

TIMBRE SEMANTIC ASSOCIATIONS VARY BOTH BETWEEN
AND WITHIN INSTRUMENTS: AN EMPIRICAL STUDY
INCORPORATING REGISTER AND PITCH HEIGHT

LINDSEY REYMORE
Arizona State University, McGill University

JASON NOBLE
University of Montreal, Montreal, Canada

CHARALAMPOS SAITIS
Queen Mary University of London, London, United Kingdom

CAROLINE TRAUBE
University of Montreal, Montreal, Canada

ZACHARY WALLMARK
University of Oregon

THE MAIN OBJECTIVE OF THIS STUDY IS TO UNDERSTAND how timbre semantic associations—for example, a sound’s timbre perceived as bright, rough, or hollow—vary with register and pitch height across instruments. In this experiment, 540 online participants rated single, sustained notes from eight Western orchestral instruments (flute, oboe, bass clarinet, trumpet, trombone, violin, cello, and vibraphone) across three registers (low, medium, and high) on 20 semantic scales derived from Reymore and Huron (2020). The 24 two-second stimuli, equalized in loudness, were produced using the Vienna Symphonic Library.

Exploratory modeling examined relationships between mean ratings of each semantic dimension and instrument, register, and participant musician identity (“musician” vs. “nonmusician”). For most semantic descriptors, both register and instrument were significant predictors, though the amount of variance explained differed (marginal R^2). Terms that had the strongest positive relationships with register include *shrill/harsh/noisy*, *sparkling/brilliant/bright*, *ringing/long decay*, and *percussive*. Terms with the strongest negative relationships with register include *deep/thick/heavy*, *raspy/grainy/gravelly*, *hollow*, and *woody*. Post hoc modeling using only pitch height and only register to predict mean semantic rating suggests that pitch height may explain more variance than does register. Results help

clarify the influence of both instrument and relative register (and pitch height) on common timbre semantic associations.

Received: February 5, 2022, accepted October 13, 2022.

Key words: timbre, pitch, cognition, language, meaning

HOW DO LISTENERS ASSOCIATE MUSICAL sound qualities with extramusical concepts and descriptions? Researchers and musicians are increasingly interested in this question (Saitis & Weinzierl, 2019), with a particular focus on semantic associations related to timbre, or timbre semantics, which refer to verbal attributes describing timbral qualities of musical sounds, such as bright, rough, or hollow (cf. Wallmark & Kendall, 2018). Recent research suggests that timbre semantic associations are far more intersubjectively consistent than has been conventionally assumed in music scholarship (e.g., Reymore, 2021; Saitis et al., 2017; Traube, 2004). However, data have only been collected for a small portion of the enormous range of timbres used in music—many different instruments, both acoustic and electronic, are commonly used in music, and furthermore, most of these instruments produce many different timbres. Thus, many questions remain unanswered about the semantic associations to which this vast timbral palette may give rise, including how timbre-semantic associations vary with pitch and/or musical instrument register.

Existing studies often use text-based analyses that characterize semantic associations at the level of individual musical instruments (e.g., Reymore & Huron, 2020; Wallmark, 2019a) or perceptual studies that measure semantic responses to a circumscribed set of stimuli, often consisting of one sound per instrument (e.g., Zacharakis et al., 2014). Valuable as such studies are, they do not account for either the huge range of timbral variation that is available *within each instrument* or the corresponding range of semantic associations. Since timbre is famously multidimensional (McAdams et al., 1995), it is no small task to determine how timbral

variation with respect to various musical parameters may be related to semantic associations across different instruments and instrument families. The enormous range of possible combinations of parameters for many instruments—dynamic level (intensity), fundamental frequency, articulation, duration, vibrato, playing technique—is a practical concern that precludes any single study from being able to thoroughly map out the full range of timbre semantic associations even for a single instrument (see Reymore, in press, for further detail). However, comparing semantic associations both within and across instruments is relevant for composers, orchestrators, musicians, music teachers, sound engineers, and the listening public. Thus, in the current study, we chose to study a group of eight instruments, focusing on timbre-semantic variability related to changes in pitch.

The classic presentation of the perceptual phenomenon of pitch separates tone chroma (pitch class) from pitch height (Shepard, 1982); the current study relates to pitch height, which is correlated to the fundamental frequency in the case of a periodic signal. As many authors have remarked (e.g., Cox, 2016; Zbikowski, 2002), to speak of pitch “height” is already to invoke an extra-auditory association, between the sensation of pitch and vertical spatial position. This is perhaps the most familiar example of how thoroughly metaphorical musical discourse often is, a tendency that finds a pronounced manifestation in the rich and diverse vocabulary used to describe timbre (see Figure 1 in Saitis, 2019). The timbres produced by musical instruments, and the terms used to describe them, often vary with pitch: for example, the lowest notes of the piano can be described as *rumbling*, *thick*, and *muddy*, while its highest notes are *tinkling*, *thin*, and *clear*. Such descriptions are common in both orchestration treatises (e.g., Adler, 2002) and scholarly writings (e.g., Fink et al., 2018).

However, comparisons of relationships between pitch and timbre across instruments are complicated by the fact that not all instruments are physically able to play the same range of notes. What is considered a high note for one instrument may be considered a low note for another: for example, the highest notes of the bassoon overlap with the lowest notes of the flute. Here, the concept of *register* becomes useful. Throughout this paper, we use the term “register” to indicate a portion of the range of a specific instrument or voice in relation to its entire available range of notes; register is relative to a given instrument’s available range and is categorical (e.g., low, middle, high). We distinguish register from “pitch height,” the term we use to refer to the pitch sensation that is correlated with fundamental frequency

and that is represented in our experiment by the use of stimuli varying by note label (e.g., “C4,” “G5”). Pitch height can be discussed independently of a specific instrument’s available range (in orchestration pedagogy, this corresponds to the concept of “orchestral register”). These two concepts—register and pitch height—are closely related, but they provide alternative operationalizations by offering different reference points (i.e., the available range of a particular instrument vs. the available range of musical notes in general).

Specific instrument registers are not defined by a fixed range of notes, but rather are related to salient changes in sound quality over the range of an instrument, which emerge from the idiosyncratic physics of the sound source (Drabkin, 2001). Descriptions of registers may invoke common terms across different instruments: for example, the low registers of the piano, clarinet, and cello may all be described as dark and rich. Previous research suggests that register may contribute timbral qualities independent of fundamental frequency—a machine learning model using Linear Discriminant Analysis (LDA) reported in Weihs et al. (2005) demonstrated highly successful classification of register with only spectral information as input (see also Thoret et al., 2021).

In designing the present study, we opted to focus on register rather than on pitch height. This choice was motivated in part by the essential role of register in compositional and orchestrational thinking; register is especially important for writing music with the practical limitations of instruments in mind. Methodologically, examining register is advantageous due to the relative nature of the concept: as described above, not all instruments can produce a single given note, but all (pitched) orchestral instruments have low, middle, and high registers. This is especially germane for comparing instruments with no registral overlap, such as the flute and the contrabassoon. Finally, while it seems intuitive that corresponding registers between different instruments should give rise to similar semantic associations, this has not yet been established empirically.

Although the study design was optimized to measure the effect of register, the data allow for secondary analysis from the perspective of pitch height. Thus, following our initial analysis of register, we also report an alternate analysis of pitch height.

Timbre, Pitch, and Register

Most of the seminal timbre psychoacoustics literature uses stimuli of the same note to control for the

perceptual influence of fundamental frequency (Grey 1977; Kendall & Carterette, 1993; McAdams et al., 1995; Zacharakis et al., 2015). Several notes have been used in previous work, most commonly E \flat 4 (e.g., Eerola, Alluri, & Ferrer, 2012; Saitis & Siedenburg 2020; Wallmark, 2019b). E \flat 4 corresponds to a fundamental frequency of approximately 311 Hz (in an equal tempered scale tuned at 440 Hz), which represents the “average” pitch in many music corpora (Huron, 2001). Although keeping the note played and other parameters such as duration and loudness constant reduces confounding variables, this paradigm suffers in ecological validity (Siddiq et al., 2018). Furthermore, in studies of instrument timbre, keeping the note played constant across stimuli may necessarily exclude certain instruments or involve the inclusion of extreme low/high notes, which can lead to non-idiomatic tone production.

Contrastingly, relatively few experiments have investigated interactions between pitch and timbre, despite evidence that distinctions are relevant to timbre processing. Differences over an octave may increase the difficulty of identifying common sound sources (Handel & Erickson, 2001; Marozeau et al., 2003), though Steele and Williams (2006) demonstrated that musicians appear to be more successful than nonmusicians at this task. Speeded classification tasks have observed interactions of pitch and timbre (Krumhansl & Iverson, 1992). An experiment reported by Allen and Oxenham (2014), moreover, observed interference effects of fundamental frequency and timbre, with results suggesting that fundamental frequency and auditory brightness are tightly related (see also Cousineau et al., 2014; Marozeau & de Cheveigné, 2007; Melara & Marks, 1990; Schubert & Wolfe, 2006; Siedenburg, 2018).

Timbral differences across instrumental registers have also been considered in experiments assessing the semantic and affective connotations of timbre. In an interlanguage semantic ratings study, Zacharakis et al. (2014) developed stimuli consisting of acoustic instrument signals of a few pitch classes spread across three octaves. They reported that fundamental frequency was negatively correlated with *mass*-related terms among English speakers, and positively correlated with *luminance*-related terms among Greek speakers. McAdams et al. (2017) investigated effects of music training, pitch register, and instrument family on affect ratings. Using isolated orchestral instrument samples of E \flat across seven octaves (E \flat 1 to E \flat 7), the researchers reported that pitch register was a significant predictor of valence, tension arousal, and energy arousal and was also involved in significant interactions in these models. Finally, Reymore (in press) used the 20-dimensional

semantic model presented in Reymore and Huron (2020) to collect ratings of recordings of two instruments, oboe and French horn, playing at three dynamics across four relative register levels. These results revealed that within-instrument timbral variability may be quite complex. While some trends between register and mean semantic ratings were linear, others followed different patterns; for example, *soft/singing* was less prevalent in the highest and lowest registers of the oboe but was especially prominent in the middle registers. Similarities between the oboe and French horn were evident within some semantic scales, but others showed different activity between the two instruments, and interactions were observed between register and dynamics for some scales. Several terms showed significant relationships to register for both instruments, including *shrill/noisy*, *rumbly/low*, and *sparkling/brilliant*.

Such semantic interactions between timbre and pitch or register have also been reported in studies of cross-sensory associations between music and color, offering converging evidence from another domain of the importance of timbre/pitch relationships. Using piano and string tones, Ward et al. (2006) found the increase of brightness with ascending pitch as a common underlying mechanism for timbre-color mappings in both sound-color synesthetes (individuals who involuntarily experience a color when hearing a certain note/sound) and non-synesthetes. A correlation of higher fundamental frequencies and brighter grayscales was also reported for non-synesthetes by Adeli et al. (2014), who examined mappings between musical instrument sounds and colored shapes (though note that pitched stimuli were filtered to share the same normalized temporal envelope, which might have compromised their timbral integrity). Reuter et al. (2018) investigated timbre-color associations in non-synesthetes using 60 orchestral sounds (ten instruments, three notes, two playing styles). Similar to our study, each instrument was represented by a low, a medium, and a high note relative to its range, but only one pitch class (E) was used in sampling notes across registers (we used two pitch classes, see “Method”). Although some instruments were significantly matched with certain colors (e.g., violin with yellow, bassoon/cello/tuba with brown), grouping the most frequent color selections for each instrument by register revealed that timbre-color mappings were dominated by the concurrence of ascending pitch and increasing brightness. Generally, higher stimuli were more often matched with lighter colors (consistent with Ward et al. 2006), as well as more saturated colors, while stimuli from the lower registers tended to be paired with darker colors. A systematic

shift of color hue towards more yellow tones at higher notes was also observed.

