Author(s): Eerola, Tuomas; Ferrer Flores, Rafael; Alluri, Vinoo

Title: Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds

Year: 2012

Version:

Please cite the original version:

Eerola, T., Ferrer Flores, 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

Timbre and Affect Dimensions: Evidence from Affect and Similarity Ratings and Acoustic Correlates of Isolated Instrument Sounds

Tuomas Eerola, Rafael Ferrer,
& Vinoo Alluri
University of Jyväskylä, Jyväskylä, Finland

Considerable effort has been made towards understanding how acoustic and structural features contribute to emotional expression in music, but relatively little attention has been paid to the role of timbre in this process. Our aim was to investigate the role of timbre in the perception of affect dimensions in isolated musical sounds, by way of three behavioral experiments. In Experiment 1, participants evaluated perceived affects of 110 instrument sounds that were equal in duration, pitch, and dynamics using a three-dimensional affect model (valence, energy arousal, and tension arousal) and preference and emotional intensity. In Experiment 2, an emotional dissimilarity task was applied to a subset of the instrument sounds used in Experiment 1 to better reveal the underlying affect structure. In Experiment 3, the perceived affect dimensions as well as preference and intensity of a new set of 105 instrument sounds were rated by participants. These sounds were also uniform in pitch, duration, and playback dynamics but contained systematic manipulations in the dynamics of sound production, articulation, and ratio of high-frequency to low-frequency energy. The affect dimensions for all the experiments were then explained in terms of the three kinds of acoustic features extracted: spectral (e.g., ratio of high-frequency to low-frequency energy), temporal (e.g., attack slope), and spectro-temporal (e.g., spectral flux). High agreement among the participants' ratings across the experiments suggested that even isolated instrument sounds contain cues that indicate affective expression, and these are recognized as such by the listeners. A dominant portion (50-57%) of the two dimensions of affect (valence and energy arousal) could be predicted by linear combinations of few acoustic features such as ratio of high-frequency to low-frequency energy, attack slope, and spectral regularity. Links between these features and those observed in the vocal expression of affects and other sound phenomena are discussed.

Received June 9, 2010, accepted November 17, 2011.

Key words: timbre, affect, emotion, instrument, monophonic

Numerous studies have investigated how individual features of music contribute to its emotional expression. These typically range from psychoacoustic features such as loudness and roughness (e.g., Leman, Vermeulen, De Voogdt, Moelants, & Lesaffre, 2005) to structural features such as mode and harmony (e.g., Gabrielsson & Lindström, 2010) and performance features such as tempo and articulation (e.g., Baraldi, De Poli, & Roda, 2006; Juslin, 2000). Composers and arrangers take great care in selecting different instruments to bring out desired characteristics and emotional colors in the musical structure (Schutz, Huron, Keeton, & Loewer, 2008). In light of this, it seems surprising that little attention has been paid to timbre in communicating emotions in music despite its role as a "major structuring force in music and one of the most important and ecologically relevant features of auditory events" (Menon et al., 2002, p. 1742). Only recently few studies have bordered the issue in the context of recognition of emotions from brief excerpts of music (Filipic, Tillmann, & Bigand, 2010; Krumhansl, 2010), although the actual qualities of the timbre with respect to emotional expression has not yet been considered. Although timbre is resistant to unambiguous definition, being often defined in comparison to what it is not (pitch, rhythm, harmony, structure), there is nevertheless a great deal known about the psychological representation of timbre (Hajda, Kendall, Carterette, & Harschberger, 1997; Caclin, McAdams, Smith, & Giard, 2008). While the exact formulation of timbre dimensions are still a matter of debate, consensus exists on which psychoacoustic aspects of sounds these three dimensions represent, namely the temporal (e.g., attack time), spectral (e.g., spectral energy distribution), and spectro-temporal (e.g., spectral flux). These dimensions

Music Perception volume 30, issue 1, pp. 49–70. issn 0730-7829, electronic issn 1533-8312. © 2012 by the regents of the university of california all rights reserved. please direct all requests for permission to photocopy or reproduce article content through the university of california press's rights and permissions website, http://www.ucpressjournals.com/reprintinfo.asp.DOI: 10.1525/mp.2012.30.1.49

have been commonly found in behavioral studies employing similarity ratings (Grey, 1977; Lakatos, 2000; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995), and have been further explored and (for the most part) corroborated in a series of meta-analytic overviews (Burgoyne & McAdams, 2008; Caclin, McAdams, Smith, & Winsberg, 2005). Several studies have also recently traced the neural underpinnings of these timbre dimensions (Caclin et al., 2008, 2006; Menon et al., 2002).

Much research on music and emotions investigates the links between perceived emotions and music (e.g., Juslin & Sloboda, 2010), but the precise definition of those emotions has been notoriously difficult to pin down, even if there is agreement over their general characteristics and subcomponents (Sloboda & Juslin, 2010; Zentner & Eerola, 2010). Here we prefer to use the term *affect*, since, according to Juslin and Sloboda, it is an ”umbrella term that covers all evaluative–or ’valenced’ (positive/negative)–states” (2010, p. 10) and the emotional expressions represented by short instrument sounds falls more appropriately within the scope of this term. Affects evoke emotional responses, but unlike basic emotions, they are not always emotional indicators themselves.

Connections between certain aspects of physical properties of timbre and emotional expression have already been found in research on expressive speech (e.g., spectral energy distribution and formant structure, Juslin & Laukka, 2003; Laukka, Juslin, & Bresin, 2005; Scherer & Oshinsky, 1977) and have recently been explored with nonverbal affect vocalizations (Belin, Fillion-Bilodeau, & Gosselin, 2008; Bradley, 2000; Redondo, Fraga, Padron, & Pineiro, 2008). The acoustical attributes of environmental sounds have also been found to affect the pleasantness and appeal of such sounds (Kidd & Watson, 2003; Maffiolo, Castellengo, & Dubois, 1999; Ohta, Kuwano, & Namba, 1999). However, whereas the identity of the sound source may not be quite as important for a musical sound as it is for an environmental one, its affective expression (whether the performance is aggressive, tender, etc.) is likely to have greater significance (Scherer, 1995; Juslin & Laukka, 2003). Timbre holds the key in both these respects.

Timbre has a history of relevance to the field of emotions in music. Scherer and Oshinsky (1977) conducted a pioneering factorial rating study with synthetic tone sequences, where timbre was one of the factors that was systematically varied through simple spectral filtering and envelope manipulation. Later, Peretz and colleagues (1998) demonstrated that emotion categories may be reliably discriminated from 250 ms segments of music, which highlights the importance of timbre since many other musical elements (e.g., harmony, structure, melody) would require a longer time frame. Similar results, though with an entirely different model of emotions, were obtained by Bigand and colleagues (2005) using a similarity rating paradigm to study the emotional distances between music excerpts that were 1 s in length. The results using these short versions were nearly identical to the inter-item distances obtained with longer excerpts (approximately 25 s long). These results show a similarity to those obtained in perceptual genre discrimination tasks where extremely short excerpts (250 ms) are used, and where timbre is thus equally important (Gjerdingen & Perrott, 2008; Schellenberg, Iverson, & McKinnon, 1999). More recently, Krumhansl (2010) explored how well musical content can be recognized from very short clips of familiar popular music (400 ms). Interestingly, the emotional content of these polyphonic and familiar excerpts was rated by the listeners as well, but it was done in a relatively cursory fashion using a forced-choice paradigm of the basic emotions. Also, there were no acoustic correlates reported for the various kinds of emotional content, but this did not detract from the overall positive result, which further underlines the importance of timbre in music recognition. Another recent study by Filipic and colleagues (2010) investigated the degree of perceptual information needed to reliably attribute familiarity and emotionality towards brief excerpts of music. Intriguingly, they found that the distinction between neutral and emotionally moving music content occurred as early as 250 ms into the segment. However, this particular study treats the broader concept of emotionality rather than specific emotions, and while acoustic measures were used, no consistent patterns were found to explain the distinction between neutral and moving sound examples.

