Psychomusicology: Music, Mind, and Brain

Creating Novel Tones From Adjectives: An Exploratory Study Using FM Synthesis

Zachary Wallmark, Robert J. Frank, and Linh Nghiem

Online First Publication, July 1, 2019. http://dx.doi.org/10.1037/pmu0000240

CITATION

Wallmark, Z., Frank, R. J., & Nghiem, L. (2019, July 1). Creating Novel Tones From Adjectives: An Exploratory Study Using FM Synthesis. *Psychomusicology: Music, Mind, and Brain*. Advance online publication. http://dx.doi.org/10.1037/pmu0000240

American Psychological Association

Psychomusicology: Music, Mind, and Brain

© 2019 American Psychological Association
0275-3987/19/$12.00

http://dx.doi.org/10.1037/pmu0000240

Creating Novel Tones From Adjectives: An Exploratory Study Using FM Synthesis

Zachary Wallmark and Robert J. Frank
Southern Methodist University

Linh Nghiem
Australian National University

Perceptual studies of timbre semantics have revealed certain consistencies in the linguistic conceptualization of acoustic attributes. In the standard experimental paradigm, participants hear timbral stimuli and provide behavioral responses. However, it remains unclear the extent to which descriptive consistency would be observed if this paradigm were reversed, that is, if participants were instructed to *create* novel timbres in response to target adjectives. Given an unfamiliar synthesis interface, would musically trained participants craft similar timbral profiles for the same familiar adjectives? In this study, we explore timbre semantics using a novel frequency modulation (FM) synthesis production task. Participants ($N = 64$) created unique timbral outputs in response to 20 common timbre descriptors drawn from orchestration treatises (e.g., *brilliant*, *dull*, *harsh*). Acoustic analyses of the resultant 1,280 signals, in conjunction with linear mixed-effects modeling and clustering analysis, indicate that participants were moderately consistent in their timbral creations. Word valence and arousal interacted to influence average spectral centroid and noisiness. Specifically, clearly positive and negative words produced significantly different acoustical profiles than more affectively neutral words. This result confirms a number of findings from the perceptual literature while offering preliminary evidence that affective dimensions of timbre semantics systematically influence sound production in an unfamiliar context.

Keywords: timbre semantics, timbre perception, FM synthesis, timbre acoustics, musical affect

Supplemental materials: http://dx.doi.org/10.1037/pmu0000240.supp

Lacking its own domain-specific terminology, timbre is often described using affectively congruent adjectives borrowed from other domains. A sound might be considered “penetrating” or “dark,” “harsh” or “melancholy.” Although the semantic norms of timbre description are more subjective than those governing other musical parameters, certain broad commonalities in descriptive practices have been noted (Kendall & Carterette, 1993a, 1993b; Lichte, 1941; Pratt & Doak, 1976; von Bismarck, 1974). For example, von Bismarck (1974) identified four common semantic structures for timbre, *full–empty*, *dull–sharp*, *colorful–colorless*, and *compact–diffused*. More recently, Alluri and Toiviainen (2010) reported that terms related to *brightness*, *activity*, and *fullness* are common to the discourse of timbre. Relatedly, Zacharakis, Pastiadis, and Reiss (2014, 2015) compared English and Greek speakers to demonstrate that timbre semantics is largely undergirded by *luminance*, *texture*, and *mass* terms. In short, timbre description appears to rely on fairly systematic and conventionalized semantic associations.

In most timbre semantics studies, researchers play participants sound signals that are manipulated along one or more categorical or continuous dimension, then record behavioral responses in the form of semantic differential scale ratings, free verbal response, adjective checklists, or other procedures (see Susini, Lemaitre, & McAdams, 2012). In addition to investigating the effect of the experimental manipulation, it is also common for researchers to examine the acoustic correlates of perceptual response, often by building predictive models explaining behavioral data by way of computationally extracted acoustic descriptors (Eerola, Ferrer, & Alluri, 2012; McAdams, Douglas, & Vempala, 2017; Wallmark, 2019a). In this way, researchers can evaluate which specific psychoacoustic variables modulate semantic response. Inferences drawn from this paradigm are unidirectional; that is, they aim to explain the perceptual effect of timbre on semantic classification. However, this method does not address the relationship between timbre acoustics and semantics from the opposite direction: How do the affective dimensions of words modulate *acoustical* response? To be sure, it remains relatively unknown whether the inverse applies to the timbre–language relationship in the act of musical creation. Given an unfamiliar sound generation interface,

Zachary Wallmark, Division of Music and Department of Psychology, Southern Methodist University; Robert J. Frank, Division of Music, Meadows School of the Arts, Southern Methodist University; Linh Nghiem, College of Business and Economics, Australian National University.

We thank Tom Tunks and Sarah Allen for early brainstorms leading to this project, Stephen McAdams and the members of the CIRMMT lab at McGill for crucial feedback on analysis and interpretation, and Jess Pinkham for help in data acquisition.

Correspondence concerning this article should be addressed to Zachary Wallmark, who is now at School of Music and Dance, University of Oregon, 1225 UO, Eugene, OR 97403. E-mail: zwallmar@uoregon.edu

would a sample of musicians—people accustomed to shaping timbre in musical performance—create fairly consistent timbral profiles to match familiar descriptive words?

To explore this question, it is first crucial to operationalize the affective mechanisms linking auditory perception to lexical access. Timbre is often considered a "sign-post for the emotions" (Boulez, 1987). Indeed, psychobiological studies indicate that the affective component of timbre perception is incorporated very early (Peretz, Gagnon, & Bouchard, 1998; Tervaniemi, Winkler, & Näätänen, 1997), and is fundamental to how we describe sound in later processing stages. Affective language is ubiquitous to the discourse of timbre: In an analysis of timbre description in orchestration texts, Wallmark (2019b) found that over one third of the corpus was purely affective (e.g., *fine*, *expressive*, *melancholic*), representing the most widespread conceptual framework for description. To investigate the acoustic correlates of timbre affect dimensions, Eerola et al. (2012) extracted acoustic data from a set of signals rated on valence and arousal scales. They found that just a couple acoustic descriptors could reliably predict affective response, especially the ratio of high-frequency/low-frequency energy in the signal, which was strongly associated with negative valence and high arousal (but see McAdams et al. [2017] for conflicting evidence). Given these consistencies, it is reasonable to hypothesize that the affective connotations of target adjectives would have an impact on the creation of novel synthetic sounds, with high-arousal adjectives (e.g., *brilliant, harsh*) tending to elicit timbral outputs with a higher spectral center of gravity relative to low-arousal words (*melancholic*, *tender*).

Affective Meaning

The affective meaning of words has often been studied using a dimensional model of the emotions (Mehrabian & Russell, 1974; Russell & Mehrabian, 1977). This approach is often thought to have originated in the influential work of Osgood, Suci, and Tannenbaum (1957), who reported that three orthogonal dimensions—*evaluation*, *activity*, and *potency*—explained about half of the total variance in semantic differential ratings, with the first factor accounting for almost 70% of common variance. Subsequent models replicated this basic tripartite dimensional structure (though with slightly different nomenclature and operationalizations; see Bakker, Van der Voordt, Vink, & De Boon, 2014). Mapping roughly onto the three factors of Osgood et al., *valence* refers to the pleasantness of the emotion evoked by a word; *arousal* is the implied activation or intensity; and *potency* (or *dominance*) is the perceived strength or degree of control implied by a word. These dimensions have been used in a number of semantic norms databases used in affective analysis of texts and discourse (Bradley & Lang, 1999; Warriner, Kuperman, & Brysbaert, 2013), as well as studies of music emotion recognition and induction (Aljanaki, Yang, & Soleymani, 2017; Eerola & Vuoskoski, 2011; Schubert, 2007).

Descriptive Adjectives in Sound Synthesis

The development of intuitive sound synthesis technologies using adjectives to control timbral parameters has long been considered something of a "holy grail" for researchers in sound design and synthesis (Carron, Rotureau, Dubois, Misdariis, & Susini, 2017). Musicians and listeners commonly conceptualize timbre according to semantic qualities, but interfaces traditionally rely on the manipulation of numerical parameters. This deficit of "semantic directness" (Seago, Holland, & Mulholland, 2004) can limit broad engagement with sound synthesis and negatively impact the creative process (Miranda, 2002). In an early attempt to remedy this gap, Ashley (1986) employed frequency modulation (FM) synthesis in a trial-and-error paradigm in which users responded with verbal descriptions to arbitrary changes in timbre to "teach" the system to associate certain settings with adjectival qualities. The SeaWave system of Ethington and Punch (1994) allowed users to shape a given "parent timbre" additively by manipulating attack, presence, and cutoff dimensions through the selection of varying degrees of adjectival properties (e.g., *plucked* attack, *resonant* presence, *damped* cutoff). The authors found, for instance, that "brightness" corresponded with increasing harmonic density.