Study Aims

The main objective of this study is to investigate how timbre semantic associations vary with instrument register and pitch height across instruments. Although several timbre studies have incorporated pitch differences into their design, registration remains undertheorized in the timbre semantics literature. Further, sample sizes of previous studies have been relatively small, diminishing the generalizability of the findings. The present experiment aims to explore the contribution of instrument and register in the generation of semantic associations involving timbre. To do so, we recruited a large global online sample of English-fluent participants to rate eight orchestral instrument sounds spanning two pitch classes (C and G) across five octaves (C2 to C7) on 20 semantic scales. Our approach is built from the understanding that timbre is a perceptual phenomenon, and that timbre is both a quality and a contributor to source identity (Siedenburg & McAdams, 2017).

This project was also motivated by the authors' involvement with the ongoing interinstitutional Composer-performer Orchestration Research Ensembles, or CORE Project, which is realized in the context of the international ACTOR Partnership (Analysis, Creation, and Teaching of Orchestration; actorproject.org/working-groups/core). CORE doubles as an educational experience for performance and composition students and as an opportunity to document and research creative processes with specific focus on timbre and orchestration. In choosing instruments as stimuli for the current study, we included instruments from the CORE ensemble: violin, bass clarinet, trombone, and vibraphone. The CORE Project continues to produce dozens of new compositions by emerging composers for an instrumentation which is common to all participating institutions; thus, the results of the current experiment, by addressing this instrumentation, provide valuable information for analyzing perceptual effects created in these pieces. By studying the variation of semantic associations with register among CORE Project instruments, we aim to: 1) contribute to the ongoing work of analyzing the resultant pieces, particularly in understanding semantic and perceptual effects; and 2) provide a resource for composers in future rounds of this project, who will be composing using the same instrumentation.

Method

PARTICIPANTS

For this online perceptual study, a general global sample of 590 adults were recruited using the Prolific platform (prolific.com). Thirty-six participants were removed due to incomplete responses (time-outs). We next assessed the data for suspected “bots” (computer program) and/or “farmers” (phony participants in server farms), which have grown increasingly prevalent in online subject pools in recent years (e.g., Chmielewski & Kucker, 2020). Multiple data quality checks were performed in this assessment, and data were flagged if answers appeared to be repeated or random; interrater correlation was low ($r < .10$); headphone check (following Woods et al., 2017) failed two consecutive times; time spent on the instructions or total experiment was impossibly low; and/or answers given in the free response attention check questions were missing or nonsensical (see “Procedure”). Based on a holistic consideration of these flagged cases, 13 participants were cut. We also cut one duplicate participant.

The remaining 540 participants (242 females, 293 males, 3 other, 2 not disclosed) were fluent in English (self-reported via Prolific) and came from a range of musical backgrounds. Using the single-question item from the Ollen Musical Sophistication Index (2006), 81% of the participants self-identified as non-musicians (“non-musician” and “music-loving non-musician”); 19% identified as musicians (“amateur,” “serious amateur,” “semiprofessional,” and “professional”; see Supplementary Material at mp.ucpress.edu for full breakdown). The average participant age was 26.9 years (range 18–71, $SD = 8.7$). Participants were paid \$4.15 USD for taking part in the study. The experiment took $M = 36$ minutes to complete ($SD = 12$ m). The study was approved by the University of Oregon Institutional Review Board.

DESIGN

In this repeated-measures design, participants listened to brief recordings of instruments playing sustained notes (approximately two seconds) and rated these sounds on 20 semantic scales. We manipulated two independent variables: register and instrument. Specifically, the stimuli included eight instruments (flute, oboe, bass clarinet, trumpet, trombone, violin, cello, vibraphone); each instrument was represented by notes from three registers (low, middle, high), for a total of 24 stimuli. Participants were shown word prompts consisting of 20 orthogonal semantic dimensions for timbre (see “Semantic Scales” below for further detail).

MATERIALS

Sound Stimuli

Stimuli came from the Vienna Symphonic Library Cube (2011), a sample library of recorded sounds played by professional musicians. Stimuli included the natural attack and were edited to 1.5s sustains plus the natural decay of the sound, which varied slightly in length among the stimuli, dependent upon the envelope. All sounds were adjusted to a matching ANSI-loudness level (American National Standards Institute) using the Genesis loudness toolbox in MATLAB. In this procedure, the loudness of each sound was first computed using the Moore model. Next, the median of all loudness values was computed across all stimuli, and finally, each sound was adjusted to this median value. Although this process helps to equalize loudness, additional variability is likely present due to differences in individual perception and participant headphones. In choosing the instruments to be rated, we began by including the four instruments that have been used in the CORE Project (see “Study Aims”): violin, bass clarinet, trombone, and vibraphone. Vibraphone sounds were bowed, rather than struck, to maintain consistency of excitation type across the instruments of the stimuli set; all other

stimuli were produced with standard playing techniques. We then added flute, oboe, trumpet, and cello in order to balance the range of the stimuli and to maximize the variability of orchestral timbres tested. We note that the vibraphone emerges as something of an outlier among our chosen instruments: the only percussion instrument, the only idiophone, the only instrument played with a technique other than its default, and the only instrument that is not a standard member of the traditional symphony orchestra. As our goals included understanding how instrument type might affect the relationship between a semantic category and register, we determined that the variability offered by the vibraphone was advantageous.

Each of the eight instruments was represented by a *low*, a *medium*, and a *high* note relative to its range. We sought to minimize the number of pitch classes used to avoid potential effects of intervals between stimuli on semantic judgments. However, use of a single pitch class precluded thorough representation of all three registers in all instruments; accordingly, two pitch classes, C and G, were used in sampling notes across registers, which allowed access to all three registers on all instruments. The interval of a perfect fifth, moreover, was deemed

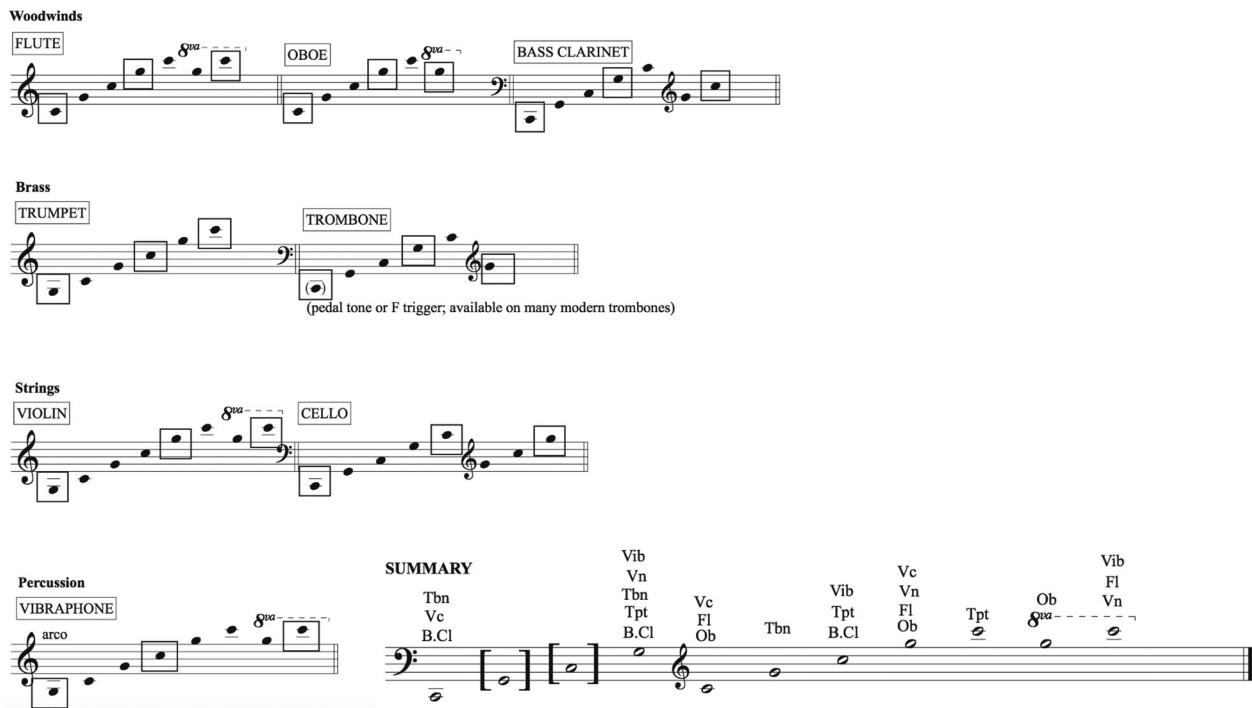


FIGURE 1. Distribution of stimuli for each instrument. Staves illustrate all available pitch classes C and G within each instrument’s range; squares indicate the notes used as stimuli. The summary staff illustrates which notes were represented by which instruments; two notes (C3 and G2) were not used as stimuli.

preferable to intervals that could confound our results with affective connotations via order effects (e.g., a minor third). The selection technique prioritized this low-medium-high distribution across individual instrument ranges. The resulting set of notes is distributed reasonably well across the orchestral ambitus, though not all notes are represented the same number of times in the stimulus set (see Figure 1).

Semantic Scales

The semantic scales used in the rating task were derived from the 20-dimensional model presented in Reymore and Huron (2020), which was built from the results of open-ended interviews and rating tasks. In that study, principal component analysis (PCA) was used to generate the model's dimensions; these 20 dimensions have proven useful in capturing timbre semantic variation (e.g., Reymore, 2021; Reymore, in press). The dimensions of the original model include varying numbers of descriptors, up to seven individual terms per dimension. To reduce cognitive load on participants and decrease the length of the current experiment, we included a maximum of three terms per scale. Table 1 contains the 20 semantic scales used in the experiment.

Given the range of sounds of the stimuli set, which included eight instruments varying in register, we first considered the possibility that some of the dimensions might display little variance and could thus be trimmed from the experiment to increase its tractability. Accordingly, we ran a pilot study with 20 participants to finalize the stimuli set and to potentially reduce the number of semantic scales. However, after considering analyses of pilot study results (correlation matrix among ratings, PCA results, and hierarchical clustering), we determined that, consistent with previous work, the original 20 scales were indeed optimal for this study.

PROCEDURE

Participants were routed from the Prolific recruitment site to GlistenIQ (Bailly, 2020), a custom platform hosted on Google Cloud Platform and created using Node.js and standard web technologies. After

consenting, participants answered demographic questions and questions about their musical background. Participants were asked to wear headphones for the duration of the experiment, and a headphone check following Woods et al. (2017) was implemented: participants judged the relative loudness of pairs of tones, where one of the tones was presented 180° out of phase across stereo channels, a task that is easy with headphones but difficult over loudspeakers due to phase-cancellation. If a participant failed the check, they were given one chance to redo the check. Next, for the purpose of familiarization with the range of stimuli, participants listened to a sound file containing 400 ms clips of all the stimuli (in a randomized sequence). Two trial ratings of sounds not included in the main study were made so that participants could adjust to using the interface. Instruction texts are included in the Supplementary Materials at mp.ucpress.edu.

