Contemporary Music Review, 2013
Vol. 32, No. 1, 99–109, http://dx.doi.org/10.1080/07494467.2013.774515

Routledge
Taylor & Francis Group

On Computer-Assisted Orchestration

Yan Maresz

In 2003, I presented to Ircam a proposal for a long-term research project on the subject of computer-assisted orchestration. The results of this research project lead to the prototype softwares, 'Orchidee' and now, 'Ato-ms', both used by numerous composers. With these systems, composers can specify a target sound and replicate it with a given, predetermined orchestra. The target sound, defined as a set of audio and symbolic features, can be constructed either by analysing a pre-recorded sound, or through a compositional process. The orchestration procedure uses large pre-analysed instrumental sound databases to offer composers a set of sound combinations. This procedure relies on a set of features that describe different aspects of the sound. Almost 10 years after the start of this project, it is time to look back at what was accomplished from the musical stand point, and to open some new perspectives on the subject, like the introduction of target descriptors for orchestral qualities, and orchestral layers.

Keywords: Orchestration; Computer-Assisted Orchestration; Dynamic Musical Orchestration; Computer-Assisted Composition

The art of orchestration is without doubt one of the hardest musical disciplines to define and transmit in a satisfactory manner. Its teaching and practice is still very much based in tradition, and the path from musical notation to acoustic realization activates an important number of variables, hardly quantifiable and remaining highly unpredictable except in classical musical styles. This may be why orchestration, a field of the purest imagination at the crossing of daring and experience, is still a highly empirical activity like no other in musical writing. In this domain, we still live from our heritage. And if some composers have surpassed it, we must admit that orchestration is still too often approached in quite an archaic manner in the age of computer music, and that a rational and scientific approach to it is still to be achieved.

Unlike other aspects of composition (harmony, rhythm, etc. …) which have been highly present in computer-assisted composition in the past 20 years, orchestration still remains quite unexplored from that point of view. Computer-aided orchestration, and sound/texture pre-calculation, had been in the mind of many composers since the early days of computer-aided composition, but in the end, most experiments and novelties in that area were made quite empirically, apart from the work on spectrum and timbre done by the composers of the spectral school like Tristan Murail or Gérard Grisey (Anderson, 2000).

In 2003, I proposed to the research teams at Ircam to investigate the possibility of making a system to address the specific problems of orchestration, or rather a particular case of orchestration, in between instrumentation and orchestration, which is an overall timbre matching or inverse instrumental synthesis. In terms of orchestration, it is a simultaneous equilibrium out of an orchestral context. Here, equilibrium is not to be understood as in dynamic balance but rather in timbral terms. It is a unique and autonomous acoustic layer. This happened after the experience of my piece, 'Metal Extensions'[1] which is an extended orchestration of 'Metallics', an earlier piece for trumpet and live electronics. This project was also linked to a very ambitious project of state-of-the-art, collaborative on-line instrumentation and orchestration treatise that I wished to put together at the time (Maresz, 2006).

While writing 'Metal Extensions', and in order to facilitate the orchestration process of the electronic part, my workflow was limited to a simple set of techniques which can be described as follows: selection of the region of sound to orchestrate from the electronic sound file, placement by hand of markers on the regions within the sound file that interested me, for a chord-sequence analysis with AudioSculpt (peaks),

Figure 1 Excerpt from *Metal Extensions, Bars 223–226*.

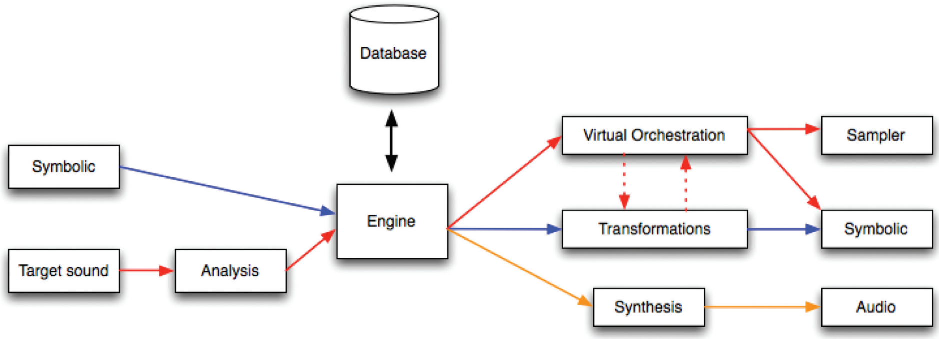

Figure 2 First Schematic Draft of the Orchestration Software.

inharmonic partial analysis on the totality of the sound file in the same programme, transcription of the given results into symbolic notation in OpenMusic and finally, realization of the final score by hand. This somehow naïve technique allowed me to approach my targets 'harmonically', and in some cases to break the more complex and noisy sounds into nodal bands by observation of the overall frequency envelopes. The choice of instruments to match the overall colours was done by analogy, in the traditional manner (Figure 1).

