded by the Portuguese foundation for science and technology FCT under grant “SFRH/BPD/115685/2016”, by the Greek State Scholarships Foundation (co-funded by the European Social Fund) under grant “Subsidy for post-doctoral researchers”, contract number: “2016-050-050-3-8116”, and by Ministerio de Economía y Competitividad of the Spanish Government under Project No. “TIN2016-75866-C3-2-R”. Part of this work has been done at Universidad de Málaga, Campus de Excelencia Internacional Andalucía Tech. The authors would like to thank Matthew Davies for kindly helping with Fig. 1.

## Appendix A. Additional information about CAMO-AIS

The following tables show the size of the subspace $\widehat{S}^r$ for each reference, the total number of possible combinations in $\widehat{S}^r$ with $M$ players, and the total number of orchestrations returned by CAMO-AIS per reference sound, which is the number of local optima found, for the four musical instrument sound databases used.

**Table A.3**
Size of the subspace $\widehat{S}^r$, total number of possible combinations in $\widehat{S}^r$ with $M = 5$ players, and the total number of orchestrations found for each reference sound for SOL.

| Reference | Subspace | Total Combinations | Orchestrations |
|---|---|---|---|
| air horn | 3871 | 7.22e+15 | 106 |
| car horn | 2205 | 4.32e+14 | 104 |
| carnatic | 4056 | 9.13e+15 | 10 |
| choir tibetan | 2656 | 1.10e+15 | 94 |
| didgeridoo | 1632 | 9.59e+13 | 62 |
| factory siren | 662 | 1.04e+12 | 23 |
| glass | 461 | 1.70e+11 | 127 |
| minimoog | 3989 | 8.40e+15 | 93 |
| musical saw | 218 | 3.92e+09 | 9 |
| purr | 2895 | 1.69e+15 | 61 |
| scream woman | 211 | 3.32e+09 | 23 |
| waterphone | 265 | 1.05e+10 | 17 |
| wind harp | 3475 | 4.21e+15 | 105 |

**Table A.4**
Size of the subspace $\widehat{S}^r$, total number of possible combinations in $\widehat{S}^r$ with $M = 5$ players, and the total number of orchestrations found for each reference sound for RWC.

| Reference | Subspace | Total Combinations | Orchestrations |
|---|---|---|---|
| air horn | 9784 | 7.46e+17 | 169 |
| car horn | 3663 | 5.48e+15 | 180 |
| carnatic | 9117 | 5.24e+17 | 190 |
| choir tibetan | 6142 | 7.27e+16 | 124 |
| didgeridoo | 4812 | 2.15e+16 | 167 |
| factory siren | 2978 | 1.95e+15 | 132 |
| glass | 528 | 3.36e+11 | 64 |
| minimoog | 7597 | 2.11e+17 | 167 |
| musical saw | 454 | 1.57e+11 | 34 |
| purr | 6430 | 9.15e+16 | 135 |
| scream woman | 452 | 1.54e+11 | 98 |
| waterphone | 1895 | 2.03e+14 | 77 |
| wind harp | 6809 | 1.22e+17 | 167 |

**Table A.5**
Size of the subspace $\widehat{S}^r$, total number of possible combinations in $\widehat{S}^r$ with $M = 5$ players, and the total number of orchestrations found for each reference sound for Phil.

| Reference | Subspace | Total Combinations | Orchestrations |
|---|---|---|---|
| air horn | 3946 | 7.95e+15 | 132 |
| car horn | 1678 | 1.10e+14 | 127 |
| carnatic | 5131 | 2.96e+16 | 209 |
| choir tibetan | 3680 | 5.61e+15 | 175 |
| didgeridoo | 1806 | 1.59e+14 | 99 |
| factory siren | 1107 | 1.37e+13 | 107 |
| glass | 335 | 3.41e+10 | 141 |
| minimoog | 3740 | 6.08e+15 | 145 |
| musical saw | 189 | 1.91e+09 | 11 |
| purr | 5709 | 5.04e+16 | 131 |
| scream woman | 335 | 3.41e+10 | 54 |
| waterphone | 640 | 8.81e+11 | 25 |
| wind harp | 3570 | 4.82e+15 | 165 |

**Table A.6**
Size of the subspace $\widehat{S}^r$, total number of possible combinations in $\widehat{S}^r$ with $M = 5$ players, and the total number of orchestrations found for each reference sound for Iowa.

| Reference | Subspace | Total Combinations | Orchestrations |
|---|---|---|---|
| air horn | 1045 | 1.03e+13 | 74 |
| car horn | 407 | 9.08e+10 | 69 |
| carnatic | 1067 | 1.14e+13 | 85 |
| choir tibetan | 721 | 1.60e+12 | 35 |
| didgeridoo | 386 | 6.96e+10 | 39 |
| factory siren | 304 | 2.09e+10 | 33 |
| glass | 65 | 8.26e+06 | 14 |
| minimoog | 926 | 5.61e+12 | 104 |
| musical saw | 50 | 2.12e+06 | 7 |
| purr | 565 | 4.71e+11 | 36 |
| scream woman | 64 | 7.62e+06 | 7 |
| waterphone | 113 | 1.40e+08 | 12 |
| wind harp | 861 | 3.90e+12 | 74 |

