

INSTRUMENT TIMBRE ENHANCES PERCEPTUAL SEGREGATION IN ORCHESTRAL MUSIC

MANDA FISCHER

*McGill University, Montreal, Canada & University of
Toronto, Toronto, Canada*

KIT SODEN, ETIENNE THORET, MARCEL MONTREY, &
STEPHEN McADAMS

McGill University, Montreal, Canada

TIMBRE PERCEPTION AND AUDITORY GROUPING principles can provide a theoretical basis for aspects of orchestration. In Experiment 1, 36 excerpts contained two streams and 12 contained one stream as determined by music analysts. Streams—the perceptual connecting of successive events—comprised either single instruments or blended combinations of instruments from the same or different families. Musicians and nonmusicians rated the degree of segregation perceived in the excerpts. Heterogeneous instrument combinations between streams yielded greater segregation than did homogeneous ones. Experiment 2 presented the individual streams from each two-stream excerpt. Blend ratings on isolated individual streams from the two-stream excerpts did not predict global segregation between streams. In Experiment 3, Experiment 1 excerpts were reorchestrated with only string instruments to determine the relative contribution of timbre to segregation beyond other musical cues. Decreasing timbral differences reduced segregation ratings. Acoustic and score-based descriptors were extracted from the recordings and scores, respectively, to statistically quantify the factors involved in these effects. Instrument family, part crossing, consonance, spectral factors related to timbre, and onset synchrony all played a role, providing evidence of how timbral differences enhance segregation in orchestral music.

Received: June 9, 2020, accepted January 24, 2021.

Key words: musical timbre, auditory stream segregation, orchestration, blend, audio descriptors

ORCHESTRATION PROVIDES AN EXCELLENT framework for investigating how timbre (sound quality) is used as a tool to shape

listeners' perceptions. Sounds that differ acoustically are organized by the auditory system into separate percepts called "auditory streams" (Bregman & Campbell, 1971). A physical sound source can produce a sequence of successive acoustic events. A stream is a psychological organization that 1) mentally represents such a sequence and 2) displays a certain internal consistency or perceptual continuity that allows the sequence to be interpreted as an integrated whole that is potentially segregated from concurrent sequences produced by another sound source (McAdams & Bregman, 1979). To examine this phenomenon, the current study used naturalistic orchestral excerpts from the symphonic repertoire to examine perceptual segregation. We are among the first, to our knowledge, to include both acoustic and score-based factors in our models that capture the dynamic interplay between a rich array of acoustic cues, such as timbre, rhythm, part crossing, and consonance. Together, these musical features compete and interact to generate the listener's perception of stream integration and segregation.

Auditory Scene Analysis: A Model for Auditory Perception

The auditory realm is composed of a rich array of acoustic properties that offer important cues about our environment. The difficulty in auditory perception, however, is that sounds originating from separate sources arrive at the eardrums simultaneously in the form of a complex pressure wave. The process of separating information coming from different sound sources, known as auditory scene analysis, groups sounds on the basis of Gestalt heuristics (Bregman, 1990, Chapters 1 & 2; Bregman & Pinker, 1978; Koffka, 1935, Chapter 4; McAdams & Bregman, 1979). A perceptual heuristic, such as the degree of similarity between successive sound events, directly affects the degree to which they will be organized into the same or separate mental representations.

This study focused on the process of auditory stream segregation (the connection of fused events into continuous event streams), where the independent streams that arise from sequential grouping are said to be

segregated from one another (Bregman, 1990, chap.2; Bregman & Pinker, 1978). Sequential grouping cues vary along a number of perceptual dimensions, such as timbre, pitch, and loudness. These attributes compete and interact to form complex and perceptually rich experiences for the listener (Bregman, 1990, Chapter 2; McAdams, 1984, 2019a; Sandell & Chronopoulos, 1997).

Timbre

Timbre, a multidimensional attribute of auditory perception, allows us to recognize and track the diverse array of sounds in our environment (McAdams, 2019b). It is comprised of both static and dynamic acoustic attributes, such as brightness, noisiness, attack-decay characteristics, etc., and has been referred to as “sound color or quality” (Wessel, 1973). In terms of auditory grouping principles, timbre is both the result of concurrent grouping and a cue for sequential grouping (McAdams, 1984). The latter effect—timbre’s effect on sequential grouping—is the focus of the current study.

In order to understand how the *degree* of timbral differentiation promotes auditory stream segregation, one can quantify the acoustic parameters that account for the perceptual structure of timbral relations. Previous multidimensional scaling studies have shown that the timbres of a set of sounds can be represented in a multidimensional space, the dimensions of which can be linked to acoustic properties (Grey, 1977; McAdams Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Miller & Carterette, 1975; Plomp, 1970). By examining the acoustic attributes present in the stimuli, such a timbre space can serve as a useful model for drawing predictive links between acoustic properties and perceptual phenomena, such as timbre-based stream segregation (Iverson, 1995; McAdams, 2019a; McAdams & Siedenburg, 2019). For example, Hartmann and Johnson (1991) found that spectral attributes affect auditory stream segregation, whereas temporal attributes do not. Although Gregory (1994) confirmed that spectral dimensions of timbre, such as the proportion of energy in lower partials, were particularly relevant for stream segregation, the duration of decays was also identified as an important factor. Furthermore, Iverson (1995) and Bey and McAdams (2003) have shown that both spectral and temporal attributes of timbre are important for perceptual segregation.

The degree to which sounds fuse perceptually or blend musically is another factor that has been shown to affect the timbre of concurrent combinations of

sounds. Concurrent grouping can facilitate the blending of two or more instruments into a perceptually new and distinct timbre. For example, Kendall and Carterette (1993) and Sandell (1995) demonstrated that timbral similarity (closer proximity in a timbre space) between two single instrument sounds forming a dyad promotes perceptual blend. Furthermore, blend has been shown to be inversely related to the overall spectral centroid of the instrument combinations involved (Sandell, 1995; Tardieu & McAdams, 2012) and is promoted when main formants are aligned, leading to greater overlap in the spectral envelope (Reuter, 2003; Lembke & McAdams, 2015).

Experimental Research Using Real Music

The study of auditory perception in real music provides the ideal framework for examining the complex perceptual processes that underlie auditory stream segregation (Deutsch, 2013; Disbergen, Valente, Formisano, & Zatorre, 2018; Ragert, Fairhurst, & Keller, 2014; Snyder, Gregg, Weintraub, & Alain, 2012; Sussman, 2005; Uhlig, Fairhurst, & Keller, 2013; Zatorre, 2013). New and innovative tools have made it possible to combine the precision of a laboratory setting with a highly naturalistic listening situation. Here, we use a high-quality orchestral simulation environment (*OrchSim*; Bouliane & Baril, n.d.), where a multitrack digital audio file is created with each instrument of the simulated orchestra in a separate track. These tracks are then shaped by the musician-programmer in terms of musical expression. McAdams and Goodchild (2018) compared these orchestral simulations to live commercial orchestral recordings of the same excerpts on a number of qualitative factors to determine their perceptual validity and degree of plausible naturalism. Their results demonstrated that although the digital simulation’s realism ratings were slightly lower than most of the commercial recordings on these scales, the ratings were well above the middle of the rating scale. This supports the validity of digital orchestration as a powerful tool for studying perceptual processes that are more readily representative of real-world contexts, while providing access to individual instrument tracks in complex mixtures.

Timbre and Auditory Stream Segregation in Orchestration

Orchestral music’s extraordinarily rich array of timbres provides a fertile ground for investigating auditory perception in a real-world framework. It is through composers’ judicious choices of instrument combinations

that the acoustic properties of the music itself may be fashioned to elicit a particular percept in the listener. Orchestration treatises, first written in the 19th century, serve as a basis for characterizing orchestration practice (e.g., Berlioz & Strauss, 1905/1948). To date, however, understanding of orchestration practice has been guided by intuition and skill alone, and developed through the study of many scores and performances. To further understand orchestration practice, the ways in which composers operate at the level of auditory organization to shape listeners' perceptions must be made explicit. A theory of orchestration may be developed by connecting orchestration practice to its underlying psychological principles (Goodchild & McAdams, 2018; McAdams & Goodchild, 2017).

The current study bridges this gap by providing perceptual principles to ground the orchestral segregation effect and demonstrating a structuring role of timbre in music. In these experiments, we use music in order to examine the perceptual processes that underlie auditory stream segregation in complex environments. Although we are specifically interested in the role of timbre in stream segregation, timbre is only one of many perceptual dimensions operating in real music, each of which interacts and competes as a cue in auditory scene analysis. Therefore, a holistic explanation of timbre in a complex scene is only complete if it is grounded in that context and if its relation to other cues is examined. We introduce a new way of addressing this issue, by including both acoustic descriptors and score-based factors computed based on orchestral excerpts. We also examine the relation between within-stream blend and between-stream segregation to determine whether single-stream perceptual coherence plays a role in clarity of segregation.

Current Study and Hypotheses

The objective of this research is to examine the effect of orchestral timbre and within-stream blend on the degree of auditory stream segregation. The aims were to: 1) validate audiovisual music analyses using purely auditory tests on the excerpts with segregated instruments or instrument groups, 2) determine if instrument family combinations are high-level cues that composers may use pragmatically in their tacit understanding and deployment of grouping cues, and 3) determine whether the strength of within-stream blend plays a role in the strength of between-stream segregation. To achieve these aims, we selected excerpts that differed in their spectral properties related to timbre and that were deemed by analysts to have two streams. Prior analyses

of orchestral movements by pairs of expert analysts were conducted with combined visual analysis of the score and aural analysis of commercial recordings. Annotated effects related to a variety of concurrent, sequential, and segmental auditory grouping principles had to be audible in the recordings. These effects included instrumental blends and segregation of individual instruments or groups of blended instruments that were perceived to constitute auditory streams of equivalent prominence (McAdams, Goodchild & Soden, 2021). The annotations were integrated into the Orchestration Analysis and Research Database (*OrchARD*). In the following, we use the term "stream" to refer to what analysts annotated as streams. To capture a broad range of degrees of perceptual segregation, we selected two-stream excerpts that varied in terms of the annotated strength of segregation between the streams (sequential grouping) and single-stream excerpts that were annotated as perceptually blended into a single stream also with varying strengths of blend (concurrent grouping). Hereafter, the designation "two-stream" refers to the former excerpts and "single-stream" to the latter.

We operationalized global timbral differences between theoretical streams in terms of instrument family. Differences in timbre between the two streams were assumed to be greater for excerpts containing a between-family instrument combination (such as trumpet vs. violin or horn, trumpet and trombone vs. violins, violas and celli) and smaller for excerpts containing a within-family instrument combination (violin vs. viola or violins and celli vs. violas and contrabasses). It should be noted, however, that woodwind instruments present a diversity of means of excitation (e.g., single or double reed, air jet). This is generally considered to result in increased acoustic heterogeneity within this class of instruments.

In the first experiment, our main hypothesis was that the degree of perceptual segregation would increase as a function of the difference in timbre between streams. Furthermore, we hypothesized that music training would elicit greater perceptual segregation for between-family excerpts compared to within-family excerpts due to a greater sensitivity to timbral differences between potential streams.

In the second experiment, we investigated how global segregation between streams is related to ratings of perceptual integration (blend) within each stream and the properties that bind the percept. It was possible that individual streams involving several instruments would not be as strongly fused as an individual instrument. Therefore, we hypothesized that the degree of

TABLE 1. Number of Stimuli Distributed Across Type of Orchestral Effect, Timbral Class, and Timbral Combination Category

Timbral Class		Within-family			Between-family			
Timbral Combination Category		S-S	W-W	A-A	S-W	S-B	W-B	O-O
Type of orchestral annotation	Two-stream	6	6	3 (2 SWB-SWB, 1 SW-B)	6 (3 S-W, 2 SW-W, 1 S-SW)	6 (3 S-B, 3 S-SB)	6 (2 W-B, 3 W-WB, 1 WB-WB)	3 (1 S-SWP, 1 SWH-B, 1 SWB-SWP)
	Single-stream	2	2	1	2	2	2	1

Note: For timbral categories S = string, W = woodwind, B = brass, A = all (S+W+B), O = other, includes harp (H) and/or percussion (P), such as celesta and xylophone. Unseparated initials indicate blended families (SW = blend of strings and woodwinds), and initials separated by a hyphen indicate combinations in different streams (SW-B = SW in one stream, B in the other stream). In the single-stream excerpts, instruments from the indicated families participate in the blend.

integration within each stream would be inversely related to global segregation, where strongly integrated individual streams would facilitate stronger and clearer separation between streams. We again compared musicians to nonmusicians to identify any group differences.

In the third experiment, the original two-stream excerpts were reorchestrated solely with string instruments. Reducing the timbral difference between streams provided a baseline for examining the contribution of other musical parameters to the segregation results in the first experiment. Comparing the original excerpts with this baseline allowed us to assess the extent to which timbre contributes to perceptual segregation in real music, independently of other musical cues.

General Method

This section contains general methods common to all three experiments, such as stimulus generation and selection, as well as behavioral, acoustic, and score-based factors included in the statistical models. Information that is specific to each experiment will appear in the relevant Method section.