Not being able to capture the quality of the timbre by analysis was obviously frustrating, since the multidimensional aspect of it was not taken into consideration. Nevertheless, I still consider this orchestration quite successful. While writing the work, I found many situations where I thought that my orchestration could have been improved, especially in the realization of complex electronic timbres. In other words, could the computer give me a higher starting point for my orchestration; a proposal closer to the sound I needed to score for a given problem. My proposal to Ircam happened at a time when many of the technologies needed to make such a system were coming to maturity: large sample databases, analysis methods, computer-aided composition environments and so on.

I started by presenting a naïve draft vision for such a system, where the user could give a sound target,[2] and receive as a result, a proposal made out of mixtures of instrumental sounds that spectrally 'sound' as much as possible like the input sound target (Figure 2). A symbolic input of a pre-existing orchestration was also included as a possible target to be processed by high-level descriptions of the desired result (i.e. transform into 'metallic', 'thin', 'granular', 'dark', etc.). These characteristics could be defined by the user or 'gradually learned' by the system as the composers tagged his results while using the software, since all composers may have different terms for their sonic palette.

Orchidée

Along with other composers and researchers, we started to narrow down this very ambitious and complex problem into smaller units, as I was compiling the sound

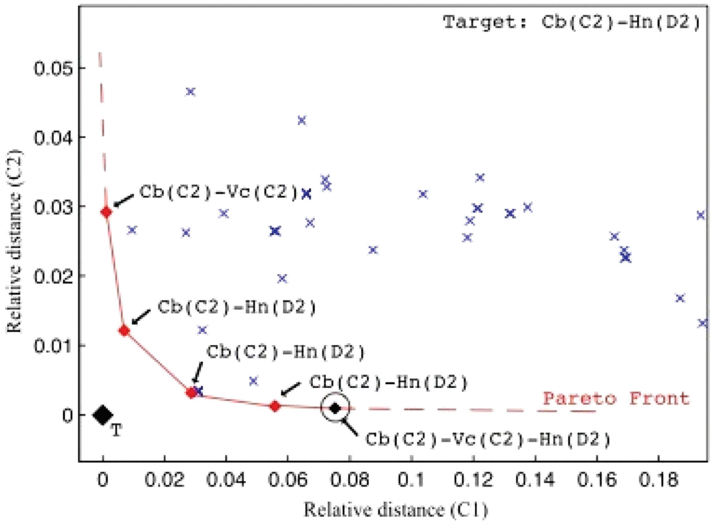

Figure 3 Pareto Front Resulting from Bi-criteria Optimization.

database of orchestral samples and extended techniques from multiple sources that would represent our orchestral knowledge.

Immediately, two main aspects had to be addressed. The first one had to do with the analysis of instrumental sound and its perception, and the second one with the explosion of the combinatorial complexity of the system. Two PhD theses were made at IRCAM on these subjects, the first one by Damien Tardieu in the Analysis/

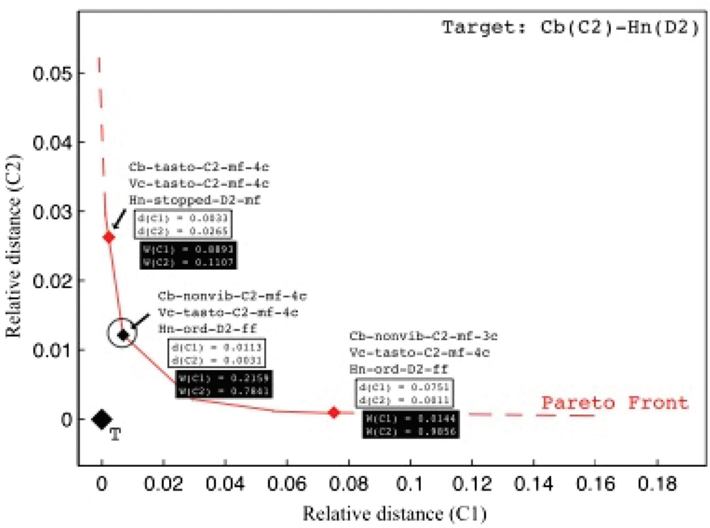

Figure 4 Further Exploration of a Combination, and Deduction of User's Preference.