## References


[1] G. Nouno, A. Cont, G. Carpentier, J. Harvey, Making an orchestra speak, in: Sound and Music Computing, Porto, Portugal, 2009.
[2] R.A. Kendall, E.C. Carterette, Identification and blend of timbres as a basis for orchestration, Contemp. Music Rev. 9 (1–2) (1993) 51–67, https://doi.org/10.1080/07494469300640341.
[3] E. Handelman, A. Sigler, D. Donna, Automatic orchestration for automatic composition, in: 1st International Workshop on Musical Metacreation (MUME 2012), AAAI, 2012, pp. 43–48.
[4] F. Rose, J.E. Hetrik, Enhancing orchestration technique via spectrally based linear algebra methods, Comput. Music J. 33 (1) (2009) 32–41.
[5] S. McAdams, B.L. Giordano, The perception of musical timbre, in: S. Hallam, I. Cross, M. Thaut (Eds.), The Oxford Handbook of Music Psychology, Oxford University Press, New York, NY, 2009, pp. 72–80.
[6] Y. Maresz, On computer-assisted orchestration, Contemp. Music Rev. 32 (1) (2013) 99–109, https://doi.org/10.1080/07494467.2013.774515.
[7] G. Carpentier, E. Daubresse, M. Garcia Vitoria, K. Sakai, F. Villanueva, Automatic orchestration in practice, Comput. Music J. 36 (3) (2012) 24–42, https://doi.org/10.1162/COMJ_a_00136.
[8] D. Psenicka, SPORCH: an algorithm for orchestration based on spectral analyses of recorded sounds, in: Proceedings of International Computer Music Conference, ICMC), 2003, p. 184.
[9] T. Hummel, Simulation of human voice timbre by orchestration of acoustic music instruments, in: Proceedings of the International Computer Music Conference, ICMC), 2005, p. 185.
[10] G. Carpentier, D. Tardieu, G. Assayag, X. Rodet, E. Saint-James, Imitative and generative orchestrations using pre-analysed sound databases, in: Proceedings of the Sound and Music Computing Conference, 2006, pp. 115–122.
[11] G. Carpentier, D. Tardieu, G. Assayag, X. Rodet, E. Saint-James, An evolutionary approach to computer-aided orchestration, in: M. Giacobini (Ed.), Applications of Evolutionary Computing, Vol. 4448 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2007, pp. 488–497.
[12] 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) (2010) 681–714.
[13] G. Carpentier, D. Tardieu, J. Harvey, G. Assayag, E. Saint-James, Predicting timbre features of instrument sound combinations: application to automatic orchestration, J. N. Music Res. 39 (1) (2010) 47–61.
[14] R. Kopiez, A. Wolf, F. Platz, J. Mons, Replacing the orchestra? - the discernibility of sample library and live orchestra sounds, PLoS One 11 (7) (2016) 1–12, https://doi.org/10.1371/journal.pone.0158324.
[15] A. Antoine, E.R. Miranda, A perceptually orientated approach for automatic classification of timbre content of orchestral excerpts, J. Acoust. Soc. Am. 141 (5) (2017), https://doi.org/10.1121/1.4988156 3723–3723.
[16] A. Antoine, E.R. Miranda, Predicting timbral and perceptual characteristics of orchestral instrument combinations, J. Acoust. Soc. Am. 143 (3) (2018), https://doi.org/10.1121/1.5035706 1747–1747.
[17] V. Alluri, P. Toiviainen, Exploring perceptual and acoustical correlates of polyphonic timbre, Music Perception, Interdiscipl. J. 27 (3) (2010) 223–242.
[18] D. Tardieu, X. Rodet, An instrument timbre model for computer aided orchestration, in: Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop, IEEE, 2007, pp. 347–350.
[19] P. Esling, G. Carpentier, C. Agon, Dynamic Musical Orchestration Using Genetic Algorithms and a Spectro-Temporal Description of Musical Instruments, Vol. 6025 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2010, pp. 371–380.
[20] J. Abreu, M. Caetano, R. Penha, Computer-aided musical orchestration using an artificial immune system, in: C. Johnson, V. Ciesielski, J. Correia, P. Machado (Eds.), Evolutionary and Biologically Inspired Music, Sound, Art and Design, Springer International Publishing, 2016, pp. 1–16.
[21] J. Grey, Multidimensional perceptual scaling of musical timbres, J. Acoust. Soc. Am. 61 (5) (1977) 1270–1277.
[22] C.L. Krumhansl, Why is Musical Timbre so Hard to Understand? Struc. Percep. Electroacoust. Sound Music 9 (1989) 43–53.
[23] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, J. Krimphoff, Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes, Psychol. Res. 58 (3) (1995) 177–192.
[24] 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) (2005) 471–482.
[25] A. Jaszkiewicz, Genetic local search for multiple objective combinatorial optimization, Eur. J. Oper. Res. 1 (137) (2002) 50–71.
[26] L. de Castro, J. Timmis, An artificial immune network for multimodal function optimization, in: Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on, vol. 1, 2002, pp. 699–704.
[27] R.M. Karp, Reducibility among combinatorial problems, in: R.E. Miller, J.W. Thatcher (Eds.), Complexity of Computer Computations, Plenum Press, New York, 1972, pp. 85–103.
[28] A. Antoine, E.R. Miranda, Towards intelligent orchestration systems, in: 11th International Symposium on Computer Music Multidisciplinary Research (CMMR), Plymouth, UK, 2015, pp. 671–681.
[29] A. Antoine, E.R. Miranda, Musical acoustics, timbre, and computer-aided orchestration challenges, in: Proceedings of the 2017 International Symposium on Musical Acoustics, Montreal, Canada, 2017, pp. 151–154.
[30] E.R. Miranda, A. Antoine, J.-M. Celerier, M. Desainte-Catherine, i-berlioz, Interactive computer-aided orchestration with temporal control, in: Proceedings of the 5th International Conference on New Music Concepts (ICNMC 2018), ABEditore, Treviso, Italy, 2018.
[31] M. Goto, H. Hashiguchi, T. Nishimura, R. Oka, RWC music database: music genre database and musical instrument sound database, in: Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR), 2004, pp. 229–230.
[32] M. Goto, Development of the RWC music database, in: Proceedings of the 18th International Congress on Acoustics, 2004, pp. 553–556.
[33] A. Camacho, J. Harris, A sawtooth waveform inspired pitch estimator for speech and music, J. Acoust. Soc. Am. 124 (3) (2008) 1638–1652.
[34] I. Borg, P.J.F. Groenen, Modern Multidimensional Scaling: Theory and Applications, second ed., Springer, 2005. Springer Series in Statistics.
[35] M. Caetano, G. P. Kafentzis, A. Mouchtaris, Y. Stylianou, Full-band quasi-harmonic analysis and synthesis of musical instrument sounds with adaptive sinusoids, Appl. Sci. 6 (5). https://doi.org/10.3390/app6050127.
[36] G.P. Coelho, F.J.V. Zuben, A concentration-based artificial immune network for continuous optimization, in: IEEE Congress on Evolutionary Computation, 2010, pp. 1–8, https://doi.org/10.1109/CEC.2010.5585919.
[37] G.P. Coelho, F.J. Von Zuben, A concentration-based artificial immune network for multi-objective optimization, in: R.H.C. Takahashi, K. Deb, E.F. Wanner, S. Greco (Eds.), Evolutionary Multi-Criterion Optimization, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 343–357.
[38] G.P. Coelho, F.O. de Frana, F.J.V. Zuben, A concentration-based artificial immune network for combinatorial optimization, in: 2011 IEEE Congress of Evolutionary Computation (CEC), 2011, pp. 1242–1249, https://doi.org/10.1109/CEC.2011.5949758.
[39] L. Jiao, Y. Li, M. Gong, X. Zhang, Quantum-inspired immune clonal algorithm for global optimization, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 38 (2008) 1234–1253.
[40] S. Cheng, Q. Qin, J. Chen, Y. Shi, Brain storm optimization algorithm: a review, Artif. Intell. Rev. 46 (4) (2016) 445–458, https://doi.org/10.1007/s10462-016-9471-0.
[41] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, in: L. Getoor, T. Scheffer (Eds.), Proceedings of the 28th International Conference on International Conference on Machine Learning, Omnipress, USA, 2011, pp. 689–696.
[42] B.O. Arani, P. Mirzabeygi, M.S. Panahi, An improved pso algorithm with a territorial diversity-preserving scheme and enhanced exploration-exploitation balance, Swarm Evolution. Comput. 11 (2013) 1–15. https://doi.org/10.1016/j.swevo.2012.12.004.

