Modeling Human Experts' Identification of Orchestral Blends Using Symbolic Information

Aurélien Antoine, Philippe Depalle, Philippe Macnab-Séguin and Stephen McAdams

Schulich School of Music
McGill University
Montreal, QC, Canada
aurelien.antoine@mcgill.ca

Abstract. Orchestral blend happens when sounds coming from two or more instruments are perceived as a single sonic stream. Several studies have suggested that different musical properties contribute to create such an effect. We developed models to identify orchestral blend effects from symbolic information taken from scores based on calculations related to three musical properties/parameters, namely onset synchrony, pitch harmonicity, and parallelism in pitch and dynamics. In order to assess the performance of the models, we applied them to different orchestral pieces and compared the outputs with human experts' ratings in the Orchestration Analysis and Research Database (orchard.actor-project.org). Using different weights for the three parameters under consideration, the models obtained an average accuracy score of 81%. These preliminary results support the initial developments. Nevertheless, future work aims to investigate the weights of each musical property and to include audio analyses to take into account timbral properties as well.

Keywords: Orchestral Blend, Computer Modeling, Perception, Music Information Retrieval

1 Introduction

Orchestration offers myriad possibilities to combine instruments and to create potential sonic effects that could not necessarily be achievable with a smaller ensemble. This musical practice is a challenging subject that is being investigated from a wide range of research perspectives. The aim of our project is to explore perceptual effects of orchestration in order to understand which are the prominent characteristics. It has been suggested that different perceptual effects can emerge from the combinations of instrumental properties, such as the blending or separation of musical instrument sounds to name but two [5]. The fusion of two or more instruments results in a blend effect in which the sounds can be perceived as belonging to a single musical event [15,14], whereas segregation happens when there is a clear separation between different concurrent sound events [11,9]. This paper focuses on the factors affecting orchestral blend and

introduces our approach for developing computer models capable of identifying groups of instruments involved in blend effects. Previous research has proposed computational models of auditory scene analysis [16] but not specifically for instrument blend effects in a musical context. Our aim was to investigate the extent to which the blend effect could be modeled using only symbolic information in order to formalize the processes involved, which would then be utilized for comparisons between computational outputs and data from perceptual experiments. Such developments could be incorporated into systems designed to perform orchestration analysis from machine-readable musical scores and also in computer-aided orchestration systems.

The remainder of this paper is organized as follows. First, we introduce the different characteristics that contribute to the creation of orchestral blends and define the properties that have been utilized in our models. Section 3 provides technical details of the implementation of the orchestral blend models. Then, we apply the models to a selection of orchestral pieces and compare the outputs with human experts' ratings in order to assess the performance of the models. The results of these analyses are then discussed in Section 5. Finally, the last section presents concluding remarks and suggests different ideas to enrich the current models.

2 Orchestral Blend

As mentioned in the previous section, orchestral blend is the perception of different sounds as being grouped into a single event. This perceptual effect is the result of the fusion of instruments properties, an important aspect in orchestration [2]. Several treatises on orchestration have discussed blending techniques and have suggested methods for combining instruments to create this effect [1,12,13]. Blending techniques can be utilized for enriching the quality of a dominating instrument by adding other instruments or for completely fusing instrument sounds to create a unique mixture [12,13,14]. These treatises usually propose guidance for selecting instruments to conceive orchestral blends. For example, Rimsky-Korsakov suggests that woodwind and string instruments tend to blend well [13].

Blend has also been investigated within perception and cognition research. It has been suggested that different acoustical properties play an important role in grouping musical events into a single stream. Following Gestalt principles, sounds evolving similarly are more likely to be grouped together [3]. For instance, onset synchrony, harmonicity, and similar changes in frequency and amplitude across successive sounds might lead the auditory system to group sound components together [10,6]. Moreover, two or more instruments with a low overall spectral centroid and with close centroids tend to blend better. Characteristics related to the the musicians' performance nuances, the spatial position of the instruments, and the room acoustics also contribute to blend effects [7,8]. In regards to musical properties linked to studies in audio perception and cognition, having instruments playing in a harmonic series, in synchrony, and with perfect paral-

lelism in pitch and dynamics often lead to a blend. An example is shown in Fig. 1, in which oboes 1 and 2 in major thirds and clarinets 1 and 2 in major thirds an octave below the oboes play together in harmonic relations (roughly harmonics 4, 5, 8, 10), in synchrony, and with perfect parallelism in pitch and dynamics. Fig. 2 shows another blend example, in which two flutes, two clarinets, and two bassoons are playing together, with two oboes coming in and out throughout the phrase.