Incorporating recent machine learning techniques with the aim of developing a system for automated synthesis, Gounaropoulos and Johnson (2006) used a neural net to train a timbre classification algorithm on a number of synthesized sounds and adjectives then set the algorithm loose classifying a hold-out sample of signals. In a related study, Kreković, Pošćić, and Petrinović (2016) used fuzzy logic algorithms to transform adjectival inputs into timbral outputs, which were then validated by musically trained raters. However, the acoustical correlates of adjectives or affective dimensions were not explored in these articles. Eschewing semantic classification, other researchers have developed graphical sound synthesis and visualization interfaces based on cross-modal correspondences between timbral parameters and dimensions of visual and tactile sensation (Giannakis & Smith, 2000; Soraghan, Faire, Renaud, & Supper, 2018). Though not explicitly linguistic, such synesthetic mappings between timbre and vision and touch are common in the everyday discourse of sound (for review, see Wallmark & Kendall, in press).

Study Aim

The present study investigates the affective mechanisms underlying the semantics of timbre in the context of synthetic sound generation. Our main goal is not to develop a fully operational semantic-based timbre control interface, but rather to use a simple synthesis patch to explore acoustic regularities in timbre creation when participants are prompted with adjectives of varying affective connotations.

This study explores the timbre–language connection by reverse engineering the standard paradigm. We ask the following question: Can musicians reliably and consistently create novel tones in an unfamiliar sonic context based on familiar verbal descriptors alone? To explore this question, we constructed an FM synthesis interface designed to allow participants to actively sculpt novel timbres in response to target adjectives. This study focuses on the synthetic sounds generated by musicians because musical training has been shown to increase perceptual acuity toward timbre (Chartrand & Belin, 2006; Siedenburg & McAdams, 2018). Twenty descriptive adjectives common to the discourse of instrument timbre were provided to 64 musically trained participants, who created "best fit" timbres for each word. Audio signals for the resulting 1,280 individual timbral creations were then analyzed using mixed-effects models and hierarchical clustering procedures to examine acoustic uniformities as a function of the affective meanings of the adjectives along valence and arousal dimensions.

Method

Participants

Sixty-four musically trained participants completed the experimental task (30 females), $M_{\text{age}} = 20.7$, $SD = 3.2$. The majority (57) were undergraduate and graduate music majors at Southern Methodist University Meadows School of the Arts, a conservatory-style music department, and the remainder (seven) were student musicians (non–music-majors) enrolled in a music appreciation course; the mean number of years musical training was 11.8 years ($SD = 3.9$). Participants received extra course credit for participation. The study was approved by the Southern Methodist University Institutional Review Board.

Stimuli

Our aim in word stimuli selection was to determine some of the most commonly used and intuitive instrumental timbre adjectives in the symphonic tradition. To do so, we analyzed two canonical orchestration treatises (Berlioz, 1882; Rimsky-Korsakov, 1933) for their use of timbre descriptions. Word stimuli consisted of the 20 most frequent descriptive terms for musical timbre in these texts: All terms pertaining to instrumental timbre were extracted manually from these sources following the procedure outlined by Wallmark (2019b). Together the books included a total frequency of 940 timbre descriptors, of which 357 represented unique tokens. This list of descriptors was then ranked according to frequency.

To determine the affective connotations of the word stimuli, we classified the most frequent adjectives according to their implied valence, arousal, and dominance using a well-established database of ~14,000 scored and validated English word stems (Warriner et al., 2013). This tripartite affective structure is based on the model of Osgood et al. (1957), and has been used in many similar semantic norms databases used in sentiment analysis, natural language processing, and corpus linguistics (Bradley & Lang, 1999). However, methodological issues in the Warriner et al. (2013) data set made dominance ratings hard to interpret (participants in that study were asked to respond to how "dominant" and "in control" *they felt* when reading each word, rather than rating the implied dominance of the words themselves). Due to this uncertainty in the operational definition, which is consistent with previous work demonstrating similar ambiguities with the dominance or potency dimension (Osgood et al., 1957), we opted to focus our analysis on word valence and arousal. This is consistent with many music psychology studies that use the circumplex model of affect (Russell, 1980; see Zentner & Eerola, 2010). Ratings for these two dimensions in Warriner et al. (2013) were originally recorded using a 1–9 Likert-type scale (negative–positive valence, low–high arousal); means and standard deviations for each of the 20 common timbre adjectives included in the Warriner et al. (2013) data set are shown in Table 1.

Additionally, adjectives were sorted by mean valence and arousal ratings into three roughly balanced ordinal categories: low, medium/neutral, and high valence/arousal. Thus, for instance, "tender" was classified as positive valence (i.e., high) and low arousal, whereas "dark" was neutral valence (i.e., medium) and medium arousal. Each category consisted of six to seven individual adjectives. An ordinal organization was selected to account for affectively borderline or ambiguous words, which we felt might be relevant in the context of timbre description. (The adjective "piercing" does not appear in the norms database; classifications for this word were determined by the authors.)

Table 1
Adjective Stimuli and Affective Ratings

Adjective	Valence ($M = 5.06$, $SD = 1.68$)	Arousal ($M = 4.21$, $SD = 2.30$)
Bright	Positive (6.84, 1.86)	Medium (5, 2.45)
Brilliant	Positive (7.5, 2.28)	High (5.95, 2.8)
Charming	Positive (7.05, 2.15)	Medium (5, 2.79)
Dark	Neutral (5.08, 1.8)	Medium (4.09, 2.43)
Dull	Negative (3.4, .94)	Low (1.67, 1.03)
Full	Neutral (6, 2.02)	Low (3.48, 2.04)
Gloomy	Negative (3.15, 1.63)	Low (3.32, 2.12)
Hard	Neutral (4.35, 1.97	Medium (4.5, 2.75)
Harsh	Negative (3.44, 1.42)	High (5.63, 2.29)
Melancholic	Negative (3.74, 1.69)	Medium (4.13, 2.56)
Mysterious	Neutral (6.05, 1.32)	High (5.45, 2.04)
Nasal	Neutral (4.26, 1.45)	Low (3.38, 1.6)
Noble	Positive (7.16, 1.95)	Medium (4.14, 2.31)
Penetrating	Neutral (5.71, 1.76)	High (6.08, 2.48)
Piercing	Negative (NA)	High (NA)
Rich	Positive (6.83, 2.31)	High (6.81, 2.04)
Rough	Negative (3.68, 1.69)	High (5.43, 1.97)
Sweet	Positive (7.77, 1.38)	Medium (4.14, 2.92)
Tender	Positive (6.47, 1.75)	Low (3.22, 2.21)
Veiled	Negative (4.14, 1.25)	Low (3.32, 1.97)

Note. NA = not available. Table displays the 20 most frequent adjectives used by Berlioz (1882) and Rimsky-Korsakov (1933) to describe instrumental timbre. Mean ratings and corresponding standard deviations for valence and arousal are shown in parenthesis (from Warriner, Kuperman, & Brysbaert, 2013).

