# Instrument Timbre Combinations Influence the Relative Prominence of Perceptual Layers in Orchestral Music

Manda Fischer
*University of Western Ontario, London, Canada*

Stephen McAdams
*McGill University, Montreal, Canada*

**The stratification of layers of differing** prominence (foreground/background) is a common technique in orchestration. Musicians heard 23 excerpts containing foreground and background layers as previously determined by music analysts. A given layer comprised either a single auditory stream of one or more blended instruments or a harmonic or rhythmic background. Two-layer excerpts had either the same, overlapping, or different instrument families (timbre class). First, musicians rated the perceived degree of segregation of musical materials in two-layer and single-stream excerpts. Second, they heard each of the two layers in isolation and then together and rated the relative prominence of the layers. Heterogeneous instrument combinations yielded the greatest difference in relative prominence, followed by overlapping and then homogeneous combinations. Acoustic and score-based descriptors were extracted to quantify their relative contribution to perceptual stratification. Timbre class and between-layer differences in timbre and dynamics played a role, providing evidence of how timbral differences enhance relative prominence in orchestral music. Perceptual segregation was positively but very weakly related to relative prominence, supporting findings that although segregation is necessary to form layers, this mechanism is separable from that which places the streams into the same representational space to allow for the assessment of their relative prominence.

*Received: December 11, 2023, accepted July 20, 2024.*

**Key words:** musical timbre, stratification, prominence, auditory stream segregation, orchestration

**The auditory environment is rich with** acoustic information that our perceptual system must sort and analyze to make sense of the auditory scene. The decoding of a musical scene is a fascinatingly complex process. What one perceives depends on the grouping of acoustic material into coherent mental representations and on the selection of information for further processing. Orchestration provides an excellent framework for investigating how timbre (sound quality) is used as a tool to organize and shape listeners' perceptions of the musical materials. The present study specifically examines the orchestral stratification effect: sounds that differ acoustically are orchestrated in such a way that the auditory system organizes them into separate layers of differing perceptual prominence. Adler (2002) mentions scoring for foreground, middleground, and background layers in his chapter on Scoring for Orchestra (p. 547). And Koechlin (1954–1959) writes about the equilibrium between sound forces realized by composers and notes that, "you can't have 'total' balance between all the parts at any one time, in the sense of equal intensity and volume. On the contrary, you have to make the most of more subdued values in the background (which, at first hearing, are only dimly perceived by less-experienced ears)." [our translation] (Vol. 3, p. 1)

## Auditory Scene Analysis and Orchestration Practice

Auditory perception organizes sounds that originate from separate sources and arrive at each eardrum simultaneously as one complex pressure wave back into separate mental representations of those sound sources (Bregman, 1990, Chapter 1; McAdams & Bregman, 1979). To make *consistent* interpretations about the environment, the auditory system uses sound source heuristics that follow Gestalt grouping principles of similarity and continuity (Bregman, 1990, Chapter 2; Bregman & Pinker, 1978; Koffka, 1935, Chapter 4). The process of auditory scene analysis relies on the assumption that sound sources emit similar sound events over time, which results in perceptual coherence and consistency (Bregman, 1990, Chapter 2; Bregman & Pinker, 1978; McAdams & Bregman, 1979). Sound events that are similar to one another are interpreted as originating from the same source and are grouped together. Sound events that are different from one another are interpreted as coming

*Music Perception*, volume 42, issue 4, pp. 311–328, issn 0730-7829, electronic issn 1533-8312.  DOI: https://doi.org/10.1525/mp.2025.2325705

from separate sound sources and are perceptually segregated (Bregman & Campbell, 1971).

The auditory system organizes sound in three different ways. Concurrent grouping integrates acoustic information into auditory events and segregates co-occurring events from different sound sources. Under appropriate conditions of harmonicity of relations among fundamental frequencies, onset synchrony, parallel changes in pitch and dynamics, and overlapping spectra, concurrent grouping can result in the blending of two or more instruments into fused events (Bregman, 1990, Chapter 3; Gianferrara, 2016; Lembke & McAdams, 2015; McAdams, 1989; McAdams et al., 2016; Sandell, 1995). Sequential grouping integrates fused events into continuous auditory streams, surface textures, and layers that have similar perceptual qualities and segregates streams, textures, and layers that have dissimilar qualities between them but similar qualities within each one. Sequential grouping can result in the integration of successive events into an auditory stream and the segregation of events into separate streams based on spectral, intensity, and spatial differences (Bregman, 1990, Chapter 2; Fischer et al., 2021; McAdams & Bregman, 1979). It can also result in a looser grouping of different musical parts into surface textures or orchestral layers, which differ in perceptual prominence and are referred to as foreground and background layers. Segmental grouping chunks groups of events into perceptual units of various sizes. Segmental grouping results in perceptual units that can be processed as patterns in short-term memory (Deliège, 1987; Deliège & El Ahmadi, 1990; Kurby & Zacks, 2008; Lalitte et al., 2004; Tillmann & McAdams, 2004). All these processes play a role in music perception and are especially relevant in the perception of orchestral music (McAdams, 2019a; McAdams et al., 2022).

The present study focuses on the stratification effect in which the process of sequential grouping gives rise to independent layers that are segregated from one another. In the musical application of scene analysis principles to the perception of orchestration, McAdams et al. (2022) make a distinction between perceptual segregation and stratification. Stream segregation involves the forming of images of the behavior of sound sources (streams). In a musical context such as polyphony, these streams are generally of equal prominence. Fischer et al. (2021) have shown that heterogeneous instrument combinations between streams yielded greater segregation than did homogeneous ones. Decreasing timbral differences of the orchestral excerpts through reorchestration to all string instruments reduced segregation ratings. Acoustic and score-based descriptors were used to statistically quantify their relative contribution to these effects. Instrument family, part crossing, consonance, spectral measures related to timbre, and onset synchrony all played a role, providing evidence of how timbral differences enhance segregation in orchestral music. Stratification involves a looser gathering of sources into groups with similar properties and musical behaviors (layers or strata), and there are differences in perceptual prominence distinguishing foreground from background. According to Goodchild and McAdams (2021), one orchestral layer can indeed contain two co-equal streams carried by different instruments in counterpoint, and those can be distinguished from a less prominent harmonic background, for example. Principles related to sequential grouping occur along perceptual dimensions of timbre, pitch, loudness, spatial location, and rhythmic independence (Bregman, 1990, Chapter 2; Bregman et al., 1990; McAdams, 1984; McAdams & Drake, 2002).

## Timbre and Auditory Grouping

Timbre is a multidimensional feature of auditory sensation that allows us to recognize and track the diverse array of sounds in our environment (McAdams, 2019b). It is composed of both static and dynamic acoustic attributes, such as brightness, noisiness, attack-decay, 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 & Siedenburg, 2019). Timbre's effect on sequential grouping is the focus in the current experiment. In accordance with Gestalt grouping principles of continuity, sound events that are similar in timbre are interpreted as coming from the same source, whereas sound events that differ in timbre are interpreted as coming from different sources. Thus, timbral differentiation induces sequential auditory grouping (McAdams, 2019a; McAdams & Bregman, 1979). Timbral differentiation between two given sounds is not simply a matter of binary distinction between similar and non-similar. Indeed, the timbres of two events produced by two sound sources (or two groups of blended sound sources) may vary in their degree of dissimilarity. Timbre dissimilarity is correlated with the degree of perceptual segregation (Bey & McAdams, 2003).

## Attention and Auditory Saliency

Auditory grouping illustrates heuristics that parse a sound mixture into separate streams or layers. These streams are then selected by attention, as what is heard also depends on how sound is filtered and preferentially processed. The resulting perception of the auditory scene is composed of mental representations of the acoustic environment that are, in part, weighted according to their saliency (Kayser et al., 2005). Once segregation has occurred and layers of musical material have formed, layer selection is a competitive process in which the most salient between two concurrent layers is the one that is most likely to attract attention (Tordini et al., 2013).