Due to processing only symbolic information, we have disregarded spectral properties and information related to performance, spatial position, and room acoustics for the first phase of this project. Instead, we have decided to process musical properties that can be estimated from a musical score. Therefore, for our models, we have estimated the onset synchrony, harmonicity, and parallelism in pitch and dynamics between instruments, recognizing that timbral aspects derivable from the audio signal often also play a role.

3 Score-based Modeling of Orchestral Blend

This project aims to model human experts' identification of orchestral blends using symbolic information. Thus, the initial stage was to define methods to retrieve the musical information from computer-readable scores before developing models for estimating orchestral blend.

3.1 Retrieving Symbolic Score Information

For computer-readable scores, we chose to work with MusicXML, a format based on the markup language XML used for representing Western musical notation. Several databases offer orchestral pieces encoded in this format. However, they sometimes contain missing or wrong information and also have different templates, which can be due to the software used for their creation. Therefore, we collaborated with OrchPlayMusic (OPM)[1], a company that offers a multichannel audio player for orchestral music and a library of high-quality simulations of orchestral pieces. First, we established a standard procedure to generate MusicXML files with consistent information. We used `MusicXML 3.1`[2] for encoding the symbolic score information. Then, several pieces were selected from the OPM Library, which contains several orchestral excerpts that have been rendered by the OPM team following a set of high-quality simulation techniques, and have been generated as MusicXML files. In order to retrieve the specific musical information required for estimating the characteristics involved in the blend effect, we programmed different functions in `Python 3.7` to process MusicXML files, following the methods further described in the next section.

[1] www.orchplaymusic.com
[2] www.w3.org/2017/12/musicxml31/

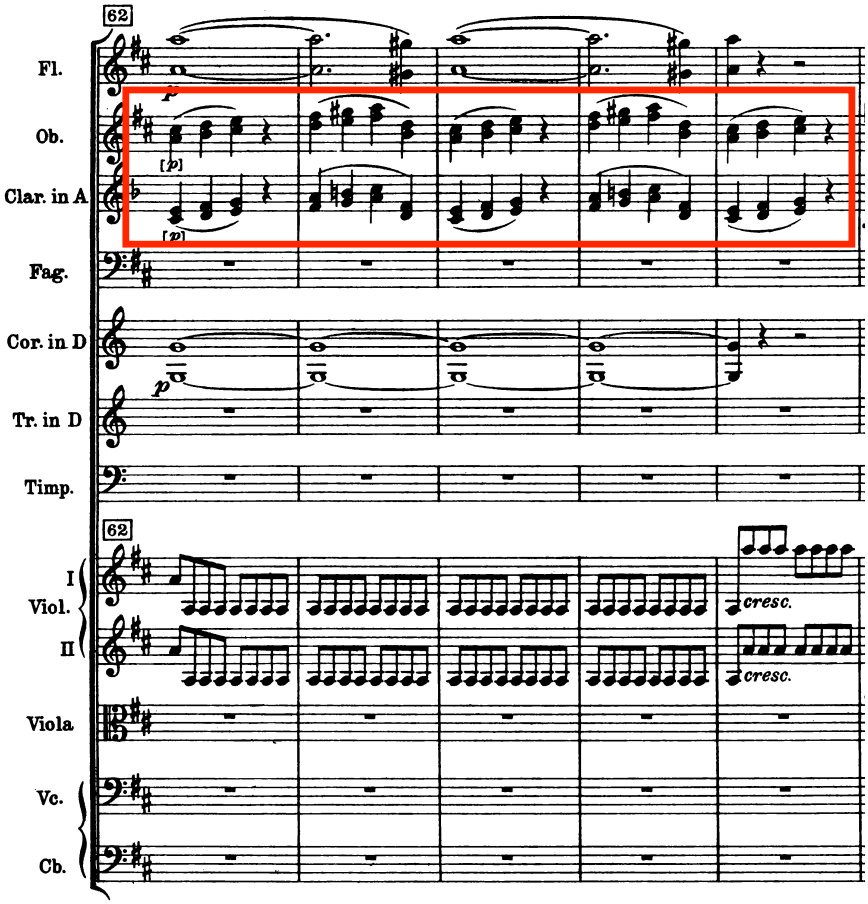

Fig. 1. Example of a blend between two oboes and two clarinets (annotated with a red box) in Mozart, Don Giovanni, mm. 62–66.