[43] Q. Long, C. Wu, T. Huang, X. Wang, A genetic algorithm for unconstrained multi-objective optimization, Swarm Evolution. Comput. 22 (2015) 1–14. https://doi.org/10.1016/j.swevo.2015.01.002.
[44] R. Denysiuk, A. Gaspar-Cunha, Multiobjective evolutionary algorithm based on vector angle neighborhood, Swarm Evolution. Comput. 37 (2017) 45–57. https://doi.org/10.1016/j.swevo.2017.05.005.
[45] C.A. Coello Coello, A short tutorial on evolutionary multiobjective optimization, in: E. Zitzler, L. Thiele, K. Deb, C.A. Coello Coello, D. Corne (Eds.), Evolutionary Multi-Criterion Optimization, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 21–40.
[46] E. Zitzler, M. Laumanns, S. Bleuler, A tutorial on evolutionary multiobjective optimization, in: X. Gandibleux, M. Sevaux, K. Sörensen, V. T'kindt (Eds.), Metaheuristics for Multiobjective Optimisation, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 3–37.
[47] H. Wang, Y. Jin, X. Yao, Diversity assessment in many-objective optimization, IEEE Trans. Cybern. 47 (6) (2017) 1510–1522, https://doi.org/10.1109/TCYB.2016.2550502.