STIMULI

As detailed in Appendix A, the stimuli consisted of 36 musical excerpts with two streams ($M_{measures} = 7.42$, $SD_{measures} = 3.86$; $M_{duration} = 14.42$ s, $SD_{duration} = 7.09$) and 12 single-stream excerpts ($M_{measures} = 5.58$, $SD_{measures} = 3.90$; $M_{duration} = 12.58$ s, $SD_{duration} = 5.22$). Some two-stream excerpts contained streams composed of multiple instrument sections that analysts considered to be integrated as a group. Therefore, two-stream excerpts contained both segregation effects between streams and integration effects within each stream.

The seven timbral category combinations used in the experiment are shown in Table 1. These combinations involved four different instrument categories (String,

Woodwind, Brass, and Other). The “All” category refers to a mixture of String, Woodwind, and Brass in a single excerpt. Furthermore, “Other” includes impulsive instruments (Harp, Celesta, and Xylophone). Harp was categorized as an impulsive instrument rather than a string instrument, due to previous research showing that pizzicato strings behave differently with respect to auditory scene analysis than do bowed strings (Lembke, Parker, Narmour, & McAdams, 2019). Furthermore, impulsive sounds (e.g., plucked or struck) do not blend as well with sustained ones (Tardieu & McAdams, 2012). Timbral category combinations were denoted as follows: [instrument family stream 1]-[instrument family stream 2], e.g., S-W for String-Woodwind. Also indicated in Table 1 is the higher-level categorization into timbral classes of within- and between-family combinations.

The set of 12 single-stream excerpts was included to anchor the ratings. Their degree of blend had been determined in a previous experiment by Gianferrara (2016; McAdams, Gianferrara, Soden, & Goodchild, 2016). The study by Gianferrara (2016) found that these excerpts in isolation from the full context had a mean blend rating of 0.46 ($SD = 0.13$) on a scale from 0 to 1. The range of blend ratings was from .21 to .61 with more excerpts on the upper end of this range. It was nonetheless expected that all of these stimuli would be more integrated than the two-stream excerpts. Timbral category combinations were denoted as follows: [instrument family or families in the single stream] = e.g., S for String or SB for String + Brass.

Stimulus selection proceeded as follows:

Step 1: Database. The stimuli were drawn from *OrchARD*, which contains annotations of concurrent, sequential, and segmental groupings in 65 orchestral movements from the Classical and Romantic repertoires. Composers and music theorists analyzed orchestral scores while listening to commercial recordings to identify orchestral effects such as integration

Allegretto ♩ = 76

FIGURE 1. An excerpt from Beethoven's Symphony No. 7, Op. 92, ii, mm. 51–58. Annotations demonstrate the segregation between two string instruments: violin 1 (stream 1 in solid box) and violin 2 (stream 2 in dashed box).

(instrumental blend) and segregation into two or more streams. Their goal was to: 1) indicate where in the music orchestral segregation and integration effects were occurring, 2) identify the instruments involved in these effects, and 3) rate the strength of perceptual segregation or integration of each section they identified on a discrete scale from 1 (*weakest*) to 5 (*strongest*) based on the recording being listened to. Segregation was defined for the annotators as: 1) involving two clearly distinguishable “voices” of equal salience, and 2) occurring between individual instruments or entirely fused instrument pairings (or instrument sections) that constitute a “virtual voice.” Blend was defined as the fusion of instrument sounds creating an augmented or emergent timbre (Sandell, 1995) over the course of a motive or phrase. Following the annotations, all information regarding the orchestral segregation and integration effects (work, movement, composer, measures of occurrence, recording, start time, end time, instruments involved, and degree of segregation/integration) were compiled into *OrchARD*.

Step 2: Selection criteria. The stimuli for the experiment were chosen based on instrument family, number of instrumental parts, annotated strength rating of degree of segregation, and duration, in order to capture a broad range of degrees of segregation. (See Appendix Table A1 for a list of the instruments present in the two-stream excerpts and Table A2 for the instruments present in the one-stream excerpts.) We attempted to find several excerpts for each combination of instrument families. However, there were not enough examples of brass vs. brass in *OrchARD* to include this category.

Step 3: Simulation in OrchSim. Based on the scores and commercial recordings of the selected excerpts, composers Félix Frédéric Baril and Denys Bouliane created high-quality multitrack digital simulations using their *OrchSim* environment (Bouliane & Baril, n.d.). The dynamics of all the instruments were balanced across tracks in the full musical context. In the stereo

simulations, each instrument was placed in its traditional spatial location on stage.

Step 4: Selection and extraction of relevant instrument tracks. The instruments specifically involved in the orchestral segregation and integration effects to be tested were then isolated from the full context (if concurrent instruments that were not part of either were originally written in the score). Spatial location of the extracted tracks was preserved. This was done using the software Logic Pro X (Apple Computer, Cupertino, CA), which enabled us to: 1) cut the sound files to the specific measures where the orchestral segregation and integration effects were occurring, 2) extract the tracks with the specific instruments involved in the orchestral effect, and 3) output a stereo mix of the selected tracks. See Figures 1–3 for example excerpts extracted and isolated from the rest of the musical context: two single-instrument streams (Figure 1), two multi-instrument streams and a reorchestration of the same excerpt for strings (Figure 2), and a blended single stream (Figure 3). To view all scores used, see Supplementary Materials (at [mp.ucpress.edu](http://132.206.14.109/supplementaryMaterials/Fischer2021/)). Stimulus excerpts are available at <http://132.206.14.109/supplementaryMaterials/Fischer2021/>.

APPARATUS

Participants were seated in an IAC model 120act-3 double-walled audiometric booth (IAC Acoustics, Bronx, NY). Sounds stored on a Mac Pro 5 computer running OS 10.6.8 (Apple Computer, Inc., Cupertino, CA) were amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA) and were presented over Dynaudio BM6a loudspeakers (Dynaudio International GmbH, Rosengarten, Germany) arranged at about $\pm 60^\circ$ facing the listener at a distance of 1.5 m. The response interface was presented on the computer screen and responses were entered with a mouse. The experimental session was run with the PsiExp computer environment (Smith, 1995). Sound

FIGURE 2. An excerpt from Brahms' Symphony No. 4, Op. 98, mm. 19–24 annotated with two multi-instrument streams (flutes and oboes vs. clarinets and bassoons). The original orchestration is in the left panel and the reorchestration for strings is in the right panel.

levels were measured with a Brüel & Kjær Type 2205 sound-level meter (A-weighting) placed at the level of listeners' ears.

DATA ANALYSIS

Several acoustic and score-based factors were used as predictors in a mixed-effects analysis of the behavioral data. We first describe these factors and then the statistical modeling procedure adopted for all three experiments.

We tested two input representations from which the audio factors were derived: an auditory representation (equivalent rectangular bandwidth: ERB; Glasberg & Moore, 1990; Moore, 1986; Moore & Glasberg, 1983) and an audio-signal representation (Short-Time Fourier Transform: STFT). Using the Timbre Toolbox in MATLAB (The Mathworks Inc., Natick, MA), 11 time-varying spectral descriptors related to timbre (Peeters, Giordano, Susini, Misdariis, & McAdams, 2011) were computed for both representations (Appendix B). For each stream, the Timbre Toolbox was used to extract time-varying descriptors within each time frame of the sound. The absolute difference was then computed between each stream for each time frame and averaged across time frames to yield one mean difference value. In order to reduce redundancy and multicollinearity, Pearson correlations were computed among the 11 descriptors across all stimuli and subjected to

hierarchical clustering analysis. A threshold for clustering was set at the same distance (roughly half of the depth) of the resulting dendrogram for all experiments. Within each cluster, one descriptor was chosen. Selection was based first on the more frequently reported descriptors in the literature and second on achieving explanatory consistency across the models for all three experiments by prioritizing having similar descriptors as input to each model. The selected input descriptors used in the mixed-effects models are listed in Table 2.

Score-based factors were derived from MusicXML representations of the scores, comparing features across streams for Experiments 1 and 3 and between instruments within streams for Experiment 2. The factors included relative synchrony of onsets, average pitch interval, proportion of voice crossing in pitch, and consonance (Experiment 2 only). Appendix C includes a detailed description of how each factor was computed.

Statistical Analyses. A mixed-effects model was performed on the behavioral data. A random intercept was included in all models in order to account for within-subject individual differences. The dependent variable was degree of segregation for Experiments 1 and 3 and degree of blend (integration) for Experiment 2. Table 2 presents the full list of predictors and the subsets that were entered into the models in each experiment. In addition to timbral class (between-family or within-family combination) and training (musician or

The image shows a musical score for a multi-instrument blend. The instruments listed are Flute 1, Flute 2, Oboe 1, Oboe 2, Bass Clarinet, Bassoon 1, and Bassoon 2. The tempo is marked 'tranquillo' and the dynamics are 'espr.' and 'p'. The music is in 4/4 time and features a melodic line with a tritone interval. The score is for a multi-instrument blend (unison and octave doublings) from Richard Strauss's 'Tod und Verklärung, Op. 24, mm. 456-458'.

FIGURE 3. A single multi-instrument blend (unison and octave doublings) from Richard Strauss's *Tod und Verklärung*, Op. 24, mm. 456–458.

nonmusician), each model input consisted of converging data types: behavioral ratings, digital audio descriptors related to timbre that were computed from the sound signal, and score-based MusicXML information. Acoustic descriptors were included in order to identify specific timbral attributes that might be particularly important for promoting perceptual segregation or integration. Reorchestration was included in Experiment 3 as a between-subjects factor in order to assess the effect of reorchestrating all excerpts to a within-family combination of string instruments (increased timbral homogeneity). In all models, the timbral combination categories were reduced to the within-family/between-family distinction. This was done because the two factors were collinear, and timbral class provided a greater reduction in the Bayesian Information Criterion (BIC) (Schwarz, 1978) compared to the null model when included with the fixed effects.

In our statistical analysis, we used model selection to identify the most parsimonious model of how acoustic

and orchestral properties affect segregation (Experiments 1 and 3) or integration (Experiment 2). Model selection started with a null model without any of the predictor variables. First, we checked to see if adding all fixed effects improved (reduced) BIC. From there, non-significant fixed effects were then selected for removal. An efficient greedy variable selection approach was used to identify a parsimonious model based on minimal BIC (Kenneth & Anderson, 2002, Chapter 2). In this iterative approach, the nonsignificant variable in the model that leads to the greatest reduction in BIC when absent is removed. This process is continued until no reduction in BIC is found (Kass & Raftery, 1995). The remaining variables constitute the final model. Type 3 tests of fixed effects are reported. Bonferroni correction was applied to post hoc contrasts in order to compensate for multiple comparisons.

Experiment 1: Segregation Ratings on Orchestral Excerpts

The objective of Experiment 1 was to examine the effect of differences in orchestral timbre on the *degree* of auditory stream segregation. Our main hypothesis is that the degree of perceptual segregation will increase as a function of the difference in timbre between streams (operationalized in terms of within- and between-family pairings). Furthermore, we hypothesize that music training will elicit greater segregation for between-family excerpts due to enhanced sensitivity to the timbral differences between potential streams.

METHOD

Participants. Data analyses were performed on 40 participants (23 female), between the ages of 18 and 37 years ($M = 23.3$, $SD = 3.9$). Twenty participants were musicians ($M_{\text{training}} = 16.0$ yrs, $SD = 3.8$) and 20 were nonmusicians ($M_{\text{training}} = 0.3$ yrs, $SD = 0.5$). A musician was defined as someone who had completed at least two years of music training at the university level (either undergraduate or graduate). A nonmusician was defined as someone who had two years or less of music training by the age of 12 and did not currently play an instrument. All listeners had normal hearing and passed a pure-tone audiometric test using octave-spaced frequencies from 125 Hz to 8 kHz that required them to have thresholds at or below 20 dB HL (ISO 389–8, 2004; Martin & Champlin, 2000). In addition to the 40 participants whose data were analyzed, three were excluded from our analysis after having completed the experiment, as they did not fit the criteria for nonmusician or musician as determined by the post-experiment

TABLE 2. Mixed-effects Model Factors (Fixed Effects)

Variable	Experiments involved	Type	Description
Random Intercept	E1, E2, E3	Factor	Factor to account for within-subject individual differences
Segregation	E1, E3	DVs	Rated degree of segregation or blend of each excerpt (0-1)
Blend	E2		
Timbral Class	E1, E2, E3	Within-subject factor	Within- (0) or between-family (1) timbral combinations
Reorchestration	E3	Between-subjects factors	Original orchestration (0) or reorchestration (1)
Training	E1, E2, E3		Nonmusician (0) or Musician (1)
Spectral Centroid	E1	Acoustic factors	Average difference in spectral centroid
Frame Energy	E1, E3		Average difference in frame energy
Spectral Crest	E1, E3		Average difference in spectral crest
Spectral Flatness	E1, E2		Average difference in spectral flatness
Spectral Skewness	E1, E2		Average difference in spectral skewness
Spectral Variation	E1, E2, E3		Average difference in spectral variation
Average Interval	E1, E3	Score-based factors	Average difference in pitch
Crossing Proportion	E1, E2, E3		Proportion of note crossings
Onset Synchrony	E1, E2, E3		Proportion of synchronous onsets
Consonance	E2		Degree of consonance in harmonic intervals

Note: Difference measures for acoustic and score-based factors were calculated between two streams in Experiments 1 and 3 and between each instrument pair in Experiment 2. For details, see Appendices B and C.

questionnaire. The questionnaire revealed that these participants had received 1–2 years of music training after the age of 12. Four other participants were excluded at the screening phase because their hearing thresholds did not meet criterion.