3.2 Estimating Orchestral Blend

The process for estimating orchestral blend from computer-readable scores is divided into different steps. First, we had to define the segmentation of the musical pieces. We decided to perform the analysis on a measure-by-measure basis, as orchestral effects most often occur over the course of at least a measure. The rationale is that performing the calculations on a shorter frame (i.e., note by note) would result in a significant increase in the amount of information to compute and compare, and it would also omit aspects related to temporal properties (i.e., parallelism). A longer analysis frame could overlook effects occurring in a single measure. Thus, a measure worth of information appeared to be an appropriate analysis time frame to start with.

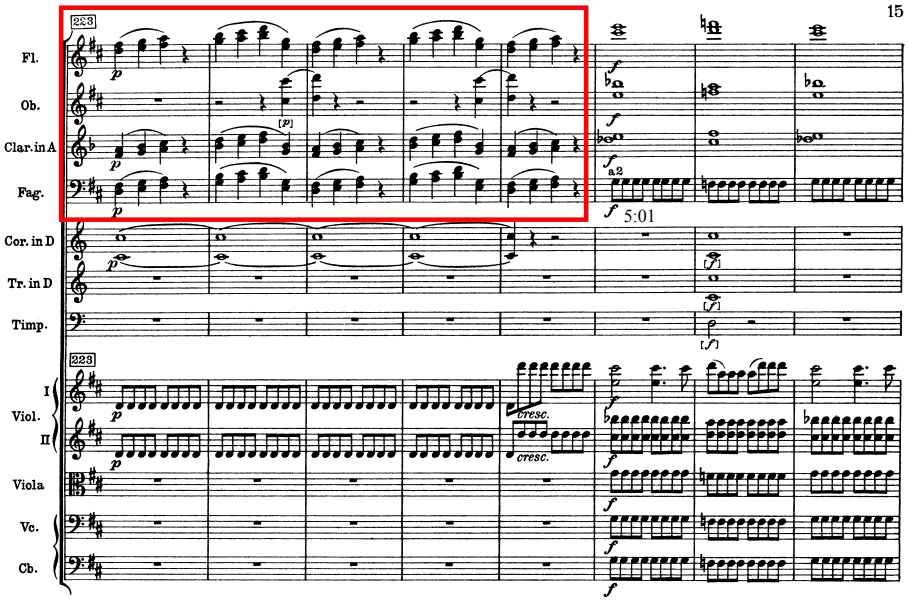

Fig. 2. Example of a blend between two flutes, two oboes, two clarinets, and two bassoons (annotated with a red box) in Mozart, Don Giovanni, mm. 223–227.

The estimations of orchestral blend are computed in three steps: first the onset synchrony, then the pitch harmonicity, and finally the parallelism in pitch and dynamics. Fig. 3 presents a diagram of the different processes for estimating orchestral blend, which are detailed below.

Onset Synchrony. The first step is to list all the onset values for each active instrument in the measure. MusicXML's note duration being represented as divisions per quarter note, it is necessary to convert the symbolic duration into time, defined in milliseconds using tempo values and note types. It is then possible to calculate the onset value for each note. Furthermore, using duration in milliseconds allows us to define a threshold for considering notes being synchronized. We set the default threshold at 30 ms, following suggestions by research on attack time [4]. Thus, notes are considered synchronized if their onset values fall within a 30-ms window. Then, the instruments sharing the most onset values are grouped together. Finally, the synchrony score is calculated with the cardinality of the intersection of the different sets of onset values. The groups of synchronous instruments are then passed to the function for pitch harmonicity calculations. Also, if there is no group of synchronous instruments, the algorithm bypasses the other functions and moves to the next measure.

Pitch Harmonicity. In the second step, this function takes as input the groups of synchronous instruments. For each onset value, the function retrieves the pitch for each active instrument and calculates the interval in semitones between the different pitches. It determines whether they are in a harmonic

series, using the lowest pitch as the root. If the instruments are not all in a harmonic series, the function lists all the different instrument pitches and checks if a tonal chord is involved, following a framework of standard Western musical chords. If no harmonic chord is found, it keeps the largest list of instrument pitches that are in a harmonic series, similar to applying a harmonic template. This step also removes instruments that share onset values but are not in the harmonic series and potentially not involved in the blend. Once the intervals for all the onset values are analyzed, the function returns a harmonicity score for the instruments that are either playing in a harmonic series or forming a tonal chord.