During the main task, participants were provided with a semantic descriptor and then listened to a set of 24 recordings of musical instruments playing sustained sounds, rating each sound according to the given descriptor. Specifically, they were prompted to rate how well each scale described the sound being played. This process was repeated for each of the 20 semantic scales; that is, all 24 sounds were rated on a single scale before moving on to the next scale. Instructions specified that when scales consisted of more than one word, participants should base their rating on the word in each set that they felt was most applicable. Ratings were made using the number keys on the keyboard on a scale from 1 ("does not describe at all/does not apply") to 5 ("describes extremely well"). The midpoint of the scale (3) was designated as "describes moderately well." Sounds were automatically loaded and played at the beginning of each new prompt, and participants were permitted to replay a sound once. A progress bar was displayed at the top of the screen to inform participants of total study duration.

The order of presentation of the scales was randomized for each participant, as was the order of audio stimuli within each scale block. Two attention check questions were employed at random over the course

TABLE 1. *Semantic Scales Used in the Experiment*

deep, thick, heavy	brassy, metallic	woody
smooth, singing, sweet	raspy, grainy, gravelly	muted, veiled
project, commanding, powerful	ringing, long decay	sustained, even
nasal, buzzy, pinched	sparkling, brilliant, bright	open
shrill, harsh, noisy	airy, breathy	focused, compact
percussive (sharp beginning)	resonant, vibrant	watery, fluid
pure, clear, clean	hollow	

Note: Scales were derived from Reymore and Huron (2020) but were limited to a maximum of three terms per scale.

of the experiment requesting that participants identify, to the best of their ability, the last instrument they heard.

Results

All analyses were carried out in R (R Core Team, 2020; version 4.0.5). Our final dataset consisted of 540 participants' ratings of the 24 instrument stimuli on 20 semantic scales. Table 2 below reports the mean and standard deviation of each stimulus that received the highest mean rating for each semantic scale. Ratings were generally in accord with expectations based on casual observation. For example, the highest rated stimulus on the *airy/breathy* scale was the flute in the middle register, while the sound rated highest on *woody* was deemed to be the cello in the low register, and the most *sparkling/brilliant/bright* sound was the oboe in the high register. Conversely, we can consider the top descriptors for each stimulus. For example, the top descriptor for oboe in the low register was *sustained/even*, for the middle register was *pure/clear/clean*, and for the high register was *shrill/harsh/noisy*.

Figure 2 illustrates mean ratings by instrument type and register. Although some terms show consistent patterns across registers for all instruments (e.g., *deep/thick/heavy*, *sparkling/brilliant/bright*), it is evident from the figure that other terms show varied types of relationships between semantic association and register,

dependent on instrument type (e.g., *brassy/metallic*, *resonant/vibrant*). For illustrations of mean ratings on register (averaged across instruments) and mean ratings on instrument (averaged across register), see the Supplementary Materials at mp.ucpress.edu.

One intriguing observation was that mean ratings of *brassy/metallic* among the 24 stimuli show the least amount of variance (range = 2.21–2.96), despite the inclusion of two brass instruments, suggesting that perceived brassiness may be a product of interactions between instruments and other parameters such as dynamics and or playing technique. Additionally, as all instruments played for approximately the same duration, it is unsurprising that there is very little variance among ratings of *sustained/even* (range = 2.86–3.74). The exception to this was the vibraphone, whose bowing technique, along with the extended ring at the end of the sound, scored lowest overall on *sustained/even*.

Finally, we can illustrate the results using radar plots (Figure 3), following the approach of Reymore (2021). In examining these plots, the vibraphone stands out as demonstrating relatively little semantic variation across registers as compared to the other instruments. By way of contrast, for example, flute, violin, and cello were more likely to be considered *shrill/harsh/noisy* in the upper register, and bass clarinet, trombone, and cello were rated as more *deep/thick/heavy* in the lower register compared to middle and high.

TABLE 2. Top Rated Instrument/Register for Each Semantic Scale, with Associated Mean and Standard Deviation

Semantic scale	Instrument	Register	M	SD
<i>airy, breathy</i>	flute	middle	3.69	1.22
<i>brassy, metallic</i>	trombone	low	2.96	1.49
<i>deep, thick, heavy</i>	cello	low	4.76	0.64
<i>focused, compact</i>	trombone	high	3.14	1.14
<i>hollow</i>	vibraphone	low	3.46	1.46
<i>muted, veiled</i>	vibraphone	low	3.49	1.35
<i>nasal, buzzy, pinched</i>	trombone	low	3.06	1.49
<i>open</i>	bass clarinet	high	3.25	1.09
<i>percussive (sharp beginning)</i>	violin	high	3.65	1.51
<i>pure, clear, clean</i>	vibraphone	high	3.78	1.37
<i>projecting, commanding, powerful</i>	cello	low	3.97	1.21
<i>raspy, grainy, gravelly</i>	cello	low	3.94	1.30
<i>resonant, vibrant</i>	cello	low	3.44	1.52
<i>ringing, long decay</i>	violin	high	3.83	1.37
<i>shrill, harsh, noisy</i>	violin	high	4.54	0.94
<i>smooth, singing, sweet</i>	flute	middle	3.85	1.15
<i>sparkling, brilliant, bright</i>	oboe	high	3.81	1.13
<i>sustained, even</i>	trumpet	medium	3.74	1.09
<i>watery, fluid</i>	vibraphone	low	3.68	1.40
<i>woody</i>	cello	low	3.35	1.46

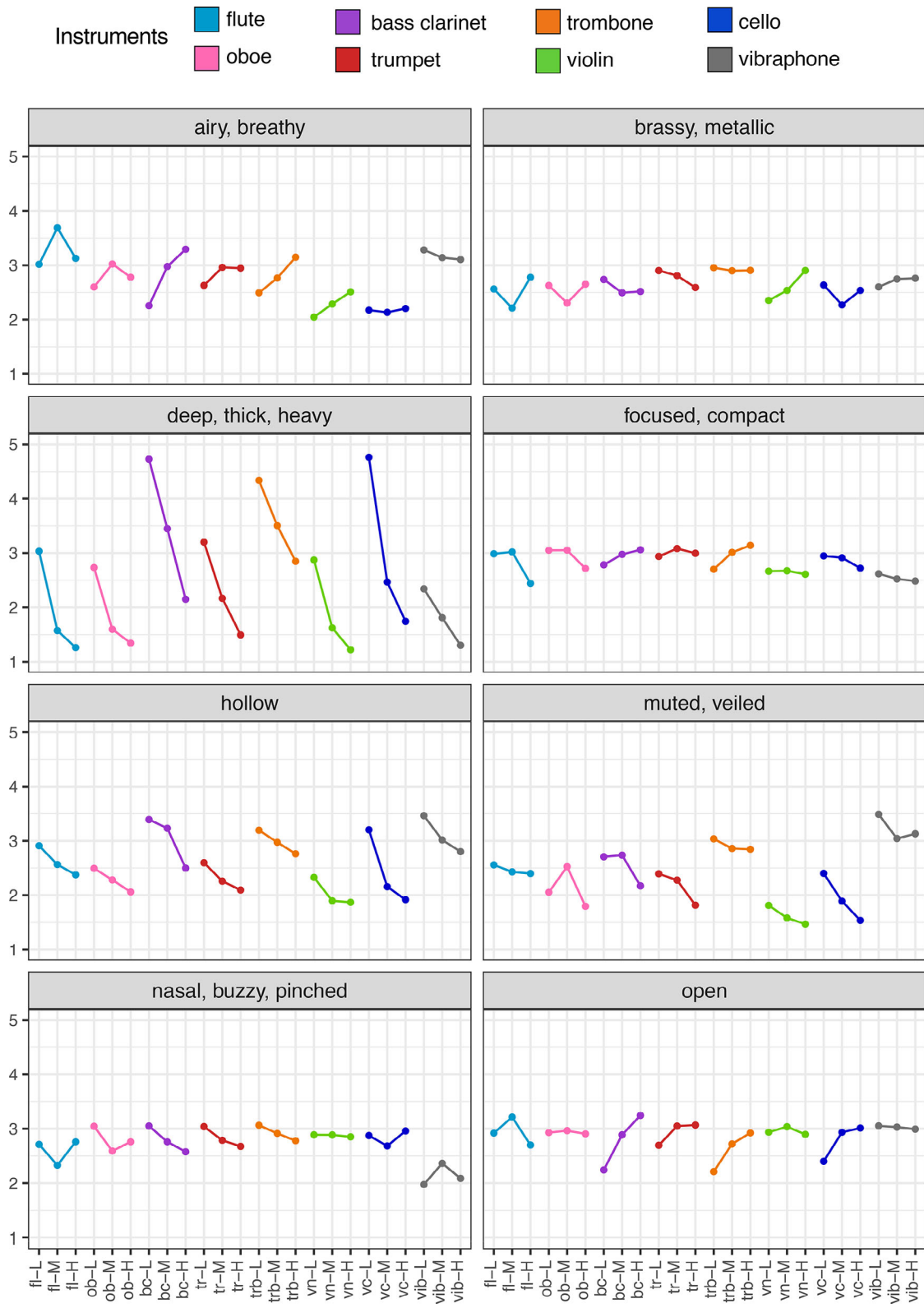


FIGURE 2. Mean ratings by instrument type and register. Within each panel, lines connect mean ratings for low, middle, and high registers of each instrument (left to right); instruments are coded by color. From left to right, instruments are flute, oboe, bass clarinet, trumpet, trombone, violin, cello, vibraphone.