Participants were recruited from the Montreal student community and were compensated 10\$ CAD/hour. All participants provided written consent, and the study was certified for ethical compliance by the McGill University Research Ethics Board II.

Stimuli. The stimuli consisted of the 36 two-stream excerpts and 12 single-stream excerpts (as annotated by expert analysts with varying degrees of segregation or blend, respectively), for a total of 48 stimuli that captured a wide range of degrees of segregation (see Appendices A1 and A2). Average sound levels for all 48 stimuli varied between 56.60 and 84.50 dB SPL (A-weighted). These are naturalistic levels for the various musical excerpts with differing dynamics.

Procedure. Before beginning the main part of the experiment, participants were shown four visual analogies in which color was used to segregate different forms. These examples served as an accessible parallel for participants to bridge their understanding of grouping concepts from the visual domain to those of the auditory domain (see Supplementary Materials, Figure S1). After viewing each visual example, participants were asked to explain in their own words, using the word “segregation,” what happened when color was introduced into the picture. They were only permitted to advance to the next visual

example once their answer clearly reflected an ease with, and understanding of, the concepts. A typical response from participants that demonstrated a clear understanding was “when the color was introduced, the shapes became segregated from one another,” whereas an answer that demonstrated that more practice was needed was, “when the color came on, I could see everything more clearly.”

Three practice trials followed. The practice excerpts were selected based on their annotated strength ratings and served as prototypical examples, capturing a broad range of degrees of segregation. They were always played in the same order.

After hearing each excerpt, participants rated the degree of segregation on a continuous horizontal scale from “integrated” (left, coded as 0) to “segregated” (right, coded as 1). They had the option to relisten to each excerpt once. Stimuli were randomly ordered across trials for each participant to minimize potential order effects. After each experiment, participants filled out a short questionnaire about their musical exposure and training. Upon completing the questionnaire, they were asked about their listening/rating strategies during the experiment. This was done in a face-to-face dialogue, to ensure that there was a clear understanding between the participant and researcher. The experiment lasted approximately one hour.

RESULTS

Our main hypothesis was that the degree of perceptual segregation would vary depending on timbral class and

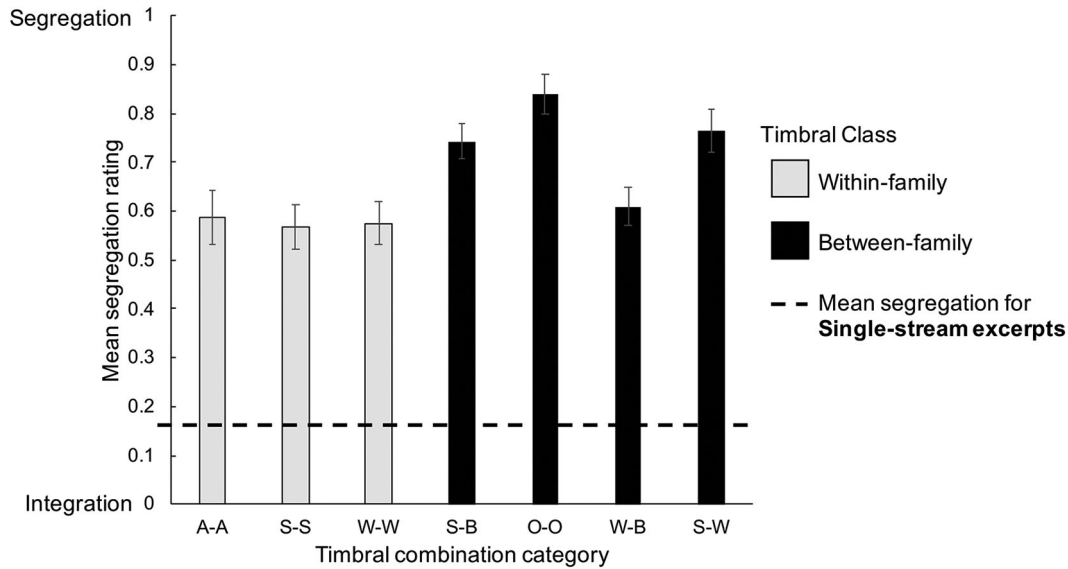


FIGURE 4. Mean segregation ratings across timbral combination categories. (A = all, S = strings, W = woodwinds, B = brass, O = other). Bars represent 95% confidence intervals.

training. Figure 4 presents the mean perceptual segregation ratings for each of the seven timbral combination categories for two-stream excerpts (bars) and the average of all single-stream excerpts (horizontal dashed line).

A low baseline rating of the average degree of segregation for single-stream excerpts ($M = 0.17$, $SD = 0.19$) indicates that the scale was well anchored. In order to assess whether the degree of segregation of single-stream excerpts varied according to timbral category, a one-way repeated measures ANOVA was performed and was significant after Greenhouse-Geisser correction for violations of sphericity, $F(3.85, 150.0) = 10.61$, $p < .001$, $\eta^2 = .214$. Bonferroni post hoc comparisons revealed that excerpts of the A category (containing strings, woodwinds, and brass) were less segregated than the other categories except SW (strings and woodwinds), and SB (strings and brass) excerpts were more segregated than all categories except S (strings) (see Figure 5).

Studying stream segregation within an ecologically valid setting requires disentangling the complex interaction between timbre and the many other perceptual attributes present in the context of real music. Therefore, in order to test the effect of timbre on the degree of perceptual segregation, while considering other musical parameters, a mixed-effects analysis was performed on the fixed effects listed in Table 2 for the two-stream excerpts.

From a hierarchical cluster analysis of intercorrelations among audio descriptors based on the ERB model,

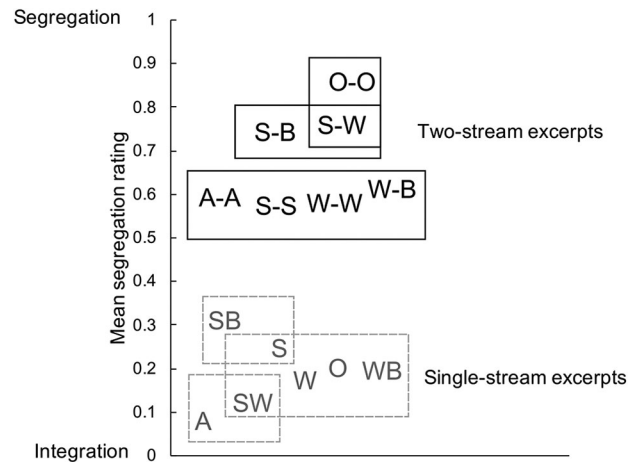


FIGURE 5. Schematic timbral category cluster representation of significant differences between means for timbral combination categories as revealed by Bonferroni-corrected contrast tests. Degree of segregation is organized along the vertical axis. Two-stream excerpts are written in black and bounded by solid boxes, whereas single-stream excerpts are written in gray and bounded by dashed boxes. The means for categories within boxes are not significantly different from one another. For example, for two-stream excerpts, S-B is not significantly different from S-W but is significantly different from O-O. Labels are spread along the x-axis for visibility, but only the position along the y-axis is relevant.

spectral centroid was selected over spectral slope, spectral spread, spectral rolloff, and spectral decrease, as it is the most cited and had the highest average correlation with the other descriptors in this cluster. Adding the

random intercept to the mixed-effects model improved BIC over the null model ($\Delta\text{BIC} = 91.0$). Adding all fixed effects (random intercept, timbral class, training, timbral class \times training, spectral centroid, frame energy, spectral crest, spectral flatness, spectral skewness, spectral variation, average interval, crossing proportion, and onset synchrony) to the mixed-effects model improved BIC over the random intercept model ($\Delta\text{BIC} = 357.3$). The efficient greedy variable selection approach led to the removal, in order, of spectral centroid ($\Delta\text{BIC} = 3.7$), frame energy ($\Delta\text{BIC} = 2.9$), training ($\Delta\text{BIC} = 1.1$), and the timbral class \times training interaction ($\Delta\text{BIC} = 0.9$). All remaining fixed effects were found to be statistically significant. There are significant individual differences among listeners as indicated by the effect of random intercept, $\mu = -.17$, $t(39) = 17.72$, $p < .0001$, $\sigma^2 = 0.009$, $z = 3.83$, $p < .0001$ (the intercept variance is different from zero). The effect of timbral class, $F(1, 1392) = 22.84$, $p < .0001$, suggests that between-family combinations elicit greater perceptual segregation than do within-family ones. Reduced overlap of several acoustic factors between streams yields greater perceptual segregation: spectral crest, $F(1, 1392) = 12.54$, $p = .0004$; spectral flatness, $F(1, 1392) = 12.01$, $p = .0005$; spectral skewness, $F(1, 1392) = 34.54$, $p < .0001$. However, greater overlap in spectral variation yields greater perceptual segregation, $F(1, 1392) = 26.87$, $p < .0001$. In terms of score-based factors, greater differences in pitch across perceptual streams yield greater perceptual segregation as evidenced by average interval, $F(1, 1392) = 37.75$, $p < .0001$. The more the streams cross each other in pitch and have synchronous onsets, the lower the segregation ratings: crossing proportion, $F(1, 1392) = 5.00$, $p = .026$; onset synchrony, $F(1, 1392) = 125.70$, $p < .0001$.

The same model-selection process was conducted with acoustic factors from the purely acoustic model (STFT) in order to compare the two models and to validate the perceptual relevance of the auditory processing (ERB) model. The results confirm that the auditory (ERB) model is preferable, as it provides the greatest reduction in BIC compared to the null model (ERB model $\Delta\text{BIC} = 525.8$ and STFT model $\Delta\text{BIC} = 462.3$), affording a richer picture of the relation between acoustic properties and perceptual segregation. The STFT model yields an identical final model in terms of the independent variables and score-based factors. However, in terms of acoustic factors, only spectral skewness and frame energy make significant contributions. This result suggests that this model is not able to account for as much timbral variance as the ERB model. Therefore, we used the ERB model as the primary input representation for the following two experiments.

The degree of segregation varied according to timbral combination category in the two-stream excerpts. Figure 5 provides a schematic representation of significant differences between means as revealed by contrasts between categories. Contrasts using least-squares (marginal) means from the final model and Bonferroni correction for multiple comparisons among timbral combination categories were computed in order to see how other musical factors in the model, aside from timbral class, affect mean segregation ratings (i.e., spectral crest, spectral flatness, spectral skewness, spectral variation, average interval, crossing proportion, and onset synchrony). In Figure 5, boxes delineate means that are not significantly different from one another. Within- and between-family categories form distinct clusters with the exception of W-B. Of the six excerpts in this category, one might be considered a within-family combination, given that both streams have both woodwinds and brass (Fl, Picc, Ob1, Ob2, Tp1 vs. Bn1, Hn1). To ensure that this did not affect our results, we excluded this WB-WB excerpt and repeated the model selection process. Doing so yielded the same final model (identical significant effects) with close quantitative agreement across all parameter estimates.

In order to explain the fact that the W-B between-family combination had a lower segregation rating on average than other between-family combinations, we ran contrasts along the seven acoustic and score-based factors from the final model structure on three groups derived from the timbral category contrast clustering: 1) W-B, 2) the remaining between-family combinations (S-W, S-B, O-O), and 3) the same-family combinations (S-S, W-W, A-A). The goal was to elucidate other parameters that may be competing and interacting with timbre's effect on segregation. The results from Bonferroni-corrected t -tests show that in comparison to the other between-family combinations W-B excerpts had less spectral overlap between streams in terms of spectral skewness, $t(1437) = -7.69$, $p < .0001$, spectral variation, $t(1437) = -6.83$, $p < .0001$, and spectral flatness, $t(1437) = -11.02$, $p < .0001$, as well as a smaller average interval $t(1437) = -12.96$, $p < .0001$, smaller crossing proportion, $t(1437) = -2.99$, $p = .009$, and greater onset synchrony, $t(1437) = 5.22$, $p < .0001$. Therefore, timbral, pitch, and rhythmic parameters may have contributed to the reduction of perceptual segregation of excerpts containing W-B combinations.

Discussion

Experiment 1 confirmed and extended the findings that timbral dissimilarity can induce stream segregation, by

demonstrating that differences in orchestral timbre can significantly contribute to segregation in real music in concert with other musical parameters. The degree of segregation varied according to timbral class in that between-family instrument combinations elicited higher degrees of segregation than did combinations of instruments from the same family with the exception of woodwind-brass combinations (all wind instruments). However, other specific timbral parameters such as spectral skewness, spectral variation, spectral flatness, as well as pitch- and rhythm-related parameters such as crossing proportion and onset synchrony, also affected mean segregation ratings perceived by listeners. Greater timbral difference, pitch distance and rhythmic asynchrony collectively promoted segregation.