Saliency is a function of context. For a given feature to be salient, it must be detectable against a background and must have the ability to attract attentional mechanisms. Two forms of saliency, intrinsic and extrinsic, exist. The former pertains to inherent properties of a given stimulus, and the latter refers to the stimulus's relation to other stimuli in a scene (Chon, 2013; Kayser et al., 2005). These two mechanisms interact to guide attention. For example, the ability to detect a cherry in a bush would depend both on the inherent red property of the fruit and on its ability to stand out relative to its dappled green, leafy background. Or in a more musical setting, a bright and strident trumpet may attract attention through its inherent saliency, but if it is embedded in an orchestral texture of other high brass instruments, it would stand out less than against the string sections. Saliency is difficult to define, as it exists between the realms of top-down (schema-driven) and bottom-up (stimulus-driven) attentional mechanisms (Tordini et al., 2013). To date, there are very few studies investigating the behavioral relevance of auditory saliency using naturalistic stimuli. Saliency, however, is critical for the representation of complex auditory scenes, such as musical scenes, and plays a prominent role in the perception of orchestral stratification.

## Timbre and Orchestral Stratification

Sound is fashioned by composers to achieve a particular sonic goal. To further understand orchestration practice and its relation to audience perception, knowledge of how composers are operating at the level of auditory organization to shape listeners' perceptions must be made explicit. A theory of orchestration may be formalized by connecting orchestration practice to its underlying psychological principles (Goodchild & McAdams, 2021). The Taxonomy of Orchestral Grouping Effects (TOGE) was recently proposed by McAdams et al. (2022) to link auditory grouping principles to orchestration practice. Orchestral stratification forms one sub-branch of the sequential grouping branch of the taxonomy.

Orchestral stratification requires: 1) the integration of sound sources into layers, as well as the segregation of different layers, and 2) the competition for attention between layers, which results in their assignment to foreground and background strata. What are the properties (timbre, pitch register, and loudness) that cause integration of sound events into strata and the segregation of different strata? What are the acoustic features that render one layer more salient than the other and that are responsible for the differential processing weight of each layer? As timbre's effect on auditory segregation and competitive layer selection is of particular focus and as saliency is a function of context, orchestration provides a fertile ground for investigating the complex interplay between perceptual grouping processes and attentional mechanisms, within the ecologically valid acoustic environment of real orchestral excerpts. It is through composers' judicious choices of instrument combinations and their realization by musicians that the acoustic properties within the music itself may be shaped to elicit a particular percept in the listener. The literature on timbre perception cited above has shown that differences in timbre promote sequential grouping (Bregman, 1990; Cusack & Roberts, 2000, 2004; Tougas & Bregman, 1985) and that certain dimensions of timbre may be more relevant for perceptual saliency (Chon & McAdams, 2013)

## Current Study and Hypotheses

The objective of this research was to use excerpts of orchestral music to test a novel conceptual framework related to the role of timbre perception in orchestral stratification. Specifically, we examined the effect of instrument timbre on the relative prominence of orchestral layers. To assess this effect, we selected excerpts that differ in their acoustic properties and that were deemed by analysts to have two layers of differing prominence (foreground and background). The analyst's annotated degree of stratification of the two layers varied across excerpts, from strongly to weakly stratified. Listeners rated the degree of perceived segregation of musical materials for two-layer excerpts and single-stream excerpts in one block of trials, and then rated the relative prominence of the layers in two-layer excerpts. We posed three hypotheses:

H1: Given the greater timbral differences between instrument families than within families, differences in prominence between the two layers would be greater for excerpts in which layers have instruments from different families between layers and smaller for excerpts containing combinations of overlapping or the same instrument families within the two layers. Although there is greater within-family timbral similarity in the brass and string families and greater dissimilarity in the woodwind and percussion families, we adopted instrument family as operationally convenient categories as a starting point, given that orchestration treatises often mention scoring for instrumental choirs of the same family in different layers (e.g., Adler, 2002, Chapter 15).

H2: Differences in instrument family combinations between layers would be reflected at the acoustic and score levels, such that the degree of relative prominence between layers would vary according to the degree of spectral, intensity, and average pitch differences between layers; prominent layers would stand out timbrally, be louder, and higher in pitch register. Higher-pitched voices have been shown to be more salient than lower-pitched ones (Fujioka et al., 2005).

H3: Greater segregation of layers would be required for a greater difference in prominence between layers.

## Method

### PARTICIPANTS

Forty-five participants took part in the experiment. Two participants did not pass the hearing test, one participant did not qualify as a musician, and two participants did not complete the study due to technical difficulties. The remaining 40 participants (22 female, 17 male, and one who preferred not to answer) between the ages of 17 and 36 years ($M = 24.2$, $SD = 4.2$) qualified as musicians with at least one year of university music training (years of university training: $M = 3.9$; $SD = 1.7$; years since start of music training: $M = 7.6$; $SD = 3.6$). They all had normal hearing (auditory thresholds at or below 20 dB HL) as verified with a pure-tone audiometric test using octave-spaced frequencies from 125 Hz to 8 kHz (ISO 389–8., 2004; Martin & Champlin, 2000). Participants provided written consent, and the McGill University Research Ethics Board II certified the study for ethical compliance.

### STIMULI

As detailed in the Appendix, the complete stimulus set consisted of 23 excerpts with foreground and background layers, each containing either a single stream with one instrument or several blended instruments, or instrument combinations forming a harmonic or rhythmic background ($M_{\#measures} = 5.4$, $SD = 2.1$; $M_{\text{duration}} = 11.0\text{s}$, $SD = 5.4$). Additionally, five single-stream excerpts were included in the segregation rating task to anchor the scale ($M_{\#measures} = 3.6$, $SD = 1.4$; $M_{\text{duration}} = 10.0\text{s}$, $SD = 4.1$). The instrumentation remained constant over the duration of each excerpt. Therefore, two-layer excerpts contained both segregation between layers and integration within each layer. Single-stream excerpts contained instrument sections that were strongly blended as determined by Gianferrara (2016; McAdams, et al., 2016).

Excerpts were made up of combinations of one to four different instrument families (Strings [bowed], Woodwinds, Brass, and Harp). For two-layer stratification excerpts, ten categories of timbral combinations were used (Appendix, Table A1). For single-stream excerpts, one to three different instrument families (Strings, Woodwinds, Brass, Harp) were combined to yield five categories of timbral combinations. Harp was categorized apart from bowed strings as an impulsive instrument, due to previous research showing that plucked strings behave differently with respect to auditory scene analysis than do bowed strings (Lembke et al., 2019). Furthermore, impulsive sounds (e.g., plucked or struck) do not generally blend as well with sustained ones (Tardieu & McAdams, 2012). Timbral combination categories are denoted with instrument family or families in each layer separated by a hyphen between layers, e.g., S-S for Strings in both layers or W-WS for Woodwinds in one layer and Woodwinds and Strings in the other layer. A complete list of timbre categories can be found in the Appendix. Also indicated in the Appendix is the higher-level categorization into timbre classes: excerpts are categorized in terms of the two layers having the same instrument family or families (within), sharing at least one family (overlapping), or having different families (between).

The five single-stream excerpts used to anchor the segregation scale had been validated in a previous experiment as having a high average degree of blend, $M_{\text{blend}} = 0.17$, $SD = 0.24$ (Fischer et al., 2021) on a scale from 0 (*totally blended*) to 1 (*not at all blended*) (Appendix, Table A2). Given that these excerpts were annotated as blend excerpts by skilled analysts, we anticipated that these integrated excerpts would be perceptually rated as less segregated on average than two-layer excerpts. Instrument family combinations were denoted as concatenated letters of family names, e.g., S for String or BS for Brass and String. Sound levels for all 28 stimuli varied between 44.6 and 77.1 dB SPL.