Using neurophysiological methods, Goydke, Altenmüller, Möller, and Münte (2004) demonstrated that violin tones representing happy or sad expression could be preattentively discriminated by listeners. And lastly, the link between timbre and affects has been observed in both infant and cross-cultural studies. The auditory system of infants is maximally sensitive to spectral slope differences (a central component in the vocal expression of affect), and this sensitivity is in the same range in both speech and music (Tsang & Trainor, 2002). Analysis of the musical features of emotions perceived in music across cultures has demonstrated the importance of timbre in addition to psychophysical dimensions such as tempo, pitch range, and rhythmic complexity (Balkwill & Thompson, 1999). Although these previous studies highlight the importance of timbre for emotional expression in both speech and music, efforts towards a more refined analysis of this connection in music have not been carried out.

Aim of the Study

The aim was to investigate the role of timbre in the perception of broad affect dimensions in music. To focus on timbre, we chose monophonic instrument sounds that would mainly vary in terms of the envelope and spectrum. This would reduce the confounding effect of other features of the music that are known to influence the perception of emotions, such as mode, tempo, register, harmony, and loudness. Whereas the monophonic sounds are seldom experienced without a context in music, they nevertheless provide examples of instances — where musicians practice their instruments — that can offer detailed information about musical expressivity.

To explore the affect structure of timbre, three experiments were designed to address the topic step by step. In the first experiment, the participants rated the perceived affects of a selection of instrument sounds. The second experiment then explored the structure of affect in these ratings by subjecting a subset of them to an emotional similarity task. In the final experiment, the affect dimensions of instrument sounds were again rated but these 105 additional instrument sounds contained systematic manipulations of the acoustic features.We decided not to focus on the basic emotion categories (happiness, sadness, fear, anger, etc.), as they would perhaps not have reflected the more subtle variations of timbre. Instead, we adopted a dimensional approach, and more precisely, the three-dimensional model of affect advocated by Schimmack and Grob (2000), which attempts to capture the core affects using the three bipolar dimensions, Valence, Energy arousal, and Tension arousal. The reason this model was chosen was because, unlike the two-dimensional variants (Russell, 1980; Thayer, 1989), it has a strong physiological basis, and accounts for empirical data from previous studies (Schimmack & Grob, 2000; Schimmack & Reisenzein, 2002), including perceived emotions in music (Eerola & Vuoskoski, 2011; Ilie & Thompson, 2006). The three-dimensional model still offers the possibility to revert back to commonly used two-dimensional models and even to collapse the affects into the basic categories (see Eerola & Vuoskoski, 2011). We also considered the intensity of affects and the listeners' preference for sounds that capture those aspects that are thought to be relevant for affective evaluation (see Kreutz, Ott, Teichmann, Osawa, & Vaitl, 2008; Rawlings & Leow, 2008).

Experiment 1: Affect Ratings of Plain Instrument Sounds

An experiment was designed to obtain affect ratings of sound examples that had timbre as their main variant characteristic. Affect dimensions were taken from the three-dimensional model of affect (Schimmack & Grob, 2000).

EXPERIMENT DETAILS

Stimuli. The stimuli were chosen to be real instrument sounds instead of artificially created sounds because of their high ecological validity. One hundred and ten instrument samples from the McGill University Master Samples (MUMS) collection (Opolko & Wapnick, 2006) were selected (see also Eerola & Ferrer, 2008). These sounds included most of the common instruments (piano, guitar, flute, clarinet, horn, oboe, etc.) as well as more exotic instruments such as Shawm and Crumhorn. A diverse selection of the different instrument families (horns, strings, woodwinds, etc.) of the MUMS collection was taken. The samples were chosen to be identical in pitch (D♯4), not only because it lies in the vicinity of average pitch (Huron, 2001) but because it allowed for a maximal overlap between the registers of instruments available in MUMS. The durations were set for all instrument sounds to 1 s with a 23 ms fade-out at the end of the each sample. The loudness was equalized manually. A full list of instrument names, their articulations, and the actual sounds are given in Appendix A and distributed online.[1]

Participants. The participants consisted of 17 females and 13 males (age *M* = 25.37, *SD* = 4.05). Thirteen percent reported having no formal music education, while 26.67% had received formal training. The rest had a mean of 11.55 years of formal music training (theory and/or practice). All of them reported having had music as a hobby for an extensive period. Music involvement was calculated according to the estimated number of hours per week spent listening to music (*M* = 11.2), together with the number of years spent playing music (*M* = 10.6).

Procedure. Participants were asked to rate the perceived affect qualities of individually presented sounds using five concepts, each represented by a bipolar, 9-point Likert scale. The words used to depict the extremes of each concept represented by a scale were: pleasant and unpleasant for the Valence scale (miellyttävä/epämiellyttävä in Finnish), awake and tired for Energy arousal (virkeä/väsynyt in Finnish), tense and relaxed for Tension arousal (jännittynyt/rentoutunut in Finnish), like and dislike for Preference, and high and low for the Intensity scale. The words were displayed in two languages (Finnish and English) and the words for three affect dimensions were taken from a previous study (Schimmack & Reisenzein, 2002). The order of sounds

[1] https://www.jyu.fi/music/coe/materials/emotion/timbre/

was random for each participant and the experiment was carried out using high-quality headphones in a quiet room.

Results

CONSISTENCY AND STRUCTURE OF AFFECTS

One participant was removed due to low inter-subject correlation (3 *SD*s below the mean inter-subject correlation). The removal resulted in an acceptable consistency in the participants' ratings (Cronbach's α = .96 for Valence, .92 for Energy Arousal, .94 for Tension Arousal, .75 for Intensity, and .93 for Preference). For Intensity ratings, the agreement between the participants was lower than the usually accepted threshold for high reliability (α > .80, McGraw & Wong, 1996) and hence this concept was discarded from all further analyses. High consistencies among the rest of the affect ratings suggest that isolated instrument sounds contain adequate cues for perceiving the emotional expression. Individual ratings were collapsed into the mean rating for each concept, due to their high agreement, and then were checked for the assumption of normality using Lilliefors test with $p < .05$ as the criterion level. No data transformations were necessary.