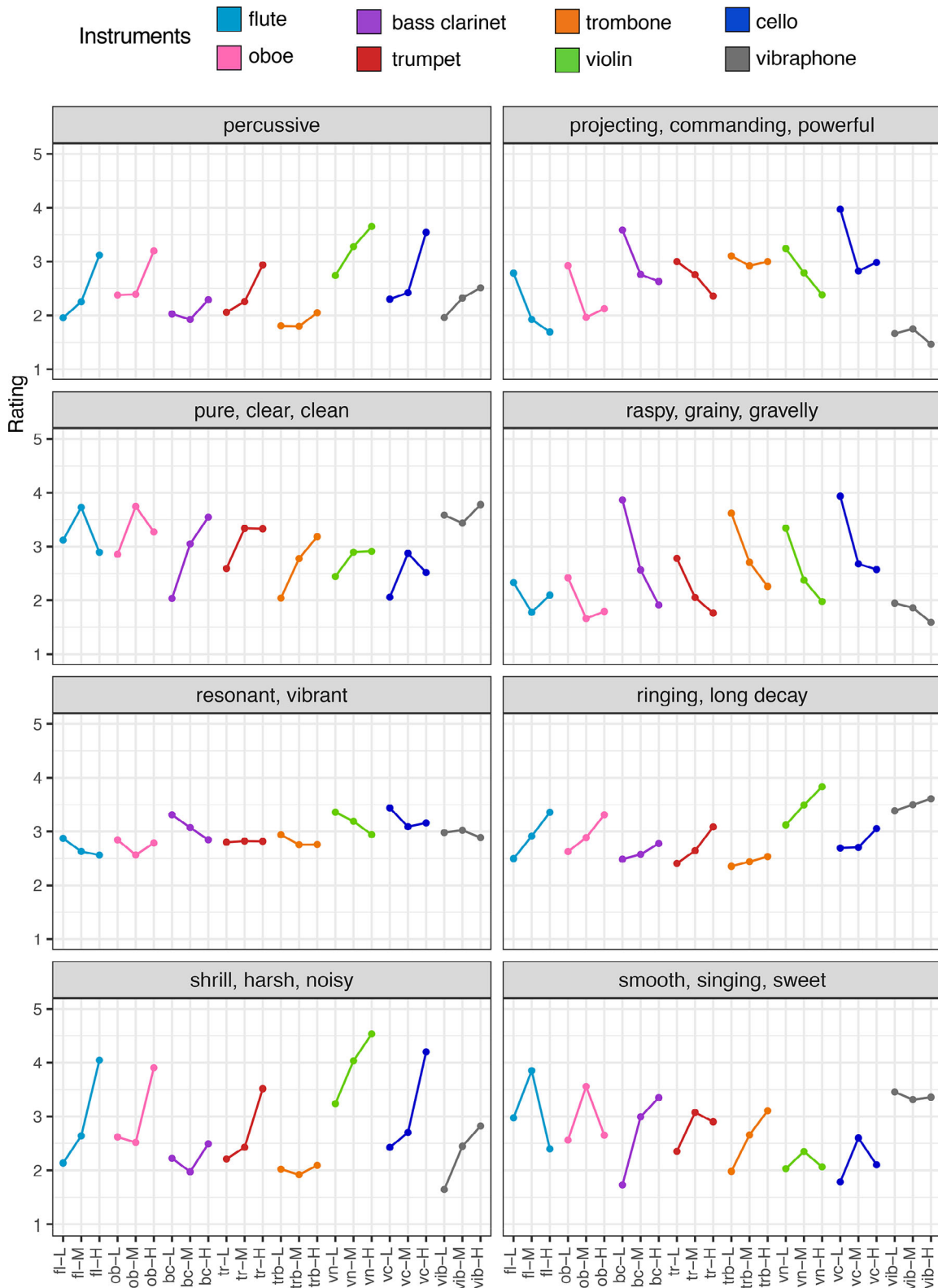


FIGURE 2. (Continued).

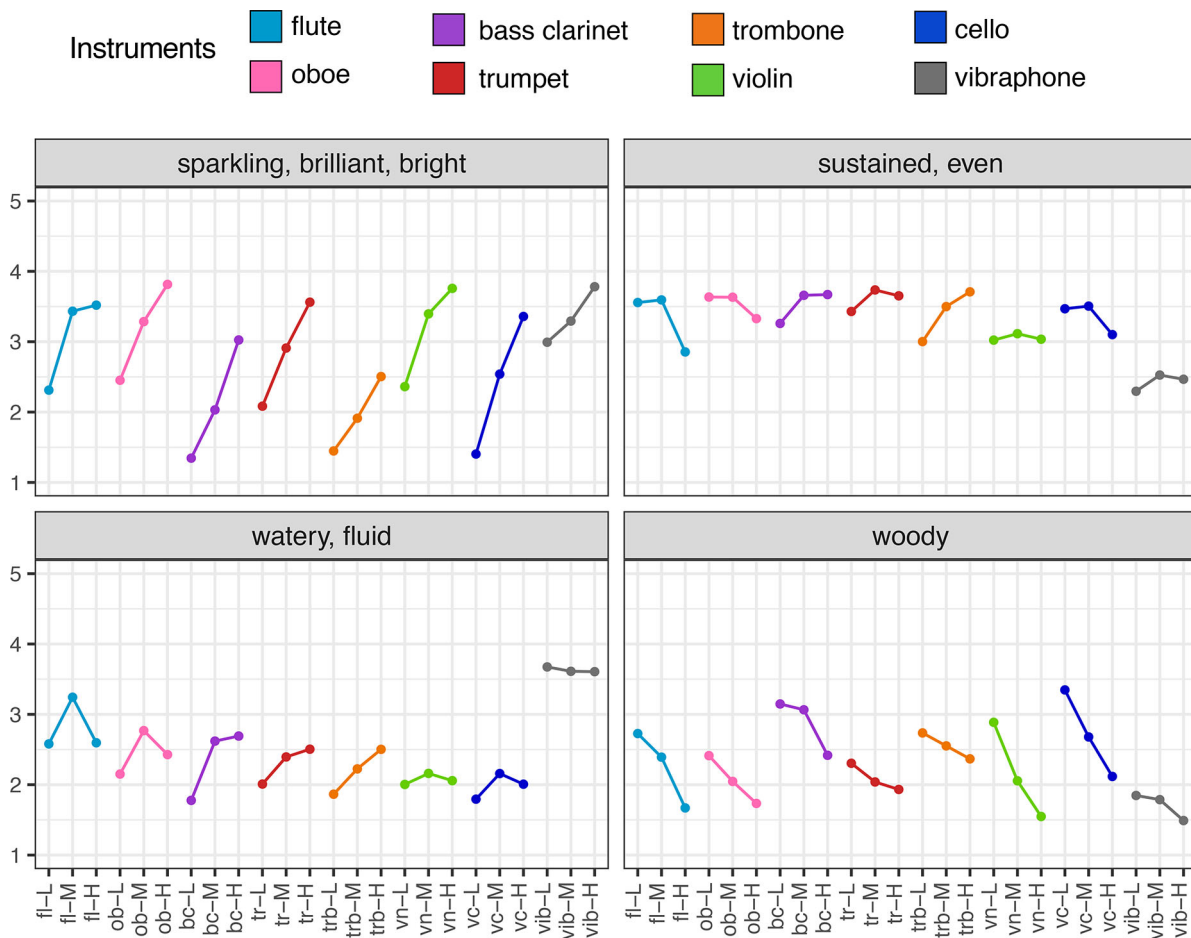


FIGURE 2. (Continued).

Analysis

EXPLORATORY CORRELATION MATRIX

First, the correlation matrix of ratings by dimension was examined to assess the independence of the 20 categories, following Reymore (2021). That is, if the 20 semantic categories are perceptually distinct for the stimuli in this study, we would expect that no two categories should demonstrate strong correlations. Indeed, no strong correlations were observed between pairs of dimensions. The matrix included 12 moderate correlations ranging in absolute value from $r = .30$ to $.43$; the other correlations were each less than $.30$. The strongest positive correlation was between *deep/thick/heavy* and *raspy/grainy/gravelly*, $r(12742) = .43$, 95% CI [.42, .45], $p < .001$, and strongest negative correlation was between *deep/thick/heavy* and *sparkling/brilliant/bright*, $r(12762) = -.42$, 95% CI [-.44, -.41], $p < .001$. These results are similar to those of Reymore (2021), in which

participants used the same 20 semantic categories to rate imagined typical sounds of 34 instruments: in that study, only 26 of the unique correlations had absolute values $\geq .30$.

HIERARCHICAL CLUSTERING

Hierarchical clustering was performed on the rating data to illustrate relationships among stimuli within the 20 semantic categories (Figure 4). This analysis was carried out using the `hclust()` function in the `stats` package, which uses Ward linkage with Euclidean distance.

A number of explanations for the semantic groupings represented in this dendrogram seem plausible. Most obviously, we might expect sounds produced by the same instrument to generate similar semantic profiles. Intriguingly, however, nearest-neighbor pairs suggested by this model are not primarily grouped by instrument (apart from the vibraphone, which stands alone in that its low, medium, and high notes are clustered together).

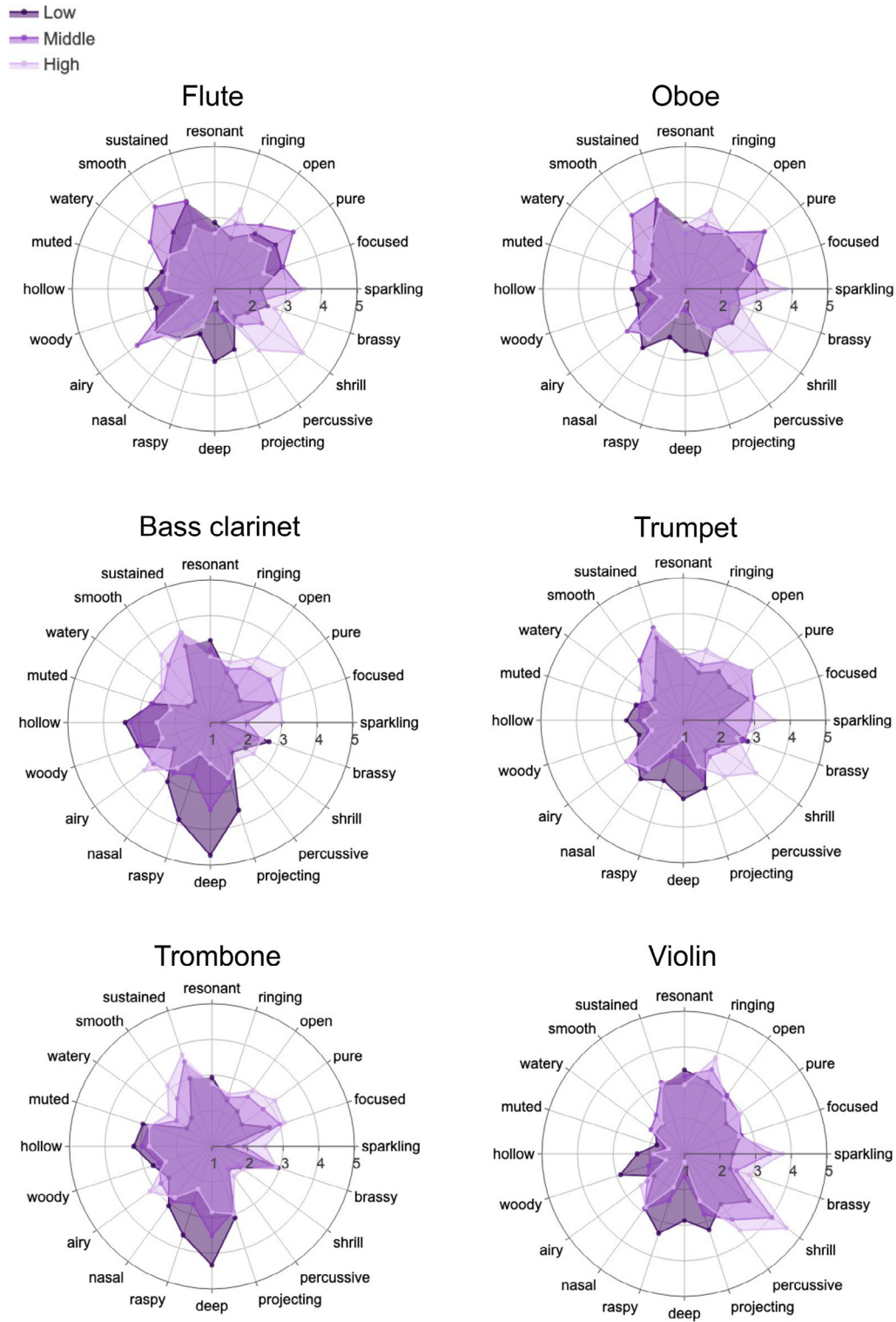


FIGURE 3. Radar plots illustrating mean semantic ratings for each instrument at three registers (high, middle, low). Mean ratings for each semantic category are represented via value on the circle radius; semantic categories are indicated by the first term (for complete categories, see Table 1). Profiles for the low register stimuli are shown in the darkest color, while profiles for the middle register are of a medium lightness, and the profiles for the high register are shown in the lightest color.