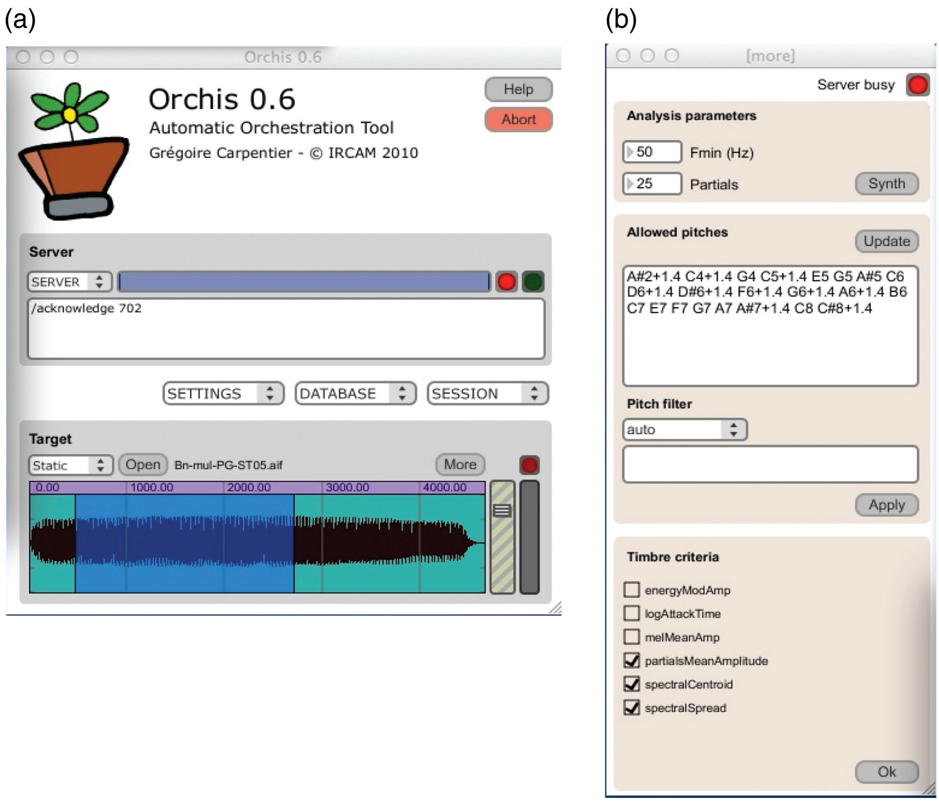

Figure 5 Orchidée/Orchis[8] Workflow.
Target selection (a) in Orchis (client for the Orchidée server) and target analysis and descriptors choices (b) to be included in the solution search procedure. Note the presence of the harmonic filter in the 'Allowed pitches section'.

Synthesis team (Tardieu, 2008), and the other by Grégoire Carpentier in the Musical Representation team (Carpentier, 2008).

By 2007, the software 'Orchidée' (Tardieu/Carpentier) was available internally at IRCAM and succeeded quite well in addressing the problem. This first system was intentionally restricted to a specific case of orchestration/instrumentation without any temporal consideration: giving a static result, in the form of a superposition of static instrumental samples, for a static harmonic target.

The system works with a certain number of audio descriptors that are extracted from the sounds (orchestra database and target) as material for combinatorial operations to match the user's target. Due to the explosion of the combinatorial complexity in the search for solutions, a few important techniques were used to reduce calculation time: a set of symbolic filters (dynamic, playing techniques, etc.) for the search engine, a harmonic filter[3] and a genetic algorithm to explore a multidimensional space of

