# From Timbre Decomposition to Music Composition

Frédéric Le Bel
Music Representation Team, IRCAM (UMR 9912 STMS), Paris, France;
Centre de recherche en informatique musicale, MUSIDANCE, Université Paris 8, Paris, France.
frdric.lebel@gmail.com | http://repmus.ircam.fr/lebel/

## Introduction

This abstract proposes to overview a work in progress that aims at developing a computer-aided-composition (CAC) approach to structuring music by means of audio clustering and graph search algorithms [Le Bel 2017]. Although parts of this idea have been investigated in order to achieve different tasks such as corpus-based concatenative synthesis [Schwartz, Beller et al. 2006], musical genre recognition [Peeters 2007] or computer-aided orchestration [Carpentier 2008] to name a few, the challenge remains to find a way of integrating these techniques into the composition process itself, not to generate material but to explore, to analyse and to understand the full potential of a given sound corpus (sound file database) prior to scoring a musical piece; being instrumental, acousmatic or mixed. As opposed to mainstream CAC tools, mostly focusing on generative methods, the following one proposes an analytical approach to structuring music based on auditory attributes and their physical correlates. Basically, the idea is to use unsupervised machine-learning to extract hidden structures from sound corpuses' features space and to translate them into musical structures at different stages of the composition; from the micro-form to the macro-form. From another perspective, the idea is to elaborate musical works based, not on patterns proliferation or similar techniques, but rather on relationships that bring together different sound entities. Consequently, the goal here is to unfold the algorithmic structure through a prototype software and to reveal how it was used to compose a recent piece of mine: '*Il ne s'agit pas de le recomposer*', for augmented string quartet and fixed media [Le Bel 2017], in order to examine the methodology, to discuss a few important co-lateral problematics, to expose the limitations of such an approach and finally to share ideas for future developments.

## Structural overview

Based on a typical unsupervised machine-learning architecture, the following algorithmic structure may be divided into three distinct stages. The first one focuses on data extraction (audio features extraction), the second one focuses on data analysis (features selection, clustering algorithm and evaluation criterion) and the third one focuses on data sorting (graph search algorithm). Essentially, the task is to deduce a musical structure from a sound file database (audio and not symbolic).

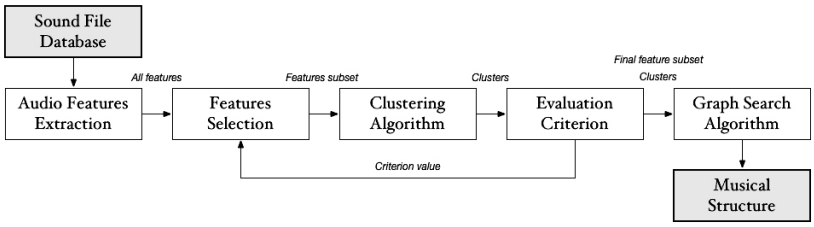


**Figure 1.** Algorithmic structure

## From timbre decomposition… (data extraction)

In the frame of this work, the audio features extraction consists of decomposing each sound file from the database by mapping the signal's short-term Fourier transform (STFT) magnitude into a lower-dimensional domain that more clearly reveals the signal characteristics. Assumed to represent specific auditory

attributes, these features inform different aspects of the temporal structure, the energy envelope, the spectral morphology and the harmonic content of sounds. From these low-level data models, the sound files are then compared, taken pairwise, on a dissimilarity/distance basis in order to generate what could be seen as a *n*-dimensional timbre-space upon which the clustering algorithm can be later applied. Briefly, the evaluation criterion for the features selection aims at maximizing the inter-clusters distances and at minimizing the intra-clusters' ones [Dy and Broadley 2004]. In this particular case, the timbre-space metaphor should be considered as an exploratory structure of dissimilarity data rather than a comprehensive perceptual coordinate system of timbre [Siedenburg, Fujinaga and McAdams 2016].

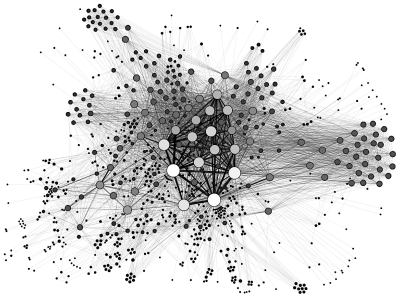

**Figure 2.** Two-dimensional timbre-space network,
where the nodes represent the sounds and the edges the distances between them.