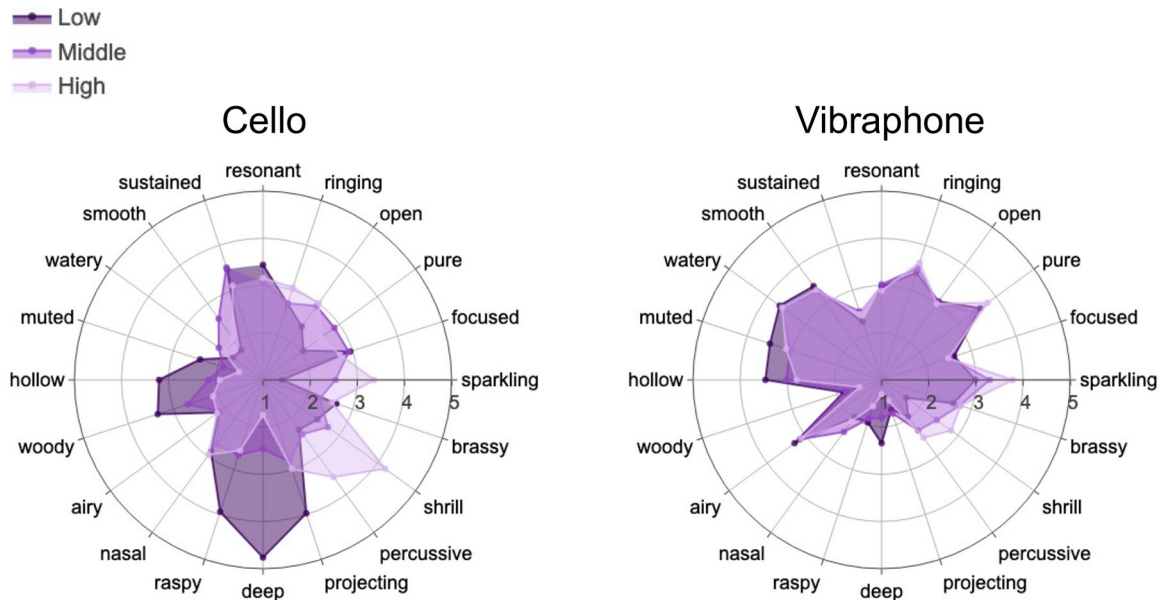


FIGURE 3. (Continued).

We might also expect to observe grouping by instrument family, with strings, woodwinds, brass, and percussion producing internally consistent and distinct semantic profiles. This also does not appear to be a strong factor, at least not according to the conventional orchestral families, as all the clusters contain representatives of at least two different families (with the exception of the vibraphone-only cluster). We also considered the groupings in relation to Hornbostel-Sachs system as used in organology. However, although all the instruments in the first cluster are aerophones, three of the five groupings also cross boundaries of the basic Hornbostel-Sachs categories, including both chordophones and aerophones. Excitation method also seems not to be a strong factor, as the bowed vibraphone is more closely clustered with the blown aerophones than with the bowed chordophones. In sum, these results are not easily explained on the basis of instrumental differences alone.

Considering pitch height, however, provides a much more satisfying interpretation of the clusters. The clusters themselves group neatly by overall orchestral register, containing between one and four notes that are always consecutive according to our overall schema (again, excepting the vibraphone-only cluster). We may consider these to be a *high* cluster (G5–C7), a *medium-high* cluster (C5–G5), a *medium-low* cluster (G3–G4), and a *low* cluster (C2). This model is not perfect, as there is a small amount of overlap: both the high cluster

and the medium-high cluster contain the note G5, indicating that there is still a role for instrumental differences. Nevertheless, the trend for semantic groupings to correspond with pitch height is clear. Relative register also appears to be a predictor, although it is difficult to disentangle the relative roles of pitch height and register. For example, the low cluster contains only the lowest notes from each of the three instruments represented. All other clusters contain mixtures of relative registers: for example, the medium-high cluster contains the high note of the bass clarinet and the middle notes of the oboe, flute, and trumpet.

EXPLORATORY LINEAR MIXED EFFECTS MODELS

To first determine the optimal modeling approach for our data, we computed exploratory models that examined relationships between ratings of each semantic dimension and instrument, register, and musical identity (“musician” vs. “nonmusician”) of the participant. As noted above, register was categorical, relative to each instrument (3 levels: “high,” “middle,” or “low”). Musician identity was derived from responses to the Ollen (2006) single-measure item as a binary category, either “musician” or “nonmusician.” Specifically, we considered instrument, register, and musical background as predictors of mean semantic ratings.

The model building process included an initial assessment of the 20 models’ random structures and consideration of the inclusion of the musical background

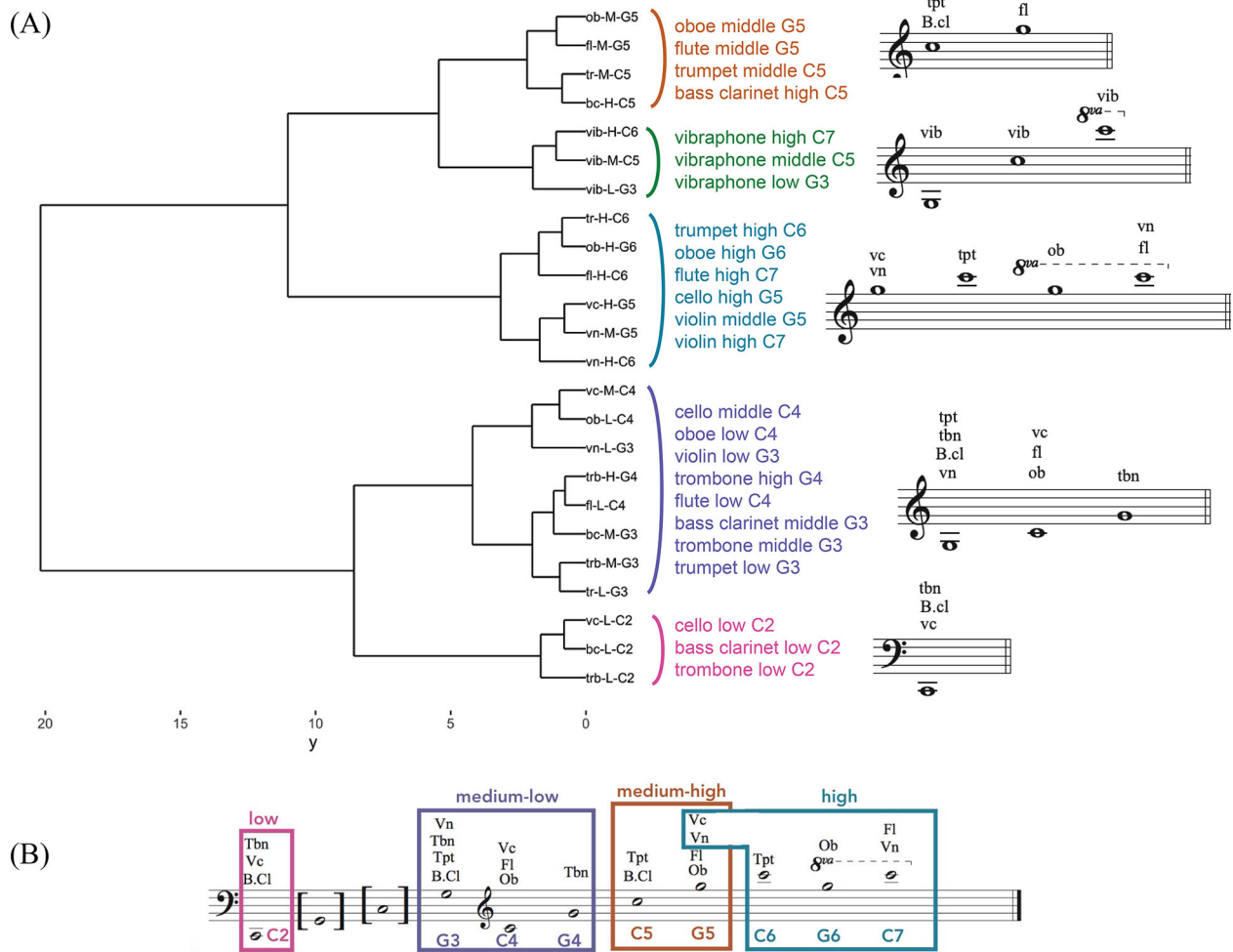


FIGURE 4. (A) shows hierarchical clustering of 24 stimuli (varied by instrument and register) from ratings of 20 semantic dimensions, showing register and pitch height. Notes and clusters are illustrated on a staff, ordered by pitch height, in (B).

variable via log likelihood ratio tests. Specifically, log likelihood ratio tests were first used to compare models with random intercepts only (for participant and stimulus) to models with a maximal random effects structure (based on the methods of McAdams et al., 2017). The maximal effects structure included random intercepts for participant with random slopes for instrument and register and random intercepts for stimulus with random slopes for musical background. We observed that the addition of random slopes significantly increased the goodness-of-fit of all 20 models, and so the maximal random structure was maintained moving forward. Next, log likelihood tests were used again, this time to determine whether musical background significantly contributed to each model. Results were mixed: while most models did not see a significant improvement in goodness-of-fit with the musical background

variable, five models did (*percussive*, *raspy/grainy/gravelly*, *airy/breathy*, *woody*, and *muted/veiled*). Consequently, the musical background variable and the random slope for musical background were included only in these five models moving forward.

Multiple linear mixed effects models for each semantic dimension were then assessed using the dredge() function from the MuMIn package (Barton, 2009), a model selection approach that compares fit among multiple model configurations. Both instrument and register were fixed for inclusion in the final model. This analysis consistently suggested the inclusion of only instrument and register as fixed effects (no interactions), based on consideration of model rankings by AIC, BIC, and AICc. This conclusion applied to all 20 models, including those five that had retained the musical background slope and variable in the previous step.

TABLE 3. Dimensions Ordered by Marginal R^2 , Along with Significance of Register and Instrument

Model (semantic scale)	Marginal R^2	Register		Instrument	
		χ^2	p	χ^2	p
<i>deep, thick, heavy</i>	.53	102.61	< .001**	84.14	< .001**
<i>sparkling, brilliant, bright</i>	.30	140.08	< .001**	110.91	< .001**
<i>shrill, harsh, noisy</i>	.29	37.19	< .001**	53.48	< .001**
<i>raspy, grainy, gravelly</i> [†]	.20	52.76	< .001**	45.55	< .001**
<i>projecting, commanding, powerful</i>	.20	33.07	< .001**	92.86	< .001**
<i>muted, veiled</i> [†]	.16	16.42	< .001**	129.56	< .001**
<i>watery, fluid</i>	.15	15.26	< .001**	110.34	< .001**
<i>percussive</i> [†]	.14	42.89	< .001**	74.34	< .001**
<i>smooth, singing, sweet</i>	.13	10.45	.005**	20.25	.005**
<i>woody</i> [†]	.13	46.63	< .001**	61.89	< .001**
<i>hollow</i>	.12	39.91	< .001**	84.53	< .001**
<i>pure, clear, clean</i>	.11	18.18	< .001**	25.44	< .001**
<i>ringing, long decay</i>	.10	40.44	< .001**	123.74	< .001**
<i>airy, breathy</i> [†]	.09	8.79	.01*	59.29	< .001**
<i>sustained, even</i>	.08	3.54	.17	51.20	< .001**
<i>resonant, vibrant</i>	.03	10.10	.006**	62.01	< .001**
<i>nasal, buzzy, pinched</i>	.03	3.43	.18	39.81	< .001**
<i>open</i>	.02	8.08	.02*	6.22	.51
<i>focused, compact</i>	.02	2.40	.30	16.74	.02*
<i>brassy, metallic</i>	.01	4.22	.12	13.26	.07

Note: Degrees of freedom for all models for Register was 2 and for Instrument was 7.