Beyond the between- and within-family distinction, the specific acoustic attributes within the instruments that make up each timbral combination category may account for the variation observed in the contrast analysis of timbral combination categories. As timbre is a multidimensional perceptual phenomenon, some dimensions of timbre may make a stronger contribution than others in promoting perceptual segregation. Analyses of the spectra of the musical instrument sounds in the orchestral excerpts confirmed this hypothesis, suggesting that spectral crest, spectral flatness, spectral skewness, and spectral variation may be particularly important dimensions of timbre in eliciting segregation. Indeed, when these difference factors and those of crossing proportion and onset synchrony are reduced, even in between-family excerpts such as W-B, segregation is significantly reduced. Our findings are also consistent with the previous literature demonstrating that onset synchrony strongly affects the perception of fusion and thus works against stream segregation (Dannenbring & Bregman, 1978). This parameter also captures rhythmic differences between annotated streams, which are certain to play a role in orchestral segregation.

The highest average segregation was found for excerpts that included impulsive instruments like harp and pitched percussion (O-O). This is expected, as these instruments specifically differ from those of other timbral categories in their attack quality (sharp/soft). This dimension of timbre was quantified in a timbre space by McAdams et al. (1995) and might serve as a highly important factor in both sequential grouping and concurrent grouping, in which differences in attack quality hinder blend and promote segregation (Bey & McAdams, 2003; Iverson, 1995; Sandell, 1995; Tardieu & McAdams, 2012).

We anticipated that there would be an effect of music training on the degree of perceptual segregation, but this factor was not significant. This result is consistent with findings that show that both musicians and non-musicians can discriminate between timbres (Peynircioglu, Brent, & Falco, 2016). Alternatively, although there is a trend suggesting that musicians are biased towards rating excerpts as more segregated compared to non-musicians, this effect may not be pronounced due to the level of task difficulty. Group differences have been shown to be modulated by task difficulty (e.g., Arroyo-Anlló, Dauphin, Fargeau, Ingrand, & Gil, 2019; Reese & Polich, 2003; Vasuki, Sharma, Ibrahim, & Arciuli, 2017). Therefore, training effects may be more relevant when task difficulty is high and timbral differences are minimized or subtle.

Experiment 2: Blend Ratings on Individual Streams

The aim of Experiment 2 was to assess the relationship between the global degree of segregation of two-stream excerpts in Experiment 1 and the degree of integration of each of its single-stream constituents. It was hypothesized that an increase in intra-stream integration would lead to greater global segregation, as it may be easier to organize and segregate clearly integrated materials.

METHOD

Participants. Data analyses were performed on 44 participants (29 female), between the ages of 18 and 35 years ($M = 22.40$, $SD = 3.15$). Twenty-one participants were musicians ($M_{\text{training}} = 15.0$ yrs $SD = 4.10$) and 23 were nonmusicians ($M_{\text{training}} = 0.10$ yrs $SD = 0.10$). Of these, nine musicians and three nonmusicians had participated in Experiment 1. Two additional participants were excluded from the analysis after completing the experiment because they did not meet the criterion for nonmusician or musician, and three others did not meet the criterion for normal hearing.

Stimuli. The stimuli consisted of 72 single-stream excerpts. These excerpts were extracted from the 36 two-stream excerpts used in Experiment 1. Logic ProX (Apple Computer, Cupertino, CA) was used to isolate the instruments that analysts indicated as belonging to each stream. Some of the blended single streams had within- or between-family combinations (see Table A.1). Average sound levels for all 72 stimuli varied between 47.70 and 82.50 dB SPL (A-weighted). Note that the A-A stimuli are classed as between-family combinations for both streams [SWB] in two excerpts

(Strauss and Verdi) and for one stream [SW] in the third excerpt (Mussorgsky). The second stream in the Mussorgsky excerpt is classed as within-family [B]

Procedure. The procedure was identical to Experiment 1, except that the anchor labels on either side of the continuous scale were “blend” to the left (coded as 0) and “no blend” to the right (coded as 1). Participants rated each excerpt for its degree of perceived *integration*, expressed in terms of blend. The term “blend” was used in the instructions in place of integration as it is the word that musicians most often use to refer to the perceptual phenomenon of integration. Blend was defined as the merging of two or more sound events into one perceived sound event. For instance, if a given type of sound event is heard playing a single melody, one hears “blend.” On the other hand, if different instruments playing independently are heard as perceptually distinct, one perceives “no blend.” Listeners were informed that they might hear something in between these two extremes as with a partially blended combination of instruments. The experiment lasted approximately one hour.

RESULTS

In this study, we calculated mean values of acoustic and score-based factors for each individual stream by taking the average difference between instrument pairs. For example, if the excerpt contained flute, oboe and violin, difference measures were taken for the three pairs Fl-Ob, Fl-Vn, and Ob-Vn and then averaged. Based on hierarchical clustering, spectral flatness was chosen over spectral spread, spectral rolloff, spectral centroid, spectral decrease, spectral slope, spectral crest, and frame energy because it made a significant contribution in Experiment 1 and had the highest average correlation within the cluster. Spectral skewness was chosen over spectral kurtosis because it made a significant contribution in Experiment 1. A consonance factor was added to the mixed-effects model for this experiment.

Adding the random intercept improved BIC compared to the null model ($\Delta\text{BIC} = 139.70$). Adding all fixed effects to the model (Table 2) improved BIC compared to the random intercept model ($\Delta\text{BIC} = 565.0$). Timbral class \times training interaction ($\Delta\text{BIC} = 3.60$), training ($\Delta\text{BIC} = 3.70$), and crossing proportion ($\Delta\text{BIC} = 3.30$) were removed in that order from the model. The random intercept’s mean and variance were both different from zero, $\mu = .24$, $t(343) = 20.93$, $p < .0001$, $\sigma^2 = 0.009$, $z = 4.05$, $p < .0001$. All remaining fixed effect factors were found to be statistically significant: The effect of timbral class, $F(1, 3118) = 73.77$, $p < .0001$, demonstrates that between-family combinations

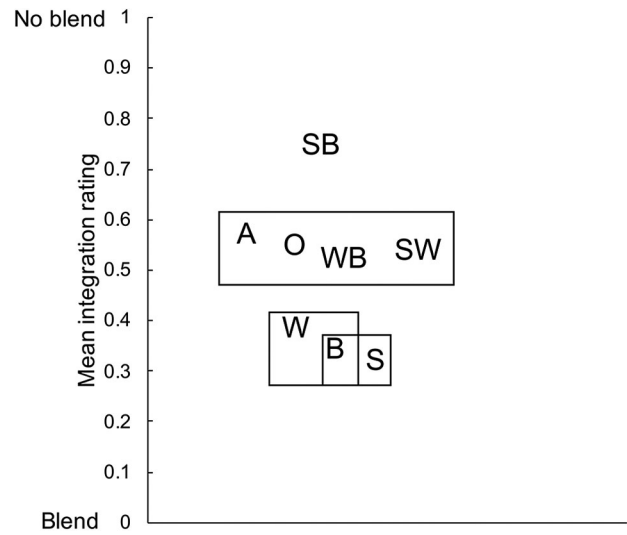


FIGURE 6. Mean blend ratings for each timbral combination category. Means that are not significantly different according to Bonferroni-corrected post hoc tests are enclosed in boxes.

were rated as less blended than within-family ones. The effect of spectral flatness reveals that greater overlap in this parameter between instrument families elicits less perceptual blend, $F(1, 3118) = 48.78$, $p < .0001$, but greater overlap in spectral skewness, $F(1, 3118) = 27.65$, $p < .0001$, and spectral variation, $F(1, 3118) = 32.04$, $p < .0001$, yield greater perceptual blend. Onset synchrony among parts within an excerpt elicits greater perceptual blend, $F(1, 3118) = 44.25$, $p < .0001$, as do more consonant pitch relations, $F(1, 3118) = 126.83$, $p < .0001$.

The mean blend results for each timbral combination category are presented in Figure 6. Boxes delineate means that are not significantly different from one another. One can see that within- and between-family categories form distinct clusters, with the string and brass combination also being significantly less blended than the other between-family combinations.

In order to explain why SB combinations were much less integrated than the other between-family categories, contrasts were computed comparing these combinations to all other between-family combinations along four factors from the final model structure: spectral flatness, spectral skewness, spectral variation, and consonance. Although significant in the final model structure, contrasts for SB combinations were not performed along the onset synchrony parameter, as there was a very strong floor effect: all pieces within the SB combination had an onset synchrony measure of 1, making it impossible to compute a regression fit. Compared to other between-

family combinations, the SB combinations contained significantly less overlap in spectral variation, $t(3166) = -2.73, p < .05$ and spectral skewness, $t(31166) = 4.73, p < .0001$, and were less consonant, $t(3166) = 4.49, p < .0001$.

Discussion

Experiment 2 served as a follow-up study in order to assess whether factors other than timbral differences per se, such as the degree of perceptual blend of the individual streams, had an effect on the global degree of segregation between streams. Although all excerpts in Experiment 1 were selected on the basis of annotated strength ratings (based on the score and listening to a commercial recording), only the 12 single-stream excerpts had been measured perceptually (Gianferrara, 2016; McAdams et al., 2016). It was assumed that in two-stream excerpts the individual streams containing multiple instruments would be totally fused and constitute a virtual voice.

The results demonstrated an effect of timbral class (instrument family combination), spectral flatness, spectral variation, onset synchrony, and consonance on blend. Note the similarity in explanatory factors between Experiments 1 and 2, although there were differences in the contributions of training and spectral crest. In addition, we included a measure of consonance, as this experiment contained single streams. Consonance is an emergent sound quality produced by two or more harmonic intervals. The use of sensory consonance and dissonance (roughness) has been argued to affect auditory stream segregation, where consonant intervals tend to blend more than dissonant ones (Wright & Bregman, 1987). The results showed a significant effect of consonance, where, as one might expect, greater consonance within the stream elicited stronger blend. This is consistent with previous work that documents the use of consonance and dissonance as a tool to shape segregation in polyphonic music as early as the 16th century (Wright & Bregman, 1987), as well as research demonstrating the role of harmonicity in perceptual fusion (McAdams, 1984).

We had hypothesized that the degree of blend in individual streams (Experiment 2) would be inversely related to the degree of inter-stream segregation measured in Experiment 1, in that the less blended the individual streams were, the more difficult it would be to organize the global material into two separate streams. The correlation between degree of global segregation of two-stream excerpts and degree of blend for single-stream excerpts was computed. Each participant's data set consisted of 36 two-stream segregation

ratings and 72 single-stream blend ratings. These were averaged over participants' data from Experiments 1 and 2, respectively. The variance in blend explained by segregation was not significant, $R^2(70) = .02, F(1, 70) = 1.71, p = .20$, suggesting that the degree of two-stream segregation and degree of single-stream blend vary independently of one another.

Many of the segregation ratings had corresponding blend ratings that were spread across the scale. Both extremely weak and extremely strong degrees of intra-stream blend are found for excerpts with strong inter-stream segregation. Alternatively, one might predict that the presence of less integrated streams in the mixture would lead to higher ratings of segregation given that the constituent instrument groups were not that blended to start with. In this case, one would expect a negative correlation between these measures, which is not the case either.

A final possibility would be that segregation between streams predicts integration within streams. Our original hypothesis assumes directionality, in that the integration of information into each single stream may precede and in turn affect global segregation between streams. However, evidence by Sussman (2005) suggests that integration of information into single streams occurs after global segregation and that single-stream constituents are affected by contextual factors. The issue of context effects was not addressed in the present study in that perceptual blend ratings might be different when hearing each stream in isolation and when hearing each stream concurrently in full context. Therefore, it remains unclear how differing degrees of segregation within each single layer are combined when heard together in context during global perceptions of segregation. This process may be more dynamic than we originally anticipated.

In Experiment 1, it was assumed that two-stream excerpts that contained multiple instruments in each stream would be totally fused and constitute a virtual voice. However, the results from Experiment 2 revealed that within-stream blend ratings were not always rated as highly blended. This raises issues related to the analysts' classification of "streams." Again, single-stream blend ratings were measured out of context, which may not entirely reflect the degree of single-stream blend when the excerpt is listened to in full context as occurred when analysts were making their annotations. More research is needed in order to further explore the interaction between integration within a single stream and segregation between streams in relation to the role that attention and context play in our perception of segregation in orchestral music.

There was no significant difference between musicians and nonmusicians. This finding is consistent with research that shows that nonmusicians have a global-processing advantage, but that this effect is solely driven by musicians' bias towards local processing (Ouimet, Foster, & Hyde, 2012; Stoesz, Jakobson, Kilgour, & Lewycky, 2007). In the current experiment, rating the degree of integration may have required participants to adopt a more synthetic or global listening strategy (Peretz, 1990; Wenhart & Altenmüller, 2019; Winkler, Denham, Mill, Bohm, & Bendixen, 2012), in which training benefits may not be as pronounced or useful. This finding highlights the importance of cognitive style and task strategy.