First we looked at the relationships between the rating scales, shown in Table 1. As in Experiment 1, correlations indicate how Preference and Valence were considered to be virtually identical by the participants, $r(108) = .97, p < .001$. Also, the two dimensions of arousal (Energy and Tension arousal) were highly collinear, $r(108) = .84$, $p < .001$, suggesting that one of the affect dimensions could be eliminated without significant loss of the overall affect structure, which resembles the findings obtained by Eerola and Vuoskoski (2011) in their comparison of emotion models using film soundtracks. However, this will be explored later in more detail using another task to minimize terminological and semantic confusions. At this stage, we are mainly interested in finding out to what degree the acoustic features can be linked to the two most independent affect dimensions, Valence and Energy arousal.

TABLE 1. Correlations Between the Affect Ratings in Experiment 1 (*N* = 110).

Concept	Valence	Energy	Tension
Energy	-.60**		
Tension	-.88**	.84**	
Preference	.97**	-.52**	-.81**

** *p* < .001, *df* = 108.

From these correlations it may be inferred that the affect space delineated by Valence and Energy arousal, shown in Figure 1, offers the least redundant portrayal of affect ratings for the sounds. A number of the instrument sounds have been labeled and circled in the Figure to help the reader to discern the pattern of affects better. These identified sounds will form the subset to be used in a follow-up experiment. It is particularly interesting to note that the instrument families (brass, strings, percussion, etc.) do not seem to relate to the observed linear relation between the affect dimensions. To qualify this observation, a one-way ANOVA with instrument family (6 levels) as the independent variable was conducted separately for the listeners' mean ratings for each of the three concepts. No differences in Valence, $F(5, 104) = 1.87$, *ns*, Energy arousal, $F(5, 104) = 1.67$, *ns*, and Tension arousal, $F(5, 104) = 1.02$, *ns*, emerged. It is of course possible that another classification of sounds (e.g., plucked, bowed, and steady-state sounds) might have yielded differences across the classes but such differences can also be analyzed by comparing the ratings with the acoustic features. Differences related to articulations will be more systematically studied in the Experiment 3.

ACOUSTIC FEATURES

Literature on the subject of acoustic features mentions several that characterize spectral, temporal, and spectrotemporal aspects of instrument sounds. As a starting point, the present study used common timbre-related computational features (Grey, 1977; McAdams et al.,

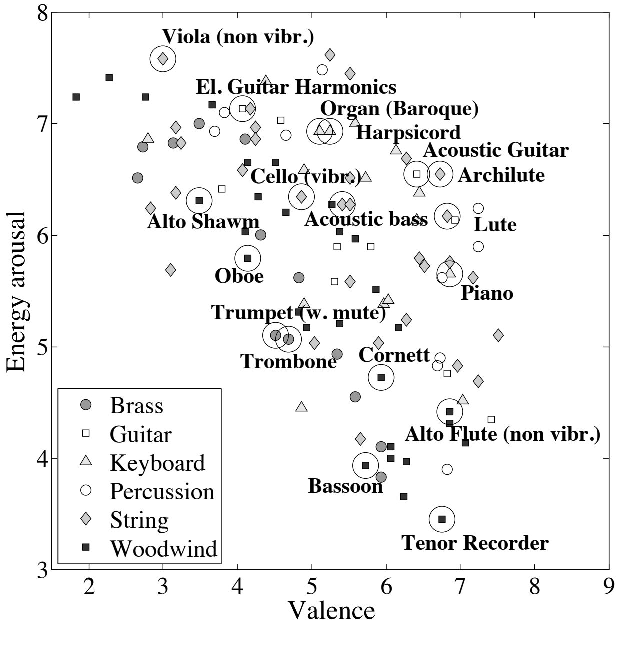

FIGURE 1. Mean ratings of valence and energy arousal of all instrument sounds in Experiment 1 (*N* = 110). The circles mark the sounds used in Experiment 2.

1995; Tzanetakis & Cook, 2002). A total of 26 such features were initially extracted to provide a sufficiently detailed yet compact account of the acoustic qualities of the 110 instrument sounds (see Table 2 for an overview).

Most of the features listed in Table 2 were gathered from the pertinent literature (cited in the table). Two additional acoustic features, however, were the Envelope Centroid and Envelope Fluctuation, which represented the centroid and standard deviation of the temporal envelope, respectively. The motivation for including these two features was to capture additional information about the temporal envelopes of the sounds, which may affect the Valence or Preference of the sounds.[2] All acoustic features were checked for the assumptions of normality using Lilliefors test, and those features that violated the assumption were transformed using a Box-Cox transformation (Box & Cox, 1964).

The feature extraction process was preceded by a trimming operation for eliminating silences at the ends of the soundfiles. The parts of the audio file with RMS energy values below 3% of the median RMS value were eliminated. All features were computed using 25 ms frames with 50% overlap. In this way, the feature set consisted of the mean of each acoustic feature across all frames. The entire analysis was carried out in the MATLAB environment using the MIRToolbox 1.2.4 (Lartillot & Toiviainen, 2007).[3]

To simplify the results and construct reliable linear models between the features and affect ratings, a pruning of the 26 acoustic features was necessary. For this, we applied Principal Components Analysis (PCA) to all 26 of the *Z*-score transformed features (extracted from the 110 instrument sounds). This yielded a seven-component solution that explained 83.4% of variance of the original matrix (using eigenvalues > 1). This number of components (seven) is also compatible with the notion of the maximum amount of predictors usable in regression modeling with 110 cases (a minimum recommendation of 10 or 20 times more observations than predictors, see Hair, Black, Babin, Anderson, & Tatham, 2006).

However, instead of relying solely on the components (PCs) as predictors, the individual acoustic features were taken into account. This is because they are more straightforward to explain than the PCs, which are less so, as a linear combination of the whole initial matrix is needed to account for them. Therefore, the selection of optimal individual features representing the seven PCs was carried out using feature selection principles as outlined by Jolliffe (see Al-Kandari & Jolliffe, 2001). In other words, the highest correlating feature for each PC was chosen (these

TABLE 2. List of Initial Acoustic Features (26 in Total).

D†	Feature	Description
T	Attack Slope	Slope of the attack portion of the sound
T	Envelope Centroid	Centroid of the temporal envelope
T	Envelope Fluctuation	Standard deviation of the temporal envelope
T	Zero-Crossing Rate	Number of time-domain zero-crossings (Tzanetakis & Cook, 2002)
S	Spectral Centroid	Geometric center of the spectrum (McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995)
S	Ratio of HF-LF energy	High-energy to low-energy ratio (Juslin, 2000)
S	Spectral Spread	Standard Deviation of the spectrum
S	Spectral Skewness	Skewness of the spectrum
S	Spectral Kurtosis	Kurtosis of the spectrum
S	Spectral Flatness	Ratio between the geometric and the arithmetic mean of the spectrum
S	Spectral Roll-off	The frequency boundary where 85% of the total power spectrum energy reside (Tzanetakis & Cook, 2002)
S	Spectral Entropy	Measure of disorder of the spectrum
S	Spectral Regularity	Degree of uniformity of the successive peaks of the spectrum, also called *Spectral Smoothness* (McAdams, Beauchamp, & Meneguzzi,1999)
S	Inharmonicity	Deviation of partials from the harmonic frequencies (Jensen, 1999)
ST	Roughness	Estimation of the sensory dissonance (Sethares, 1998)
ST	Spectral Flux	Change between the consecutive spectral frames (McAdams et al., 1995)
ST	Sub-Band No. 1-10 Flux	Spectral flux within particular frequency bands (Alluri & Toiviainen, 2010)

†S = Spectral, T = Temporal, ST = Spectro-Temporal Domain.