[†] indicates that random slope for the musician variable was included in the final model. * $p < .05$; ** $p < .01$.

Although the musical background variable was removed from the final models for these five dimensions, the musical background slope was still included.

Register and Instrument

Statistical significance for these final linear mixed effects models was determined through Type III Wald chi-squared tests. Results are shown in Table 3. For most models, main effects of both register and instrument were significant. *Brassy/metallic* was exceptional in that neither register nor instrument were significant predictors of scale performance. Several models indicated significant main effects of one variable but not the other: for example, while register was significant for *open*, instrument was not; conversely, while instrument was significant for *focused/compact*, *nasal/buzzy/pinched*, and *sustained/even*, register was not. Conditional and marginal R^2 values for the final models were calculated using the `r.squaredGLMM()` function from the `MuMIn` package (Barton, 2009). The semantic scales in Table 3 are listed in order of marginal R^2 values (that is, R^2 for fixed effects only) from highest to lowest. Terms that showed the strongest positive associations with register include *shrill/harsh/noisy*, *sparkling/brilliant/bright*, *ringing/long decay*, and *percussive*; terms with the strongest negative associations with register include *deep/*

thick/heavy, *raspy/grainy/gravelly*, *hollow*, and *woody* (see Figure 5).

Register vs. Pitch Height

In addition to asking the question of how register, a variable relative to each instrument, affects semantic associations with timbre, we can also consider how pitch height affects the same semantic associations. Recall that the term “register” refers only to the position of a given note within the overall available range on a given instrument. Because instruments vary greatly in range, the same note may be in the low register of one instrument but in the high register of another. Our principal interest in register came in part from the possibility that certain timbral qualities may be associated with register; that is, the highest and lowest notes an instrument is capable of playing may share particular semantic associations. The results of the exploratory modeling described above suggest that this is the case.

However, as we saw earlier in the hierarchical cluster analysis, another question remains as to what extent explanations of semantic variance from register overlap with explanations from pitch height and whether one of these factors explains more variance than the other. Thus, we focus in our post hoc analysis on critically comparing pitch height and register. Two models were

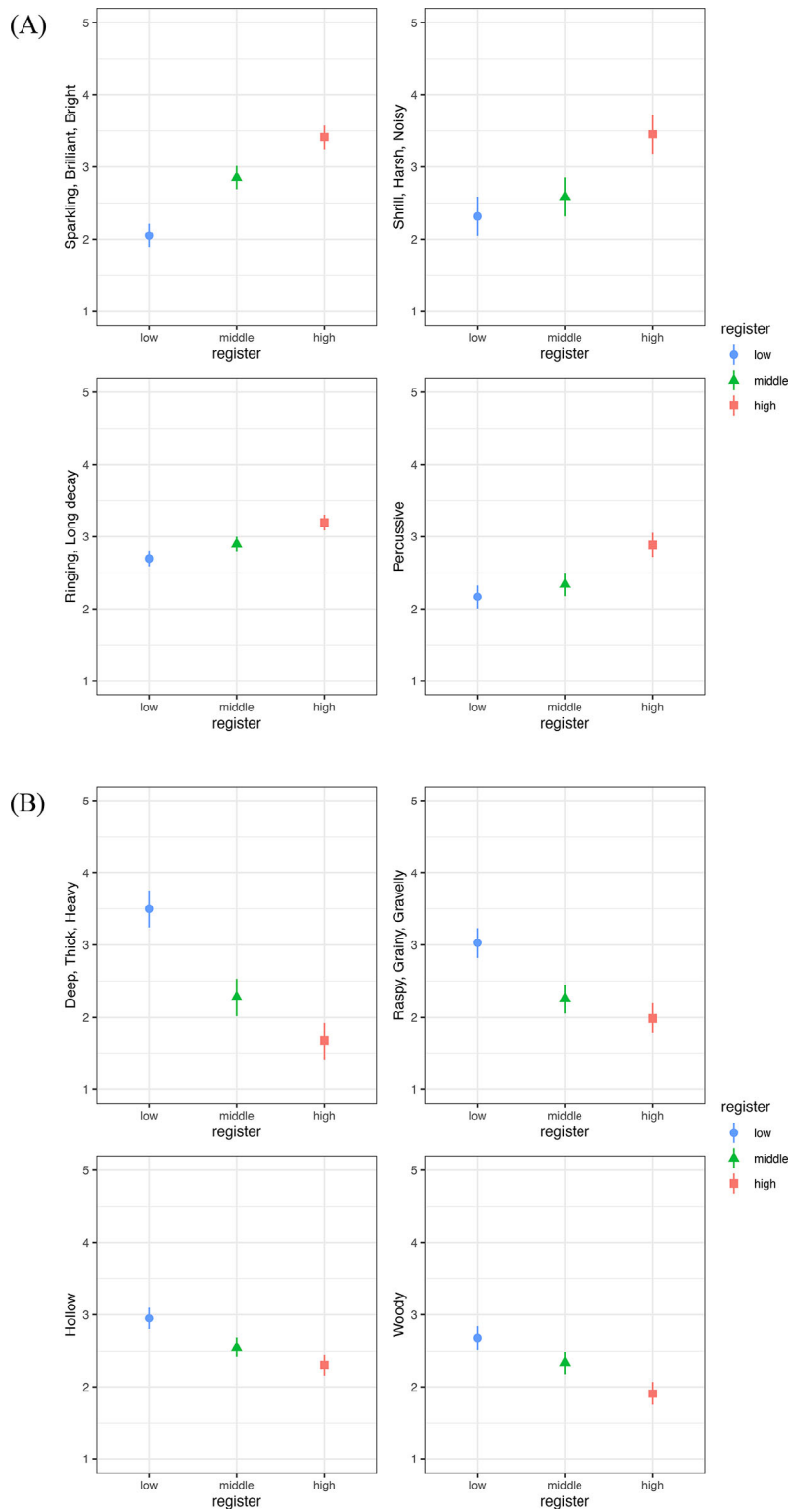


FIGURE 5. Example estimated marginal means and 95% confidence intervals for (A) four semantic scales with the strongest positive associations with register (*sparkling/brilliant/bright*, *shrill/harsh/noisy*, *ringing/long decay*, and *percussive*); (B) four semantic scales with the strongest negative associations with register (*deep/thick/heavy*, *raspy/grainy/gravelly*, *hollow*, *woody*). Graphs of estimated marginal means for register and for instrument for all 20 terms can be found in the Supplementary Materials at mp.uccpress.edu.

considered for each semantic scale, one using only pitch height as a fixed effect and one using only register as a fixed effect. Both types of models included the random effects structures determined through the initial analysis; however, the pitch height models included pitch height as a random slope, whereas the register models included register as a random slope. As with register, pitch height was modeled as a categorical variable. Recall that we designed the experiment to principally consider register; across stimuli, registers were represented equally (that is, each instrument was represented once at each of three relative levels). However, notes were not distributed evenly across stimuli due to the instruments used; although this distribution is not ideal for modeling, all models nonetheless converged successfully.

Log likelihood tests showed that all pitch height-only models demonstrated significantly better goodness-of-fit than register-only models (all p values < .001). We also compared marginal R^2 values between the two types of models. Table 4 reports marginal R^2 (fixed effects only) for models that demonstrated a significant effect for either pitch height, register, or both. The marginal R^2 for pitch height models was consistently higher than that of the register models, suggesting that although both register and pitch height are related to timbre semantic associations, pitch height explains more variance in semantic ratings than does register. It should be noted that the pitch height variable included nine categories, whereas the register variable

included three; the additional degrees of freedom may contribute to the increase in R^2 .

Discussion

The present experiment examined the effects of instrument and register on timbre semantic associations. Analyzing semantic ratings data from a large global sample, we found, for the first time, that most of the 20 ubiquitous semantic categories for timbre vary with register and instrument. Exceptions include *brassy/metallic*, which did not significantly vary with either register or instrument; *sustained/even*, *nasal/buzzy/pinched*, and *focused/compact*, which varied with instrument but not with register; and *open*, which varied with register but not with instrument. However, considering marginal R^2 as a measure of effect size, it is apparent that magnitude varies greatly among semantic categories. This measure is particularly important in the interpretation of the results. For example, it is apparent that for those cases described above in which instrument, register, or both were not significant, the marginal R^2 values are low (all under .10). In other cases, as with *airy/breathy* and *resonant/vibrant*, both register and instrument were significant, yet R^2 is again less than .10. The secondary analyses comparing pitch height-only and register-only models also provide insight into the effect size with respect to pitch height and register specifically (in comparison to the instrument-register models reported in the primary analysis, where R^2 values may include overlapping variance). In other words, although our primary analysis demonstrated significant relationships between register and semantic category for 16 of the 20 models, the effect size for many of these categories is relatively small. Indeed, the secondary analysis reveals that register alone explains 5% or more of the variance for only 8 of the 20 categories. Among these, four categories stand out as most relevant for both register and pitch height: *deep/thick/heavy*, *sparkling/brilliant/bright*, *shrill/harsh/noisy*, and *raspy/grainy/gravelly*.

In interpreting our results, one should also keep in mind that our a priori model selection method resulted in models with no interaction terms, as interactions did not significantly improve model fit. However, examination of the mean ratings by both instrument and register (Figure 1) demonstrates that there may often be at least potential mild interaction between instrument and register, in that not all instruments always change on a given category with respect to register in the same manner. Thus, the results of the exploratory modeling reported here suggest which semantic categories

TABLE 4. Marginal R^2 for Pitch Height-Only vs. Register-Only Mixed Effects Models for the 13 Semantic Scales with R^2 Values Equal to or Greater than .05 for Either Pitch Height or Register

Semantic Scale	Marginal R^2	
	Pitch height only	Register only
<i>deep, thick, heavy</i>	.54	.29
<i>sparkling, brilliant, bright</i>	.29	.16
<i>shrill, harsh, noisy</i>	.20	.12
<i>raspy, grainy, gravelly</i>	.20	.08
<i>projecting, commanding, powerful</i>	.13	.05
<i>woody</i>	.12	.05
<i>pure, clear, clean</i>	.09	.05
<i>percussive</i>	.09	.05
<i>smooth, singing, sweet</i>	.08	.04
<i>hollow</i>	.08	.04
<i>muted/veiled</i>	.06	.03
<i>ringing, long decay</i>	.06	.02
<i>watery/fluid</i>	.05	.02

demonstrate global effects of register on semantic rating, but this does not necessarily mean that each individual instrument always follows the same pattern. Furthermore, it is also possible that terms that did not reach significance or that explain little variance globally (such as *nasal/buzzy/pinched* or *open*) might actually demonstrate variation by register for certain individual instruments but not for others. For example, register explained very little overall variance for ratings of *open*. However, observation of rating means (Figure 1) suggests that *open* hardly varied at all with register for oboe, violin, and vibraphone but did show notable variation for other instruments, such as the bass clarinet and trombone.