Experiment 3: Segregation Ratings on Reorchestrated Excerpts

The goal of this experiment was to provide a baseline measure of segregation by reducing timbral differences between streams. This baseline would allow us to evaluate the extent to which all other parameters contribute to the perception of segregation over and above timbre. Excerpts were reorchestrated and resimulated to be composed entirely of string instruments (violin, viola, cello, double bass). To assess the unique effect of timbre on segregation, we compared the results from Experiment 1 to those from Experiment 3. It should be noted, however, that 1) there were still perceptual differences in timbre between the instruments of the bowed-string family, and 2) within a given instrument, differences in pitch register were also accompanied by timbral differences (McAdams & Goodchild, 2017).

METHOD

Participants. Data analyses were performed on 40 participants (25 female), between the ages of 18 and 29 years ($M = 23.48$, $SD = 3.05$). Twenty participants were musicians ($M_{\text{training}} = 15.80$ yrs, $SD_{\text{training}} = 4.50$ yrs) and 20 were nonmusicians ($M_{\text{training}} = 0.10$ yrs, $SD_{\text{training}} = 0.20$ yrs). Two of the musicians had participated in Experiments 1 and 2. Two other listeners did not meet the criterion for normal hearing.

Stimuli. The stimuli consisted of 35 of the two-stream excerpts and the 12 single-stream excerpts from Experiment 1 for a total of 47 stimuli. One excerpt from Experiment 1 (Smetana, *Ma Vlast*, Die Moldau, mm. 218–228) was not included in Experiment 3, as it would have required a double digital string orchestra and was thus not technically feasible with the *OrchSim* system. The eight original, two-stream excerpts in Experiment 1

that were already composed for strings remained unaltered for Experiment 3. During the reorchestration, score-based information, such as pitch, rhythm, phrasing, and tempo fluctuations were kept constant and woodwind and brass instruments were substituted with string instruments of similar register (e.g., flute by violin, trombone by cello) (see Figure 2). Furthermore, if nonstring instruments in the excerpt were playing in unison with string instruments, these instruments were simply removed and the sound level of the string part was boosted, if necessary, to compensate for overall orchestral balance. In one excerpt (Vaughan Williams, Symphony No. 8, IV, mm. 87–90), the harp was kept, but its level was reduced to compensate. The 12 single-stream excerpts in Experiment 1 were used to anchor the scale but were presented in their original orchestration so that the anchoring would be similar between Experiments 1 and 3. Average sound levels across an excerpt for all 47 stimuli varied between 48.40 and 75.80 dB SPL (A-weighted).

Procedure. The procedure, instructions, and interface were identical to those in Experiment 1. The experiment lasted approximately one hour.

RESULTS

As mentioned earlier, our main hypothesis for Experiment 1 was that the degree of perceptual segregation would depend on timbral combination category and music training. This experiment sought to provide a baseline effect of reduced timbral cues on the degree of segregation.

We hypothesized that the degree of segregation would be significantly reduced for more timbrally homogeneous excerpts compared to timbrally heterogeneous excerpts. If instrument family can be operationalized in terms of global timbral classes, then reorchestrating the original excerpts to contain solely string instruments should significantly lower segregation ratings. We maintained the classification of excerpts as within- and between-family from Experiment 1 even though they are all within-family in Experiment 3. To compare Experiments 1 and 3, the factor “reorchestration” was included in the model as a between-subjects effect.

Based on hierarchical clustering of acoustic factors for the 35 reorchestrated two-stream excerpts, spectral skewness was chosen over spectral flatness and spectral variation because of these three factors, which contributed significantly in Experiments 1 and 2, spectral skewness had the highest correlation with the other descriptors in the cluster. Adding the random intercept improved BIC compared to the null model ($\Delta\text{BIC} =$

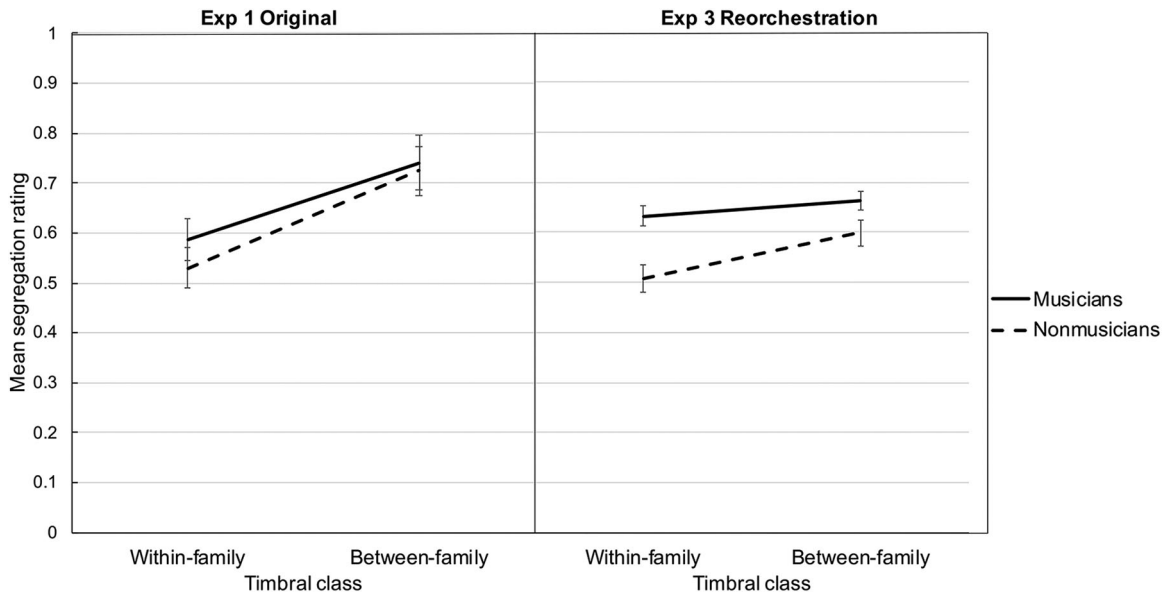


FIGURE 7. Mean segregation ratings according to timbral class and training for original and reorchestrated excerpts. Error bars represent 95% confidence intervals.

184.40). Adding fixed effects (Table 2) improved BIC compared to the random intercept model ($\Delta\text{BIC} = 711.70$). Mean frame energy ($\Delta\text{BIC} = 4.20$), the timbral class \times reorchestration \times training interaction ($\Delta\text{BIC} = 4.10$), the reorchestration \times training interaction ($\Delta\text{BIC} = 3.20$), and spectral crest ($\Delta\text{BIC} = 2.50$) were removed in that order. The random intercept's mean and variance were both different from zero, $\mu = -.03$, $t(76) = 24.21$, $p < .0001$, $\sigma^2 = 0.009$, $z = 5.06$, $p < .0001$. All remaining fixed effects were statistically significant. Figure 7 depicts mean segregation ratings for the excerpts in Experiments 1 (original) and 3 (reorchestrated) for musicians and nonmusicians according to previously categorized within- and between-family combinations (timbral class). We will focus on the independent variable factors involving reorchestration given that this is the main aim of this experiment. The main effect of reorchestration, $F(1, 2714) = 14.64$, $p = .0001$, reveals that original excerpts (Experiment 1) were rated as more segregated than reorchestrated ones (Experiment 3). However, as expected, this effect depends on its interaction with timbral class, $F(1, 2714) = 27.18$, $p < .0001$, given that the difference in perceived segregation between within- and between-family excerpts was more pronounced for original compared to reorchestrated ones. In terms of acoustic and score-based factors, greater overlap in spectral skewness led to lower perceived segregation, $F(1, 2714) = 74.80$, $p < .0001$. Greater

differences in pitch between the two streams elicited greater perceived segregation, $F(1, 2790) = 26.93$, $p < .0001$. A greater proportion of part crossing and greater onset synchrony between streams both yielded lower perceived segregation: crossing proportion, $F(1, 2714) = 28.54$, $p < .0012$; onset synchrony, $F(1, 2714) = 190.89$, $p < .0001$.

Discussion

The goal of Experiment 3 was to provide a baseline effect with limited timbral cues for the degree of segregation. The *OrchSim* technology manipulated the timbre difference between streams by reorchestrating the original timbrally heterogeneous excerpts with only string instruments. All other musical characteristics aside from timbre remained identical to the original excerpts, allowing us to specifically reduce the effect of timbre on segregation.

When comparing Experiment 1 to Experiment 3, the results confirm that the reorchestration significantly reduced timbral differentiation between streams. This finding demonstrates that timbre, operationalized in terms of instrument family combinations, contributes to perceptual segregation over and above nontimbral cues.

The fact that spectral skewness contributed to the final model in Experiment 3 highlights that fine-grained timbral differences related to spectral shape still

exist even after excerpts are reorchestrated to a homogeneous instrument family of similar orchestral register. That timbral class meaningfully contributed to the final model in Experiment 3 suggests that other musical features may covary with instrument family, such that reorchestrated excerpts may still vary systematically across the musical materials originally assigned to these timbral classes. Furthermore, subtle differences in timbre between violins, violas, celli, and contrabasses, still exist, even when playing at the same pitch. And importantly, timbre covaries with pitch register in any given instrument, and registral differences still contain timbre differences in addition to pitch differences (McAdams & Goodchild, 2017). These two factors make it impossible to *completely* isolate a timbral effect in real instrumental music. Therefore, timbre cannot be completely captured by our operational definition.

The results regarding score-based factors in the model are consistent with the previous two experiments, demonstrating that onset synchrony is an important cue for segregation. In addition, the proportion of part crossing was found to be statistically significant, in that greater part crossing was negatively related to segregation ratings. This finding is supported by previous work by Hartmann and Johnson (1991), who showed that the number of times musical lines cross one another in pitch reduces listeners' judgments of segregation. Just as in Experiment 1, the average pitch interval between streams was found to be a strong contributor to the models. This supports the notion that pitch distance serves as an important cue for perceptual segregation of orchestral music (Bregman & Pinker, 1978; Deutsch, 2013). As we have noted, however, register changes are also accompanied by timbral changes.

The interaction between timbral class and reorchestration indicates that there is a greater difference between timbral classes (within- and between-family) for original excerpts compared to reorchestrated ones. The main effect of timbral class, therefore, is likely driven by the original excerpts (Experiment 1).

The main effect of training may be due to musicians' heightened timbral sensitivity or adoption of an analytic listening strategy during the task. These findings support previous research showing that music training gives participants an advantage for "hearing out" individual streams (Başkent & Gaudrain, 2016; Beauvois & Meddis, 1997; Bey & McAdams 2003; Chandrasekaran & Kraus, 2010; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Zendel & Alain, 2009). Başkent and Gaudrain (2016), for example, found a strong advantage in musicians for speech-on-speech perception, which may have been due to better segregation acquired through music

training. This result supports findings that show that both musicians and nonmusicians can discriminate between timbres (Peynircioğlu, Brent, & Falco, 2016), but that perceptual exposure with feedback, such as music training, can further enhance perceptual sensitivity to timbral differences (Bates, Peynircioğlu, & Brent, 2019).

The training \times timbral class interaction, revealed by musicians' higher segregation ratings for within-family timbral combinations compared to nonmusicians, suggests that musicians' show a benefit in using subtle timbral cues to segregate musical content. This effect is driven by the reorchestrated excerpts (Experiment 3), as this interaction is not significant when rating the original excerpts (Experiment 1). Interestingly, benefits associated with music training may be modulated by task difficulty. Specifically, musicians and nonmusicians may both utilize clear timbral cues (Peynircioğlu et al., 2016), but when the timbral distance between streams is reduced and the cues become more subtle, musicians may be more adept at utilizing them. This idea is considered in further detail in the General Discussion section.

General Discussion

The aim of this paper was to examine the effect of timbre on the degree of perceptual segregation and the role of within-stream blend on between-stream segregation in musical excerpts drawn from the symphonic repertoire. It was hypothesized that composers use instrument-family combinations as a high-level tool to shape listeners' perceptions because they provide timbral dissimilarities that contribute to stream segregation. In addition, we wanted to assess whether certain timbral properties may be particularly important for segregation and whether musicians and nonmusicians differ in terms of their ability to perceive segregated streams. In order to capture the richness of a real musical context, nontimbral cues (acoustic and score-based) were also included in all statistical models to see how timbre contributes to perceptual segregation over and above these other cues (reinforcing them or competing with them). In line with orchestrators' intuitions, the results confirm that heterogeneous instrument combinations yield greater perceptual segregation than homogeneous ones.

We also tested whether the degree of blend within individual streams (Experiment 2) would be inversely related to the degree of inter-stream segregation measured in Experiment 1. We did not find a significant correlation between the two measures. This result

suggests that the degree of within-stream blend may not directly influence the perception of segregation between streams. Future work should examine a wider range of within-stream degrees of perceptual blend in order to test whether the relationship between the two measures is nonlinear. Furthermore, perceptual segregation was measured in a richer musical context compared to perceptual blend of single streams that was measured in isolation. Both perceptual measures could be assessed in the same musical context by manipulating attention. These potential avenues of future research may capture the dynamic interplay between within-stream blend and between-stream segregation.

The third experiment reorchestrated the excerpts with string instruments only and confirmed that overall perceived segregation across excerpts was lower when timbral differences between streams were reduced, demonstrating that timbre, in terms of instrument-family combinations, also contributes to perceptual segregation in real music. Most importantly, significant contributions of certain score-based and acoustic factors to the statistical models allowed us to pinpoint the musical properties that underly this effect.