Experimental Interface

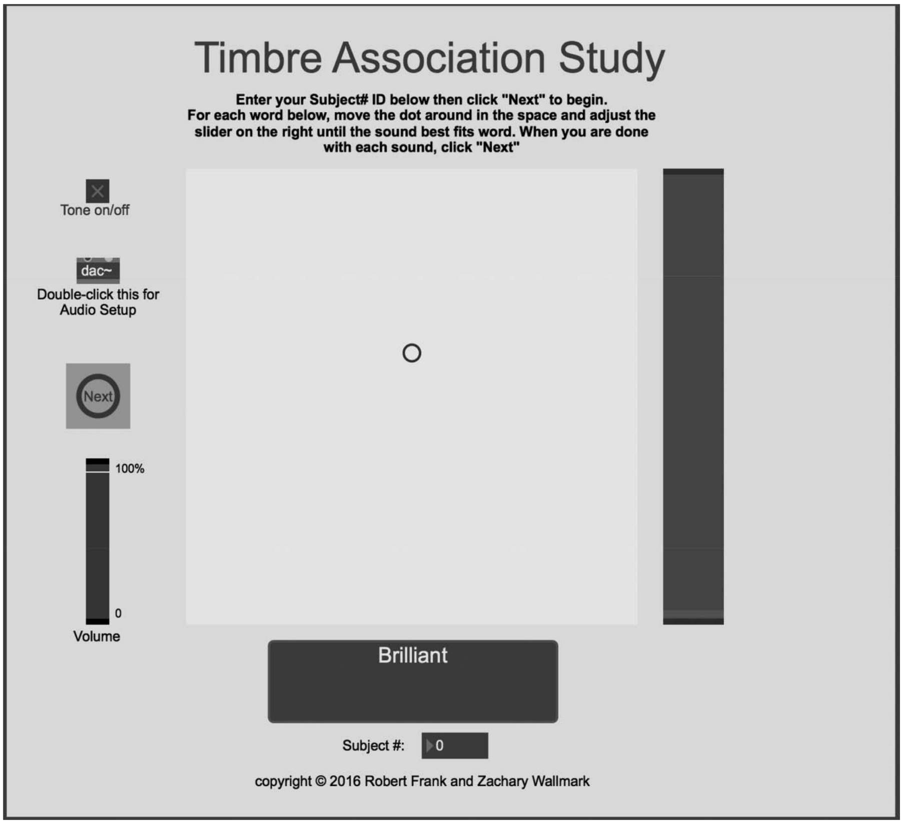
The testing instrument used a simple FM generator programmed in the Max/MSP environment. FM, developed and outlined by Chowning (1977), alters the frequency of one oscillator (carrier) via the output of a second oscillator (modulator). The ratios between the frequencies of these two oscillators, or the *harmonicity ratio*, causes distortion of the waveform and extreme variance in timbre via resultant harmonic frequencies symmetrically both above and below the carrier frequency (i.e., sidebands). In this study, lower harmonicity ratios (0–1) were chosen to limit the range, as wider variance was determined in pilot testing to be too overwhelming for participants.

One strength of the FM method using the present "sweet spot" ratios is that it is perceptually nonlinear but also relatively intuitive to navigate; that is, the interaction of carrier and modulator results in a wide range of sometimes unpredictable sonic qualities that do not map onto any monotonic perceptual scales within the 2D space, while also constraining the more wildly variant options characteristic of FM synthesis (Ashley, 1986). This nonlinearity was preferable here to mitigate demand effects and encourage more exploration of the available space by musician participants (e.g., if corners represented the perceptual extremes of an affectively relevant linear acoustic scale, such as spectral centroid, we might expect participants to respond to "extreme" words, once these associations are learned, simply by dragging the cursor to the corners). The interface is shown in Figure 1.

The object generated a pure sine wave signal. Input variables from the interface were chosen following aural experimentation to provide as much variety of timbre while retaining continuity. The final interface model used the following input specifications: carrier frequency fixed at 440 Hz; *x*-axis harmonicity ratio (scaled from 0 to 1, linear, left to right); and modulation index fixed at 0.2. (Additional details about the FM synthesis interface and an example of the patch itself can be found in the online supplementary materials; see Figure S1 in the online supplemental materials.) The output of the FM generator was then fed directly into a resonant bandpass filter with a center frequency controlled by the *x*-axis (100–2100 Hz, left to right) and a Q controlled by the *y*-axis (0–28, bottom to top; Figure S2 in the online supplemental materials). The addition of this filter had the effect of sweeping the center frequency from left to right, and narrowing and broadening the filter width (Q) from bottom to top. Additionally, the slider to the right of the *x*/*y* space controlled a distortion amplifier applied to the normalized signal, amplifying it linearly from 100% (no change) to 499% overdrive, resulting in heavy clipping distortion. This signal was then normalized using the Max/MSP "clip~" object to constrain the max/min output level to a consistent 100% before sending the output to the participant.

Participants were able to toggle the audio on/off using the click box in the upper right of the interface (see Figure 1). The small circle in the main field was able to be moved via mouse click-drag motions, with audio settings updated every 100 ms. The slider on the right was also click-drag assignable with continuous real-time updates of the data. Participants were instructed not to adjust the volume slider during the experimental task. Adjectives were presented below the interface; after participants created a timbre that best matched each word using mouse click-drag motions, they selected "Next," which reset to an X value of 50%, Y value of 60%, and slider value of 100% (null). "Next" also advanced participants to the proceeding word until all 20 were completed. The data sent from the main *x*/*y* space resulted in values from 0 to 1,000 on each axis, and 0 to 499 (%) for the slider. These coordinate values were then saved to the data files.

The harmonicity ratio has inherent consonant nodes at *x*-axis (prescaled) values of 0 (no modulation), 500 (1:2), and 1,000 (1:1), and lesser audible nodes at 333 (1:3), 250 (1:4), and 200 (1:5), and their symmetrical nodes at 660 (2:3), 750 (3:4), and 800 (4:5), where sidebands and difference tones of those ratios are generated. The additional resonant bandpass filter was used to reduce the similarity of these basic ratio nodes. The *y*-axis filtering would have slight reinforcement of the carrier at data values (prescaled) of 170 (1:1), 390 (2:1), 610 (3:1), and 830 (4:1). Lesser perceptible nodes that resonate at first-level sidebands occur at data values of 60, 280, 500, and 670.



Timbre Association Study

Enter your Subject# ID below then click "Next" to begin.
For each word below, move the dot around in the space and adjust the slider on the right until the sound best fits word. When you are done with each sound, click "Next"

copyright © 2016 Robert Frank and Zachary Wallmark

Figure 1. Study interface.

Procedure

To counterbalance for potential order effects, three versions of the Max/MSP protocol were created, each with a different randomly ordered presentation of the adjectives. Participants were randomly assigned in equal numbers to the different versions.

The study took place in a quiet room, and participants recorded their responses on iMac computers listening through Bose SoundTrue headphones. To familiarize them with the experimental interface, prior to beginning they were given a brief training session (~2 m) in which they were encouraged to explore the timbral space and adjust the computer volume to a comfortable listening level. We then instructed them that they would be presented with different adjectives on the screen; their task would be to manipulate the timbral environment to find a quality of sound that "best fits" the given word. They could take as much time as they wanted for this task. None of the participants reported any difficulty or confusion afterward with the experimental task. The duration of the experiment was ~15 min.

Results

Exploratory Acoustic Descriptor Analysis

We created WAV sound files (44.1 kHz sampling rate) from the coordinate data for acoustic analysis. Twenty-two low-level acoustic descriptors were computationally extracted from the 1,280 audio files using MIRtoolbox 1.6.1 in the MATLAB environment (Lartillot & Toiviainen, 2007). Because there were no temporal features due to continuous audio playback, only spectral characteristics of the signals were computed. These standard spectral parameters are displayed in Table 2. In addition to timbre descriptors available in the MIRtoolbox, we calculated "subband flux," or the average frequency fluctuation in 10 octave-scaled bands of the signal (Alluri & Toiviainen, 2010): This frequency-segregated index of spectrotemporal change was found to be perceptually relevant in semantic judgments of timbre (Alluri & Toiviainen, 2012; Eerola et al., 2012; Wallmark, 2019a). Zero-cross was trimmed due to insufficient variability across the signals, leaving 21 acoustic variables, including RMS power (root mean square) as a measure of total signal energy and a rough proxy for perceived loudness.

To reduce the number of acoustic variables to a more manageable number, a principal component analysis (PCA) with varimax rotation was conducted on the whole set of standardized variables for the 1,280 output signals. Sampling adequacy for the PCA was confirmed (Kaiser–Meyer–Olkin index = .84). The procedure generated three latent acoustical components with eigenvalues greater than 1 that together explained 72% of variance. As shown in Table 3, PC1 (30%) was characterized by strong loadings on subbands 9 and 10, rolloff, brightness, centroid, flatness, entropy, roughness, and inharmonicity, suggesting that the first PC relates to strength in high-frequency components and a relatively noisy spectrum. For convenience, we can label PC1 the "Intensity" factor. PC2 (23%) was driven by loadings on spectral fluctuation in the middle, musical frequency range of 100–6400 Hz (hereafter, the "Flux" factor). Finally, PC3 (19%) was characterized primarily by RMS strength and brightness, with strong negative loadings on skewness and kurtosis ("Loudness" factor). These three uncorrelated PCs were used to derive optimally weighted linear combinations of the acoustic descriptors for use as factor scores (−1 to 1) in the subsequent analysis (see Table S1 in the online supplemental materials for mean factor scores organized by word).