To explore in more detail how the semantic-register relationship can vary among instruments, consider Figure 2. We can see that some terms showed directional effects of register that were clearly consistent across all eight instruments, including *deep/thick/heavy*, *sparkling/brilliant/bright*, *woody*, *hollow*, and *ringing/long decay*. However, even among terms that demonstrated overall significance, we can observe that not all instruments behaved in similar ways. For example, for *smooth/singing/sweet*, the bass clarinet and trombone showed increased mean ratings with higher register (positive relationship). However, flute, oboe, trumpet, violin, and cello were all deemed to be most *smooth/singing/sweet* in their middle registers but less so in both the low and high register (convex relationship). The bowed vibraphone was exceptional in that it showed almost no difference with respect to this semantic category in relation to register. As a second example, although most instruments decreased on ratings of *raspy/grainy/gravelly* as register increased (negative relationship), the flute and the oboe were least *raspy/grainy/gravelly* in their middle registers (concave relationship).

Thus, it may be useful to consider the categories that did not demonstrate significant relationships with register in order to refine our understanding of how these relationships may or may not vary by instrument. For example, all instruments showed little variation on *brassy/metallic* with respect to register; however, flute, oboe, and cello demonstrated concave relationships, with the middle register rated as least *brassy/metallic*. The trumpet decreased on *brassy/metallic* with register (negative relationship), while the violin increased (positive relationship). Although these effects are small, they nonetheless may be compositionally useful in some circumstances. Previously, we noted that mean *brassy/metallic* ratings demonstrated the least amount of variation among the stimuli and suggested that this semantic category may be more closely related to parameters

other than pitch, such as dynamics or playing technique. Another possible contributing factor to this lack of variation might be differences in the understanding between musicians and nonmusicians of the terms “brassy” and “metallic” as applied to timbre. Note that the semantic category was derived from studies involving professional musicians, whereas our current participants were largely nonmusicians, and our musician participants were nearly all amateurs. It is unclear, for example, to what extent knowledge of an instrument’s material influences ratings of relevant terms, including *brassy/metallic* and *woody*.

In general, more research is needed to understand potential differences in how musicians and nonmusicians apply semantic terms to timbre. Our consideration of this question in the current experiment is to some extent limited by the imbalance between musicians and nonmusicians in our sample: note that as our research question did not involve differences between these groups, we did not aim to recruit equal numbers. For this reason, musician identity was included in the exploratory modeling procedure as a potential fixed effect. Although the dredge() step of the procedure ultimately did not recommend including musician identity in the final models, earlier steps showed that the addition of musician identity as a predictor significantly improved model fit for 5 of the 20 semantic categories. This suggests that music training may play a role in timbre semantics; however, this preliminary result requires direct testing in future work.

It should be noted that because of potential instrument-register differences, our global results from the models may be in part a product of the particular set of eight instruments that were chosen for the experiment and the extent to which they demonstrated similar trends to the other instruments in the group. That is, they do not necessarily represent global trends across all musical instruments. However, the eight instruments used here are diverse in instrument family and range, providing a solid representation of typical Western orchestral instruments. In future research, it will be necessary to expand the palette of instrumental sounds to determine whether our results generalize across a broader range of musical timbres.

Collecting data online necessarily introduces variability that may affect results, including variation in the quality of listening devices used by participants. To mitigate this, we aimed to recruit a large sample size and incorporated a headphone check at the beginning of the experiment. Although this does not provide control over headphone quality, our results show significant, systematic differences in ratings among stimuli for most

of the semantic scales, suggesting that consistent responses can be obtained with variation in headphone quality. Furthermore, as discussed below, a portion of our data replicates data collected in Reymore (in press), a study that was carried out in a sound booth with quality headphones and controlled volume. Both studies collected ratings on the same 20 semantic categories for the oboe across registers, and results are remarkably consistent despite the differences in platform (laboratory vs. online) and participant pool (music majors vs. mixed, primarily nonmusicians).

Reymore (in press) collected ratings across four registers and three dynamic markings for the oboe and French horn. The three dimensions in the current study with the highest marginal R^2 values (*deep/thick/heavy*, *sparkling/brilliant/bright*, and *shrill/harsh/noisy*) also appeared to be among the most important dimensions for both the oboe and the horn in the previous study. Several other dimensions overlap as important for both instruments in that study and for the assessment across eight instruments in the current study, including *raspy/grainy/gravelly*, *smooth/singing/sweet*, *open*, and *resonant/vibrant*. Reymore (in press) also shows converging evidence that registral effects on semantic associations can be instrument-specific; for example, the equivalent dimensions to *projecting/commanding/powerful* and *watery/fluid* were shown to be significantly affected by register for the horn, but not for the oboe, while *woody* and *muted/veiled* were significantly associated with register for the oboe, but not for the horn. Interestingly, the category including the term “nasal” did not demonstrate significance across the eight instruments in the current study but did so for the in press study, for both oboe and horn. More specifically, however, ratings for the oboe are similar across both studies, with the lowest ratings for *nasal/buzzy/pinched* in the middle registers, suggesting reliability across studies and further suggesting that, as described earlier, it may be the case that we did not observe a global significant effect (across instruments) in the current study for *nasal/buzzy/pinched* in part because different instruments vary on these terms in different ways with respect to register.

Hayes et al. (2022) further demonstrate that effects of register on timbre-semantic associations can be context-specific. Presented with a frequency modulation (FM) synthesizer, sound designers were asked to create a sound that has more or less of a given semantic attribute (e.g., is rougher or less rough) than a played reference. Participants were then asked to rate the created sound against the reference on additional semantic attributes (e.g., how much thicker or less thick it is).

Reference sounds were presented at three registers, but no significant relationship was found between stimulus register and the semantic ratings. It is possible that the use of comparative versus absolute (Reymore, in press; Zacharakis et al., 2014; this study) ratings effectively rendered any registral effects null. On the other hand, their lack might be the result of the specific characteristics of FM synthesis. The introduction of sidebands both above and below the fundamental frequency of the carrier operator during the synthesis task might have falsely implied a lower or higher pitch/register, likely due to interference between pitch height and auditory brightness. In other words, in certain contexts perceptual interactions between pitch and timbre may obscure semantic interactions between the two attributes.

How do our findings compare to the written discourse of orchestral timbre? A number of best-fit adjectives in our study correspond to patterns gleaned from a corpus of orchestration treatises (Wallmark 2019a). For example, in the current study, the mid-register flute was rated highest for *smooth/singing/sweet* and *airy/breathy*, while “sweet” and “breathy” are among the most common descriptors for the flute across the orchestration corpus. Low trombone was the most *nasal/buzzy/pinched* (“nasal” is a top descriptor of the trombone in the corpus), and low cello the most *deep/thick/heavy* (“deep” is a common cello adjective). Besides these few convergences, the present findings differ considerably from the corpus: in fact, none of the top 10 corpus descriptors for the oboe, trumpet, and violin were replicated in the current study. Perhaps this lack of convergence is the result of the specific way in which register was defined in the current study (Wallmark, 2019a, only examined descriptions at the level of individual instruments); further, treatises were the product of professional musicians, while the present study drew upon ratings from a general global sample.

A more detailed examination that would account for semantics in relation to different groupings of instruments (for example, by orchestral family or Hornbostel-Sachs category) remains elusive at this point and will require more data in future studies. A related question pertains to the extent to which instrument recognition may affect semantic judgments, particularly regarding semantic categories that are nominally associated with an instrument’s material (i.e., *woody* and *brassy/metallic*). Although we did not directly measure participant instrument recognition ability, data collected from the attention check questions suggest that between one half and three quarters of our participants were able to recognize instruments to some extent. However, the results do not show evidence that knowledge of an instrument’s

material were driving ratings. For example, we did not find instrument to be a significant predictor of ratings of *brassy/metallic*. From Figure 2, it is apparent that in some cases, we observed the opposite of what would be expected if ratings were based on knowledge of the instrument material. For example, all three registers of the trombone and the high register of the violin received about equal ratings on *brassy/metallic*, and the high register of the trombone was judged to be woodier than the high register of the violin. As another example, the low register of the flute was considered to be woodier than the low register of the oboe. However, it remains plausible that prior knowledge may affect ratings, possibly more so with respect to the more typical middle register sounds of an instrument, which are likely more easily identifiable. The extent to which instrument knowledge affects semantic judgments may also vary by semantic category; such questions provide an intriguing topic for future research.

The respective contributions of relative register and pitch height also remain to be clarified, though our comparative modeling demonstrates that models using pitch height rather than instrument register to predict semantic ratings explain more of the variance and offer a significantly improved model fit. In general, the pitch height-only models tended to explain around twice as much variance as did the register-only models. This is an intriguing finding that challenges the intuitive assumption that corresponding registers between instruments would be more semantically similar than corresponding notes. This finding contributes to developing understanding of the concept of *metatimbres* (Soden, 2020), collections of timbres related by shared attributes—in this case, pitch—that may lead them to be grouped perceptually and semantically. Metatimbres may also be organized around other shared attributes such as attack quality, spectral composition, dynamic level/contour, and so forth. Specifically, results from the comparative modeling suggest that it is possible that pitch height may be a greater influence on the perceptual organization of metatimbres than register or instrument. As noted above, however, this post hoc comparative approach has limitations, in that the study was designed using stimuli equally distributed by register but not pitch height, and that the additional degrees of freedom in the pitch height-only models have the potential to increase R^2 . Thus, we understand our finding of the relative strength of pitch height to provide a hypothesis for future experiments, rather than a complete account. Future studies should explore the many possibly relevant parameters for metatimbre by directly testing the relative influences of pitch height, register, and instrument.

Finally, acoustical modeling of listeners' semantic patterns will help further delineate the nature of semantic interactions between timbre and pitch height/register. For instance, in analyzing spectral envelopes from 50 sustained orchestral instruments sampled across their entire range, Siedenburg et al. (2021) found complex statistical dependencies between both pitch and register and two key correlates of timbre perception, namely the position of the envelope, as modeled by the spectral centroid (Caetano et al., 2019) and spectral shape.

Conclusion

Overall, our results suggest that both instrument type and fundamental frequency have crucial roles in determining a wide range of timbre-semantic associations. While this may seem intuitive and in keeping with the descriptions found in orchestration textbooks, it represents an advancement in the field of timbre semantics research, which has tended to exclude register from the discussion. Further, our study is the first of its kind with the statistical power to detect small effects. Across our group of eight varied instruments, most semantic categories demonstrated a significant relationship with both register and instrument. For some terms, register explained more of the total variance in semantic ratings, especially *deep/thick/heavy*, *sparkling/brilliant/bright*, *shrill/harsh/noisy*, and *raspy/grainy/gravelly*. Other categories, such as *nasal/buzzy/pinched* and *sustained/even*, seemed to be driven more by instrument (and likely other factors) than register. Results also demonstrate that there can be differences in the types of relationships that instruments have with register; for example, the relationship between register and *raspy/grainy/gravelly* is concave for the flute, with middle register rated lower than the high and low registers, but linear for the bass clarinet, where ratings decrease as register increases.