SCORE-BASED FACTORS

Across all three studies, score-based results showed an effect on segregation and integration of average pitch interval between streams, part crossing, and onset synchrony, as well as an effect on within-stream blend of the consonance of harmonic pitch intervals. Furthermore, contrast analyses in Experiment 1 revealed that pitch and rhythmic parameters may compete or interact with timbral cues during a perceptual segregation task. Furthermore, in Experiment 2, contrast analyses also demonstrated that consonance may compete or interact with timbral cues in a perceptual blend task.

These findings demonstrate the complexities that arise in natural listening environments, in which factors, such as onset synchrony and consonance, which promote concurrent grouping and perceptual fusion, compete with the absence of part crossing, which promotes sequential grouping and auditory streaming (Bregman & Pinker, 1978; Wright & Bregman, 1987).

ACOUSTIC FACTORS

One of the primary contributions of this study is identifying the audio descriptors underlying the perception of segregation and integration in orchestral music. Specifically, acoustic analyses demonstrated that spectral variation, flatness, skewness, and crest may be particularly important correlates of timbre for auditory organization. These findings support those of Cusack and

Roberts (2004), in which differences in spectral variation were shown to elicit stronger perceptual segregation. Furthermore, our findings support previous research that found that differences related to spectral content are most important for the segregation of tone sequences (Gregory, 1994; Hartmann & Johnson, 1991; Singh & Bregman, 1997).

Research has not yet directly examined the role of spectral flatness and spectral crest in auditory stream segregation. However, these dimensions of timbre have been perceptually validated and found to be relevant for auditory perception (Koelsch et al., 2013; Laurier, 2011). Interestingly, spectral flatness has been found to be important for identifying musical phrases (Olsen, Dean, & Leung, 2016), and spectral crest has been implicated in emotional processing (Koelsch et al., 2013). Experiments 1 and 3 revealed a statistically significant effect of spectral skewness on segregation ratings. This correlate of timbre has been perceptually validated and found to be relevant for the classification of consonants (Forrest, Weismer, Milenkovic, & Dougall, 1988). Taken together, the acoustic analyses suggest that these acoustic cues are relevant for auditory segregation in orchestral music, which supports the notion that certain acoustic attributes of timbre predominate over others in relation to perceptual segregation.

MUSIC TRAINING

There was a main effect of training in Experiment 3, but not in Experiment 1. Additionally, there was a slight difference between musicians and nonmusicians in Experiment 1, but the effect was too weak to be considered reliable. However, the effect in both experiments was close enough to suggest that training did not significantly vary between the two experiments, as evidenced by the non-significant training \times reorchestration interaction. Task difficulty may moderate this group effect. The effect of training in Experiment 3, due to musicians' enhanced timbral sensitivity or analytic listening strategy, may be especially useful and pronounced when timbral cues are more subtle (higher task difficulty). Therefore, both groups may be able to use timbral cues in the original excerpts to segregate, but only musicians may have an advantage in the reorchestrated excerpts when timbral distance is minimized and cues are subtler. Further research is needed to tease these possibilities apart.

Similarly, a main effect of training was observed in Experiment 3 but not in Experiment 2, where participants rated the degree of integration or "blend." This finding highlights the importance of cognitive strategy on perceptual segregation of real music, in which

bottom-up and top-down factors interact (Judd, 1979; Kayser, Petkov, Lippert, & Logothetis, 2005; Tordini, Bregman, Ankolekar, Sanholm, & Cooperstock, 2013; van Noorden, 1975). It is possible that musicians adopt a more analytic strategy and that nonmusicians adopt a more global processing strategy (Ouimet et al., 2012; Stoesz et al., 2007). In the current study, when the task requires participants to segregate information, the difference in mean segregation ratings between musicians and nonmusicians are pronounced. However, when task demands are different and participants are required to integrate information, the difference between the two groups is minimized, as an analytic strategy is made less useful. Different cognitive strategies have also been found to be supported by separate neural correations. For example, perceptual segregation has been shown to recruit the planum temporale, whereas perceptual integration has been shown to recruit the inferior parietal sulcus (Ragert et al., 2014). Furthermore, the left planum temporale has been shown to be more enlarged in musicians than in nonmusicians (Schlaug, Jäncke, Huang, & Steinmetz, 1995). Thus, the results are consistent with previous findings that demonstrate that nonmusicians have a global-processing advantage, but that this effect is driven by musicians' bias towards local processing (Ouimet et al., 2012; Stoesz et al., 2007).

More research is needed in order to further explore the interaction between music training and segregation by musical timbre. An interesting direction of future research would be to investigate how trained composers and conductors differ from noncomposing musicians and from nonmusicians in their ability to segregate individual streams from a given orchestral excerpt.

Conclusions

A limitation of the current study pertains to the selection of the stimuli. First, the stimuli were selected based on specific criteria. The streams in two-stream excerpts were judged by analysts to be of equal perceptual prominence. These excerpts were selected to vary in terms of annotated strength rating (1–5) and the timbral combination categories involved. Single-stream excerpts were also required to vary in terms of timbral combination categories. Due both to the strict criteria and to what was available in *OrchARD*, the stimuli were not distributed evenly across timbral categories.

In sum, these findings provide a psychological basis for understanding how composers use timbral cues to shape listeners' perceptions in concert with other musical parameters. In addition, this study provides empirical evidence that timbral differences can be

operationalized in terms of instrument-family combinations. This certainly serves as an intuitive high-level tool that composers use to shape listeners' perceptions. These results substantiate the findings of past research that timbre plays a significant role in the process of auditory stream segregation and extends them to a corpus of orchestral excerpts. The primary contribution of this study is to lay the groundwork for creating a model that quantifies the weighted contribution of acoustical properties within orchestral timbres and score-based musical features, which together interact with training and cognitive strategy to affect segregation between instrument pairings within a musical framework. This model is schematized in Figure 8. Standardized beta coefficients for each factor demonstrate how the relative prominence of different cues varies across the different stimulus contexts (Table 3). Onset synchrony is particularly strong in Experiments 1 and 3 and less so in Experiment 2. The average inter-stream interval and part crossing, while both significant contributors, are of lesser importance. Among the timbral audio descriptors, spectral skewness plays a more important role in the perception of segregation (Experiments 1 & 3) than it does for blend (Experiment 2). Spectral flatness and variation are primarily effective in Experiments 1 and 2 in which multiple instrument families are involved and not at all in the case of the more homogeneous orchestration with strings in Experiment 3. In summary, the importance of spectral shape in general is key in the orchestral segregation effect in diverse orchestrations in concert with the more pitch-based factors. As these effects demonstrate basic principles of auditory organization, the findings could extend to musical contexts (genres and cultures) outside the Western orchestra. Future research could test the weighted contributions of perceptual and cognitive cues on perceptual segregation and how these relative contributions might change according to musical context.

Composers are truly masterful at shaping sound to elicit a particular percept in the listener. The goal of the current study was to bridge the gap between composers' intuitions and empirical research by connecting example-based orchestration practices to their underlying psychological principles. By using a novel and ecologically valid simulated orchestra and by combining behavioral, acoustic, and symbolic musical-score data, we have proposed a model for timbre's role in a controlled real-world musical context. By quantifying timbre's contribution and interaction with other musical features, we lay the groundwork for better understanding how general processes of auditory scene analysis, such as segregation, affect how music is composed and

TABLE 3. Standardized Beta Coefficients for Each Factor Across Experiments

Factor	Experiment 1	Experiment 2	Experiment 3
Timbral Class (TC)	.30	.34	.05
Reorchestration (R)	—	—	-.12
Training (T)	Removed	Removed	.40
TC X R	—	—	.33
T X R	—	—	Removed
TC X T	Removed	Removed	.17
Spectral Crest	.10	—	Removed
Spectral Flatness	.11	-.13	—
Spectral Skewness	.17	.08	.19
Spectral Variation	-.16	.12	—
Average Interval	.17	—	.12
Crossing Proportion	-.07	Removed	-.12
Onset Synchrony	-.29	-.11	-.22
Consonance	—	-.20	—

Note: “—” indicates that the factor was not included in the initial fixed effects model. Factors that were removed by the greedy variable selection process are indicated. Factors that were always removed by this process are not included in the table. Degree of segregation (Experiment 1 & 3) = Higher values indicate greater segregation; Degree of integration (Experiment 2) = Higher values indicate greater segregation; Timbral Class = Between-family (1) - Within-family (0); Training = Musician (1) - Nonmusician (0); Reorchestration = Reorchestrated (1) - Original (0). Positive values indicate that the first term has a higher degree of segregation (Experiments 1 & 3) or lower degree of blend (Experiment 2) than the second term. Negative values indicate that the second term has a higher degree of segregation or lower degree of blend than the first term. Acoustic factors: Higher values indicate greater separation in spectral overlap; Crossing Proportion: Higher values indicate more part crossing; Onset Synchrony: Higher values indicate greater synchrony; Consonance: Higher values indicate greater consonance.

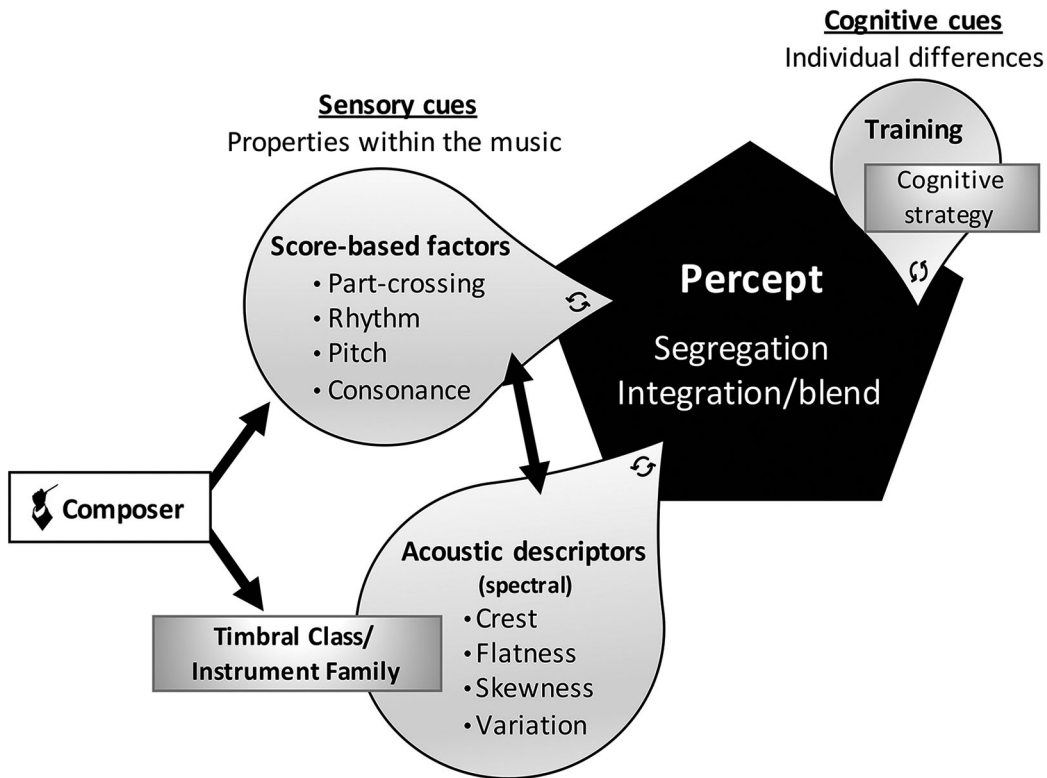


FIGURE 8. Final model: A framework for the contributions of low-level sensory cues (timbral and score-based) and cognitive cues on the perception of segregation and integration in orchestral music.

listened to. Surprisingly, we also found that the degree of blend within an auditory stream was not significantly related to the degree of segregation between streams. Taken together, this research serves as a basis for developing a set of perceptual principles to scientifically ground the orchestral segregation effect and for understanding the processes that underlie auditory musical perception more generally.

Author Note

This work was supported by grants from the Canadian Natural Sciences and Engineering Research Council (RGPIN 2015-05280), the Fonds de recherche du

Québec—Société et culture (2017-SE-205667), and a Canada Research Chair (950-223484) awarded to Stephen McAdams. The authors thank Bennett K. Smith for programming the experimental interface, Nathaniel Condit-Schultz for help with the computation of score-based factors, Félix Frédéric Baril for reorchestrating and creating the orchestral simulation of the excerpts in Experiment 3, and OrchPlayMusic, Inc. for granting us access to the original simulated orchestral excerpts.