## …To musical structure (data sorting)

Following an extensive data analysis stage (the features selection and the clustering algorithm), the graph search consists of finding path(s) within the timbre-space that could suggest way(s) of sequencing the different clusters of sounds and their components in a chronological order. Considering that the resulting network may be seen as a complete, undirected and weighted graph, multiple approaches could be used but spanning trees were investigated first. More precisely, the minimum spanning tree (MST) seemed to provide an interesting premise accordingly with the timbre-space's underlying dissimilarity data structure. A MST is a subset of the edges of an undirected and weighted graph that connects all the nodes together with the minimum possible total edge weight without any cycles. In other words, it is a tree (not a path) whose sum of edge weight is as small as possible [Graham and Hell 1985]. Hence, the MST does not represent an ordered sequence of clusters but rather some sort of an optimized road map through the timbre-space from which multiple paths may be drawn. However, the clusters and their components remain connected in a way that the global similarity is maximized. The Kruskal algorithm [Kruskal 1956] was used in the frame of this work in order to obtain structures such as the following one.

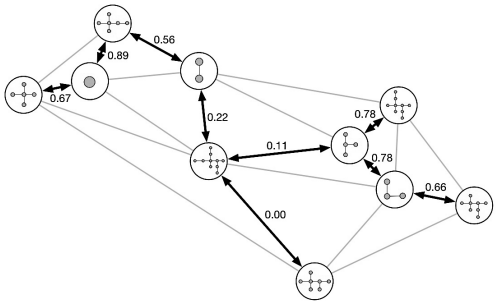


**Figure 3.** Two-layer MST extracted from the timbre-space, where the big nodes represent the clusters of sounds, the small nodes represent the sounds and the edges on both layer represent the transition/emission probabilities.

The previous two-layer MST already suggests a rather clear musical structure, or a particular definition of it, but yet remains to be expressed in the time domain. For that, the model is translated into a hidden markov model (HMM) for which the edges, originally distance based, are converted into transition and emission probabilities respectively for each layer (global structure and local structures) according to the following principle: $1\text{-}d_i/\text{argmax}(d)$. In other words, the closer or the more similar are two sounds, or two clusters of sounds, higher is the probability to transit from one to the other and vice-versa. Finally, a customized polyphonic step sequencer is used to articulate the resulting probabilistic model on a timeline and let the musical structure be heard. My last piece: '*Il ne s'agit pas de le recomposer*', for augmented string quartet and fixed media, is entirely based on these principles and was composed as a proof of concept in order to assess, as objectively as possible, the artistic potential of such an approach to composition.

For more details: http://repmus.ircam.fr/lebel/from-timbre-decomposition-to-music-composition

## References

Carpentier, G. (2008). Approche computationelle de l'orchestration musicale: Optimisation multicritère sous contraintes decombinaisons instrumentales dans de grandes banques de sons. *Thèse de doctorat*, Université Paris VI – Pierre et Marie Curie, Paris, France.

Dy, J. G. and Broadley, C. E. (2004). Feature Selection for Unsupervised Learning. *Journal of Machine Learning Research 5*, 845-889.

Graham, R. L., Hell, P. (1985). On the History of the Minimum Spanning Tree Problem. *Annals of the History of Computing*, Vol. 7, No. 1, 43-57.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, Vol. 7, 48-50.

Le Bel, F. (2017). Structuring Music by Means of Audio Clustering and Graph Search Algorithms. *Proceedings of the Journées d'Informatique Musicales 2017*, Paris, France.

Le Bel, F. (2017). Il ne s'agit pas de le recomposer, for augmented string quartet and fixed media. *TANA String Quartet*, Premiered in Lille, France.

Peeters, G. (2007). A Generic System for Audio Indexing: Application to Speech/Music Segmentation and Music Genre Recognition. *Proc. Of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France.

Schwartz, D., Beller, G. et al. (2006). Real-time Corpus-based Concatenative Synthesis with Catart. *Proc. Of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada.

Siedenburg, K., Fujinaga, I. and McAdams, S. (2016). A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology. *Journal of New Music Research*, DOI:10.1080/09298215.2015.113273