The preparation and selection of the stimulus excerpts was a multi-step process that proceeded as follows.

*Step 1: Database*

The stimuli were drawn from the Orchestration Analysis and Research Database (*OrchARD*, n.d.), 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 auditorily identify orchestral effects such as integration (instrumental blend), segregation into two or more streams, and stratification into foreground and background layers. The goal was to: 1) indicate where in the music auditory grouping effects were occurring in the music, 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. Stratification was defined for the annotators as: 1) involving the grouping of instrumental parts into orchestral layers that resulted in clearly distinguishable groups of different prominence, and 2) occurring between individual instruments or entirely fused instrument pairings (or instrument sections) that constitute a "virtual voice." Integration was defined as the fusion of instrument sounds creating an augmented or emergent timbre (Sandell, 1995) over the course of a motive or phrase. Each annotation was determined separately by two analysts who then compared analyses and came to a consensus on the final annotation. All information regarding the orchestral stratification and integration effects (work, movement, composer, measures of occurrence, recording, start time, end time, instruments involved, and degree of stratification/integration) were compiled into *OrchARD*.

*Step 2: Selection Criteria*

The stimuli for the experiment were chosen from the 545 strafication annotations in *OrchARD* based on instrument family, number of instrumental parts, duration, and annotated strength rating of degree of stratification to capture a broad range of degrees of stratification. We attempted to find several excerpts for each combination of instrument families. However, there were not enough examples with brass in both layers 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 with one track per instrument using their *OrchSim* environment (Bouliane & Baril, n.d.). The dynamics of all instruments were balanced across tracks in the full musical context. In the stereo simulations, each instrument is 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 the effect were originally written in the score). The 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 and 2 for example excerpts extracted and isolated from the rest of the musical context: two multi-instrument layers and a blended single stream, respectively.

### PROCEDURE

The experiment consisted of three blocks. The first block contained three practice trials to familiarize participants with the interface, tasks of rating the degree of segregation and prominence, and describing which musical features informed their prominence ratings. The second block required participants to rate the degree of perceptual segregation of each excerpt. All 28 stimuli (23 stratified and 5 integrated) were randomly ordered across participants to minimize potential order effects. In the third block, participants rated each of the 23 stratified excerpts in random order for the relative perceptual prominence of the two layers and explained their rating qualitatively using an empty text box. After the experiment, participants filled out a short questionnaire about their music training and listening habits.

*Task 1: Degree of Segregation Rating*

Perceived segregation was measured to test whether the relative prominence between layers was related to their degree of segregation. The PsiExp software (Smith, 1995) was programmed to present a horizontally configured continuous scale on the computer screen with anchors ranging from "strongly integrated" (left) to "strongly separated" (right). After listening to an excerpt, participants rated the degree of perceptual


FIGURE 1. [Color online] An annotated excerpt from Haydn's Symphony 100, ii, mm. 17–23, with WS-S stratification between string and woodwind instruments (SW) in the foreground layer (flutes and violins, red box) and lower string instruments (S) in the background layer (violas, celli, basses, dashed blue box).


FIGURE 2. An annotated excerpt from Vaughan Williams Symphony 8, iv, mm. 81–86, with a WSH integrated stream composed of a unison blend of clarinet (transposing down a whole tone), harp, and viola.

segregation by dragging the slider along the continuum of the scale. The slider only appeared on the scale when the participant clicked along the continuum to minimize any potential bias with an initial placement of the slider. Listeners were only permitted to listen to each excerpt once. Once they were satisfied with their rating, they clicked on the button "ok" to move on to the next excerpt.

### *Task 2: Relative Prominence Rating*

The second task concerned the relative prominence of the two layers. Prominence was measured to identify particularly prominent musical features or acoustic features that would render a given layer more salient than the other. Specifically, this task was used to test whether the perceptual prominence would increase as a function of differences between layers in known acoustic correlates and musical parameters implicated in previous research as being relevant for saliency, such as spectral centroid (Chon & McAdams, 2013) or pitch register (Fujioka et al., 2005). A qualitative description section was also included following each prominence rating, but those data will not be considered in this paper.

In each trial, participants heard four things before making their rating. First, they heard the excerpt with the entire set of instruments (full context) (Figure 3a). To hear the following sequence, they were required to actively indicate that they were ready by clicking on the button "Start listening" (Figure 3b). Then participants successively heard a portion of the instrument set (Layer A) (Figure 3c), the other portion of the entire

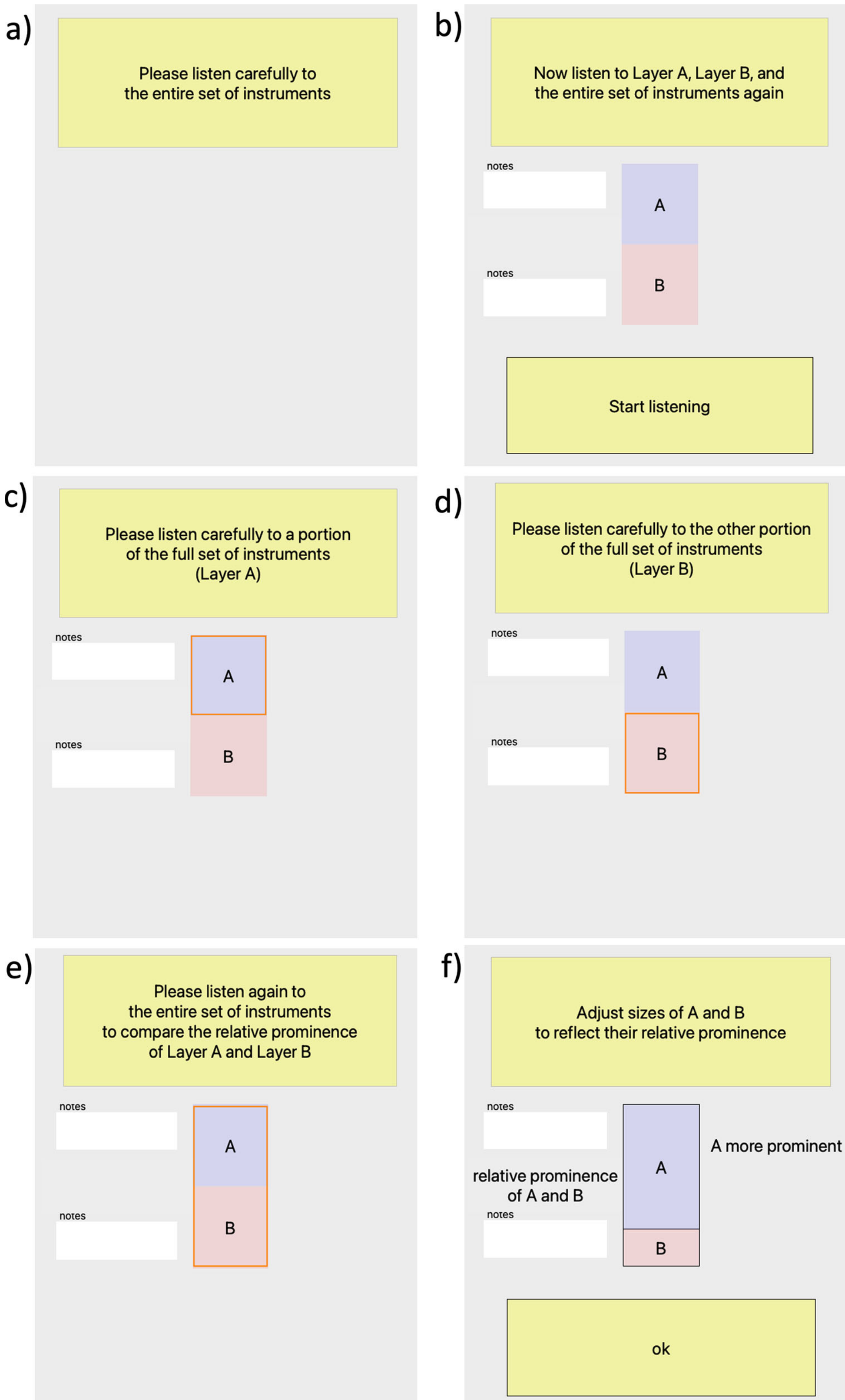


FIGURE 3. [Color online] Screenshot of different phases of the experimental interface in the relative prominence block.