Correspondence concerning this article should be addressed to Stephen McAdams, Schulich School of Music, McGill University, 555 rue Sherbrooke ouest, Montréal, QC, Canada H3A 1E3. E-mail: stephen.mcadams@mcgill.ca

References

- ARROYO-ANLLÓ, E. M., DAUPHIN, S., FARGEAU, M. N., INGRAND, P., & GIL, R. (2019). Music and emotion in Alzheimer's disease. *Alzheimer's Research and Therapy*, *11*(1), 69. <https://doi.org/10.1186/s13195-019-0523-y>
- BASKENT, D., & GAUDRAIN, E. (2016). Musician advantage for speech-on-speech perception. *Journal of the Acoustical Society of America*, *139*(3), EL51–EL56.
- BATES, K. B., PEYNIRCIOĞLU, Z. F., & BRENT, W. (2019). Learning to discriminate new timbres. *New Ideas in Psychology*, *55*, 1–7. <https://doi.org/10.1016/j.newideapsych.2019.04.001>
- BEAUVOIS, M. W., & MEDDIS, R. (1997). Time decay of auditory stream biasing. *Perception and Psychophysics*, *59*(1), 81–86. <https://doi.org/10.3758/BF03206850>
- BERLIOZ, H., & STRAUSS, R. (1948). *Treatise on instrumentation*. (T. Front, Trans.). Edwin F. Kalmus. (Original work published 1905)
- BEY, C., & McADAMS, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 267–279.
- BOULIANE, D., & BARIL, F. F. (n.d.). OrchPlayMusic Library, <https://orchplaymusic.com>, last accessed October 18, 2019.
- BREGMAN, A. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.
- BREGMAN, A., & CAMPBELL, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244–249.
- BREGMAN, A., & PINKER, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, *32*(1), 19–31.
- CHANDRASEKARAN, B., & KRAUS, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*(2), 236–246. <https://doi.org/10.1111/j.1469-8986.2009.00928.x>
- CUSACK, R., & ROBERTS, B. (2004). Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. *Hearing Research*, *193*(1–2), 95–104. <https://doi.org/10.1016/j.heares.2004.03.009>
- CUTHBERT, M., & ARIZA, C. (2010). *Music21: A toolkit for computer-aided musicology and symbolic music data*. Paper presented at the Proceedings of the 11th International Society for Music Information Retrieval Conference, Utrecht.
- DANNENBRING, G. L., & BREGMAN, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex waves. *Perception and Psychophysics*, *24*(4), 369–376.
- DEUTSCH, D. (2013). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 183–248). Elsevier.
- DISBERGEN, N. R., VALENTE, G., FORMISANO, E., & ZATORRE, R. J. (2018). Assessing top-down and bottom-up contributions to auditory stream segregation and integration with polyphonic music. *Frontiers in Neuroscience*, *12*, 121. <https://doi.org/10.3389/fnins.2018.00121>
- FORREST, K., WEISMER, G., MILENKOVIC, P., & DOUGALL, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*(1), 115–123. <https://doi.org/10.1121/1.396977>
- GIANFERRARA, P. (2016). *Perceptual data for a relational database on orchestration: The case of timbral blend* (Unpublished honours thesis). McGill University, Montreal, Canada.
- GLASBERG, B. G., & MOORE, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138.
- GOODCHILD, M., & McADAMS, S. (2018). Perceptual processes in orchestration. In E. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre*. Oxford University Press.
- GREGORY, A. H. (1994). Timbre and auditory streaming. *Music Perception*, *12*(2), 161–174.

- GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5), 1270–1277. <https://doi.org/10.1121/1.381428>
- HARTMANN, W. M., & JOHNSON, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, 9(2), 155–183.
- ISO 389–8. (2004). *Acoustics-Reference zero for the calibration of audiometric equipment-Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones* (Tech. Rep). Geneva, Switzerland: International Organization for Standardization.
- IVERSON, P. (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4), 751–763.
- JUDD, T. (1979). Comments on Deutsch's musical scale illusion. *Perception and Psychophysics*, 26(1), 85–92. <https://doi.org/10.3758/BF03199863>
- KASS, R. E., & RAFTERY, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.2307/2291091>
- KAYSER, C., PETKOV, C. I., LIPPERT, M., & LOGOTHETIS, N. K. (2005). Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology*, 15(21), 1943–1947. <https://doi.org/10.1016/j.cub.2005.09.040>
- KENDALL, R. A., & CARTERETTE, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review*, 9(1), 51–67.
- KENNETH, P. B., & ANDERSON, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. Springer.
- KOELSCH, S., SKOURAS, S., FRITZ, T., HERRERA, P., BONHAGE, C., KÜSSNER, M. B., & JACOBS, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage*, 81, 49–60. <https://doi.org/10.1016/j.neuroimage.2013.05.008>
- KOFFKA, K. (1935). *Principles of gestalt psychology*. Harcourt, Brace & World.
- LAURIER, C. (2011). *Automatic classification of musical mood by content-based analysis* (Unpublished doctoral dissertation). Universitat Pompeu Fabra, Barcelona, Spain.
- LEMBKE, S.-A., & McADAMS, S. (2015). The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. *Acta Acustica united with Acustica*, 101(5), 1039–1051. <https://doi.org/10.3813/AAA.918898>
- LEMBKE, S.-A., PARKER, K., NARMOUR, E., & McADAMS, S. (2019). Acoustical correlates of perceptual blend in timbre dyads and triads. *Musicae Scientiæ*, 23(2), 250–274. <https://doi.org/10.1177/1029864917731806>
- MALMBERG, C. F. (1918). The perception of consonance and dissonance. *Psychological Monographs*, 25(2), 93–133. <https://doi.org/10.1037/h0093119>
- MARTIN, F. N., & CHAMPLIN, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, 11(2), 64–66.
- McADAMS, S. (1984). The auditory image: A metaphor for musical and psychological research on auditory organization. In R. Crozier & A. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 289–323). Elsevier Science B.V. (North-Holland).
- McADAMS, S. (2019a). Timbre as a structuring force in music. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 211–243). Springer International Publishing.
- McADAMS, S. (2019b). The perceptual representation of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 23–57). Springer International Publishing.
- McADAMS, S., & BREGMAN, A. (1979). Hearing musical streams. *Computer Music Journal*, 3(4), 26–43.
- McADAMS, S., GIANFERRARA, P., SODEN, K., & GOODCHILD, M. (2016). Factors influencing instrumental blend in orchestral excerpts. In *Proceedings of the 14th International Conference on Music Perception and Cognition*, San Francisco, CA.
- McADAMS, S., & GOODCHILD, M. (2017). Musical structure: Sound and timbre. In R. Ashley & R. Timmers (Eds.), *Routledge companion to music cognition* (pp. 129–139). New York: Routledge.
- McADAMS, S., & GOODCHILD, M. (2018). Perceptual evaluation of the recording and musical performance quality of DOSim and OrchSim orchestral simulations compared to professional recordings. *MPCL Technical Report MPCL-001*. Retrieved from https://mcgill.ca/mpcl/files/mpcl/mcadams_2018_mpcl.pdf. Last accessed May 28, 2020.
- McADAMS, S., GOODCHILD, M., & SODEN, K. (2021). *A taxonomy of perceptual effects of orchestration related to auditory grouping principles*. Manuscript in preparation.
- McADAMS, S., & SIEDENBURG, K. (2019). Perception and cognition of musical timbre. In P. J. Rentfrow & D. J. Levitin (Eds.), *Foundations of music psychology: Theory and research cognition* (pp. 71–120). MIT Press.
- McADAMS, S., WINSBERG, S., DONNADIEU, S., DE SOETE, G., & KRIMPHOFF, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58, 177–192.
- MILLER, J. R., & CARTERETTE, E. C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America*, 118(1), 471–482.
- MOORE, B. C. (1986). Parallels between frequency selectivity measured psychophysically and in cochlear mechanics. *Scandinavian Audiology, Supplementum*. 25, 139–152.

- MOORE, B. C., & GLASBERG, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74, 750–753.
- OLSEN, K. N., DEAN, R. T., & LEUNG, Y. (2016). What constitutes a phrase in sound-based music? A mixed-methods investigation of perception and acoustics. *PLoS ONE* 11(12): e0167643. <https://doi.org/10.1371/journal.pone.0167643>
- OUIMET, T., FOSTER, N. E. V., & HYDE, K. L. (2012). Auditory global-local processing: Effects of attention and musical experience. *Journal of the Acoustical Society of America*, 132, 2536–2544. <https://doi.org/10.1121/1.4747009>
- PARBERY-CLARK, A., SKOE, E., LAM, C., & KRAUS, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, 30, 653–661. <https://doi.org/10.1097/AUD.0b013e3181b412e9>
- PEETERS, G., GIORDANO, B. L., SUSINI, P., MISDARIIS, N., & MCADAMS, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, 130(5), 2902–2916. <https://doi.org/10.1121/1.3642604>
- PERETZ, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain: A Journal of Neurology*, 113(4), 1185–205. <https://doi.org/10.1093/brain/113.4.1185>.
- PEYNIRCIOĞLU, Z. F., BRENT, W., & FALCO D. E. (2016). Perception of blended timbres in music. *Psychology of Music*, 44(4), 625–639. <https://doi.org/10.1177/0305735615578313>
- PLOMP, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397–410). Sijthoff.
- RAGERT, M., FAIRHURST, M. T., & KELLER, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLoS ONE*, 9(1), e84085. <https://doi.org/10.1371/journal.pone.0084085>
- REESE, C., & POLICH, J. (2003). Alcoholism risk and the P300 event-related brain potential: Modality, task, and gender effects. *Brain and Cognition*, 53(1), 46–57. [https://doi.org/10.1016/S0278-2626\(03\)00202-1](https://doi.org/10.1016/S0278-2626(03)00202-1)
- REUTER, C. (2003). Stream segregation and formant areas. In R. Kopiez, R. Lehmann, A. C. Wolther, & C. Wolf (Eds.) *Proceedings of the 5th triennial ESCOM conference* (pp. 329–331). Osnabrück, Germany: epOs-music
- SANDELL, G. J. (1995). Roles for spectral centroid and other factors in determining “blended” instrument pairings in orchestration. *Music Perception*, 13(2), 209–246.
- SANDELL, G., & CHRONOPOULOS, M. (1997). Perceptual constancy of musical instrument timbres; Generalizing timbre knowledge across registers. In A. Gabrielsson (Ed.), *Proceedings of the third triennial ESCOM conference* (pp. 222–227). Uppsala, Sweden: Uppsala University.
- SCHLAUG, G., JÄNCKE, L., HUANG, Y., & STEINMETZ, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, 267(5198), 699–701. DOI: 10.1126/science.7839149
- SCHWARZ, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, 6, 461–464.
- SINGH, P. G., & BREGMAN, A. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America*, 102(4), 1943–1952.
- SMITH, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. In *Society for Music Perception and Cognition Conference’95*. Berkeley, CA: University of California, Berkeley.
- SNYDER, J. S., GREGG, M. K., WEINTRAUB, D. M., & ALAIN, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology*, 3(15). <https://doi.org/10.3389/fpsyg.2012.00015>
- STOESZ, B., JAKOBSON, L., KILGOUR, A., & LEWYCKY, S. (2007). Local processing advantage in musicians: Evidence from disembedding and constructional tasks. *Music Perception*, 25, 153–165. <https://doi.org/10.1525/MP.2007.25.2.153>
- SUSSMAN, E. S. (2005). Integration and segregation in auditory scene analysis. *Journal of the Acoustical Society of America*, 117(3), 1285–1298. <https://doi.org/10.1121/1.1854312>
- TARDIEU, D., & MCADAMS, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*, 30(2), 117–128.
- TORDINI, F., BREGMAN, A., ANKOLEKAR, A., SANHOLM, T. E., & COOPERSTOCK, J. R. (2013). *Toward an improved model of auditory saliency*. Paper presented at the 19th International Conference Auditory Displays (ICAD2013), Łódź, Poland.
- UHLIG, M., FAIRHURST, M. T., & KELLER, P. E. (2013). The importance of integration and top-down saliency when listening to complex multi-part musical stimuli. *Neuroimage*, 77, 52–61. <https://doi.org/10.1016/j.neuroimage.2013.03.051>
- VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* (Unpublished Ph.D. dissertation). Eindhoven University of Technology, Eindhoven, The Netherlands.
- VASUKI, M. P. R., SHARMA, M., IBRAHIM, R. K., & ARCIULI, J. (2017). Musicians’ online performance during auditory and visual statistical learning tasks. *Frontiers in Human Neuroscience*, 11, 114. <https://doi.org/10.3389/fnhum.2017.00114>
- WENHART, T., & ALTENMÜLLER, E., (2019). A tendency towards details? Inconsistent results on auditory and visual local-to-global processing in absolute pitch musicians. *Frontiers in Psychology*, 10, 1664–1078. <https://doi.org/10.3389/fpsyg.2019.00031>.
- WESSEL, D. L. (1973). Psychoacoustics and music: A report from Michigan State University. *PACE: Bulletin of the Computer Arts Society*, 30, 1–2.