Parallelism in Pitch and Dynamics. The final step looks at the evolution of the pitches and the dynamics over the course of a measure. The function estimates whether the different instruments are playing in parallel. For each instrument, it lists the note sequence by examining if the next note is higher ($+1$), lower (-1), or the same (0) in pitch as the initial note of the measure, which is set to 0. Then, it compares each element of the different lists of note sequences and adds 1 if for each note they all have the same value (i.e. $+1$, -1 or 0) and 0 if at least one is different. The resulting score is then divided by the number of elements in the list, giving us a proportion that is then used for the parallelism score. A similar procedure is applied for the dynamics, where the function examines whether the instrument is playing the notes harder, softer, or at the same dynamic, and then calculates the parallelism.

Output Decision. Once the three properties have been calculated, their corresponding scores are averaged and compared to a defined threshold. If the average score is above the threshold, the group is output as a potential blend. If it is below, the group is ignored and the next group is tested, or the program moves to the next measure if there is no other group of instruments to analyze. A threshold defined as 100 would mean that all the instruments in the group would have perfect synchrony, harmonicity, and parallelism in pitch and dynamics. Lowering the output threshold would allow for more flexibility in the calculations of the musical characteristics and would tolerate deviation from the theoretical rules. This would also account for the different strengths of the blend effect. Furthermore, the properties are set as having the same weights in the calculations. For instance, synchrony is as important as parallelism and harmonicity.

For each measure, the program lists the group(s) of blended instruments with their scores, if a blend has been detected. Nevertheless, orchestral effects can happen over several measures. Thus, we also apply a post-blend analysis function in order to find groups of blended instruments that span consecutive measures. It compares the list of instruments in two neighboring measures and groups them if all the instruments are in both measures. The grouping continues until the instruments are not present in the next measure. The blend is then listed as happening from measure a to measure b, with the names of the instruments involved in the effect. Using the example shown in Fig. 1, the blend occurs from measures 62 to 66, with oboes 1 and 2 and clarinets 1 and 2 playing in every

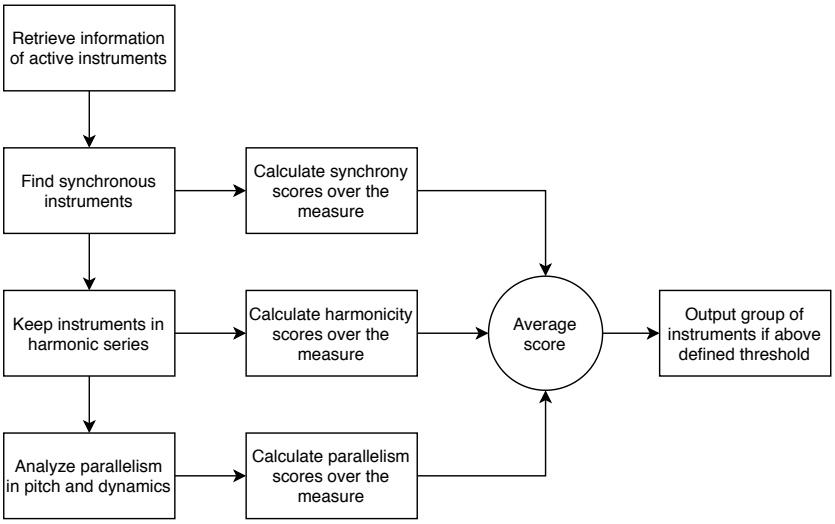

Fig. 3. Synopsis of the orchestral blend estimation algorithm.

measure and returned as blending instruments. Here, the model would return the group of instruments (oboe 1-2 and clarinet 1-2) and specifies that the effect starts at measure 62 and finishes at measure 66.