Modeling the Association Between Affective Meaning and Acoustic Factor Scores

To explore the relationship between semantic structure and acoustics, we next used a linear mixed-model approach (West, Welch, & Galecki, 2006) to determine whether affective meaning had any systematic effect on the three acoustic factors described above. Lexicon-based studies of semantic orientation typically approach affective meaning in text as either gradations along a scale of polarity (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011)—for example, from *very negative* to *very positive*—or as a classification problem, where affective words are placed into a single nominal or ordinal category of best fit (e.g., see the popular binary lexicon of Liu, 2015). Hence, two parallel versions of this analysis were calculated using the nlme package in R (Pinheiro et al., 2018): one corresponding to the interval version of the semantic data (Warriner et al., 2013), the other to ordinal (three levels: low, medium, high). For the ordinal analysis, *F* tests (Type III sum

Table 2
Acoustic Descriptors

Descriptor	Definition
Subband flux (10 regions)	Spectrotemporal fluctuation within 10 frequency bands (Alluri & Toiviainen, 2010)
Zero-cross rate	Number of signal changes per unit of time
Rolloff	Frequency threshold below which 95% of energy is contained
Brightness	Proportion of total spectral energy above 1500 Hz
Centroid	Center of spectral energy distribution
Spread	Standard deviation of spectral energy
Skewness	Asymmetry of spectrum
Flatness	Wiener entropy of signal
Kurtosis	Flatness of spectrum around mean
Entropy	Shannon entropy of signal
Irregularity	Degree of variation between successive spectral peaks over time
Roughness	Sensory dissonance averaged through time
Inharmonicity	Frequency deviation of partials from ideal harmonic series

Table 3
Principal Component Analysis (PCA) on Acoustic Descriptors

Acoustic descriptor	PC1 (30%)	PC2 (23%)	PC3 (19%)
Subband 10 (12.8–22 kHz)	**.86**	.27	.21
Flatness	**.81**	—	.31
Centroid	**.80**	—	.52
Rolloff	**.79**	—	.51
Roughness	**.78**	—	.37
Subband 9 (6.4–12.8 kHz)	**.74**	.47	.20
Entropy	**.72**	.23	**.61**
Brightness	**.67**	—	**.66**
Inharmonicity	**.65**	—	—
Subband 5 (400–800 Hz)	—	**.96**	—
Subband 4 (200–400 Hz)	—	**.94**	—
Subband 6 (800–1.6 kHz)	.20	**.89**	—
Subband 7 (1.6–3.2 kHz)	.32	**.80**	—
Subband 8 (3.2–6.4 kHz)	.52	**.67**	—
Subband 3 (100–200 Hz)	.41	**.66**	—
RMS power	.29	.20	**.81**
Skewness	−.35	—	**−.88**
Kurtosis	—	—	**−.87**
Subband 2 (50–100 Hz)	.42	.27	—
Irregularity	—	—	.21

Note. PC1 = Intensity; PC2 = Flux; PC3 = Loudness; RMS = root mean square. Loadings <.20 are omitted; loadings >.60 are displayed in bold. Descriptors with no loadings greater than .20 have been omitted.

of squares) and significance levels were estimated using the car package in R (Fox & Weisberg, 2010). The models treated valence and arousal ratings (as well as two-way interactions) as fixed effects, and participant variability as a random intercept (see Equation S2 in the online supplemental materials for additional details).

In the interval version of the PC1 analysis, we found a significant main fixed effect of valence, $b = 0.19$, $SE = 0.07$, $t(1149) = 2.68$, $p = .008$, but not of arousal, $b = 0.16$, $SE = 0.08$, $t(1149) = 1.95$, $p = .05$. More importantly, the interaction between valence and arousal was significant, $b = -0.03$, $SE = 0.02$, $t(1149) = -2.17$, $p = .03$. Moving to the PC2 and PC3 models, however, we found that affective meaning did not appear to have any effects on acoustic factor scores (all main fixed effects and interactions, $p > .05$). Thus, although word semantics appear to have exerted a systematic effect on relative strength of high-frequency components, inharmonicity, auditory roughness, and so on, it did not relate to spectral fluctuation and signal strength. However, the adjusted R^2 was low for all the models (PC1 $R^2 = .03$, PC2 $R^2 = .03$, PC3 $R^2 = .02$), indicating that, although statistically significant, the semantic variables failed to predict most of the variation in acoustic responses.

Alternatively, we investigated the effect of ordinally conceived semantic attributes on the resultant three acoustic factors, again using linear mixed-effects models, with the expectation that results would by and large conform to the interval analysis. (Because we compare mean differences in acoustic factor scores between semantic categories in this analysis, we report here only analysis of variance output and Tukey-corrected post hoc comparisons.) For PC1, a significant main fixed effect of valence was found, $F(2, 1208) = 8.8$, $p = .0002$, but the main effect of arousal was nonsignificant, $F(2, 1208) = 2.7$, $p = .07$. Moreover, valence and arousal demonstrated a borderline significant interaction, $F(4, 1208) = 2.4$, $p = .048$, as plotted in Figure 2a. As is clear given the main fixed effects, the interaction is driven primarily by valence: Positive and negative (i.e., low and high valence) words are associated with elevated PC1 scores, and do not differ much from one another (positive M score = 0.19, 95% confidence interval [CI; 0.08, 0.30] vs. negative $M = 0.08$, 95% CI [−0.03, 0.19]), mean Tukey-corrected difference = −0.11, 95% CI [−0.25, 0.04], $p = .3$. However, the PC1 scores for positive and negative words were significantly higher than the scores for neutral words (medium-valence M score = −0.30; 95% CI [−0.41, −0.20]). Post hoc Tukey comparison of the mean PC1 difference between positive vs. neutral valence gives $M = 0.50$, 95% CI [0.36, 0.64], $p < .001$; similarly, the comparison between negative vs. neutral gives $M = 0.39$, 95% CI [0.25, 0.53], $p < .001$.

However, especially among medium and high-arousal words, the neutral valence, somewhat affectively ambiguous words (e.g., "dark," "full," and "mysterious") suppressed PC1 scores relative to the bipolar valence categories (neutral valence/medium arousal $M = -0.40$, 95% CI [−0.57, −0.22], neutral valence/high arousal $M = -0.39$, 95% CI [−0.56, −0.22]). Post hoc Tukey pairwise comparisons indicated a significant difference in the PC1 scores of neutral/medium and neutral/high words compared with all other categories except neutral/low, $p < .05$. In other words, acoustical features associated with the "Intensity" factor (PC1) were sculpted by participants to be more prominent when adjectives were either clearly positive *or* clearly negative, but were lower when adjectives were somewhere in between these affective poles.

For the PC2 model, no significant main fixed effects or interactions were found (at $p < .05$). Confirming the continuous model, then, this indicates that spectral fluctuations were not affected by the semantic implications of the target adjectives.

Finally, PC3 exhibited significant main fixed effects of both valence, $F(2, 1208) = 13.32$, $p < .0001$, and arousal, $F(2, 1208) =$

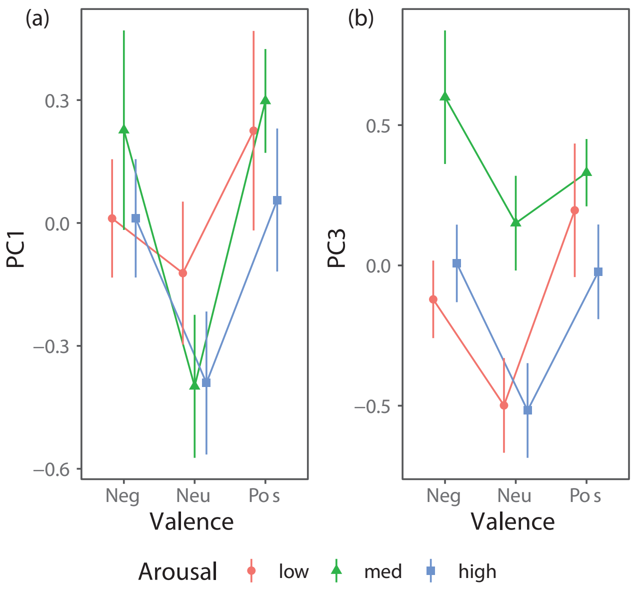


Figure 2. Interactions between valence and arousal in (a) PC1 and (b) PC3. Error bars: 95% confidence interval. PC1 = Intensity; PC3 = Loudness. See the online article for the color version of this figure.