[2] The MFCCs were also tried out but were excluded due to their low correlations with the perceptual dimensions and interpretational difficulties.

[3] In addition to analyzing the sounds as a whole, separate descriptors for different parts of the envelope, i.e., the onset and steady state, were extracted. However, the difference between the features extracted from the steady state and the entire sound was marginal (±.02 in correlation coefficients between the acoustic features and the participants' ratings) and thus the entire sound files were utilized in the analysis.

TABLE 3. Correlation of Selected Seven Acoustic Features with the Principal Components.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
Variance explained	33.4%	15.8%	12.8%	7.3%	6.5%	4.5%	3.3%
Attack Slope	-.47	.50	-.10	.46	**-.29**	-.06	.28
Envelope Centroid	.39	-.51	.33	**-.56**	-.00	-.13	-.06
Ratio of HF-LF energy	**.89**	-.02	-.12	-.01	.10	.12	.11
Spectral Skewness	-.06	**-.67**	.44	.08	-.49	.16	.15
Spectral Regularity	-.34	.05	.32	-.53	-.07	**-.40**	-.05
Spectral Flux	-.00	.32	.51	.44	.09	.13	**-.28**
Sub-Band No. 6 Flux	.30	-.21	**.69**	.25	.38	-.01	.04

correlations are marked in bold in Table 3). This operation resulted in seven acoustic features, roughly divisible into *temporal* (Attack slope and Envelope Centroid), *spectral* (Ratio of high-frequency to low-frequency energy, Spectral Skewness, Spectral Regularity), and *spectro-temporal* (Spectral Flux and Sub-Band No. 6 Flux). These features are not only representative and compact, but also nearly orthogonal. This is because they are based on orthogonal PCs, and the mean absolute correlation between the features is low, $r(108) = .23, p < .05$.

RESULTS

Acoustic correlates of the affect ratings. The correlations between the acoustic features and affect ratings were first visualized in order to discover any nonlinear relationships or outliers in the variables, but no such trends or outliers were found. Instead, high linear relationships between the acoustic features and the ratings were visible. For instance, $r(108) = -.74, p < .001$, between Valence and Ratio of high-frequency to low-frequency energy and $r(108) = -.42, p < .001$, between Energy and Spectral Regularity exemplify these interrelationships. An example is shown in Figure 2 and all correlations are shown in Table 7.

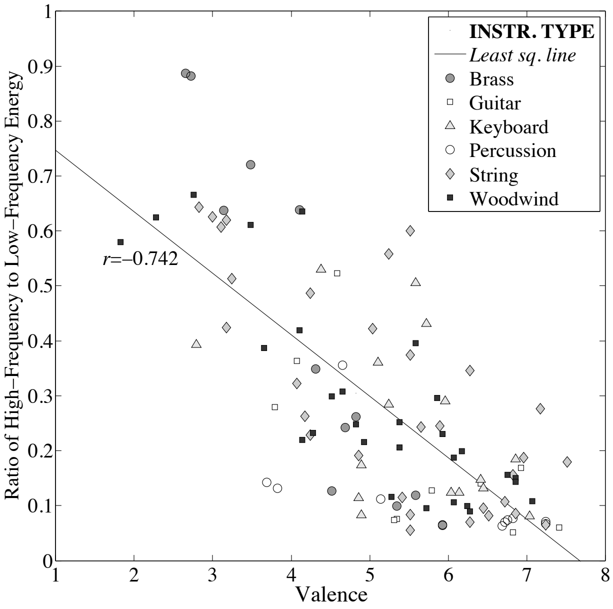

FIGURE 2. Mean ratings of valence (*x*-axis) and ratio of high-frequency to low-frequency energy measurements (*y*-axis) of all instrument sounds in Experiment 1 ($N = 110$).

To investigate whether linear combinations of the features could explain the ratings of two affect dimensions, multiple regression analyses were conducted for each affect dimension. Robust regression was chosen as the form of multiple regression, due to its resilience against outliers and distributional problems (Street, Carroll, & Ruppert, 1988). The multiple regression used the seven aforementioned individual acoustic features. The results are displayed in Table 4, where the normalized regression coefficients are given for the regression models, together with the individual acoustic features. The overall explanation rate of the models was good (> 50% variance is explained), and all the regression models were significant at the $p < .001$ level, where $F(7, 102) > 40.00$.

A high proportion of Valence ratings were explained (approximately 60% of variance) using mainly just three of the seven acoustic features. The significant normalized beta coefficients that represent these features in Table 4 suggest that positively valenced sounds have a high envelope centroid. In other words, they are not percussive, but more likely to be sustained sounds, and contain more energy in the lower frequencies compared to high frequencies. Energy arousal ratings could also be explained (≈ 60% of variance) using another combination of the three acoustic features. Sounds that are energetic tend to have fast attacks, an emphasis on the early part of the envelope and with the dominant distribution of energy in the high frequencies. The internal cross-validations (5-fold) of the models indicate only minor decreases (0 and 2% for Valence and Energy arousal) in the fit of the models, suggesting that the pattern is stable within this particular set of data. We will later explore to what degree

TABLE 4. Summary of Regression Analysis in Experiment 1 (N = 110).

	Valence	Energy
	R^{2adj}	R^{2adj}
Prediction rate	.58	.59
	β	β
Attack Slope	.04	.28**
Envelope Centroid	.33*	-.50***
Ratio of HF-LF energy	-1.09***	.91***
Spectral Skewness	-.28*	-.01
Spectral Regularity	-.01	.01
Spectral Flux	-.15	-.04
Sub-Band No. 6 Flux	.03	.11

* $p < .05$, ** $p < .01$, *** $p < .001$

these particular regression models can predict another set of data, by conducting a follow-up experiment to gain a more conservative estimate of the generalizability of the models.

DISCUSSION

The results of the regression analysis are interesting in terms of the previous studies. For example, Scherer and Oshinsky (1977) observed that the activity dimension correlated with the number of harmonics and the sharpness of the envelope. This is consistent with the results of the present study, in which the relevant measures are ratio of high-frequency to low-frequency energy, spectral centroid, and attack slope. Pleasantness, in their study, was linked to a low number of harmonics and sharp envelope, although a direct comparison is difficult as their study was a factorial one, with synthetic stimuli and manipulation of several other features (contour, pitch height, volume, tempo) that may have contributed to the ratings. Nevertheless, some parallels can be drawn. In a speech and emotion study by Laukka, Juslin, and Bresin (2005), Valence was connected with ratio of high-frequency to low-frequency energy in the same (negative) fashion as in our study, and activation was positively linked with ratio of high-frequency to low-frequency energy and attack slope, again consistent with the results of the current study. Also, articulation differences (Attack Slope and Envelope Centroid) in happy and sad emotional expressions have been documented in a number of expressive performance studies, where staccato articulation is related to more active emotions such as happiness and anger, whereas legato articulation is typically used for tender and sad emotions (Bresin & Friberg, 2000; Gabrielsson & Lindström, 1995; Juslin, 1997, 2000).