instrument set (Layer B) (Figure 3d), and then the entire set of instruments again (full context Layers A and B) (Figure 3e). Empty text boxes adjacent to each label, A and B, were activated, as soon as the participant clicked "Start listening." Feedback from pilot participants indicated that the second block of the experiment regarding perceptual prominence was of high cognitive demand. Thus, optional note boxes were implemented to reduce the possibility of errors in remembering which musical materials corresponded to A and B. They were told that they could write any sort of memory aid for themselves in the text boxes, to help them keep track of what was assigned to each arbitrary label (A and B). In addition, they were told that use of the text boxes was optional. The assignment of what analysts had deemed as foreground and background layers to A and B labels was randomized for each trial.

A vertically configured, continuous scale (coded 1 at the bottom and 0 at the top) divided into two portions was presented to participants (Figure 3f). To indicate the degree of relative prominence of each layer, participants were instructed to drag the horizontal middle line dividing the two portions either up or down to adjust the relative area of each section. The participant could drag the horizontal line down to make A bigger (slider value range = 0.55–1), indicating the degree of prominence of Layer A, the horizontal line up to make B bigger (slider value range = 0–0.45), or the horizontal line close to the middle to make A and B approximately the same size (slider value range = 0.45–0.55). Labels appeared to the right of this box indicating "A more prominent," "B more prominent," or "About the same," respectively, to help participants to view their rating categorically.

### *Instructions and Practice Trials*

Prior to the main part of the experiment, participants underwent an extensive and pre-scripted verbal session in order align their understanding of the three main concepts: separation, integration, and musical feature. For the sake of clarity, the word "separation" was used in place of the word "segregation." In addition, as the goal of the experiment was to test perceptually the theoretical ratings given by the analysts, particular caution was taken to ensure that the words "foreground" and "background" were never mentioned to participants. Listeners were asked to describe in their own words what they thought "separation," "integration" (Task 1 degree of segregation), "prominence," and "musical feature" (Task 2 degree of relative prominence) were. Once their answers satisfactorily demonstrated a clear grasp of these four concepts, they were shown visual analogies. As previous literature has shown that auditory grouping principles follow similar rules to that of Gestalt grouping principles related to visual perception (Bregman, 1990, Chapter 1; Iverson, 1995; Koffka, 1935, Chapter 10), examples from the visual domain were included in the instruction stage of this experiment to bridge their understanding of grouping and salience concepts from the visual domain to those in the auditory domain. Participants were only permitted to advance to the next visual example once their answer clearly reflected an ease with and understanding of the concepts.

A short practice session followed, to ground the visual analogies in auditory examples and to allow participants a chance to navigate the interface. Three practice excerpts were selected based on their annotated strength ratings and served as prototypical examples that captured a broad range of degrees of segregation. The practice excerpts were always presented to the listener in the same order, from most separated to most integrated. They comprised Verdi's *La Traviata,* Prelude, mm. 29–36 (Woodwinds vs. Strings; Stratification strength = 5), Schubert's *Symphony No. 8,* Movement 1, mm. 31–35 (Woodwinds vs. Brass; Stratification strength = 1), and Ravel's orchestration of Mussorgsky's *Pictures at an Exhibition,* Movement 10, mm. 1–2 (Woodwinds vs. Strings; Integration strength = 5). The range of stimuli in the practice trials was chosen to minimize context effects and intra-subject variability (Anderson, 1974, pp. 231–232) related to the calibration of the rating scale by giving participants the necessary information to build a frame of reference.

### APPARATUS

Participants were seated in an IAC model 120act-3 double-walled audiometric booth (IAC Acoustics, Bronx, NY). Sounds were 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 $\pm 60°$, facing the listener at a distance of 1.8 m. The experimental session was run with the PsiExp computer environment (Smith, 1995). Sound levels were measured with a Brüel & Kjær Type 2205 sound-level meter (A-weighting) placed at the level of the listener's ears.

### DATA ANALYSIS

Several acoustic and score-based measures were used as predictors in a mixed-effects analysis of the behavioral data. We first describe these predictors and then discuss the statistical modeling approach adopted.

Fischer et al. (2021) validated the perceptual significance of the equivalent rectangular bandwidth (ERB) auditory processing model (Glasberg & Moore, 1990; Moore, 1986; Moore & Glasberg, 1983) over the purely acoustic short-time Fourier transform (STFT) model as an input representation for computing acoustic descriptors. They showed that it can account for more timbral variance in relation to perceptual segregation. Therefore, we used the ERB model to calculate the acoustic descriptors. Using the Timbre Toolbox (Peeters et al., 2011) in MATLAB (The Mathworks Inc., Natick, MA), we derived 11 time-varying spectral descriptors related to timbre. For each layer, the Timbre Toolbox was used to extract the descriptors within each time frame of the sound. The relative difference (foreground layer value minus background layer value) was then computed between the layers for each time frame and averaged across time frames to yield one mean difference value. Positive values were associated with larger values of the acoustic descriptor in the foreground layer compared to the background layer. To reduce redundancy and multicollinearity, Pearson correlations were computed among the 11 descriptors across all stimuli. A threshold of .80 was set and from among those that were highly correlated (> .80) only one descriptor was chosen. Priority was given to descriptors that were found to be perceptually relevant in Fischer et al. (2021). The selected acoustic descriptors used in the mixed-effects models are listed in Table 1.

One score-based predictor was derived from musicXML representations of the scores used for the simulation by comparing features across layers. The calculation of the average pitch interval between the musical materials in the two layers for each excerpt proceeded as follows.

1. Rests were included.
2. Each layer was sliced into every possible triplet 64th-note position to ensure that the pitch averaging scores reflected the durations of notes.
3. Every instrument in each layer was paired with every instrument in the other layer, and the difference between their pitches at every slice was computed. The average of absolute pitch differences was taken across each triplet 64th-note slice and the average value for the slices was computed to get the average interval between layers.

A mixed-effects model was applied to the behavioral data. A random intercept was included in the model to account for within-subject individual differences. The dependent variable was the degree of prominence of the foreground compared to the background layer. Table 2 presents the full list of predictors and the subsets that were entered into the model. In addition to timbre class (between-family, overlapping-family, or within-family combinations), each model input consisted of converging data types: digital audio descriptors related to timbre computed from the sound signal, and score-based information derived from the MusicXML representation. Acoustic descriptors were included to identify specific loudness and timbral attributes that might be important for promoting perceptual prominence. All variable values were mean centered and normalized to compare the relative effects of each variable on the dependent variable.

We used model selection to identify the most parsimonious model of how acoustic and orchestral properties affect prominence. 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) the Bayesian Information Criterion (BIC) (Schwarz, 1978). From there, nonsignificant fixed effects were then selected for removal. An efficient greedy variable selection approach was used to identify a parsimonious model based on minimal BIC (Burnham & 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 (Raftery, 1995). The remaining variables constitute the final model. Type III (Wald) tests of fixed effects are reported. Bonferroni correction was applied to post hoc contrasts to compensate for multiple comparisons.