5.9, $p = .003$. The interaction between valence and arousal was likewise significant, $F(4, 1208) = 3.39$, $p = .009$, and tells a story similar to the PC1 model (Figure 2b). Positive and negative words were not much different from one another, and together prompted significantly higher PC3 "Loudness" scores relative to affectively neutral words (respectively, positive $M = 0.17$, 95% CI [0.06, 0.27] and negative $M = 0.16$; 95% CI [0.06, 0.27] vs. neutral $M = -0.29$, 95% CI [−0.38, −0.19]). Post hoc Tukey pairwise comparison between positive vs. neutral valence showed an average difference of $M = 0.46$, 95% CI [0.32, 0.60], $p < .001$, and the difference between negative vs. neutral valence was $M = 0.45$, 95% CI [0.31, 0.59], $p < .001$; by way of contrast, positive and negative valence were basically the same, $M = 0$, 95% CI [−.0.13, 0.13], $p = .99$. Somewhat inexplicably, however, the medium arousal words ($M = 0.36$, 95% CI [0.25, 0.47]) showed significantly higher PC3 scores compared with high/low arousal words (high $M = -0.18$, 95% CI [−0.27, −0.08], low $M = -0.14$, 95% CI [−0.25, −0.03]). Post hoc tests of the difference between high versus medium arousal gave $M = -0.53$, 95% CI [−0.67, −0.40], $p < .001$, and low versus medium arousal showed $M = -0.50$, 95% CI [−0.65, −0.35], $p < .001$, but no difference between positive and negative valence was found, $M = 0.04$, 95% CI [−.0.10, 0.17], $p = .86$. Given the well-established link between loudness and perceived arousal (Dean, Bailes, & Schubert, 2011; Leman, Vermeulen, De Voogdt, Moelants, & Lesaffre, 2005), this interaction is somewhat difficult to interpret. Nonetheless, like in the PC1 model, the acoustic descriptors subsumed under PC3 appeared to be sensitive to the clearer poles of the valence dimension compared with words in the middle. In sum, semantic variables interacted to modulate the "Loudness" factor in addition to the more straightforwardly spectral attributes indexed by PC1.

Exploratory Cluster Analysis

In the linear mixed-effects models (LMM) analysis, we were interested in the association between affective features of the words and acoustics. We also wanted to know how distant each individual word is from one another given both the semantic and acoustic features; in other words, we were interested in the relationship among these words themselves. We next applied statistical clustering methods to partition the words in such a way that those in the same clusters are closer to each other than words in different clusters, according to a predefined distance criterion (see James, Witten, Hastie, & Tibshirani, 2013). We clustered the words based on a combination of all five semantic (valence and arousal) and acoustic variables (PC1, PC2, PC3). The LMM model (interval version) in the previous section suggested that only a small amount of variability in acoustic variables can be explained by the semantic variables, so clustering based on *all* semantic and acoustic variables is more informative than clustering based on semantic or acoustic variables separately.

To define a distance between any pair of words, we used the interval version of the semantic variables (the word "piercing" is not included in the analysis due to missingness of interval data for the semantic variables). The primary challenge in defining the distance between any two words was that the acoustic variables associated with each word were created by 64 participants; therefore, the variation within these participants also needed to be taken into account. Hence, we followed the approach recommended by Yeung, Medvedovic, and Bumgarner (2003), who developed and tested clustering methods for data with repeated measurements. More specifically, we used an average linkage hierarchical clustering algorithm combined with standard-deviation weighted Euclidean distance (see R code in online supplementary materials). First, we denote PC_{ijr} as the values of the *j*th factor scores on the word *i* measured by the *r*th subject, $i = 1 \ldots, 19$, $j = 1,2,3$ and $r = 1, \ldots 64$. Then we calculated the average factor score (Equation 1) and variance (Equation 2) for each word, as given below

$$D_{ij} = \frac{1}{64}\sum_{r=1}^{64} PC_{ijr} \tag{1}$$

$$\sigma_{ij}^2 = \frac{1}{64}\sum_{j=1}^{64} (PC_{ijr} - D_{ij})^2 \tag{2}$$

The *SD*-weighted Euclidean distance between word *i* and word *k* is given in Equation 3:

$$d_{ik} = \sqrt{\sum\nolimits_{j \in \{1,2,3,V,A\}} \frac{(D_{ij} - D_{kj})^2}{\sigma_{ij}^2 + \sigma_{kj}^2}}, i = 1, \ldots, 19; k = 1, \ldots, 19; i \neq k \tag{3}$$

Finally, in implementing the hierarchical clustering method, we needed a metric to compute the distance between two clusters. The distance between cluster C_1 and cluster C_2 is defined as the average of all the distances between every pair of words (*i,k*) with word *i* belonging to C_1 and word *k* belonging to C_2. In this approach, each word begins as its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters and continuing until there is just a single cluster. Based on this clustering procedure, the optimal number of clusters in the data set is five, which was chosen based on the elbow method (see scree plot, Figure S4 in the online supplemental materials). The corresponding dendrogram is presented in Figure 3.

The clustering analysis based on both semantic and acoustic variables revealed a number of insights into the relationship between semantic and acoustic data. The highest level bifurcation of branches is structured by valence, with more positive words on the left and more negative words on the right. Positive words form three clusters. At the farthest distance, *full* forms its own branch. The next cluster to the right consists of *noble*, *tender*, *sweet*, *bright*, and *charming*, positively valenced words of low or middling arousal. Finally, the third cluster consists of positive, high arousal words (*brilliant*, *rich*, *mysterious*, and *penetrating*). Interestingly, *brilliant* and *rich*, verbal attributes that Kendall and Carterette (1991) reported to be perceptually dissimilar, form their own subcluster here, indicating a close similarity between semantic and acoustic structure of these two words.

Moving to the more negative branch on the right side of the dendrogram, we see that the word *dull* forms its own cluster, separating from the other negative words at a fairly high level. The rest of the negative words form three subclusters. Somewhat surprisingly, *harsh* and *melancholic* cluster together, despite differing in arousal. *Gloomy*, *nasal*, and *veiled*—low arousal negative words—make up the next subcluster. Next, *rough*, *dark*, and *hard* clustered together; as these are fairly semantically divergent, this subcluster suggests commonalities in acoustic profiles for these three words.

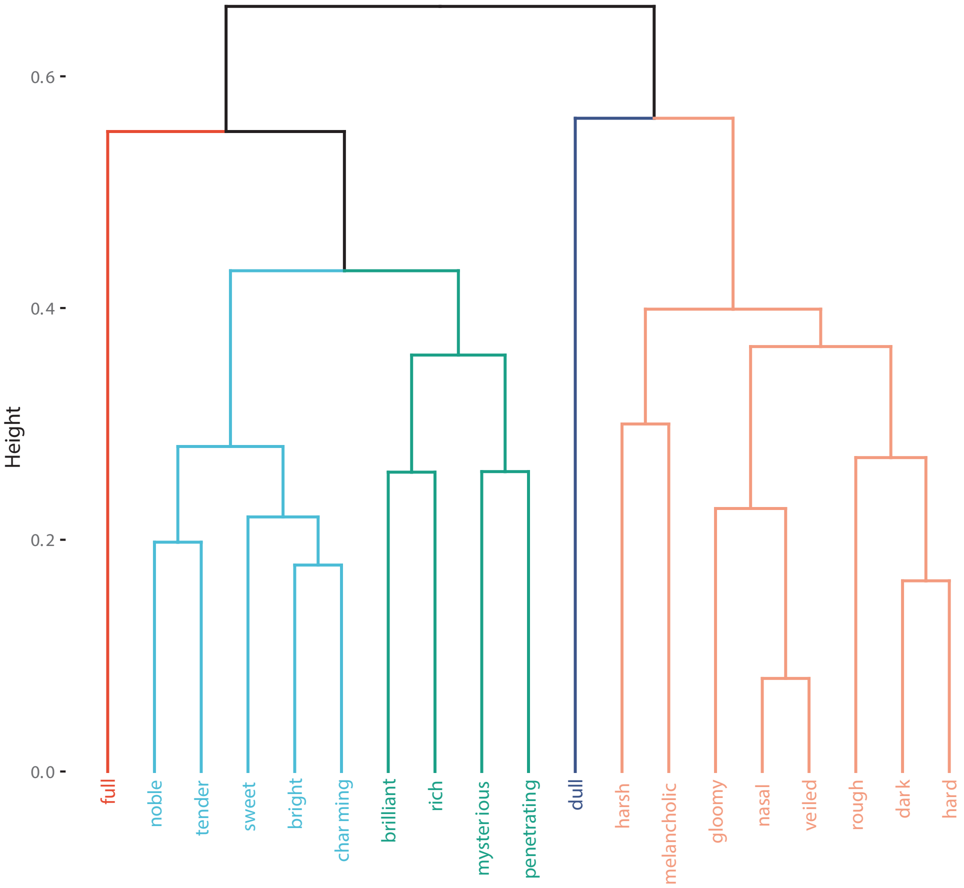


Figure 3. Dendrogram from the average linkage hierarchical clustering algorithm for the clustering based on two semantic variables (valence, arousal) and three acoustic variables (PC1, PC2, and PC3). See the online article for the color version of this figure.