It is perhaps worth noting that previous studies of emotions in music related to timbre have used a considerably richer set of stimuli. These stimuli typically have been varied in several different dimensions (F_0, speech rate, tempo, mode, dynamics, etc.) in comparison with the isolated instrument sounds in this experiment that had identical pitch height and loudness. Considering the paucity of the available information for the listeners, it is remarkable that in the present study such clear patterns of results could be observed. However, it may be premature to draw conclusions about a general pattern in the acoustic correlates of affects since the actual structure of affects (2 to 3 dimensions and intensity) was applied to the data on purely theoretical grounds. A more robust follow-up experiment was therefore devised to explore what kind of affect dimensions would be most appropriate to describe the affective qualities of the instrument sounds.

Experiment 2: Structure of Affect Using Emotion Similarity Rating Task

Since Valence, Energy arousal, Tension arousal, and Intensity operated in a collinear fashion in Experiment 1, we felt that a separate, non-theory-driven investigation of the structure of affects provided by the sounds was needed. An emotion similarity rating task would identify the critical dimensions of affects without resorting to any explicit labelling of the affects. This method has been used successfully in past music and emotion research as a means to uncover the dimensions of emotions for complete, polyphonic excerpts of music (Bigand et al., 2005; Vieillard et al., 2008) and also to reveal the emotional processes involved in the mental representation of music (Barrett & Fossum, 2001; Russell, 1980) and perception of affects (e.g., facial perception, Hamann & Adolphs, 1999).

EXPERIMENT DETAILS

Stimuli. The eighteen instrument sounds used in Experiment 1 were used in Experiment 2. The sounds were sampled by taking randomly two sounds from Experiment 1 using an evenly spaced grid (3 × 3, defined by 33.3% and 66.6% percentiles in the data) overlaid to the valence - energy arousal space. These chosen examples are identified with circles in Figure 1.[4] For a similarity rating experiment, this resulted in 153 paired comparisons using a single item pairing order.

Participants. The participants consisted of four female and nine male music students (mean age = 26.21

[4] Acoustic bass, acoustic guitar, alto flute with vibrato, alto shawm, archlute, organ (baroque plenum), trumpet with bucket mute, bassoon, cello (vibrato), electric guitar harmonics, harpsicord (8 stop), lute (renaissance 8 course), piano, oboe, tenor recorder, treble cornett, trombone (tenor, muted), viola (non vibrato). Reference numbering to the Experiment 1 sounds is also noted in Appendix A.

SD = 3.16), all of whom were extensively trained in music (> 10 years of music training, including proficiency with several instruments).

Procedure. The participants were asked to rate the affective similarity of each pair of sounds on a similarity scale of 1 to 9, where 1 indicated the minimum and 9 the maximum. The two sounds were separated by 800 ms of silence. The order of sounds was individually randomized and the experiment was carried out in a sound isolated booth and presented on a computer.

RESULTS

Underlying affect structure of sounds. High inter-rater consistency in the similarity ratings was observed (Cronbach's α = .93). The individual similarity ratings were converted into individual distance matrices by subtracting the ratings from 10 and reorganizing them into matrix shapes. These individual matrices were subjected to individual multidimensional scaling (SMACOF, De Leeuw & Mair, 2009), which yielded 2, 3, and 4 dimensional solutions with a reasonable stress (.27, .11, and .05, respectively).[5] To compare the resulting scaling solution with the results obtained from the ratings of the two affect dimensions (Valence and Energy arousal) for the same sounds in Experiment 1, procrustes analysis was carried out to rotate, scale, and translate the scaling solution in an optimal way with the coordinate positions of the two affect dimensions (Cox, 2001). This analysis yielded highly significant symmetric correlation between the solutions, *r*(16) = .86, *p* < .001 (estimated using 1000 permutations). In effect, the multidimensional scaling solution was simply rotated 40 degrees clockwise to obtain the maximal fit between the two two-dimensional solutions.

The rotated two-dimensional solution (Figure 3) is directly comparable with the ratings for Valence and Energy arousal for the same sounds (the highlighted markers in Figure 1). Similarities between the scaling solution and affect ratings are clearly evident, since the extremes of the X-axis show the same instrument sounds in both representations (viola, alto shawm on the left; lute and tenor recorder on the right). Equally, the positions of the sounds in both the *y*-axes portray underlying similarities. Trombone, cornett, alto flute, and bassoon are among five of the lowest scoring sounds in both datasets, while viola and electric guitar harmonics are among the top five highest scores. There are also some differences between the datasets, such as the acoustic bass, which was situated in the middle of the Valence-Energy affect space, and yet on the extreme right of the *x*-axis in the scaling solution. The differences are probably explained by the fact that the two datasets are based on different methodologies and subsets of data. Perhaps the pair-wise comparison method nevertheless emphasizes a higher similarity between the acoustic bass and other plucked string instruments.

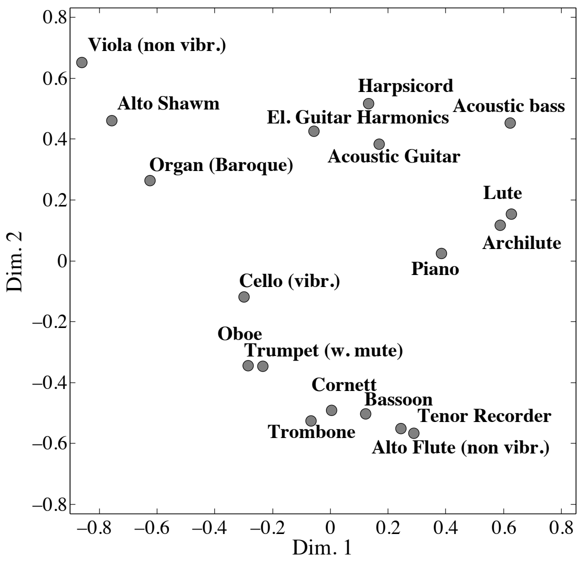

FIGURE 3. Multidimensional scaling solution of individual similarity ratings of instrument sounds.