- WINKLER, I., DENHAM, S., MILL, R., BOHM, T. M., & BENDIXEN, A. (2012). Multistability in auditory stream segregation: A predictive coding view. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1591), 1001–1012. <https://doi.org/10.1098/rstb.2011.0359>
- WRIGHT, J. K. & BREGMAN, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*, 2(1), 63–92. <https://doi.org/10.1080/07494468708567054>
- ZATORRE, R. J. (2013). Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science*, 342(6158), 585–589. <https://doi.org/10.1126/science.1238414>
- ZENDEL, B. R., & ALAIN, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21, 1488–1498. <https://doi.org/10.1162/jocn.2009.21140>

Appendix A

TABLE A1. Two-stream Excerpts Annotated by Analysts as Containing Two Blended Streams

Timbral combination category	Timbral class	Stream combination	Excerpt	Instruments in stream 1	Instruments in stream 2	Annotated strength
S-S	Within	S-S	Beethoven, Symphony 7, II, 51-58	Vn1 [S]	Vn2 [S]	4
		S-S	Berlioz, <i>Symphonie Fantastique</i> , IV, 33-40	Vn1&2, Va [S]	Va, Vc, Cb [S]	4
		S-S	Bruckner, Symphony 6, II, 113-117	Vn1 [S]	Vc [S]	5
		S-S	Schubert Symphony 9 III, 198-205	Vn1 [S]	Vc [S]	3
		S-S	Smetana, <i>The Bartered Bride</i> , Overture, 31-35	Vn1 [S]	Vn2 [S]	3
W-W	Within	S-S	Vaughan Williams, <i>The Lark Ascending</i> , 82-90	Vn solo [S]	Vn1 [S]	5
		W-W	Brahms, Symphony 4, IV, 18-24	Fl&2, Ob1&2 [W]	Cl&2, Bn&2 [W]	3
		W-W	Mendelssohn, Symphony 3, II, 32-39	Fl1&2, Ob1&2 [W]	Cl1 [W]	3
		W-W	Mendelssohn, Symphony 3, IV, 176-179	Fl1 [W]	Ob1 [W]	2
		W-W	Mussorgsky (orch. Ravel), <i>Pictures at an Exhibition</i> , III, 7-8	Fl1-3, Cl1 [W]	Bn1 [W]	2
A-A	Between	SW-B	Vaughan Williams, Symphony 8, I, 144-150	Ob1 [W]	Cl1 [W]	5
		SWB-SWB	Vaughan Williams, Symphony 8, IV, 69-71	Fl1, Ob1, Cl1 [W]	Bn1&2 [W]	3
		SWB-SWB	Mussorgsky (orch. Ravel), <i>Pictures at an Exhibition</i> , II, 76-81	Picc1, Fl1&2, Ob1-3, Cl&2, Vn1&2, Va [SW]	Tb3, Tu [B]	5
		SWB-SWB	Strauss, <i>Tod und Verklärung</i> , 433-437	Fl3, Tp3, Vn1&2 [SWB]	BCl, Bn1&2, Hn3&4, Vc, Cb [SWB]	5
		SWB-SWB	Verdi, <i>Aida</i> , Dance, 53-56	Fl1&2, Picc, Ob1&2, Cl1&2, Hn1-4, Tp1&2, Vn1&2 [SWB]	Bn1&2, Tb1-3, BTb, Va, Vc, Cb [SWB]	4
W-B	Between	W-B	Beethoven, Symphony 7, II, 119-122	Cl1 [W]	Hn2 [B]	4
		W-B	Mahler, Symphony 1, I, 290-295	Fl1, Ob1, Cl1 [W]	Hn1 [B]	3
		W-WB	Schubert, Symphony 9, III, 345-360	Fl1&2, Ob1&2 [W]	Cl1&2, Bn1&2, Hrn1, ATb, Tb [WB]	2
		W-WB	Schubert, Symphony 8, I, 31-35	Fl1, Ob1, Hn1&2 [WB]	Fl2, Ob2, Cl1&2, Bn1&2 [W]	1
		W-WB	Vaughan Williams, Symphony 8, I, 150-154	Fl1&2, Ob1&2, Bn2 [W]	Cl1&2, Bn1, Hn1&2, Tp1&2 [WB]	3
B-S	Between	S-SB	Vaughan Williams, Symphony 8, II, 100-102	Fl1, Picc, Ob1&2, Tp1 [WB]	Bn1, Hn1 [WB]	4
		S-SB	Borodin, <i>In the Steppes of Central Asia</i> , 40-53	Hn2, Vn1 [SB]	Va, Vc [S]	1
		S-B	Debussy, <i>La Mer</i> , III, 137-139	Tp1 [B]	Vn1 [S]	4
		S-SB	Mahler, Symphony 1, IV, 277-281	Hn1&2, Va, Vc [SB]	Vn1&2 [S]	4
		S-B	Mendelssohn, Symphony 3, II, 260-266	Hn1&2 [B]	Vn1 [S]	3
S-SB	Between	S-SB	Schubert, Symphony 9, I, 594-603	Vn1&2, Va [S]	Tp1&2, ATb, Tb, BTb, Vc, Cb [SB]	5
		S-B	Smetana, <i>Ma Vlast</i> , Die Moldau, 218-228	Hn1-4, Tb1-3, [B]	Vn1&2, Va, Vc, Cb [S]	4

(continued)

TABLE A1. (continued)

Timbral combination category	Timbral class	Stream combination	Excerpt	Instruments in stream 1	Instruments in stream 2	Annotated strength
W-S	Between	W-S	Berlioz, <i>Symphonie Fantastique</i> , IV, 49-61	Bn1&2 [W]	Vn1, Vn2, Va, Vc, Cb [S]	5
		S-SW	Mozart, <i>Don Giovanni</i> , Overture, 129-133	Vn1&2, Va, Vc, Cb [S]	Fl1&2, Ob1&2, Cl1&2, Bn1&2 [SW]	2
		W-SW	Schubert, <i>Symphony 9</i> , IV, 543-558	Fl1, Cl1 [W]	Ob1&2, Vn1&2, Va, Vc [SW]	4
		W-S	Sibelius, <i>Symphony 2</i> , II, 200-202	Fl1&2, Ob1&2 [W]	Vn1, Vc [S]	3
		W-SW	Vaughan Williams, <i>Symphony 8</i> , IV, 59-66	Ob1&2, Vn1&2, Vc [SW]	Fl1, Picc, Cl1&2, Bn1&2 [W]	5
O-O	Between	S-SW	Verdi, <i>La Traviata</i> , Prelude, 29-36	Cl1, Bn1, Vc [SW]	Vn1 [S]	5
		SWB-SWP	Mussorgsky (orch. Ravel), <i>Pictures at an Exhibition</i> , II, 86-98	Fl1-3, Ob1-3, Cl1&2, Hp1, Vn1&2, Va, Xylo [SWP]	Bn1&2, Hn1&3, Tb3, Vc, Cb [SWB]	4
		S-SWP	Mussorgsky (orch. Ravel), <i>Pictures at an Exhibition</i> , XIV, 121-123	Picc 1, Fl1&2, Ob1&2, Cl1&2, Vn1&2, Xylo [SWP]	Va, Vc [S]	2
		SWH-P	Vaughan Williams, <i>Symphony 8</i> , IV, 87-95	Ob1, Cl1, Hp1, Va, Vc [SWH]	Csta [P]	4

Note: Annotated segregation strength is taken from *OrchARD*. S = string, W = woodwind, B = brass, A = all (SWB), O = other (including H = harp, P = percussion). Unseparated initials in square brackets indicate blended families (SW = blend of strings and woodwinds), and initials separated by a hyphen indicate combinations in different streams (SW-B = SW in one stream, B in the other stream).

TABLE A2. Single-stream Excerpts Annotated As a Blended Passage

Timbral combination category	Excerpt	Instruments	Annotated strength	Mean blend rating
S	Hadyn, <i>Symphony 100</i> , III, 50-56	Vn2, Va	2	.41
	Bruckner, <i>Symphony 6</i> , I, 209-216	Vn2, Va, Vc, Cb	3	.58
W	Strauss, <i>Tod und Verklärung</i> , 452-454	Fl1&2, Ob1&2, BCl, Bn1&2	4	.40
	Schubert, <i>Symphony 9</i> , IV, 543-558	Fl1, Cl1	4	.56
A	Verdi, <i>Aida</i> , Dance, 45-53	Picc, Fl1, Ob1&2, Cl1&2, Tp1&2, Vn1&2	3	.61
WB	Mussorgsky (orch. Ravel), <i>Pictures at an Exhibition</i> , II, 61-63	Fl1-3, Ob1&2, EH, Cl1&2, BCl, Bn1&2, CBn, Hn1&2	4	.35
	Debussy, <i>La Mer</i> , I, 71-72	Bn1&2, Hn1-4	3	.56
SB	Mussorgsky (orch. Ravel), <i>Pictures at an Exhibition</i> , II, 68-69	Hn1-4, Tp1, Vn1&2, Va, Vc	2	.61
	D'Indy, <i>Choral Varié</i> , 70-74	Hn1-4, Tp1, Cb	3	.21
SW	Mussorgsky (orch. Ravel), <i>Pictures at an Exhibition</i> , X, 1-2	EH, Cl1&2, BCl, Bn1&2, Vn1&2, Va, Vc, Cb	5	.52
	Mozart, <i>Don Giovanni</i> , Overture, 23-26	Fl1&2, Vn1	4	.31
O	Vaughan Williams, <i>Symphony 8</i> , IV, mm. 81-86	Cl1, Hp1, Va	5	.39

Note: Average blend rating (0-1) is from Gianferrara (2016). (See Table A1 for abbreviations.)

Appendix B

Acoustic Descriptor Dictionary

TABLE B1. Acoustic Factors Defined in Terms of Peeters et al. (2011)

Audio descriptor	Description
Spectral Centroid	The spectral center of gravity. The ‘brightness/darkness’ of a sound.
Spectral Spread	The spread of the spectrum around its mean value.
Spectral Skewness	A measure of asymmetry of the spectrum around its mean value.
Spectral Kurtosis	The degree of flatness of a spectrum around its mean value.
Spectral Slope	Linear regression over the spectral amplitude values of each frequency bin (STFT) or auditory channel (ERB).
Spectral Decrease	The average set of slopes between the k th partial and the first partial.
Spectral Rolloff	Frequency below which 95% of signal energy is contained.
Spectral Variation	The amount of change in the spectrum over time.
Frame Energy	The sum of the squared amplitudes of each frequency bin (STFT) or auditory channel (ERB).
Spectral Flatness	Tonal vs. noise content of the spectrum: Ratio of geometric and arithmetic means of the spectrum.
Spectral Crest	Tonal vs. noise content of the spectrum: Ratio of the maximum value and arithmetic mean of the spectrum.

Appendix C

Score-based Factor Calculations

For all experiments:

1. MusicXML scores are input to music21 (Cuthbert & Ariza, 2010).
2. MIDI pitch values are averaged for chords.

Experiments 1 and 3: Inter-stream calculations

These calculations are computed between a pair of streams for each excerpt.

Onset synchrony:

1. Create a composite rhythm combining all instruments in an annotated stream.
2. The proportion of common onsets between streams is the onset synchrony score.

Pitch-related measures: Average Interval and Crossing Proportion

1. Rests are included.
2. Each stream is sliced into every possible triplet 64th-note position. This assures that the pitch averaging scores reflect the durations of notes.
3. Every instrument in each stream is paired with every instrument in the opposite stream, and the difference between their pitches at every slice is computed.
 - a. Average Interval: The average of absolute pitch differences is taken across each triplet 64th-note slice and the average value for the slices is

computed to get the average interval between streams.

- b. Crossing Proportion: For each triplet 64th-note slice, we compute the proportion of slices in which any intervals are negative: Call this value x . If x is greater than 0.5, this means that stream 1 contains pitches higher than stream 2 more than half the time. We report $(1-x)$ as the cross proportion, as these represent places where stream 2 contains notes higher than stream 1. If x is less than 0.5, it means stream 2 is higher than stream 1 most of the time, so we just report x directly as the cross proportion.

Experiment 2: Intra-stream calculations

These calculations are computed separately for each of the two annotated streams in each excerpt.

Onset Synchrony

1. Only slices where at least one instrument articulates an onset are used.
2. For each stream, we identify places where any of the instruments are tacit (not playing):
 - a. Instruments are tacit until they first play an onset.
 - b. Instruments are tacit immediately after the last onset they play, until the end of the excerpt.
 - c. Instruments are tacit in any span in which they rest while two or more onsets occur in other instruments.

1. For each slice, we count the total number of instruments that are not tacit—this is the number of “active instruments” at any time.
2. For each slice, the number of instruments playing an onset is counted.
3. Any place where the number of active instruments is the same as the number of onsets is a synchrony point.
4. Divide the number of synchrony points by the total number of (nontacit) slices to obtain the onset synchrony score.

Crossing Proportion

1. In this case, we again only use slices where at least one instrument articulates an onset.
2. For all pairs of instruments within the stream, compute the interval between them at each onset, and ask whether this interval is < 0 , > 0 , or $= 0$ and call this the “direction.”
3. For each of these directions (each representing one pair of instruments), identify all the places where the direction changes from one slice to the next.

(Note that since unison is one direction category, changing from unison to positive/negative, or the reverse, is also considered a voice crossing).

4. Count all the places across all instrument pairs where the voices cross positions. Divide this number by the total number of interval pairs between all voices: This assures that the crossing proportion value is between 0 and 1.

Consonance:

1. Again, divide the score into triplet 64^{th} slices, and keep rests, assuring that the average accounts for the duration of notes, not just the number of onsets.
2. For each pair of instruments within the stream, calculate the interval between them, and weight the intervals according to Malmberg’s (1918) set of consonance scores.
3. Take the average consonance score across each slice, and then the average across these slice-averages to get the consonance score for the entire stream.