Taken together, this analysis tells us that semantic and acoustic patterns largely converge in this data set: the three acoustic PCs are associated primarily with the valence dimension, with arousal playing a differentiating role at lower levels. This result resonates with the LMM analysis in affirming the centrality of word valence and arousal in affecting timbral features.

Discussion

The present study explored the influence of affective semantic dimensions on the creation of synthetic timbres. One novel component of this study was conceptual: We inverted the structure of most timbre semantic studies to investigate the effect of verbal dimensions on sounds, as opposed to vice versa. Another novel aspect was methodological, including the development of an FM synthesis interface for testing semantic–timbral associations (Ashley, 1986; Ethington & Punch, 1994; Miranda, 2002). Our aim was to probe whether varying affective dimensions of target timbre adjectives (valence and arousal) would exert a systematic effect on the acoustic outputs produced by musically trained participants instructed to create tones that best fit each adjective.

Results from acoustic, linear mixed-effect models, and cluster analyses converge on a few main findings. First, PCA of computationally extracted acoustic attributes from the 1,280 signals revealed three orthogonal acoustic PCs that together accounted for 72% of total variance. Increases in PC1 (“Intensity”), which accounted for 30% of this total, are associated with increasing high-frequency energy, spectral rolloff, spectral fluctuations in the highest bands of the spectrum (>6.4 kHz), entropy, inharmonicity, and auditory roughness. Using LMM to predict PC1 factor scores from affective meaning data for each adjective (in interval and ordinal versions), we found a statistically significant interaction between valence and arousal. PC1 was higher for both positive *and* negative words (e.g., *brilliant, harsh*) compared with words that fell in the middle of the valence scale (*dark*, *mysterious*). This effect was most pronounced in the medium and high-arousal conditions. Put differently, PC1 exhibited something of a bimodal distribution along the valence dimension: Positive and negative words provoked higher acoustic intensity, but the neutral words, which are by definition more affectively ambiguous and contextually determined, led to more “neutral” sounds (i.e., tones with less perceptually salient spectral components). This would seem to run counter to the implied bipolarity of this dimension, in which we could expect to see PC1 scores either increasing or decreasing linearly from negative to positive valence. Rather, LMM results may suggest a more binary affective association between word valence and PC1, or perhaps the interaction of two orthogonal valence dimensions, positive affect and negative affect (Watson, Clark, & Tellegen, 1988).

This result is consonant with much of the literature in affirming the affective salience of high-frequency energy and spectral noisiness. However, it sheds new light on how these acoustic components may relate to the valence dimension of timbre semantics. For example, Eerola et al. (2012) found that the ratio of high to low-frequency energy was the most significant acoustic predictor of both negative valence and high energy arousal. A similar result was obtained by Wallmark, Iacoboni, Deblieck, and Kendall (2018), who found that spectral centroid and an inharmonic spectral distribution are associated with high arousal and negative valence in both isolated instrument tones and brief samples of popular music. Notably, however, other studies have shown an association between spectral centroid and *positive*, high arousal affect (McAdams et al., 2017). Our findings would appear to harmonize these seemingly contradictory results in showing that degree of affective *polarity* is most clearly associated with this important basket of acoustic descriptors, not necessarily one side of the valence coin or the other.

There are at least two plausible interpretations for this result. Words rated as close to neutral in the Warriner et al. (2013) data set may have provoked a greater degree of variability in PC1 between participants, thus leading to less consistent and more middle-of-the-road mean factor scores than words at the twin poles of the valence scale. However, if this were true, we would expect to see wider 95% CIs associated with neutral words compared with the other categories; as shown in Figure 2, this is not the case. Alternatively, neutral words may have consistently struck participants as more affectively lukewarm than the poles, and responded by suppressing the more intense spectral options available from the interface.

Acoustic attributes captured in PC1, moreover, have long been linked in perceptual studies to some of the specific terms included in the adjective set. In particular, the cross-modal adjectives *bright* and *brilliant* are commonly associated with timbres with high spectral centroid, whereas timbres with a lower spectral center are described using opposing adjectives (e.g., *dark*; Alluri & Toiviainen, 2012; Beauchamp, 1982; Schubert & Wolfe, 2006; Wallmark, 2019a; Wessel, 1979; Zacharakis et al., 2014). Our result would seem to basically confirm this association from the other direction: Participants crafted sounds with more strength in higher frequencies in response to both positive and negative high-arousal words such as *brilliant* and *rough*. This semantic link has also been established in a number of other adjective-controlled synthesis systems (Ethington & Punch, 1994).

The affective connotations of the spectral parameters captured in PC1 have been interpreted through the framework of embodied music cognition (Leman, 2007). In this account, high-arousal, typically negatively valenced bodily states are related to physiological changes in vocal production that accentuate high-frequency components and noisiness (Johnstone & Scherer, 1999; Juslin & Laukka, 2003; Scherer & Oshinsky, 1977). Although these "push effects" (Scherer, Johnstone, & Klasmeyer, 2003) originate in vocal expression, in some cases they may also provide important state cues (Huron, 2001) informing our perceptual response to instrumental timbre (Tsai et al., 2010). For example, Juslin and Laukka (2003) suggested that instrumental sound (theoretically including synthesized sound) can function as a "superexpressive voice" by mimicking certain crude acoustic features of affective vocal expression. Although the timbral properties of our FM patch were not actually modulated by any physical push effects—that is, participants did not need to exert greater physical exertion to achieve qualities of timbre often associated with arousal—it is possible that such learned bioacoustic correlations informed participants' interaction with the interface, guiding them toward regions of the nonlinear space that seemed affectively congruent with the target adjectives. Crucially, moreover, it is likely that these general associations between spectral disposition and affect dimensions have come to influence the conventional lexicon for musical timbre, which is reflected (in the context of Western art music, at least) in the discourse common to instrumental timbre as promulgated in orchestration texts (Wallmark, 2019b). Our findings suggest that, at least to a degree, the perceptual association between the acoustic attributes of PC1 and timbre semantics that has been documented in many previous studies may generalize among musicians to the creation of novel timbres using an unfamiliar synthesis interface.

Moving to the other two PCs: The association between affective structure and PC2, which corresponded mainly to spectral fluctuations between 100 and 6400 Hz, was not related to the affect dimensions in either of our parallel mixed-effects models. However, PC3, the "Loudness" factor that is positively associated with RMS energy and spectral brightness (and negatively with spectral skewness and kurtosis) exhibited a significant interaction between valence and arousal (in the ordinal analysis only). This result accords with the patterns in the PC1 models: Particularly in the high-arousal condition, words occupying the clear poles of the valence scale prompted higher PC3 values than the more neutral words. The role of arousal in this interaction, though, is difficult to interpret. RMS energy, which has the strongest loadings on this factor, is typically correlated with perceived loudness, and loudness is associated with high arousal (Dean et al., 2011; Leman et al., 2005); it is difficult to explain the fact, therefore, that the medium arousal condition had the highest PC3 scores across the three valence levels. To be clear, RMS energy is not a timbral property; with strongest loadings on this variable, PC3 is more closely associated with amplitude differences than spectral attributes (hence the shorthand, "Loudness" factor), and is therefore governed by a different set of psychoacoustic principles (for review, see Moore, 2014). Unexpected arousal results from PC3 models may indicate that loudness plays a complex and contextually variant role linking semantics to sounds in this task.