**TABLE 1.** *Acoustic Descriptors Selected for Model Input and Defined in Terms of Peeters et al. (2011)*

| Audio descriptor | Description |
|---|---|
| Frame Energy | The sum of the squared amplitudes of each auditory channel (ERB). |
| Spectral Crest | Tonal vs. noise content of the spectrum: Ratio of the maximum value and arithmetic mean of the spectrum. |
| Spectral Flatness | Tonal vs. noise content of the spectrum: Ratio of geometric and arithmetic means of the spectrum. |
| Spectral Skewness | A measure of asymmetry of the spectrum around its mean value. |
| Spectral Variation | The amount of change in the spectral envelope over time. |

**TABLE 2.** *Factors and Covariates Used in the Analysis of Prominence Ratings*

| Type | Variable | Description |
|---|---|---|
| Dependent variable | Prominence | Rating on an analog scale of the degree of relative prominence between layers, where 0 = background most more prominent and 1 = foreground most prominent. |
| Independent variable | Timbre Class | Within- (0), overlapping- (1), or between-family (2) combinations |
| Acoustic covariates | Frame Energy | Average difference in frame energy |
| | Spectral Crest | Average difference in spectral crest |
| | Spectral Flatness | Average difference in spectral flatness |
| | Spectral Skewness | Average difference in spectral skewness |
| | Spectral Variation | Average difference in spectral variation |
| Score-based covariate | Average Interval | Average difference in pitch |

*Note*: Difference measures for acoustic and score-based covariates were calculated between the two layers as foreground value minus background value.

## Results

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

We calculated the difference in BICs between two consecutive models (ΔBIC) to select the best model. A model with a BIC value that is lower than the previous model by 10 or more typically indicates very strong evidence that the current model is a better representation of the data, whereas a difference of 6–10 indicates strong evidence, and a difference of 2–6 indicates positive evidence (Raftery, 1995).

Adding the random intercept to the mixed-effects model improved BIC over the null model (ΔBIC = 94.0). Adding all fixed effects (random intercept, timbre class, spectral skew, frame energy, spectral crest, spectral flatness, spectral skewness, spectral variation, and average interval) to the model improved BIC over the random intercept model (ΔBIC = 189.0). The efficient greedy variable selection approach led to the removal of average pitch interval (ΔBIC = 3.7) followed by the removal of spectral crest (ΔBIC = 2.5), spectral skew (ΔBIC = 1.2), and spectral flatness (ΔBIC = 2.5). All remaining fixed effects were statistically significant (timbre class, frame energy, and spectral variation). In line with our second hypothesis, higher values of the acoustic descriptors in the foreground layer yielded greater perceptual prominence: frame energy, $\beta = .20$, $F(1, 876) = 53.10$ $p < .0001$, and spectral variation, $\beta = .06$, $F(1, 876) = 4.76$, $p = .029$. The intercept was also significant, $p < .0001$, suggesting that its variance was reliably different from zero.

### TIMBRE-CLASS CONTRIBUTIONS TO THE RELATIVE PROMINENCE BETWEEN LAYERS

Our first hypothesis was that the degree of perceptual prominence would vary depending on timbre class. Figure 4 presents the mean perceptual prominence ratings for each of the three timbre classes (within-, overlapping-, and between-instrument family combinations) for the two-layer excerpts. Recall that prominence ratings were measured on an analog scale of the degree of relative prominence between layers, where 0 = background most prominent and 1 = foreground most prominent. Equal prominence of the two layers would yield a rating in the vicinity of 0.5. Bonferroni post hoc comparisons within the main effect of timbre class, $F(1, 876) = 61.43$, $p = .0001$, demonstrate that between-family combinations yielded the greatest perceptual foreground prominence ($M = 0.74$, $SE = 0.008$), followed by overlapping-family combinations ($M = 0.68$, $SE = 0.006$), and within-family combinations ($M = 0.63$, $SE = 0.006$). These results suggest that differences in timbre, operationalized in terms of instrument-family combinations, yield an increase in perceptual asymmetry between layers, with the foreground rated by listeners as most prominent when instrument family combinations are heterogeneous between layers (between-family combination).

To further refine this analysis, Figure 5 provides a schematic representation of significant differences between means as revealed by contrasts between the ten timbre categories of instrument-family combinations. Combinations within boxes delineate means that are not significantly different from one another. Within-, overlapping-, and between-family categories generally formed distinct clusters. However, post hoc contrasts revealed two exceptions: 1) W-S instrument family combinations were less prominent and different from the other between-family instrument combination (WB-S and S-W), and 2) S-S (within-family

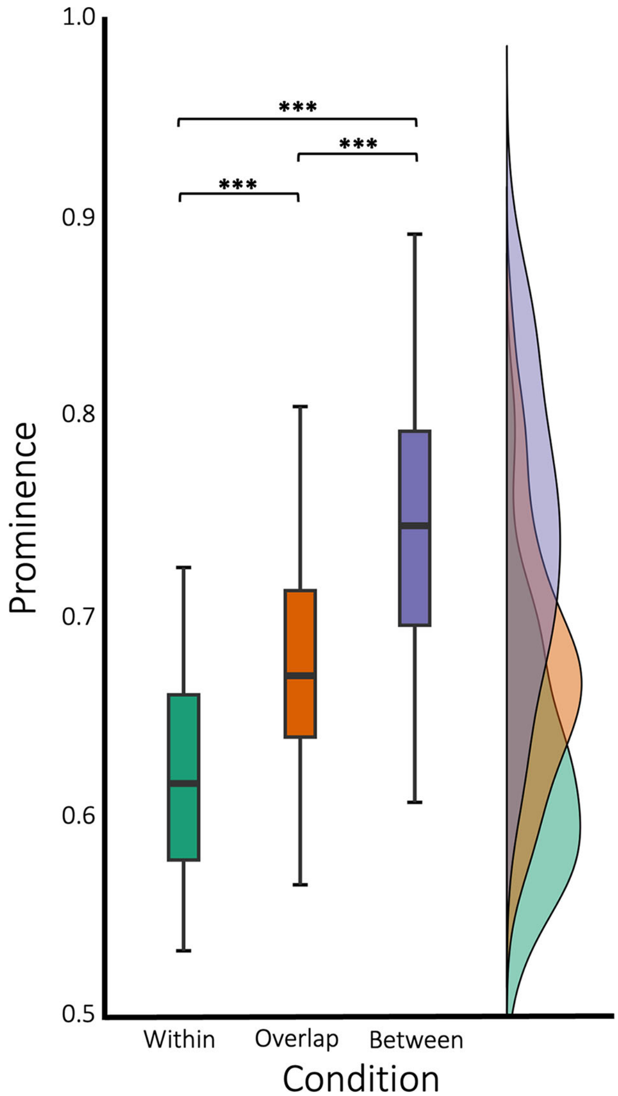


**FIGURE 4.** [Color online] Average prominence ratings of the timbre classes. *** $p < .0001$.

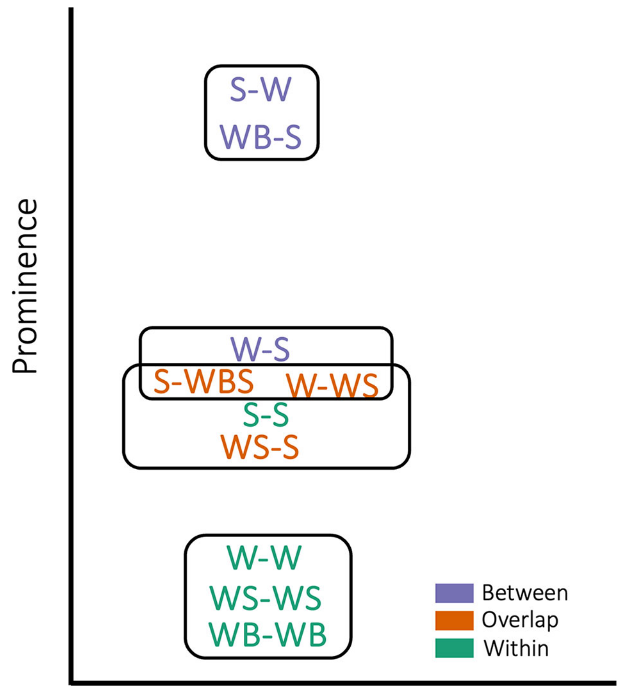


**FIGURE 5.** [Color online] Estimated means for the ten timbre categories of stratified excerpts. Boxes contain means that are not significantly different from one another.