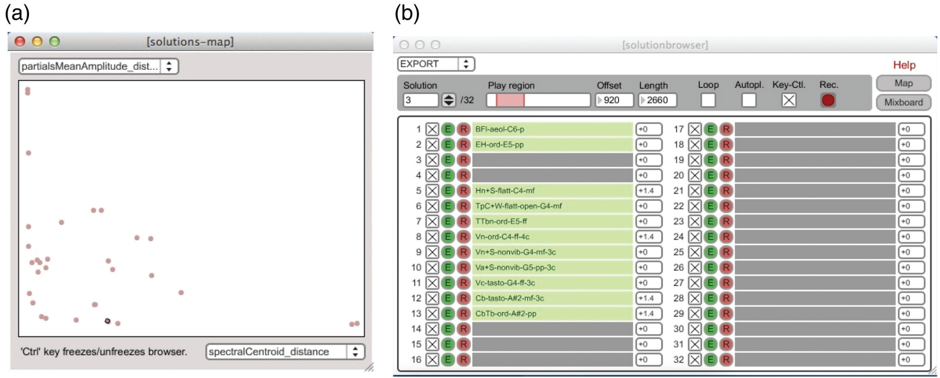

Figure 6 Solutions Displayed in a Pareto Front Representation in Orchis (Here, Two Features: Partial Mean Amplitudes and Spectral Centroid) and the Individual Sample Components for the Chosen Solution.

solutions (Tardieu, Carpentier, & Rodet, 2007). In other words, there is no best solution, but rather an optimal solution in with regards to one or another feature of the sound according to its position in a *Pareto front*[4] (Figures 3 and 4).

The harmonic filter and the Pareto front techniques, while allowing us to move forward, were also taking us away from two very important aspects: the use of non harmonic sounds and the idea that there is no best solution in the absolute. I still think that there is a best solution somewhere and that we do need a system that limits the number of results the user has to browse (since the number of results increases with the number of audio features chosen by the user) (Figures 5, 6 and 7).

Ato-ms

This system has then subsequently evolved into a completely new software, 'Ato-ms' (Abstract Temporal Orchestration), written by Philippe Esling, as the result of a third thesis on the subject at Ircam in the Musical Representation team (Figure 8).

Many developments were undertaken and included in this software. The most important are the management of time, and the use of an optimal warping algorithm, based on a combination of multiobjective and time-series matching algorithms (Esling & Agon, in press). Another useful feature is the possibility of creating abstract targets by designing envelopes for any audio feature.

Both systems, despite their use with great success by many composers, are not completely satisfactory from a musical standpoint. 'Orchidée' has no time dimension and only allows the 'treatment' of sounds of harmonic nature. It can be very useful for imitating the sound of an instrument by other instruments, or to approach the timbralic signature of a more complex sound. But in the case of a non-harmonic, or worse, noisy target, the system is inoperative. The harmonic paradigm is a brick wall that has to be overcome.

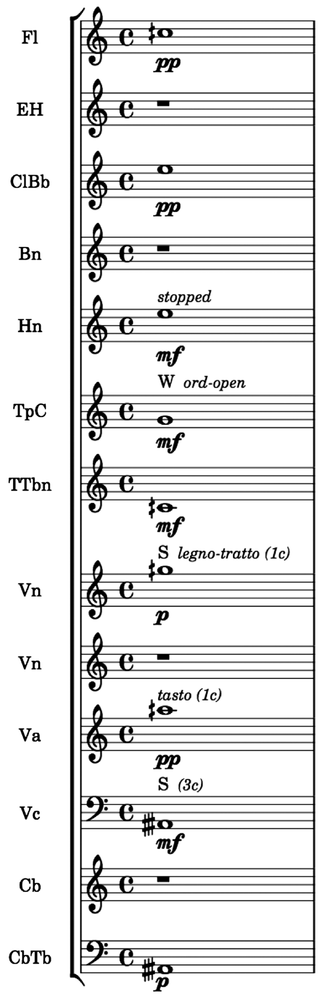

Figure 7 A Result from Orchidée, Automatically Exported into Lilypond. Here, a Bassoon Multiphonic Orchestrated for a Small Ensemble. The Score is in C.

'Ato-ms' has proven to be efficient in finding time profiles for any audio feature within large database of audio samples and provides timelines for an orchestration simulation. But in the normal use of the software, the resulting orchestrations suffer from a lack of quality in their timbral accuracy. Also, it is no longer possible to get simple 'Orchidée-like' static results for a static target unless by getting around 'Ato-ms' time-series matching algorithms. Lastly, both systems only address a particular case of an orchestration problem, timbre matching. They do not yet consider yet the orchestration concept of orchestral layers that can be seen as the crucial perspective on the sonic space of the mixed instrumental sounds.