In addition to modeling the effects of the semantic dimensions on the separate acoustic factor scores, we performed a cluster analysis to explore the perceptual distances between individual words in groupings based on a combination of both affective and acoustic data. To do so, we implemented a novel hierarchical clustering algorithm that uses *SD*-weighted Euclidean distances to account for repeated measurements (Yeung et al., 2003). Our analysis revealed two main patterns (see Figure 3). First, when clustering our 20 target adjectives on the two semantic variables and three acoustic variables, we found a dichotomous split by word valence. This suggests that valence drives meaning structure of this small set of representative timbre adjectives more than arousal, which generally agrees with previous studies. Indeed, studies in affective semantics since the 1950s (Mehrabian & Russell, 1974; Osgood et al., 1957; Russell & Mehrabian, 1977) have shown that valence is typically the most salient of the major affect dimensions, so this result is certainly not without precedent. However, incorporating timbral components into this interpretation, this result tells us that positively and negatively valenced words also tended to generate acoustical profiles that approximated this same basic distinction. Interestingly, then, in contrast to the LMM results showing a significant difference between the poles and the more valence-neutral words, the clustering analysis produced a basically dichotomous solution.

Next, arousal and the three acoustic PCs added nuance and affective granularity to this binary portrait by differentiating words into five total clusters (three positive, two negative) as well as a number of subclusters. For instance, *dull*, though still ultimately clustering with the other negative words, bifurcates from the others at a high level; this suggests that both semantically and acoustically, *dull* occupies a somewhat singular position among these word stimuli. Similarly, *full* forms a cluster of its own among the positive words. Together, this indicates that these two adjectives reflect distinctive affective and acoustic profiles. (By way of contrast, *sweet–bright–charming* are closely related, as are *noble–tender*.) Closer distances at lower levels of the positively valenced left side of the dendrogram reflect similarities in arousal and acoustic factor scores: For example, higher arousal positive words (*brilliant, rich, penetrating*) are grouped into a separate cluster from lower arousal positive words (*noble, tender, sweet*). By way of contrast, with the exception of *dull*, negative words formed just one large cluster (with eight members). The clustering procedure also revealed a couple of subgroupings that are challenging to interpret: *mysterious–penetrating* formed its own subcluster, for example, and *harsh–melancholic* were found to be closely connected. Notwithstanding a couple of odd connections, however, the coherence of this clustering analysis indicates that word valence and arousal relate to acoustics in a way that is largely complementary and mutually reinforcing.

A number of limitations to the present study must be stated in conclusion. We specifically focused this study on the timbral creations of classically trained musicians. Our rationale was threefold: First, we wanted to investigate the interrelation of semantics and timbre among a population with relatively consistent exposure to the same representative sample of timbre adjectives. Second, musicians are regularly accustomed to adjusting their sound in performance according to semantic cues, whether through rehearsal, instruction, discourse with fellow musicians, or solitary practice. Third, musical training has been shown to affect behavioral responses to timbre (Chartrand & Belin, 2006; McAdams et al., 2017; Siedenburg & McAdams, 2018). For these reasons, the experimental task was relatively intuitive and natural for musicians. Yet this design decision leaves open the possibility that the present results are exclusive to trained musicians and would fail to generalize to a nonmusician population (though see Filipic, Tillmann, and Bigand [2010] for evidence that musical training has a marginal effect on affective responses). In future research, it will be necessary to expand the participant population to address whether such consistencies in novel sound generation apply beyond trained musicians.

Additionally, we believe it would be advantageous to use a larger set of adjective stimuli in future paradigms to decrease the variance associated with each individual word. It could also be interesting to examine adjectives from more recent treatises: The words used here were selected from Wallmark (2019b) prior to the completion of that study (which ultimately included 11 orchestration texts), at which time only Berlioz and Rimsky-Korsakov had been fully analyzed. Also, it is possible that some adjectives were more easily represented in our continuous playback interface than others. It would further be interesting to derive adjectives from sources outside of the orchestral tradition. Although our goal with the present design was to deliberately limit ecological validity to evaluate whether familiar timbre terms translated into an unfamiliar medium, this decision—in concert with a nonlinear interface—arguably led to ambiguity among some participants (although no participants reported such difficulty or confusion after completing the task).

Finally, in future studies, it may be profitable to imagine other kinds of intuitive multidimensional synthesis interfaces for such a task, and to expand beyond FM synthesis, particularly to additive models (Ethington & Punch, 1994; Gounaropoulos & Johnson, 2006). Linking affective semantic models more explicitly to theories of musical embodiment, moreover, it could be interesting to create a vocal version of this basic task; for example, asking participants to sing a uniform pitch in a manner that best fits target adjectives (Parise & Pavani, 2011).

Conclusion

The convergent results presented in this study suggest that musically trained participants learned to "play" an unfamiliar synthetic interface by locating regions corresponding to the affective implications of target adjectives. Specifically, valence and arousal systematically interacted to influence PC1 ("Intensity") and PC3 ("Loudness"): Words at the poles of the valence scale (i.e., clearly positive or clearly negative) led to higher average acoustic factor scores (and were not significantly different from one another), whereas words in the middle of that scale led to lower scores. Taken together, valence appeared to be the most important affective dimension in predicting acoustical response, as indicated by both linear mixed-effects modeling and a clustering analysis. The clustering analysis produced a basically dichotomous valence structure, with certain individual words (e.g., *dull*, *full*) standing out from the others in the semantic–acoustic space. In sum, musician participants were fairly consistent in mapping the affective dimensions of words onto the acoustic outputs they create. This finding adds to the contemporary discourse of timbre semantics by demonstrating that affective verbal prompts can systematically influence sound production, just as listening to timbre has been shown to elicit systematic verbal judgments (for reviews, see Saitis & Weinzierl, 2019; Wallmark & Kendall, in press). Beyond semantics, moreover, this study is in line with other recent research in demonstrating the affective significance of timbre in the generation of musical meaning (Fink, Latour, & Wallmark, 2018; Noble & McAdams, 2018).

References

Aljanaki, A., Yang, Y.-H., & Soleymani, M. (2017). Developing a benchmark for emotional analysis of music. *PLoS ONE, 12,* e0173392. http://dx.doi.org/10.1371/journal.pone.0173392

Alluri, V., & Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception, 27,* 223–242. http://dx.doi.org/10.1525/mp.2010.27.3.223

Alluri, V., & Toiviainen, P. (2012). Effect of enculturation on the semantic and acoustic correlates of polyphonic timbre. *Music Perception, 29,* 297–310. http://dx.doi.org/10.1525/mp.2012.29.3.297

Ashley, R. (1986). A knowledge-based approach to assistance in timbral design. In *Proceedings of the 1986 International Computer Music Conference* (pp. 11–16). Den Haag, the Netherlands: Royal Conservatory.

Bakker, I., Van der Voordt, T., Vink, P., & De Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Current Psychology, 33,* 405–421. http://dx.doi.org/10.1007/s12144-014-9219-4

Beauchamp, J. (1982). Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society, 30,* 396–406.

Berlioz, H. (1882). *A treatise on modern instrumentation and orchestration* (M. C. Clarke, Trans.). London, United Kingdom: Novello, Ewer and Co.

Boulez, P. (1987). Timbre and composition—Timbre and language. *Contemporary Music Review, 2,* 161–171. http://dx.doi.org/10.1080/07494468708567057

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings* (Technical Report No. C-1). Gainesville, FL: University of Florida, NIMH Center for Research in Psychophysiology.

Carron, M., Rotureau, T., Dubois, F., Misdariis, N., & Susini, P. (2017). Speaking about sounds: A tool for communication on sound features. *Journal of Desert Research, 15,* 85–109. http://dx.doi.org/10.1504/JDR.2017.086749

Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters, 405,* 164–167. http://dx.doi.org/10.1016/j.neulet.2006.06.053

Chowning, J. M. (1977). The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society, 21,* 526–534. Retrieved from https://www.jstor.org/stable/23320142

Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: An experimental investigation. *PLoS ONE, 6,* e18591. http://dx.doi.org/10.1371/journal.pone.0018591

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,* 49–70. http://dx.doi.org/10.1525/mp.2012.30.1.49

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39,* 18–49. http://dx.doi.org/10.1177/0305735610362821

Ethington, R., & Punch, B. (1994). SeaWave: A system for musical timbre description. *Computer Music Journal, 18,* 30–39. http://dx.doi.org/10.2307/3680520

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin and Review, 17,* 335–341. http://dx.doi.org/10.3758/PBR.17.3.335

Fink, R., Latour, M., & Wallmark, Z. (Eds.). (2018). *The relentless pursuit of tone: Timbre in popular music*. New York, NY: Oxford University Press.