The dimensions were interpreted by first correlating the axis coordinates for each sound with their respective affect ratings from Experiment 1, then with their acoustic features. The first result implies that the dimensions of the rotated scaling solution could equally represent valence and energy arousal, as the first dimension correlated highly with the valence ratings for each excerpt, *r*(16) = .83, *p* < .001, while the second dimension positively correlated with the energy arousal ratings, *r*(16) = .89, *p* < .001. The third dimension did not correlate with any of the ratings collected in Experiment 1, *r*(16) < |.16| for all. Bigand and colleagues (Bigand et al., 2005) tentatively interpreted this third dimension as kinetic since it could be connected to melodic movement, and more generally to body postures, gestures, and movement. Because the added variance of this

[5] Since the sample size is small (*N* = 13), we explored the reliability of the MDS solution by leaving out one participant across all possible candidates, and calculating the correlation between this new solution and the one obtained with 13 participants. This analysis yielded high correlations (*r[16]* = .997 and .985 for dimensions 1 and 2). Similar analysis was repeated by leaving out two participants (78 combinations), which resulted again in correlations, *r*(16) = .995, *p* < .001, and *r*(16) = .98, *p* < .001. Omitting randomly three participants for 1,000 times, the correlations were still at the level, *r*(16) = .99, *p* < .001, and *r*(16) = .98, *p* < .001, for dimensions 1 and 2). From this additional analysis it can be concluded that the initial solution offered is robust.

dimension was relatively small (≈ 10%) in both studies and neither Bigand nor we provide additional data to characterize the semantic qualities of this dimension, we can only speculate about it. In their analysis of short excerpts of Indian music, Alluri and Toiviainen (2010) found *fullness* to best represent the third factor of the perceptual dimensions, a characteristic already suggested by Von Bismarck (1974). Fullness was described by such bipolar scales as *empty-full* and *compact-scattered* and was connected to fluctuations in the lower end of the spectrum. This is, in our opinion, a closer match with the third dimension in the present study than the kinetic dimension in Bigand's study, which seemed to capitalize on melodic movements not present in our study using identical pitch heights.

We believe this interpretation of the dimensions is justified, since the fit of all alternative formulations in a procrustes analysis resulted in a lower fit between the two coordinate solutions. Energy and tension arousal correlated only moderately with all possible combinations of the dimensions in the scaling solution (between .41 and .80). In addition, Valence is more compatible with previous studies of emotion that have been conducted in a purely music context (Bigand et al., 2005; Vieillard et al., 2008). In both of these studies, Arousal was associated with the second dimension, as it is here.

When the dimensions were compared to the acoustic features of the examples, the first dimension was found to correlate highly with the Ratio of HF-LF energy, $r(16) = .–82$, $p < .001$. Meanwhile the second dimension correlated both with the Spectral skewness, $r(16) = –.71$, $p < .001$ and to a lesser extent with Spectral flux, $r(16) = -.58$, $p < .05$. Both of these results are to be expected, as the analyses in Experiment 1 showed that these particular acoustic features already correlated with the rating dimensions obtained from a larger set of data. The third dimension is more difficult to explain, since it correlated only moderately with Spectral flatness, $r(16) = .45$, $p = .059$. Finally, the fourth dimension, which adds another 6% to the explained variance in the similarity ratings, correlated best with Sub-Band Flux No. 4 ($r(16) = .61$, $p < .01$). It seems imprudent however to come up with labels for the third and fourth dimensions based on these analyses, since the patterns within these higher dimensions remain somewhat unclear, with only a minor portion of variance accounted for.

DISCUSSION

The emotional similarity task based on instrument sounds allowed us to uncover an underlying affect space without resorting to any postulated semantic concepts. Individual multidimensional scaling analysis revealed a familiar affect structure based on the two dimensions of Valence and Energy Arousal. This interpretation was supported by maximal correlations of these dimensions to the ratings in a procrustes analysis between all possible permutations of the three dimensions originally used for rating and scaling. A simplification of the affect space into two dimensions (Valence and Energy arousal) was therefore justifiable on the grounds that it keeps the model parsimonious, with the focus of research on the precise nature of the link between perceived affects in music and its physical attributes.

To investigate whether we could validate and replicate the results observed in the previous two experiments, and also extend the scope of the findings, a third experiment was created.

Experiment 3: Affect Ratings of Manipulated Instrument Sounds

For the third experiment, we turned our attention to the Vienna Symphonic Library (VSL[6]), which is a colossal (550 GB) collection of sampled classical music instrument excerpts, performed by high-level musicians using a wide variety of articulations, in all registers, and at several dynamic levels. Our smaller set of high-quality sounds explored more subtle differences between the timbres than in Experiment 2. Two variant versions of the sounds were made by altering the apparent source of the sounds. This was done by using different dynamic levels, and by filtering the sounds in terms of one of the main features contributing to affective evaluations in Experiment 2; namely, ratio of high-frequency to low-frequency energy.

EXPERIMENT DETAILS

Stimuli. A total of 105 instrument samples representing different sections of the symphony orchestra were chosen as stimuli from the Vienna Symphonic Library (VSL). The library includes sounds played at different dynamic levels, as well as in different registers. The samples were then split into three subsets of 35 samples each. Subset 1 was chosen from the forte dynamic level. These included 11 unique instruments (violin, cello, trombone, trumpet, bassoon, flute, oboe, marimba, clarinet, horn, vibraphone) with up to seven articulations (plain, staccato, vibrato, legato, sforzato, marcato, and pizzicato; see Appendix B for details). Based on the results of Experiment 1, we predicted that sustained sounds, in other words, those with a higher Envelope Centroid (plain, vibrato, legato) would lead to higher ratings of Valence and lower ratings of Energy. Articulations with a faster Attack Slope (sforzato, marcato, and staccato) would also be rated higher on Energy. However, more subtle

[6] http://www.vsl.co.at

differences within the percussive type of articulations (staccato, sforzato, and marcato) or sustained articulations (legato, vibrato, and plain) were difficult to predict beforehand.

The instruments were chosen to maximize compatibility with the previous experiment, and the pitch for all was set to D♯4. This time however, only 9 of the 35 sounds were 1 s long (0.8-1.0 s to be precise). The maximum was 2 s long, and most of them were exactly this length (the median of sound duration for the entire set was 1.99).

Subset 2 comprised of the same 35 sounds but were taken from a different initial dynamic level (mezzo-forte) in the sample library. Nevertheless, they were equalized to the same level as the Subset 1 sounds. The result of producing sounds at a lower dynamic level should have resulted in fewer higher harmonics, less spectral fluctuation, and less alterations in the temporal envelope (Pitt & Crowder, 1992; Strong & Clark, 1976). This should have particularly affected instruments such as the trumpet and cello, as these features of timbre are known to be a substantial part of their sound (e.g., Fletcher & Tarnopolsky, 1999).

Finally for Subset 3, we wanted to take the same 35 sounds from Subset 1, and make a straightforward alteration to the spectrum to see whether such a manipulation of ratio of high-frequency to low-frequency energy modified the affect ratings in the predicted direction.

For this purpose, a two-pole IIR filter with a resonant frequency at 2,000 Hz was applied to each sound. The difference equation for this filter can be seen below. It describes what the filter output signal is in relation to its input signal, at any given point in time.

$$y(n)=.05^{*}x(n)+1.78^{*}\,y(n–1)–.86^{*}\,y(n–2) \qquad (1)$$

where $y(n)$ refers to the output and $x(n)$ to the input, at a given time n. The coefficients were calculated according to the transfer function of a two-pole IIR filter (Rocchesso, 2004) with an altered gain factor G of 0.05 and a bandwidth of 1,000 Hz.

$$H(z)=\frac{G}{1-2r\cos\omega_0 z^{-1}+r^2z^{-2}} \qquad (2)$$

$$r=1-\frac{bandwidth(radians)}{2} \qquad (3)$$

The magnitude response can be viewed in Figure 4.