combinations) were moderately stratified and yielded prominence ratings akin to overlapping instrument combinations. To quantify and better understand these results, we first grouped the prominence rating data—least-squares (marginal) means from the final model—into two groups: W-S vs. the other two between-family timbre categories and S-S vs. the three other within-family timbre categories. We then compared these two groups along the acoustic and score-based predictors that were present in the final model. All contrasts were Bonferroni-corrected for multiple comparisons. The goal of this analysis was to see how the other musical predictors in the model, aside from timbre class (i.e., frame energy and spectral variation), contribute to mean prominence ratings.

We contrasted the W-S values of the two acoustic predictors derived from the final model structure with the values for WB-S and S-W to elucidate parameters that may be reducing the prominence ratings of W-S excerpts. To perform the contrast, the W-S excerpts are dummy coded as 1, and the rest of the between-family combinations are dummy coded as 0. The contrast performed is 0 – 1, so a positive value indicates a lower value for W-S of a given acoustic descriptor. The results from the Bonferroni-corrected *t*-tests revealed that compared to other between-family combinations the foreground layer in W-S excerpts had lower frame energy, $t(318) = 3.04$, $p = .0025$ ($M_{W\text{-}S} = 1.84$ vs. $M_{other\ between} = 2.76$), and lower spectral variation, $t(318) = 4.70$, $p < .0001$ ($M_{W\text{-}S} = -0.003$ vs. $M_{other\ between} = 0.009$).

In a similar approach, we contrasted the four acoustic predictors of S-S with the aggregate values of the other within-family combinations. The results from the Bonferroni-corrected *t*-tests revealed that in comparison to other within-family combinations the foreground in S-S excerpts had higher frame energy, $t(278) = -15.56$, $p < .0001$ ($M_{S\text{-}S} = 3.95$ vs. $M_{other\ within} = -0.75$), and lower spectral variation, $t(278) = 9.39$, $p < .0001$ ($M_{S\text{-}S} = -0.02$ vs. $M_{other\ within} = -0.003$).

Therefore, frame energy and spectral variation appear to have contributed to the lower prominence ratings in W-S and higher ratings in S-S excerpts compared to their other timbre-class counterparts.

#### RELATION OF SEGREGATION BETWEEN LAYERS TO FOREGROUND PROMINENCE

To evaluate our third hypothesis, we ran a repeated-measures correlation (Bakdash & Marusich, 2017) to determine the degree to which the separation is related to the relative prominence between layers. Repeated-measures correlation is used to determine the common within-individual association for paired measures assessed on two or more items for multiple individuals to avoid biased results due to a violation of independence and/or differing patterns between participants versus within participants. Perceptual segregation was positively related to relative prominence, $r_m(879) = .14$, $p < .0001$, but the effect was small, explaining less than 2% of the variance. This result suggests that composers may use a combination of musical features that capitalize on perceptual principles to maintain voice independence and shape perceptions of prominence. It is worth noting that, as one might expect, mean segregation ratings for single-stream excerpts ($M = 0.640$, $SE = 0.015$) were much lower than mean ratings for all of the two-layer excerpts ($M = 0.079$, $SE = 0.001$), $t(39) = 33.10$, $p < .001$, Cohen's $d = 5.233$.

## Discussion

The goals of this study were: 1) to assess how different combinations and contrasts of instrument families (represented by our timbre-class factor) contribute to the relative prominence between orchestral layers and 2) to determine the role of acoustic and score-based predictors as well as 3) the perceived degree of segregation of the layers in prominence differences between the layers. We had participants rate two-layer stratification excerpts for their degree of perceptual segregation and for the degree of relative prominence between layers. The degree of relative prominence was measured to investigate the acoustic properties that account for the differential processing weight of each layer (i.e., foreground and background), thus rendering a given layer more salient than the other. We hypothesized that the degree of relative prominence would be higher for layers that contain salient acoustic correlates and musical parameters. The final linear mixed effects model for prominence ratings included timbre class (within-, overlapping-, and between-family combinations) as a factor and the audio descriptors spectral variation (timbre-related) and frame energy (loudness-related descriptor) as predictors. The degree of segregation was measured to determine whether it was related to prominence ratings, the hypothesis being that greater segregation would allow for greater differences in prominence to emerge.

#### INSTRUMENT-FAMILY COMBINATIONS

Our first hypothesis was confirmed: the degree of overlap of instrument families played a strong role in prominence. Globally, foreground layers were rated as more prominent than background layers in all excerpts, confirming the perceptions of the analysts who prepared the initial annotations for OrchARD. Foreground layers were rated as more prominent in between-family excerpts followed by overlapping-family and within-family excerpts. Timbral differences as conceived of in terms of instrument family distinctions play a central role in the prominence differences that emerge. However, there were two timbre categories that go against this global tendency. Woodwind foregrounds against string backgrounds (W-S) were not as prominent as combined woodwinds and brass against a string background (WB-S) or strings against a woodwind background (S-W). Indeed, the W-S category was not significantly different from two timbre categories in the overlapping class and the S-S within-family category. Further, the S-S category had a more prominent foreground than did any of the three other within-family combination categories. The acoustic predictors underlying these results, spectral variation and frame energy, provide an explanation for these exceptions.

#### ACOUSTIC AND SCORE-BASED PREDICTORS

Our second hypothesis concerning the role of acoustic and score-based predictors in the perception of prominence was partially confirmed. Differences in pitch register between layers don't seem to have played a role for these excerpts. However, one timbre-related descriptor and one loudness-related descriptor had strong effects. Foreground spectra tend to have more spectral variation over time in addition to being louder. These predictors also explain why the W-S foregrounds were rated as less prominent than the WB-S and S-W foregrounds and why S-S foregrounds were more prominent than the other within-family categories.

It is interesting to note a certain coherence between this experiment on orchestral stratification and experiments on blend and segregation. Fischer et al. (2021, Experiment 2) found that in single stream excerpts with several instruments playing together, within-family combinations blended more strongly than between-family combinations, and greater overlap in spectral skewness and spectral variation and reduced overlap in spectral flatness yielded greater blend of instrumental parts (in addition to score-based predictors of onset synchrony and consonant pitch relations among parts). In their Experiment 1, which measured perceived segregation in two-stream excerpts, between-family excerpts yielded greater segregation than within-family excerpts, had greater overlap in spectral variation, and had reduced overlap in spectral crest, spectral flatness, spectral skewness, and average pitch, as well as lower onset synchrony and a lower degree of part crossing in pitch. Interestingly, frame energy was included in the full model but removed during model selection. In the present experiment on stratification and relative prominence of layers, higher values of spectral variation and frame energy were correlated with higher foreground prominence ratings. These acoustic predictors thus appear to play various roles in different sequential grouping processes in the perception of orchestration: stream segregation and stratification of orchestral layers. Moreover, the results indicate that the auditory salience space encompasses multiple dimensions, confirming previous findings that spectral shape and loudness contribute to auditory salience in natural soundscapes (Huang & Elhilali, 2017) and extending their implication to the stratification of musical materials.

