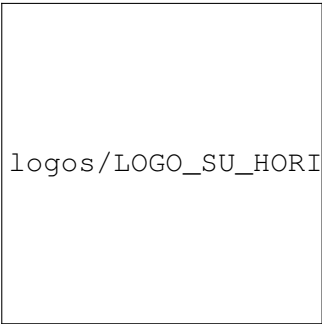



report



logos/LOGO_SU_HORIZ_SEUL.jpg



logos/Logo_TeÀleÀcom_Pagos.jpg



logos/IBGAM.CP.jpg

Internship Report
Written in August 2022

Automatic effect recognition and configuration for timbre reproduction

Intern:

Alexandre D'HOOGE

alexandre.dhooge@atiam.fr

Supervisor:

Gaëtan HADJERES

gaetan.hadjeres@sony.com

March 1st — August 31st

Key-words — audio effects, music information retrieval, differentiable signal processing, sound matching, computer music;

Mots-clés — effets audio, traitement de l'information musicale, traitement du signal différentiable, reproduction sonore, musique assistée par ordinateur.



logos/ATIAM_logo_ROND_RVB.jpg



logos/logoCSL512.png

January 15, 2025

Introduction

Many perceptual phenomena are contained in the score, intentionally planned or not by the composer. A blend is a phenomenon that occurs in orchestral music when several voices fuse together to generate a new timbre^{<empty citation>} This happens when a listener cannot distinguish the different voices, that is to say the different instruments, that are playing at the same time. A blend can be intentional or not, the aim of this project is to detect such phenomena in orchestral music scores.

However, symbolic music computational analysis is a vast and complex topic that requires a study on the concept of music scores, voices and streams. To analyse symbolic music, and more precisely to compare parts within a score, we first need to define those terms.

Scores are the written representation of music. They are a way to communicate music to musicians and to listeners. We will use MusicXML files as scores. MusicXML is a format that allows to represent music scores in a computer-readable format. Voices in a score is not a concept as simple as it seems. Cambouropoulos in 2008 defined three different ways to define voices in a score. Voices can be simply defined as the sound sequence produced by a single source (an instrument or an individual choral voice). However, voices can also be defined in a perceptual way, that is to say voices would be delimited by the sound sequences the listeners perceive as distinct. Finally, voices can be separated based on harmonic theory relative to music composition. With this definition, voices are delimited by the musical structure of the score^[1].

We chose to adopt the term stream, that is an assemble of parts is played by different instruments grouped by perceptual principles. These definitions are extremely basic but are necessary approximations to be able to extract features for a machine learning model. Using these definitions, a blend can be defined as a phenomenon that occurs when a stream can not be broken down into parts by a listener.

In this work we focused on concurrent grouping, starting from the lowest level of the taxonomy, classifying blends and non-blends. This grouping effect arises from the basic technique of combining instruments at the unison or at particular pitch intervals, usually implying rhythmic synchrony and parallelism in pitch. However, as we will see, not all couplings result in blending of the component instrument. This first analysis oriented us in our choice of features and machine learning model^{<empty citation>}

Indeed, music can take an infinite number of shapes but an algorithm or machine learning model needs a finite number of inputs^{<empty citation>} We sorted features into four categories using two questions: is the feature related to a single part or the comparison of two parts? Is the feature computed on a single bar or is it a global characteristic of the instrument? All the features were computed on scores taken from the OrchARD dataset¹.

Our project is not the first to tackle the problem of identifying blends in orchestral scores, algorithms and experiments have already been developed to detect blends in scores. The algorithm presented by Aurélien Antoine and al in 2021 used synchronicity, harmonicity and parallelism^{<empty citation>} However, we think many more features could be decisive for this problem such as timbre characteristics on instruments^{<empty citation>} Moreover, machine learning models have not been used yet to detect blends in scores while we are convinced it is a promising approach.

Therefore, we decided to build a machine learning model taking as input a score, computing several features and returning a prediction (0/1) for each couple of parts and for each bar of the score. This model is composed of a classifier (add classifier type when chosen) and a hierarchical clustering algorithm to print out a dendrogram tree of the voices' proximity on each bar.

¹<https://orchard.actor-project.org/about/>

Contents

1	Data retrieval, feature extraction and feature selection	4
2	Model selection and first iterations	4
3	Model improvement and further iterations	4
4	Conclusion	4

1 Data retrieval, feature extraction and feature selection

As presented before, the data was extracted from the OrchARD dataset. It is composed of 67 movements' scores written by various composers such as Debussy, Berlioz, Beethoven and many others. They are accompanied by an annotation table describing every motive - among which blends - that occurs in the score. In total, 2796 blends were annotated. We have access to the bar where the blend starts and the bar where it ends, the strength of the blend, the instruments involved and other information. The annotations are such that we could retrieve for each bar and each couple of parts if the pair is blending or not on this bar; these are the labels our model will be trained on. We didn't take in account the annotated strength of the blends and left it for further studies.

Now that we know what we wanted the model to predict, we needed to extract significant features. Features are sorted in 4 categories:

1. Global Features computed on single parts
2. Global features computed on couples of parts
3. Features computed on single parts for each bar
4. Features computed on couples of parts for each bar

(presentation beamer plan structuré + exemple musique au début récupérer de TOGE voir readme)

2 Model selection and first iterations

3 Model improvement and further iterations

4 Conclusion

ouverture sur le fait que certains facteurs importants ne peuvent pas être pris en compte (attack time, room acoustics, dynamics <https://timbreandorchestration.org/writings/timbre-lingo/blend>)

References

- [1] A. Vaswani *et al.*, *Attention Is All You Need*, Aug. 2023. arXiv: [1706.03762](#). Accessed: Oct. 15, 2024.