4 Comparison between Score-Based Models and Human Experts' Identification

In order to test and assess the performance of the score-based modeling of orchestral blend, we decided to apply the models to orchestral excerpts and compare the output with blends that have been identified by human experts. Therefore, we decided to use annotations taken from the Orchestration Analysis and Research Database (Orchard)[3], which contains numerous annotations of orchestral effects derived from the analysis of several orchestral pieces, with the majority spanning the period from 1787 to 1943. However, the selection of pieces to perform the testing was restrained to what is available in both the OPM library and the Orchard database. Furthermore, only parts of some of the orchestral pieces are included in the OPM library. In the end, we utilized the MusicXML files of the following 5 orchestral excerpts:

- Berlioz - Symphonie Fantastique - IV (mm. 1–77)
- Mozart - Don Giovanni - Overture (mm. 1–284)
- Haydn - Symphony 100 - II (mm. 1–70)
- Mussorgsky - Pictures at an Exhibition - II (mm. 57–109)
- Mussorgsky - Pictures at an Exhibition - XIII (mm. 1–22)

Table 1. Blend-detection score using an output decision threshold set at 100 for a blend in Mozart, Don Giovanni, mm. 62-66, shown in Fig. 1.

Measure number	Human experts	Model	Score (ratio)
62	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	1.0
63	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	1.0
64	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	1.0
65	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	1.0
66	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	Oboe 1, Oboe 2, Clarinet 1, Clarinet 2	1.0

Table 2. Blend-detection score using an output decision threshold set at 80 for a blend in Mozart, Don Giovanni, mm. 223-227, shown in Fig. 2.

Measure number	Human experts	Model	Score (ratio)
223	Flute 1, Flute 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	Flute 1, Flute 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	1.0
224	Flute 1, Flute 2, Oboe 1, Oboe 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	Flute 1, Flute 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	0.75
225	Flute 1, Flute 2, Oboe 1, Oboe 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	Flute 1, Flute 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	0.75
226	Flute 1, Flute 2, Oboe 1, Oboe 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	Flute 1, Flute 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	0.75
227	Flute 1, Flute 2, Oboe 1, Oboe 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	Flute 1, Flute 2, Clarinet 1, Clarinet 2, Bassoon 1, Bassoon 2	0.75

The comparison process was designed to evaluate if the models were retrieving the same groups of instruments as the ones labeled as part of a blend by the human experts. For example, Table 1 details the results of the comparison for a blend identified in the Overture to Mozart's Don Giovanni (shown in Fig. 1). Here, using an output decision threshold set at 100, the models have output the same group of instruments (oboe 1-2 and clarinet 1-2) as the human experts for the five measures, obtaining a score of 4/4. Table 2 details the results for the blend shown in Fig. 2, also taken from Don Giovanni. Here, the output decision threshold was set at 80. Note that the models have output the same instruments as the ones labeled by the experts for measure 223 (i.e., flute 1-2, clarinet 1-2, and bassoon 1-2), resulting in a score of 6/6. However, for the measures 224 to

[3] `https://orchard.actor-project.org`

Table 3. Summary of the average ratio scores (in %) and number of blends for each and across all orchestral pieces.

	Average ratio scores (in %)			
Musical pieces	**Output decision threshold = 100**	**Output decision threshold = 80**	**Output decision threshold = 60**	**Number of blends**
Berlioz - Symphonie Fantastique - IV	49.57	81.53	77.81	8
Mozart - Don Giovanni	41.29	91.04	91.34	17
Haydn - Symphony 100 - II	71.60	88.93	89.88	7
Mussorgsky - Pictures at an Exhibition - II	70.48	70.64	70.64	17
Mussorgsky - Pictures at an Exhibition - XIII	35.91	77.75	76.20	7
Across all pieces	53.77	81.98	81.17	56

227, the models missed the oboe 1-2, thus, obtaining a proportional score of 0.75 (6/8).

A similar comparison process was applied to the whole testing set. We ran the models using output decision thresholds set at 100, 80, and 60, in order to investigate if more flexibility in the calculations would improve the accuracy or not. Table 3 proposes a summary of the average proportion scores, expressed in %. The performances for the three different output thresholds are listed for each orchestral excerpt and across all the pieces, along with the number of blends for each piece. Note that the best average score was obtained using an output decision threshold set at 80, which resulted in an accuracy of 81.98% for the 56 blends labeled across all the orchestral pieces.

5 Discussions

From the results detailed in the previous section, it is clear that using only the standard rules, represented by defining the output decision threshold at 100, is not sufficient. We note a significant improvement between the output decision thresholds set at 100 and at 80, while the difference between 80 and 60 is small. As detailed in Table 3, the average accuracy score across all the pieces using the threshold at 100 was 53.77%, whereas it was over 81% with thresholds set at 80 and 60. For the symbolic information from Mozart - Don Giovanni and Haydn - Symphony 100, movement II, the models performed better with the lowest threshold. In regards to the two movements of Mussorgsky - Pictures at an Exhibition, the models did not improve when the output decision threshold was set at a lower value. The performance even decreases when setting the threshold at 60 for Berlioz - Symphonie Fantastique, movement IV, suggesting that more flexibility in the rules may have created confusion in the models.