### SEGREGATION BETWEEN LAYERS

Our third hypothesis was only very weakly confirmed. Segregation ratings on excerpts were significantly correlated with prominence ratings. However, the segregation ratings explained less than 2% of the variance. This result supports other findings that demonstrate that although segregation is necessary to group sequential auditory information, a separate mechanism must place the streams into the same representational space to allow for the assessment of the relationship between them (Uhlig et al., 2013), prominence in our case. As such, musicians may rely on prioritized integrative attention, which incorporates both stream segregation and integration to assess the salience relationship between concurrent parts. Given that the segregation-rating block was presented before the prominence-rating block, it is possible that the former influenced the latter. However, given the exceedingly weak relation, we conclude that any such effect is minimal.

### FURTHER CONSIDERATIONS

We used a highly naturalistic musical context to study the effect of timbre in perceptual prominence. It should be noted, however, that this ecologically valid framework limited the selection of the stimuli. Due both to strict criteria to directly test the stratification effect in its simplest form and to what was available both as analyzed in the Orchestration Analysis and Research Database and as simulated by OrchPlay Music with their OrchSim system, the stimuli were not distributed evenly across timbre categories. More specifically, stratification excerpts had to contain two layers, each with either one auditory stream or a harmonic-rhythmic background, and be stable in their instrumentation over the duration of the excerpt. In addition, these excerpts had to vary in terms of annotated strength rating (1–5) and timbre combination categories involved (within: WB-WB, W-W, WS-WS, S-S; overlapping: WS-S, W-WS, S-WBS; between: W-S, WB-S, S-W). Integrated excerpts had to be strongly blended (Gianferrara, 2016; McAdams et al., 2016) and were required to vary in terms of timbre combination categories.

Another consideration is that the current study focused on examining intrinsic saliency derived from the stimulus itself. It is possible, however, that participants' ratings of perceptual prominence may have been affected by the presentation of each layer separately between the full context presentations. The latter approach was needed to define each to-be-contrasted layer. Future studies could investigate the role of context in the perception of relative prominence between layers.

### CONCLUSION

Composers are truly masters at shaping sound to elicit a particular percept in the listener. Their knowledge of orchestration practice is both astounding, yet implicit, as it is guided by experience and skill alone. As a result, the goal of the current study is to bridge this gap by connecting example-based orchestration practice as presented in orchestration treatises to its underlying psychological principles. Our findings highlight the significance of timbre class and between-layer differences in spectral variation and frame energy in the relative prominence of perceptual layers in orchestral music and provide additional evidence for the multidimensionality of the auditory salience space. This research serves as a basis for future studies, with the aim of developing a set of perceptual principles to scientifically ground the perceptual effects of orchestration practice.

## Author Note

Both authors contributed equally to this work and share first authorship.

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), the Canadian Social Sciences and Humanities Research Council (895-2018-1023), and a Canada Research Chair (950-223484) awarded to Stephen McAdams. The authors thank Bennett K. Smith for programming the experimental interface, Kit Soden for help with choosing the excerpts, OrchPlayMusic, Inc. for granting us access to the original simulated orchestral excerpts, and Félix Frédéric Baril for extracting the excerpts from the full orchestral context.

*Correspondence concerning this article should be addressed to* Stephen McAdams (stephen.mcadams@mcgill.ca).

## References

Adler, S. (2002). *The study of orchestration* (3rd ed.). W. W. Norton and Company.

Anderson, N. (1974). Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2, pp. 215–298). Academic Press.

Bakdash J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456

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. https://doi.org/10.1037/0096-1523.29.2.267

Bouliane, D., & Baril, F. F. (n.d.). *OrchPlayMusic Library*. Last accessed 6/12/2023 from https://orchestraplayer.com

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press. https://doi.org/10.7551/mitpress/1486.001.0001

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244–249. https://doi.org/10.1037/h0031163

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology*, *44*(3), 400–413. https://doi.org/10.1037/h0084255

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, *32*(1), 19–31. https://doi.org/10.1037/h0081664

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. Springer.

Chon, S. H. (2013). *Timbre saliency, the attention-capturing quality of timbre* [Doctoral dissertation, McGill University]. https://doi.org/10.1121/1.4708879

Chon, S. H., & McAdams, S. (2013). Timbre saliency vs. timbre dissimilarity – What is the relationship? *Journal of the Acoustical Society of America*, *133*(5), 3449(A). https://doi.org/10.1121/1.4806106

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception and Psychophysics*, *62*(5), 1112–1120. https://doi.org/10.3758/BF03212092

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

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, *4*(4), 325–360. https://doi.org/10.2307/40285378

Deliège, I., & El Ahmadi, A. (1990). Mechanisms of cue extraction in musical groupings: A study of perception on *Sequenza VI* for viola solo by Luciano Berio. *Psychology of Music*, *18*, 18–44. https://doi.org/10.1177/0305735690181003

Fischer, M., Soden, K., Thoret, E., Montrey, M., & McAdams, S. (2021). Instrument timbre enhances perceptual segregation in orchestral music. *Music Perception*, *38*(5), 473–498. https://doi.org/10.1525/mp.2021.38.5.473

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience*, *17*(10), 1578–1592. https://doi.org/10.1162/089892905774597263

Gianferrara, P. (2016). *Perceptual data for a relational database on orchestration: The case of timbral blend* [Cognitive Science Honours Thesis, McGill University].

Glasberg, B. G., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2),103–138. https://doi.org/10.1016/0378-5955(90)90170-T

Goodchild, M., & McAdams, S. (2021). Perceptual processes in orchestration. In E. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 497–524). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190637224.013.10

Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *Journal of the Acoustical Society of America*, *141*(3), 2163–2176. https://doi.org/10.1121/1.4979055

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]. 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. https://doi.org/10.1037/0096-1523.21.4.751

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

Koechlin, C. (1954–1959). *Traité de l'orchestration: En quatre volumes* [Treatise on orchestration: In four volumes]. M. Eschig.

Koffka, K. (1935). *Principles of Gestalt psychology*. Harcourt, Brace and World.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Lalitte, P., Bigand, E., Poulin-Charronnat, B., McAdams, S., Delbe, C., & D'Adamo, D. (2004). The perceptual structure of thematic materials in *The Angel of Death*. *Music Perception*, *22*(2), 265–296. https://doi.org/10.1525/Mp.2004.22.2.265

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 (2019). Acoustical correlates of perceptual blend in timbre dyads and triads. *Musicæ Scientiæ*, *23*(2), 250–274. https://doi.org/10.1177/1029864917731806

Martin, F. N., & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, *11*(2), 64–66. https://doi.org/10.1177/108471380300700302

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. https://doi.org/10.1016/S0166-4115(08)62356-0

McAdams, S. (1989). Segregation of concurrent sounds. 1: Effects of frequency-modulation coherence. *Journal of the Acoustical Society of America*, *86*(6), 2148–2159. https://doi.org/10.1121/1.398475

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. https://doi.org/10.1007/978-3-030-14832-4_8

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. https://doi.org/10.1007/978-3-030-14832-4_2

McAdams, S., & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal*, *3*(4), 26–43. https://www.jstor.org/stable/4617866

McAdams, S., & Drake, C. (2002). Auditory perception and cognition. In H. Pashler & S. Yantis (Eds.), *Stevens' handbook of experimental psychology: Vol. 1. Sensation and perception* (3rd ed., pp. 397–452). Wiley. https://doi.org/10.1002/0471214426.pas0110

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* (p. 134). Causal Productions.

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* (pp. 71–120). MIT Press.