The two-pole IIR filter was applied to the sounds in Subset 1 to create Subset 3 (35 sounds). The result of this filtering operation was an increase in the values of ratio of high-frequency to low-frequency energy for most examples due to its peak being in the vicinity of 2,000 Hz. Apart from increasing the ratio of high-frequency to low-frequency energy, the filtering also led to alterations in the spectrum above 2,000 Hz due to the high attenuation slope following the peak in magnitude response.

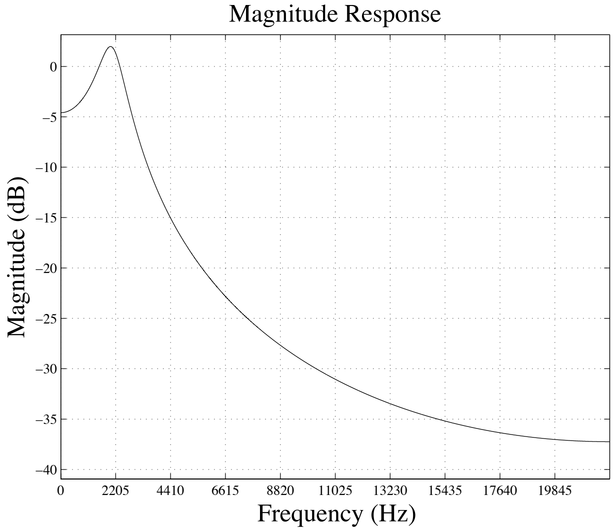

FIGURE 4. Magnitude response of the filter used to increase ratio of high-frequency to low-frequency energy in Subset 3.

We hypothesized that this manipulation would lead to decreased Valence and increased Energy arousal ratings due to the importance and the direction of the ratio of high-frequency to low-frequency energy coefficients received in Experiment 1. It is also possible that the attenuation of the high-frequencies (above 2,000 Hz) also counteracts some of the predicted effects since absence of high-frequency energies may also be interpreted as imparting distance cues of the sound source. To verify that our two manipulations would actually create differences in terms of acoustic features, we conducted a one-way ANOVA on ratio of high-frequency to low-frequency energy across the three subsets, which resulted in significant differences, $F(2, 102) = 18.79, p < .001$, between all of them in post-hoc tests, and mean values of .44 (95% confidence intervals .37–.51), .30 (.23–.36), and .58 (.52–.64) for the Subsets 1, 2, and 3, respectively. The subsets also differed in terms of Spectral Skewness, $F(2, 102) = 30.92$, $p < .001$, Spectral Regularity, $F(2, 102) = 8.27$, $p < .001$, and Spectral Flux, $F(2, 102) = 5.64, p < .01$. These features all showed the lowest mean values in Subset 2 (except Spectral Regularity, operating in a reverse fashion), which was just as predicted (Pitt & Crowder, 1992; Strong & Clark, 1976). In other words, the sounds generated with lower dynamics should prove to be less chaotic for most of the spectro-temporal features. In total, Experiment 3 had 105 sounds that were equalized in terms of loudness, using peak RMS value normalization and careful subjective evaluation. A 23 ms fade-out at the end of each sample was also made to prevent any abrupt termination of the sound.

Participants. The participants consisted of 14 females and 17 males (age $M = 26.35$, $SD = 6.04$). Only 6.45% of them had no formal music education, and 16% described themselves as being trained formally for more than the half of their lives. The rest had a mean of 8.06 years of formal music training.

Procedure.Experiment procedures (task, instructions, scales, and adjectives) were identical to those used in Experiment 1.

RESULTS

Acoustic correlates of affects. As in Experiment 1, similar pre-processing operations were carried out, and the same seven acoustic features were used. The only exception was Ratio of high-frequency to low-frequency energy, due to alterations caused by the IIR filter described in the previous section. In Experiment 3 the cut-off frequency for calculating Ratio of high-frequency to low-frequency energy was increased from 1,500 Hz to 2,000 Hz. The behavioral ratings were first screened to eliminate outliers using the same criteria as in Experiment 1, ultimately resulting in the removal of four participants owing to their low inter-subject correlations (3 *SD*s from the group center). After this, the inter-rater reliabilities were $\alpha = .88$ for Valence, .93 for Energy arousal, .92 for Tension, .85 for Preference, and .83 for Intensity. Again, a reasonable consensus existed for most of the affect dimensions, which suggests that even this limited set of sounds was enough to create affective connotations for most listeners. Ratings of preference and intensity were also less homogeneously distributed across participants and these concepts were not used in the analyses. The pattern of correlations for these four concepts, displayed in Table 5, was almost identical to the ones obtained in Experiment 1 (see Table 1), corroborating earlier general findings.

Again, correlations between the acoustic features and the ratings of affect dimensions revealed moderate correlation coefficients mainly for Ratio of high-frequency to low-frequency energy, $r(103) = -.55$, $p < .001$ for Valence, $r(103) = .45$, $p < .001$ for Energy but also with other features such as Sub-Band No. 6 Flux (see Table 7 for all the correlations between acoustic features and ratings in both experiments). Before looking at the similarities between these results and those of Experiment 2 (Table 6), we will first replicate the analysis of the ratings by constructing linear combinations of the acoustic features, using robust regression and cross-validation.

TABLE 5. Correlations Between the Ratings in Experiment 3 (*N* = 105).

	Valence	Energy	Tension
Energy	-.45**		
Tension	-.66**	.84**	
Preference	.93**	-.23*	-.50**

* $p < .01$, ** $p < .001$, $df = 103$.

TABLE 6. Summary of Regression Analysis in Experiment 3 (*N* = 105).

	Valence	Energy
	R^{2adj}	R^{2adj}
Prediction rate	.54	.74
	β	β
Attack Slope	-.17*	.32***
Envelope Centroid	.71***	-.42***
Ratio of HF-LF energy	-.52***	.71***
Spectral Skewness	-.30***	-.30**
Spectral Regularity	.03	-.13
Spectral Flux	.81***	-.09
Sub-Band No. 6 Flux	-.75***	.33**

* $p < .05$, ** $p < .01$, *** $p < .001$

In these analyses, robust and predictive models could be built to explain variance in the participants' ratings. More specifically, Valence ratings could be explained (≈ 55% variance explained) using a linear combination of the features. Although the important features contributing significantly to the regression equation were slightly different to those in Experiment 2, the overall pattern was similar: positively valenced sounds tended to have a slower attack and higher Envelope Centroid, both again showed higher positive valence when instruments had long decay and low attack slopes. In line with Experiment 2 (and Experiment 1 for that matter), a high Ratio of HF-LF energy was negatively associated with Valence. The measurements of flux also affected the Valence ratings, the overall Spectral Flux in a positive fashion and the flux within a selected sub-band in a negative fashion. This is somewhat harder to interpret since the two features correlate, $r(103) = .40$, $p < .001$, whereas in the robust regression, the two variables are related to different aspects of the residual variance. If we explore this in more detail, we can compare the unique contribution of these variables within the regression by using the squared semipartial correlations. Thus, by partialing out the contribution of all the other six features, we notice that the Spectral Flux has almost twice the unique contribution ($sr^2 = .196$) to that of the Sub-Band No. 6 Flux ($sr^2 = .106$). Thus, pleasant sounds also have a temporally dynamic spectrum, possibly linked to vibrato, but not to the frequency region around 800-1,600 Hz, which is the frequency region of the Sub-Band No. 6 Flux.