Some limitations of this initial implementation have emerged from the comparison process. For instance, in the blend shown in Fig. 2, the models missed

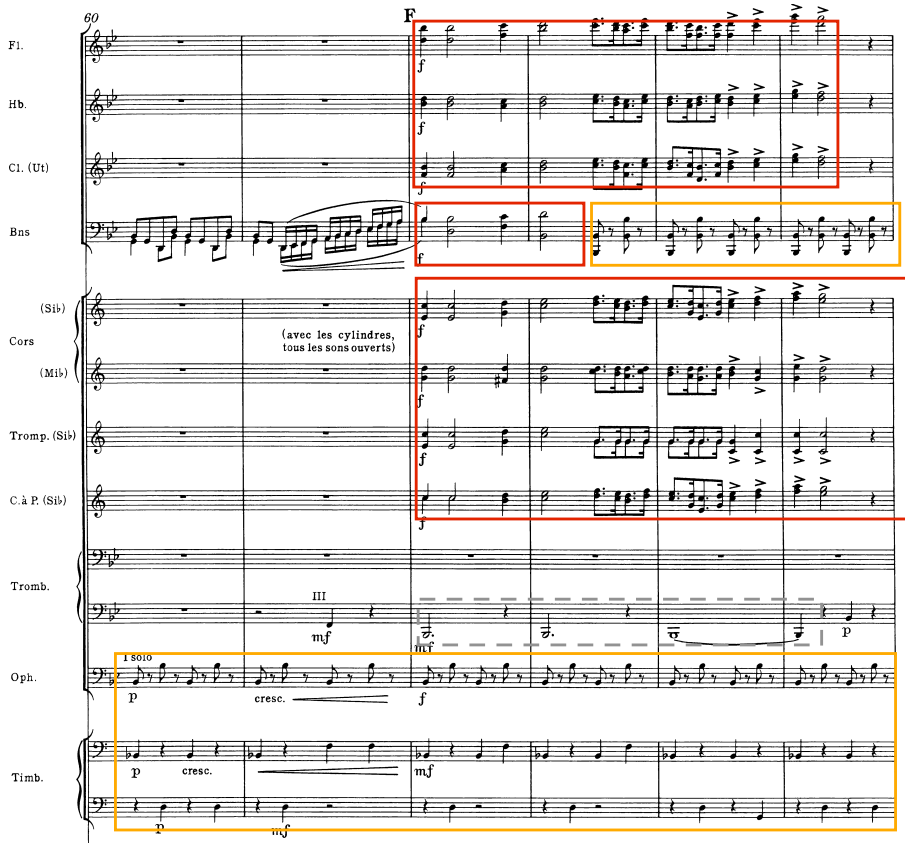

Fig. 4. Example of a blend between a tuba, two timpani, and four bassoons later in the phrase (annotated with an orange box), mm. 60–65, and a blend between two flutes, two oboes, two clarinets, four bassoons, four horns, two trumpets, and two cornets (annotated with a red box), mm. 62–65, in Berlioz, Symphonie Fantastique IV

the oboe 1-2 in the measures 224 to 227. This is due to them playing one note in each measure, and thus, having a low score for onset synchrony (1 common onset value out of 4 with all the other instruments) as well as for the parallelism properties. The current implementation is not able to retrieve instruments involved in a blend and playing sporadically compared to the rest of the blended instruments. Another limitation is illustrated with Fig. 4. Here, the four bassoons switch from one group of blend (annotated with a red box) to an other (annotated with an orange box) in the middle of the measure 63. The models have grouped the bassoons with the flutes, oboes, clarinets, horns, trumpets, and cornets instead of with the tuba and timpani for the measure 63 and 64. Due to performing the analysis on a measure-by-measure basis, the models cannot notice a change of blended group if it occurs within a measure. Furthermore, given that the pitch harmonicity function is based either on harmonic series of

semitone intervals or on a succession of tonal chords, if instruments play notes that are do not follow this framework, they would be discarded. This could be another reason that the models did not output all of the instruments involved in an orchestral blend.

