INTRODUCTION

This poster 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 a tool 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, acoustic or mixed. As opposed to mainstream CAC approaches, the following offers a new approach to audio features extraction, clustering algorithm and evaluation criterion. The second one focuses on data extraction (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).

FROM TIMBRE DECOMPOSITION...

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 model above should be considered an exploratory structure of dissimilarity data rather than a comprehensive perceptual coordinate system of timbre [Siedenburg, Fujinaga and McAdams 2016].

A SPECTROMORPHOLOGICAL APPROACH

As described earlier, the features selection and the evaluation criterion phases may seem relatively straight forward but it appears to be one of the most difficult and most crucial step of the process for the resulting timbre space to have a consistent psychoacoustic meaning. Indeed, many different strategies can be found in both the machine learning and the psychological literature but the propositions yet remain either task specific, simply incomplete or require a lot of human inference [Lagrange et al. 2015]. Because of that, different empirical approaches were deployed in order to serve adequately the exploratory purposes of the framework and to appear to be particularly efficient. Being somehow not too generic but also not too specific, that is the spectromorphological approach (Smulders 1986). In this section, the averaged relative specific loudness (RLS) [Peeters 2004] of each pair of sounds is taken to assess their similarity.

Figure 1. Algorithmic structure

FROM TIMBRE DECOMPOSITION...

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

THE SEQUENCER

Based on a typical step sequencer, the timbre space exploration can be done in real-time manually or autonomously by controlling or pre-setting various parameters such as the tempo, the number of steps per bars (one bar corresponding to one cluster through a transition period), the length of each steps (one step corresponding to one sound file or the emission period), the amount of rhythmic variation, the velocity and the envelope of each sound file, and the polyphonic density (maximum number of superimposed voices) among others. The spatialisation is done by assigning each sounds to one of 16 evenly spaced source positions on a circular plane in order to give a sense of spatial imprint to each cluster.

CONCLUSIONS

Many questions related to sound perception remain open despite solutions are put forward. Among them, the temporal dimension should be further investigated in order to have a better understanding on the effect of time (sound durations) through perception for example. The CAC approach may even be considered as a solution to these challenges. Another one is the method for measuring the similarity itself. Using more than one approach simultaneously (magnitude, orientation and dependency) may be a fairly good solution but the problem of interpreting the results accurately remains open. More specifically when comparing a sound and its re-synthesised version [Le Bel 2017]. Indeed, up to three distinct stages of space implied by the dimensionality reduction and their impact on the shape of clusters. Although the Euclidean space seems to be well suited for achieving such tasks in general, this question is another one that should be further investigated from a perceptual angle. Another one is related to the graph exploration. Considering that the context of this work is art oriented, the graph search algorithms should be further investigated from a perceptual angle rather than an optimization one in order to exploit the full potential of these tools into the creative process. In this sense, these algorithms should be further evaluated for their musical potential rather than for their efficiency. In other words, the question is about the kind of musical structure the various graph search algorithms may lead to. This approach also comes with certain limitations. The first one is about using raw audio signal as input. Contrary to mainstream CAC approaches, it may be an advantage but it is also its main disadvantage because the quality of the output is directly correlated to the quality of the input but also because the whole process depends on it. Another limitation is related to the use of low-level audio features. Although the resulting space of variables may quickly become very complex and give the impression of covering a very large array of sounds, the results remain interpretable on a low-level basis only, meaning that no aesthesetical nor emotional effects may be considered using such a approach. Then, the clustering method is itself another notable limitation. Based on unsupervised learning, the method does not provide any information on the clusters components other than a similarity index. In other words, the results are simply quantitative and not qualitative. Finally, the complexity of this tool may be a limitation itself because an ‘unequipped’ user may end spending a lot of time understanding the multiple parameters of this approach.

REFERENCES


Peeters G. (2004). A large set of audio features for sound description (similarity and classification) in CUIDADO project. unpublished version 1.0 (23 avril), Ircam.


FOR MORE INFORMATION

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