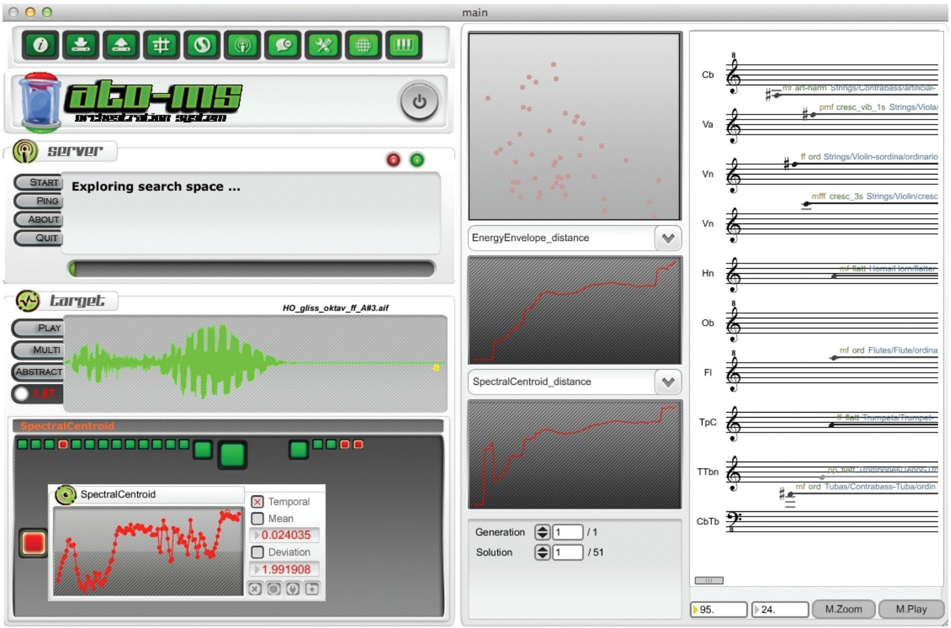

Figure 8 Ato-ms Interface for Automatic Orchestration.

In the past, much progress was made on the efficiency of algorithms for orchestration to include better temporal representations and assure better combinatory. But there is now a great need to work more on consolidating the core of the system, which is to provide good, convincing and reliable orchestration proposals. The most important task that needs to be dealt with is the definition of the audio descriptors (or their combination) required to address the problem of analysis for certain categories of sounds that must be included in the system's instrumental knowledge. This concerns multiple sounds, inharmonic sounds, multiphonics, percussive sounds, noisy sounds and nodal sounds. Since the system refers to a sample database representing instrumental knowledge in order to calculate the solutions, we have no sounds other than monophonic and harmonic ones in our results for the moment. To this must be added the need for better, more robust addition functions used to predict the mixture of descriptors.

From the listener's standpoint, there is also a great need to understand the purpose for which the rendering of a solution is calculated by the system. Indeed, we use rather close microphone position samples in our database, but the effect of orchestration is also really the effect of the fusion of timbres at a certain listening distance. The rendering of solutions should take this into account in the computation, and be user accessible for its 'conditions' parameters.[5] Others aspects that need to be addressed are the prediction of the emergence phenomena (especially in the attack/resonance orchestration models[6]), and a better way to explore the space of solutions in a multi-objective situation.

Orchestral Layers and Orchestral/Qualities Target Descriptors

The architecture we have been working on so far actually addresses only one particular case of an orchestration problem which is more an 'instrumentation' problem. It is a problem that tends to be more or less local, vertical in its conception and does not necessarily involve a time dimension except for an overall amplitude envelope.

But a second aspect, that would be more of an 'orchestration' one, would deal simultaneously with the problem of instrumental mixture and its behaviour in time, but also in its 'depth' in relation to another layer of sound by acting and defining orchestral categories. Indeed, high-level orchestration is the art of combining simultaneous, differentiated sound layers, each layer with its own identity and specific parameters depending on the musical context.

To achieve this, we need to introduce descriptors of orchestral 'qualities' that could be used as targets or modification parameters for results, thus allowing them to interact with one another. Orchestrating often implies using such descriptions, like homogeneity, thickness, transparency, etc., and each of these has an effect on the relationship with the other musical layers present. Composers and orchestrators have a good practice and knowledge of this. For example: similar sounds, in terms of thickness and intensity, will blend into a simultaneous equilibrium. Differences in equilibrium form different sound layers, and their position relative to one another depends on their transparency which depends on their relative wideness and intensity. Sounds both 'thick (voluminous, or wide) and weak' tend to place themselves in the background, and 'thin and intense' sounds tend to come to the foreground. These 'empirical formulas' date back to the orchestration treatise of Charles Koechlin from 1941 (Koechlin, 1954).