McAdams, S., Soden, K., & Goodchild, M. (2022). A taxonomy of orchestral grouping effects derived from principles of auditory perception. *Music Theory Online*, *28*(3). https://doi.org/10.30535/mto.28.3.6

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. https://doi.org/10.1121/1.389861

OrchARD (n.d.). *The Orchestration Analysis and Research Database*. Last accessed 6/12/2023 from https://orchard.actor-project.org

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.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception*, *13*(2), 209–246.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, *6*, 461–464. https://doi.org/10.1214/aos/1176344136

Smith, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. *Society for Music Perception and Cognition Conference*.

Tardieu, D., & McAdams, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*, *30*(2), 117–128. https://doi.org/10.1525/Mp.2012.30.2.117

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(5), 1131–1142. https://doi.org/10.1037/0278-7393.30.5.1131

Tordini, F., Bregman, A., Cooperstock, J. R., Ankolekar, A., & Sandholm, T. (2013). Toward an improved model of auditory saliency. In *Proceedings of the International Conference on Auditory Display* (pp. 189–196). http://www.zueit.eletel.p.lodz.pl/icad2013/paper/24_S6-4_Tordini.pdf

Tougas, Y., & Bregman, A. (1985). Crossing of auditory streams. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(6), 788–798. https://doi.org/10.1037/00961523.11.6.788

Uhlig, M., Fairhurst, M. T., & Keller, P. E. (2013). The importance of integration and top-down salience when listening to complex multi-part musical stimuli. *Neuroimage*, *77*, 52–61. https://doi.org/10.1016/j.neuroimage.2013.03.051

Wessel, D. L. (1973). Psychoacoustics and music: A report from Michigan State University. *PACE: Bulletin of the Computer Arts Society*, *30*, 1–2.

## Appendix

### COMPOSITION AND CLASSIFICATION OF STRATIFIED AND INTEGRATED MUSICAL EXCERPTS

Stratification excerpts contain two layers, each with either an integrated stream or a harmonic or rhythmic background, whereas integrated excerpts contain a single stream. S = strings, W = woodwinds, B = brass, H = harp. Conjoined letters indicate more than one family in each stream (WS = woodwinds and strings). Hyphens separate the families in the two layers (W-WS = woodwinds in the foreground layer, woodwinds and strings in the background layer).

**TABLE A1.** *Instrument Family Combination Type, Timbre Category, Musical Excerpt, and Instruments Involved in the Annotator-Defined Foreground and Background Layers in Stratified Excerpts*

| Type of Orchestral Effect | Instrument Family Combination (Timbre Class) | Timbre Category | Musical Excerpt | Annotated Foreground Instruments | Annotated Background Instruments |
|---|---|---|---|---|---|
| Stratification | Within | S-S | Haydn, *Symphony No.93,* I, mm. 21-24 | violins 1 & 2 | viola, cello, bass |
| | | | Mozart, *Symphony No. 40,* I, mm. 1-9 | violins 1 & 2 | viola, cello, bass |
| | | W-W | Haydn, *Symphony No.100,* II, mm. 8-16 | oboes 1 & 2 | clarinets 1 & 2, bassoons 1 & 2 |
| | | | Vaughan Williams, *Symphony No. 8,* II, mm. 85-90 | flute 1, piccolo, oboes 1 & 2 | clarinets 1 & 2, bassoons 1 & 2 |
| | | WB-WB | Tchaikovsky, *Symphony No. 4,* I, mm. 7-10 | flutes 1 & 2, oboes 1 & 2, clarinets 1 & 2, trumpets 1 & 2 | bassoons 1 & 2, horns 1, 2, 3 & 4, trombones 1, 2 & 3, tuba |
| | | WS-WS | Mussorgsky, *Pictures at an Exhibition,* IV, mm. 41-44 | flutes 1 & 2, clarinets 1 & 2, violin 1 | oboe 1, English horn, bass clarinet, bassoons 1 & 2, violin 2, viola, cello, bass |
| | | | Mussorgsky, *Pictures at an Exhibition,* IV, mm. 39-41 | flutes 1 & 2, violin 1 | clarinets 1 & 2, bass clarinet, bassoons 1 & 2, violin 2, viola, cello, bass |
| | Overlapping | S-WBS | Verdi, *La Traviata,* Prelude, mm. 18-24 | violin 1, viola, cello | flute 1, oboe 1, clarinet 1, bassoon 1, horns 1, 2 & 3, violin 2, bass |
| | | W-WS | Mussorgsky, *Pictures at an Exhibition,* IV, mm. 45-47 | oboe1, saxophone | clarinets 1 & 2, bassoons 1 & 2, cello |
| | | | Smetana, *Ma Vlast, Die Moldau,* mm. 185-186 | flutes 1 & 2 | clarinets 1 & 2, violins 1 & 2, viola, cello |
| | | WS-S | Rimsky-Korsakov, *Three Russion Themes,* Overture, mm. 50-52 | clarinet 1, viola | violins 1 & 2 |
| | | | Beethoven, *Symphony No.3,* III, mm. 7-14 | violin 1, oboe 1 | violin 2, viola, cello |
| | | | Haydn, *Symphony No.100,* II, mm.1 -7 | flute 1, violins 1 & 2 | viola, cello, bass |
| | | | Haydn, *Symphony No.100,* II, mm. 17-23 | flute 1, violins 1 & 2 | viola, cello, bass |
| | | | Beethoven, *Symphony No.3,* III, mm. 25-28 | flute 1, oboe 1, bassoon 1, violin 1 | violin 2, viola, cello, bass |
| | Between | S-W | Vaughan Williams, *The Lark Ascending,* II, mm. 79-82 | solo violin | oboe 1, clarinet 2, bassoons 1 & 2 |
| | | W-S | Mendelssohn, *Symphony No. 3,* II, mm. 8-16 | clarinet 1 | violins 1 & 2, viola, cello |
| | | | Tchaikovsky, *Nutcracker Suite, Dance of the reeds,* mm. 3-6 | flutes 1, 2 & 3 | viola, cello, bass |
| | | | Tchaikovsky, *Nutcracker Suite,* Overture, mm. 16-20 | flute 1 | violin 2, viola |
| | | | Tchaikovsky, *Nutcracker Suite,* Overture, mm. 20-24 | clarinet 1 | violin 2, viola |
| | | | Mahler, *Symphony No.1,* I, mm. 7-9 | oboe 1, bassoons 1 & 2 | violins 1 & 2, viola, cello, bass |
| | | | Mahler, *Symphony No.1,* I, mm. 18-21 | piccolo 1, oboe 1, English horn, bass clarinet | violins 1 & 2, cello, bass |
| | | WB-S | Debussy, *La Mer,* I, mm. 12-16 | English horn, trumpet 1 | cello, bass |

**TABLE A2.** *Timbre Category, Musical Excerpt, and Instruments Involved in Integrated Excerpts*

| Type of Orchestral Effect | Timbre Category | Musical Excerpt | Instruments Present in Stream |
|---|---|---|---|
| Integration | W | Strauss, *Tod und Verklärung,* mm. 456-458 | flutes 1 & 2, oboes 1 & 2, bass clarinet, bassoons 1 & 2 |
| | WB | Mussorgsky, *Pictures at an Exhibition,* II, mm. 61-63 | flutes 1, 2 & 3, oboes 1 & 2, English horn, clarinets 1 & 2, bass clarinet, bassoons 1 & 2, contrabassoon, horns 1 & 2 |
| | BS | Mussorgsky, *Pictures at an Exhibition,* II, mm. 68-69 | horns 1, 2, 3 & 4, trumpet 1, violins 1 & 2, viola, cello |
| | WS | Mozart, *Don Giovanni* Overture, mm. 23-26 | flutes 1 & 2, violin 1 |
| | WSH | Vaughan Williams, *Symphony No. 8,* IV, mm. 81-86 | clarinet 1, viola, harp |