Fox, J., & Weisberg, H. S. (2010). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Giannakis, K., & Smith, M. (2000). Auditory-visual associations for music compositional processes: A survey. *Proceedings of the International Computer Music Conference,* 12–15. Retrieved form http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1.8905&rep=rep1&type=pdf

Gounaropoulos, A., & Johnson, C. (2006). Synthesising timbres and timbre-changes from adjectives/adverbs. In F. Rothlauf et al. (Eds.), *Applications of evolutionary computing* (pp. 664–675). Berlin, Germany: Springer. http://dx.doi.org/10.1007/11732242_63

Huron, D. (2001). *Toward a theory of timbre*. Paper presented at the 12th Annual Conference of Music Theory, Midwest, Cincinnati, OH.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-7138-7

Johnstone, T., & Scherer, K. R. (1999). The effects of emotions on voice quality. In J. J. Ohala et al. (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 2029–2032). San Francisco, CA: ICPhS.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814. http://dx.doi.org/10.1037/0033-2909.129.5.770

Kendall, R. A., & Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception, 8,* 369–404. http://dx.doi.org/10.2307/40285519

Kendall, R. A., & Carterette, E. C. (1993a). Verbal attributes of simultaneous wind instrument timbres: I. Von Bismarck's adjectives. *Music Perception, 10,* 445–467. http://dx.doi.org/10.2307/40285583

Kendall, R. A., & Carterette, E. C. (1993b). Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's orchestration. *Music Perception, 10,* 469–501. http://dx.doi.org/10.2307/40285584

Kreković, G., Pošćić, A., & Petrinović, D. (2016). An algorithm for controlling arbitrary sound synthesizers using adjectives. *Journal of New Music Research, 45,* 375–390. http://dx.doi.org/10.1080/09298215.2016.1204325

Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. Proceedings of the 10th International Conference on Digital Audio Effects, 237–244.

Leman, M. (2007). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/7476.001.0001

Leman, M., Vermeulen, V., De Voogdt, L. D., Moelants, D., & Lesaffre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research, 34,* 39–67. http://dx.doi.org/10.1080/09298210500123978

Lichte, W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology, 28,* 455–480. http://dx.doi.org/10.1037/h0053526

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge: New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139084789

McAdams, S., Douglas, C., & Vempala, N. N. (2017). Perception and modeling of affective qualities of musical instrument sounds across pitch registers. *Frontiers in Psychology*. Advance online publication. http://dx.doi.org/10.3389/fpsyg.2017.00153

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge, MA: MIT Press.

Miranda, E. R. (2002). *Computer sound design: Synthesis techniques and programming* (2nd ed.). Oxford, United Kingdom: Focal Press.

Moore, B. C. J. (2014). Development and current status of the "Cambridge" loudness models. *Trends in Hearing*. Advance online publication. http://dx.doi.org/10.1177/2331216514550620

Noble, J., & McAdams, S. (2018). Meaning beyond content: Extra musical associations are plural but not arbitrary. In R. Parncutt & S. Sattmann (Eds.), *Proceeding ICMPC15/ESCOM10* (pp. 389–394). Graz, Austria: Center for Systematic Musicology, University of Graz.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Parise, C. V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research, 214,* 373–380. http://dx.doi.org/10.1007/s00221-011-2836-3

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68,* 111–141. http://dx.doi.org/10.1016/S0010-0277(98)00043-2

Pinheiro, J., Bates, D. M., DebRoy, S., Sarkar, D., Heisterkamp, S., & Van Willigen, B. (2018). *nlme: Linear and nonlinear mixed effects models (Version 3.1–137)*. Retrieved from https://CRAN.R-project.org/package=nlme

Pratt, R. L., & Doak, P. E. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration, 45,* 317–328. http://dx.doi.org/10.1016/0022-460X(76)90391-6

Rimsky-Korsakov, N. (1933). *Principles of orchestration* (E. Agate, Trans.). New York, NY: Edwin F. Kalmus.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. http://dx.doi.org/10.1037/h0077714

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11,* 273–294. http://dx.doi.org/10.1016/0092-6566(77)90037-X

Saitis, C., & Weinzierl, S. (2019). The semantics of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. Popper, & R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 119–149). New York, NY: Springer.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). New York, NY: Oxford University Press.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1,* 331–346. http://dx.doi.org/10.1007/BF00992539

Schubert, E. (2007). The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music, 35,* 499–515. http://dx.doi.org/10.1177/0305735607072657

Schubert, E., & Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica, 92,* 820–825.

Seago, A., Holland, S., & Mulholland, P. (2004). A critical analysis of synthesizer user interfaces for timbre. In A. Dearden & L. Watt (Eds.), *Proceedings of the 18th British HCI Group Annual Conference* (Volume 2, pp. 105–108). Bristol, United Kingdom: Research Press International.

Siedenburg, K., & McAdams, S. (2018). Short-term recognition of timbre sequences: Music training, pitch variability, and timbral similarity. *Music Perception, 36,* 24–39. http://dx.doi.org/10.1525/mp.2018.36.1.24

Soraghan, S., Faire, F., Renaud, A., & Supper, B. (2018). A new timbre visualization technique based on semantic descriptors. *Computer Music Journal, 42,* 23–36. http://dx.doi.org/10.1162/comj_a_00449

Susini, P., Lemaitre, G., & McAdams, S. (2012). Psychological measurement for sound description and evaluation. In B. Berlund, G. B. Rossi, J. T. Townsend, & L. R. Pendrill (Eds.), *Measurement with persons: Theory, methods, and implementation areas* (pp. 227–253). New York, NY: Psychology Press.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37,* 267–307. http://dx.doi.org/10.1162/COLI_a_00049

Tervaniemi, M., Winkler, I., & Näätänen, R. (1997). Pre-attentive categorization of sounds by timbre as revealed by event-related potentials. *Neuroreport, 8,* 2571–2574. http://dx.doi.org/10.1097/00001756-199707280-00030

Tsai, C.-G., Wang, L.-C., Wang, S.-F., Shau, Y.-W., Hsiao, T.-Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception, 27,* 209–222. http://dx.doi.org/10.1525/mp.2010.27.3.209

von Bismarck, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica, 30,* 147–172.

Wallmark, Z. (2019a). Semantic crosstalk in timbre perception. *Music and Science, 2,* 1–18. http://dx.doi.org/10.1177/2059204319846617

Wallmark, Z. (2019b). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music, 47,* 585–605. http://dx.doi.org/10.1177/0305735618768102

Wallmark, Z., Iacoboni, M., Deblieck, C., & Kendall, R. A. (2018). Embodied listening and timbre: Perceptual, acoustical, and neural correlates. *Music Perception, 35,* 332–363. http://dx.doi.org/10.1525/mp.2018.35.3.332

Wallmark, Z., & Kendall, R. A. (in press). Describing sound: The cognitive linguistics of timbre. In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre*. Advance online publication. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780190637224.013.14

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45,* 1191–1207. http://dx.doi.org/10.3758/s13428-012-0314-x

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3,* 45–52. http://dx.doi.org/10.2307/3680283

West, B. T., Welch, K. B., & Galecki, A. T. (2006). *Linear mixed models: A practical guide using statistical software* (2nd ed.). Boca Raton, FL: CRC Press. http://dx.doi.org/10.1201/9781420010435

Yeung, K. Y., Medvedovic, M., & Bumgarner, R. E. (2003). Clustering gene-expression data with repeated measurements. *Genome Biology, 4,* R34. http://dx.doi.org/10.1186/gb-2003-4-5-r34

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception, 31,* 339–358. http://dx.doi.org/10.1525/mp.2014.31.4.339

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2015). An interlanguage unification of musical timbre: Bridging semantic, perceptual, and acoustic dimensions. *Music Perception, 32,* 394–412. http://dx.doi.org/10.1525/mp.2015.32.4.394

Zentner, M. R., & Eerola, T. (2010). Self-report measures and models. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 187–221). Oxford, United Kingdom: Oxford University Press.

Received October 11, 2018
Revision received March 20, 2019
Accepted May 31, 2019 ■