With the introduction of orchestral qualities we could simply start with a symbolic target, such as the notes of a chord, define our orchestra and request different types of orchestrations by acting on these orchestral quality descriptors which become 'targets'; like, for example, by requesting a result favouring maximum *homogeneity* and the greatest possible *wideness*, or a *diaphanous*, completely opposed result. And one orchestration solution could be further transformed by acting gradually on commonly used quality descriptors such as:

- opaque (or dense) transparent (both with simultaneous action on thickness and intensity)
- thin–fat (wide, voluminous)
- weak–intense
- dark–brilliant
- homogeneous–heterogeneous
- granulous–smooth
- pure–rich–saturated
- blurred–focused

Some of these qualities are interrelated or are difficult to explicit explain in non-musical terms,[7] and a great deal of work is still to be done in this area.

Notes

[1] *Metal Extensions* (2001), Edition DURAND (D. & F. 15449), Metallics (1995), Edition DURAND (D. & F. 14806) on compact disk 'Yan Maresz, Compositeurs d'aujourd'hui', IRCAM, Ensemble Intercontemporain. *Accord Universal.*

[2] The concept of the *audio sound target* is still very important to me in the field of computer-assisted orchestration, even in time-varying situations. I think that's because the sounds I choose to become 'orchestration problems' already bear in them all the qualities that interest me musically, both in their timbre properties and their temporal shapes. It is somewhat similar to the assumption of acousmatic or concrete music, in which, while recording sound objects, the emphasis is on the close relationship between the producing gesture and the resulting sound, which, interrelated, interact to form true morphological paths. Acts of real discovery through the combination of sound materials from different origins are not made easily by an exhaustive combinatorial method in which the composer analyses and retains the results that suit him. I feel it is the same with hand-made abstract target descriptions, at least in the field of orchestration, as we are then drifting more towards the shape-matching exploration of sample databases. But audio targets are not the only possible ones; symbolic targets, as in the case of pre-existing orchestrations, could potentially be included, but it is still a perspective.

[3] Harmonic filtering allows to the restraint of the search space by automatically deriving the note used from the partials analysis

[4] A Pareto front (or Pareto frontier) is a framework for *partially* evaluating a set of actions with multidimensional outputs assuming a weak desirability partial ordering. In our orchestration context, it refers to the impossibility of evaluating all instrumental combinations for a given target sound. It is useful for reducing a set of candidates prior to further analysis given perceptive criteria.

[5] Mainly: hall type, listener distance and sitting plan of the instruments.

[6] Some work on this subject was done by Damien Tardieu at Mc Gill University in Tardieu and McAdams (2012).

[7] Koechlin himself, talking about volume and intensity: 'Let us note that in that matter, two elements are involved: the sound volume and intensity. I think it is unnecessary to define these two terms, and besides, I would be embarrassed doing so. Any musician knows what is a fat sound, an intense sound'. For Koechlin, volume is to be understood as 'volume in space', the sound's perceived size and not its loudness.

[8] *Orchis* is an Orchidée client written in Max-Msp by Grégoire Carpentier.

References

Anderson, J. (2000). A provisional history of spectral music. *Contemporary Music Review*, *9*(2), 7–22.

Carpentier, G. (2008). Approche computationnelle de l'orchestration musicale: Optimisation multicritère sous contraintes de combinaisons instrumentales dans de grandes banques de sons (Unpublished doctoral dissertation, IRCAM and University of Pierre & Marie Curie (UPMC), Paris, 2008).

Esling, Ph., & Agon, C. (in press). Multiobjective time series matching for audio classification and retrieval. *IEEE Transactions on Speech Audio and Language Processing*.

Koechlin, C. (1954). *Traité de l'orchestration*. Paris: Max Eschig.

Maresz, Y. (2006, November). Pour un traité d'orchestration au XXIe siècle, in L'Étincelle: Le journal de la creation à l'Ircam. Retrieved December 2012, from http://etincelle.ircam.fr/652.html

Tardieu, D. (2008). *Modèles d'instrumentspour l'aide à l'orchestration*. Paris: IRCAM and University of Pierre & Marie Curie (UPMC).

Tardieu, D., Carpentier, G., & Rodet, X. (2007). *Computer-aided orchestration based on probabilistic instruments models and genetic exploration*. Proceedings of International Computer Music Conference, Copenhagen, Denmark.

Tardieu, D., & McAdams, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*, *30*(2), 117–128.