In addition to the specific findings reported above, we take the present study to be a proof of concept that sound sources, in particular musical instruments, are not necessarily the optimal “units” of timbre semantics in research and creative practice. Common descriptions such as “the oboe is nasal” and “the trumpet is brilliant” are partial accounts, representing semantic nuclei associated with prototypical subsets of the timbres that such instruments are capable of producing. Since each instrument represents “a constrained universe of timbres” (McAdams & Goodchild, 2017, p. 129), each also represents a constrained universe of potential semantic associations, with vast potential for meaningful variation within a single instrument and

even vaster potential for meaningful correspondences and interactions with other instruments. Composers and musicians have long tapped into these potentials: the challenge before us as researchers is to come to an explicit understanding of mechanisms that have been implicitly applied for centuries. In this study, we hope to have provided a model of analysis that may be productively applied to this end, which may help to mitigate the significant challenges posed by the complexity of the many-to-many mappings underlying semantic associations, the advanced cognitive processes involved in relating attributes of sound to extramusical domains, and the large degree of inter-subjective variation bound to be encountered in any area of musical interpretation. We anticipate that the results, especially the timbral profiles provided by mean ratings of semantic descriptors, will be useful in composition and orchestration as well as for music analysis, both in practice and in pedagogy. They may also be informative for future research in timbre perception—for example, in informing stimuli selection and providing a foundation for conceptualizing

relationships between timbre-semantic associations and register across instruments. Results also offer a unique opportunity to begin analyzing the relative contributions of pitch height and register in timbre semantics.

Author Note

The authors would like to express their thanks to the Analysis, Creation, and Teaching of Orchestration (ACTOR) Partnership for supporting data collection with ACTOR Strategic Project Funding.

Zachary Wallmark is co-founder of Glisten Labs, LLC, a venture that hopes to make commercially available the GlistenIQ experiment design platform, a developmental version of which was used free of charge in this research.

Correspondence concerning this article should be addressed to Lindsey Reymore, School of Music, Dance and Theatre, Arizona State University, 50 E Gammage Parkway, Tempe, AZ, USA 85281. E-mail: lreymore@asu.edu

References

- ADELI, M., ROUAT, J., & MOLOTCHNIKOFF, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Frontiers in Human Neuroscience* 8, 352. <https://doi.org/10.3389/fnhum.2014.00352>
- ADLER, S. (2002). *The study of orchestration*. W.W. Norton.
- ALLEN, E. J., & OXENHAM, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *Journal of the Acoustical Society of America*, 135(3), 1371–1379. <https://doi.org/10.1121/1.4863269>
- BAILLY, C. D. (2020). *GlistenIQ*. [Computer software]. Glisten Labs.
- BARTON, K. (2009). *Mu-MIn: Multi-model inference*. R Package Version 0.12.2/r18. <http://R-Forge.R-project.org/projects/mumin/>
- CAETANO, M., SAITIS, C., & SIEDENBURG, K. (2019). Audio content descriptors of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 297–333). Springer.
- CHMIELEWSKI, M., & KUCKER, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4), 464–473. <https://doi.org/10.1177%2F1948550619875149>
- COUSINEAU, M., CARCAGNO, S., DEMANY, L., & PRESSNITZER, D. (2014). What is a melody? On the relationship between pitch and brightness of timbre. *Frontiers in Systems Neuroscience*, 7, 127. <https://doi.org/10.3389/fnsys.2013.00127>
- COX, A. (2016). *Music and embodied cognition: Listening, moving, feeling, and thinking*. Indiana University Press.
- DRABKIN, W. (2001). Register. In *Grove Music Online*. <https://doi.org/10.1093/gmo/9781561592630.article.23072>
- EEROLA, T., FERRER, R., & ALLURI, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, 30(1), 49–70. <https://doi.org/10.1525/mp.2012.30.1.49>
- FINK, R., LATOUR, M., & WALLMARK, Z. (Eds.). (2018). *The relentless pursuit of tone: Timbre in popular music*. Oxford University Press.
- 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>
- HANDEL, S., & ERICKSON, M. L. (2001). A rule of thumb: The bandwidth for timbre invariance is one octave. *Music Perception*, 19(1), 121–126. <https://doi.org/10.1525/mp.2001.19.1.121>
- HAYES, B., SAITIS, C., & FAZEKAS, G. (2022). Disembodied timbres: A study on semantically prompted FM synthesis. *Journal of the Audio Engineering Society*, 70(5), 373–391. <https://doi.org/10.17743/jaes.2022.0006>
- HURON, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, 19(1), 1–64. <https://doi.org/10.1525/mp.2001.19.1.121>

- KENDALL, R. A., & CARTERETTE, E. C. (1993). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives. *Music Perception, 10*, 445–467. <http://dx.doi.org/10.2307/40285583>
- KRUMHANSL, C. L., & IVERSON, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 739–751. <https://psycnet.apa.org/doi/10.1037/0096-1523.18.3.739>
- MAROZEAU, J., & DE CHEVEIGNÉ, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *Journal of the Acoustical Society of America, 121*(1), 383–387. <https://doi.org/10.1121/1.2384910>
- MAROZEAU, J., DE CHEVEIGNÉ, A., MCADAMS, S., & WINSBERG, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America, 114*(5), 2946. <https://doi.org/10.1121/1.1618239>
- MCADAMS, S., DOUGLAS, C., & VEMPALA, N. N. (2017). Perception and modeling of affective qualities of musical instrument sounds across pitch registers. *Frontiers in Psychology, 8*, 153. <https://doi.org/10.3389/fpsyg.2017.00153>
- MCADAMS, S., & GOODCHILD, M. (2017). Musical structure: Sound and timbre. In R. Ashley & R. Timmers (eds.), *The Routledge companion to music cognition* (pp. 129–139). Routledge.
- 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*(3), 177–192.
- MELARA, R. D., & MARKS, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics, 48*(2), 169–178. <https://doi.org/10.3758/BF03207084>
- OLLEN, J. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* [Dissertation]. Ohio State University.
- R CORE TEAM (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- REYMORE, L. (in press). Variations in timbre qualia with register and dynamics in the oboe and French horn. *Empirical Musicology Review*.
- REYMORE, L. (2021). Characterizing prototypical musical instrument timbres with Timbre Trait Profiles. *Musicae Scientiae, 26*(3), 648–674. <https://doi.org/10.1177/2F10298649211001523>
- REYMORE, L., & HURON, D. (2020). Using auditory imagery tasks to map the cognitive linguistic dimensions of musical instrument timbre qualia. *Psychomusicology, 30*(3), 124–144. <https://doi.org/10.1037/pmu0000263>
- REUTER, C., JEWANSKI, J., SAITIS, C., CZEDIK-EYSENBERG, I., SIDDIQ, S., & OEHLER, M. (2018). Colors and timbres – Consistent color-timbre mappings at non-synesthetic individuals. In *Proceedings of the 34. Jahrestagung der Deutschen Gesellschaft für Musikpsychologie: Musik im audiovisuellen Kontext*. German Society for Music Psychology.
- SAITIS, C. (2019). Beyond the semantic differential: Timbre semantics as crossmodal correspondences. In *Proceedings of the 14th International Symposium on CMMR* (pp. 338–345). International Symposium on Computer Music Multidisciplinary Research.
- SAITIS, C., FRITZ, C., SCAVONE, G. P., GUASTAVINO, C., & DUBOIS, D. (2017). Perceptual evaluation of violins: A psycholinguistic analysis of preference verbal descriptions by experienced musicians. *Journal of the Acoustical Society of America, 141*(4), 2746–2757. <https://doi.org/10.1121/1.4980143>
- SAITIS, C., & SIEDENBURG, K. (2020). Brightness perception for musical instrument sounds: Relation to timbre dissimilarity and source-cause categories. *Journal of the Acoustical Society of America, 148*(4), 2256–2266. <https://doi.org/10.1121/10.0002275>
- SAITIS, C., & WEINZIERL, S. (2019). The semantics of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 119–149). Springer.
- SCHUBERT, E., & WOLFE, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica United with Acustica, 92*(5), 820–825.
- SHEPARD, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review, 89*(4), 305–333.
- SIDDIQ, S., REUTER, C., CZEDIK-EYSENBERG, I., & KNAUF, D. (2018). Towards the physical correlates of musical timbre(s). In *Proceedings of ICMPC15/ESCOM10* (pp. 411–415). ICMPC15/ESCOM10.
- SIEDENBURG, K. (2018). Timbral Shepard-illusion reveals ambiguity and context sensitivity of brightness perception. *Journal of the Acoustical Society of America, 143*(2), EL93–EL98. <https://doi.org/10.1121/1.5022983>
- SIEDENBURG, K., & MCADAMS, S. (2017). Four distinctions for the auditory “wastebasket” of timbre. *Frontiers in Psychology, 8*:1747. <https://doi.org/10.3389/fpsyg.2017.01747>
- SIEDENBURG, K., JACOBSEN, S., & REUTER, C. (2021). Spectral envelope position and shape in sustained musical instrument sounds. *Journal of the Acoustical Society of America, 149*(6), 3715–3726. <https://doi.org/10.1121/10.0005088>
- SODEN, K. (2020). *Orchestration combinations and transformations in operatic and symphonic music*. [Doctoral dissertation]. McGill University.
- STEELE, K. M., & WILLIAMS, A. K. (2006). Is the bandwidth for timbre invariance only one octave? *Music Perception, 23*(3), 215–220. <https://doi.org/10.1525/mp.2006.23.3.215>

- TRAUBE, C. (2004). *An interdisciplinary study of the timbre of the classical guitar*. [Doctoral Dissertation]. McGill University.
- THORET, E., CARAMIAUX, B., DEPALLE, P., & MCADAMS, S. (2021). Learning metrics on spectrotemporal modulations reveals the perception of musical instrument timbre. *Nature Human Behaviour*, 5(3), 369–377. <https://doi.org/10.1038/s41562-020-00987-5>
- VIENNA SYMPHONIC LIBRARY GMBH (2011). *Vienna Symphonic Library*. Available online at: <http://vsl.co.at>
- WALLMARK, Z. (2019a). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music*, 47(4), 585–605. <https://doi.org/10.1177/0305735618768102>
- WALLMARK, Z. (2019b). Semantic crosstalk in timbre perception. *Music and Science*, 2, 1–18. <https://doi.org/10.1177/2059204319846617>
- WALLMARK, Z., & KENDALL, R. A. (2018). Describing sound: The cognitive linguistics of timbre. In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 578–608). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190637224.013.14>
- WARD, J., HUCKSTEP, B., & TSAKANIKOS, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, 42(2), 264–280. [https://doi.org/10.1016/S0010-9452\(08\)70352-6](https://doi.org/10.1016/S0010-9452(08)70352-6)
- WEIHS, C., REUTER, C., & LIGGES, U. (2005). Register classification by timbre. In C. Weihs & W. Gaul (Eds.), *Classification—the ubiquitous challenge* (pp. 624–631). Springer.
- WOODS, K. J., SIEGEL, M. H., TRAER, J., & MCDERMOTT, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, 79(7), 2064–2072.
- ZACHARAKIS, A., PASTIADIS, K., & REISS, J. D. (2014). An inter-language study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception*, 31, 339–358. <https://doi.org/10.1525/mp.2014.31.4.339>
- ZACHARAKIS, A., PASTIADIS, K., & REISS, J. D. (2015). An inter-language unification of musical timbre: Bridging semantic, perceptual, and acoustic dimensions. *Music Perception*, 32, 394–412. <https://doi.org/10.1525/mp.2015.32.4.394>
- ZBIKOWSKI, L. M. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford University Press.