Although our models have accurately output almost 82% of blended instruments on average across 56 blends, the results detailed in Section 4 indicate that processing only symbolic information is not enough to thoroughly model the orchestral blend effect.

6 Conclusion and Future Directions

In this paper, we have presented our approach for modeling human experts' identification of orchestral blends using symbolic information from computer-readable scores. Our partnership with OrchPlayMusic has allowed us to get standard, precise, and consistent MusicXML files of orchestral pieces from which to process symbolic score information. We based our models on the evaluation of three musical characteristics suggested by previous research on orchestral blend: onset synchrony, pitch harmonicity, and parallelism in pitch and dynamics, as described in Section 3. In order to evaluate the performance of the models, we decided to compare their outputs with blends labeled by human experts. Therefore, we selected and processed five orchestral pieces that had been previously analyzed by musical experts and also generated as MusicXML files by the OPM team. As detailed in Section 4, the models achieved an average accuracy score of 81% across all the pieces.

Preliminary results support the initial developments and suggest that estimations based on symbolic information can account for a significant part in modeling orchestral blends. However, further investigation is required to overcome the current limitations discussed in Section 5. For instance, tuning the weights of the different calculations could be an aspect to consider, as perhaps onset synchrony is a more prominent characteristic than pitch harmonicity. The use of supervised machine learning techniques combined with a large set of labeled blend examples could potentially aid in addressing this question. Moreover, spectral, temporal, and spectrotemporal audio properties also contribute to the blend effect, as mentioned in Section 2. Therefore, future work also aims to include audio analyses to take into account timbral characteristics as well.

References

1. Berlioz, H.: Grand traité d'instrumentation et d'orchestration modernes. Henry Lemoine, Paris, France (1844)
2. Blatter, A.: Instrumentation and orchestration. Schirmer Books, New York, NY, 2nd edn. (1997)
3. Bregman, A.S.: Auditory scene analysis: The perceptual organization of sound. MIT press, Cambridge, MA (1990)
4. Bregman, A.S., Pinker, S.: Auditory streaming and the building of timbre. Canadian Journal of Psychology/Revue canadienne de psychologie 32(1), 19 (1978)

5. Goodchild, M., McAdams, S.: Perceptual processes in orchestration. In: Dolan, E.I., Rehding, A. (eds.) The Oxford Handbook of Timbre. Oxford University Press, New York, NY (2018)
6. Kendall, R.A., Carterette, E.C.: Identification and blend of timbres as a basis for orchestration. Contemporary Music Review 9(1-2), 51–67 (1993), https://doi.org/10.1080/07494469300640341
7. Lembke, S.A.: When timbre blends musically: Perception and acoustics underlying orchestration and performance. Ph.D. thesis, McGill University, Montreal, QC, Canada (2014)
8. Lembke, S.A., Parker, K., Narmour, E., McAdams, S.: Acoustical correlates of perceptual blend in timbre dyads and triads. Musicae Scientiae pp. 1–25 (2017), https://doi.org/10.1177/1029864917731806
9. McAadams, S.: Musical timbre perception. In: Deutsch, D. (ed.) The psychology of music, pp. 35–67. Academic Press, San Diego, CA, 3rd edn. (2013)
10. McAdams, S.: The auditory image: A metaphor for musical and psychological research on auditory organization. In: Crozier, W.R., Chapman, A.J. (eds.) Cognitive Processes in the Perception of the Art, pp. 289–323. Elsevier, North-Holland, Amsterdam (1984)
11. McAdams, S., Bregman, A.S.: Hearing musical streams. Computer Music Journal 3(4), 26–43 (1979)
12. Piston, W.: Orchestration. WW Norton, New York, NY (1955)
13. Rimsky-Korsakov, N.: Principles of orchestration. Dover Publications, New York, NY, 1st edn. (1964)
14. Sandell, G.J.: Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. Music Perception: An Interdisciplinary Journal 13(2), 209–246 (1995), https://doi.org/10.2307/40285694
15. Sandell, G.J.: Concurrent timbres in orchestration: a perceptual study of factors determining "blend". Ph.D. thesis, Northwestern University, Evanston, Illinois, United States of America (1991)
16. Wang, D., Brown, G.J. (eds.): Computational auditory scene analysis: Principles, algorithms and applications. Wiley-IEEE Press, New York, NY (2006)