Over 70% of variance in Energy arousal ratings could be explained by the regression. Energetic sounds were characterized by sharp attacks and bursts of energy in the initial parts of the envelope. Spectrally energetic sounds tended to have a dominant proportion of energy in the high-frequency regions and have particular dynamic fluctuation of the spectrum within the 800-1,600 Hz region, which is consistent with the results of Experiment 2. It is worth noting, when speaking of the internal validity of the models, that the cross-validation of the two regression models results in only minor decreases (1-2%) in prediction rates.

Before attempting to draw connections between our results and those of past studies, we shall first explore in more detail manipulations in the dynamics and Ratio of HF-LF energy that were made, as well as the articulation differences within the sounds used in Experiment 3.

Results of acoustic manipulations and articulation styles on affect ratings. The differences across the three subsets (35 sounds in each) and the articulation styles (7) within the subsets will be explored using traditional analyses of variance. A visualization of the three affect ratings across the subsets is shown in Figure 5, implying notable differences in the ratings as a function of subset.

Separate two-way repeated measure ANOVAs were conducted for each concept using Subset (3 levels) and Articulation (7 levels) as within-subjects factors. For Valence, significant main effects of Subset, $F(2, 46) = 24.99$, $\eta^2 = .52$, $p < .001$, and Articulation, $F(6, 138) = 18.38$, $\eta^2 = .44$, $p < .001$, were evident. There was also a significant factor interaction, $F(12, 276) = 9.35$, $\eta^2 = .29$, $p < .001$. A post-hoc comparison (Tukey, adjusted for multiple tests) showed that Subset 2 was the one that differed significantly from the other subsets ($p < .001$). We hypothesized that Subset 3 would show lower Valence ratings due to the increase in Ratio of HF-LF energy but this was not borne out by the analysis. Although the Valence ratings are somewhat lower in Subset 3 than in Subset 1, the difference is minor and not significant ($p = .73$). In line with our predictions, Subset 2 sound had elevated levels of Valence in comparison with Subset 1, as Subset 2 had the intention of being phenomenally softer and smoother, due to their lower initial dynamics. These differences were also evident in the acoustic summaries of the subsets. The Articulation differences in Valence ratings, demonstrated graphically in the upper panel of Figure 6, indicated mainly that the string sounds, particularly the pizzicato ones, were the most favorably rated in terms of Valence and the rapid onset sounds were at this end of the continuum. A planned contrast between the sustained (plain, legato, vibrato) and impulse-type envelope sounds (pizzicato, staccato, sforzato, marcato) was not significant ($Z = 1.31$, $p = .19$) despite the significant differences between pizzicato, staccato/sforzato/marcato, and plain/legato/vibrato articulations.

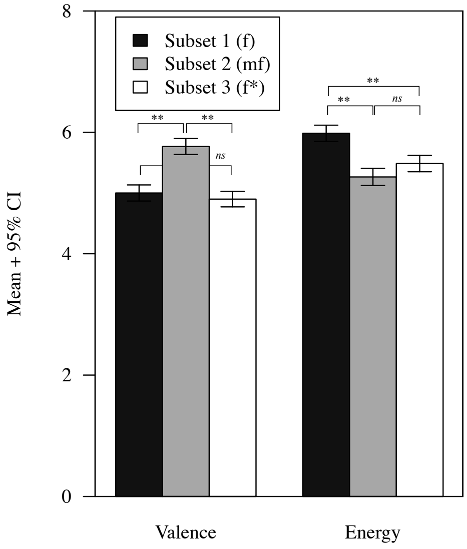

FIGURE 5. Mean ratings across the three subsets.

For Energy arousal, another two-way repeated ANOVA yielded a significant main effect for Subset, $F(2,46) = 32.48$, $\eta^2 = .59$, $p < .001$, Articulation, $F(6, 138) = 45.54$, $\eta^2 = .66$, $p < .001$, and Interaction, $F(12, 276) = 13.35$, $\eta^2 = .37$, $p < .001$. According to the predictions, the Energy ratings should have been lower in Subset 2 and higher in Subset 3 when compared to Subset 1. Only the first hypothesis received support, since the Energy arousal ratings were in fact significantly lower in Subset 2 than in Subset 1 ($p < .001$ in Tukey contrasts), while Subset 3 received significantly lower ratings than Subset 1 ($p < .001$). The articulation differences in Energy arousal

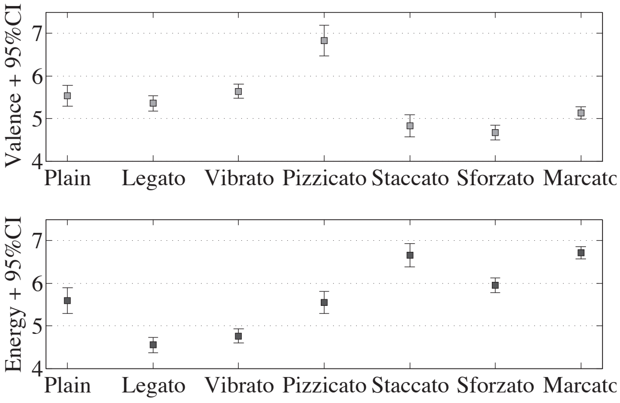

FIGURE 6. Mean ratings across articulation styles (*N* = 105).

were mostly as predicted, since we hypothesized that the impulse-type envelopes (staccato, sforzato, pizzicato, marcato) would lead to higher ratings of Energy arousal than those with sustained envelopes (plain, legato, vibrato), and this was supported by the statistically significant planned contrast ($Z = 11.90, p < .001$).

To summarize these findings, Subset 2 sounds were initially generated at a lower dynamic level that resulted in longer attacks and generally more energy on lower frequencies and less time-varying spectra, as was witnessed when features from Subsets 1 and 2 were compared. Although the dynamic level of these sounds was normalized to the level of other subsets, the ratings indicated that listeners were treating these sounds as being softer and hence, as originally intended, more pleasant, less energetic, and more relaxed: trends that are familiar from the studies of expressive performance (e.g., Bresin & Friberg, 2000; Juslin, 2000). Moreover, the findings concerning articulations produced interpretable differences related to the attack slope and the shape of the temporal envelope (for both impulse and sustained types) that have been characterized in terms of staccato-legato articulations in many of the expressive performance studies (e.g., Bresin & Friberg, 2000; Gabrielsson & Lindström, 1995; Juslin, 1997, 2000), and particularly, in a study of an expressive performances using only single piano notes (Baraldi et al., 2006).

The ratio of high-frequency to low-frequency energy manipulations in Subset 3 did not, however, have the predicted effects of decreasing Valence and increasing Energy Arousal. There are several possible explanations for this. First, the Ratio of high-frequency to low-frequency energy manipulation may have been too subtle. Acoustically measured Ratio of high-frequency to low-frequency energy values of Subset 3 were significantly higher than those for Subset 1 but this is not a guarantee of its perceptual salience. The filtering was performed for the entire spectrum, which perhaps led to other, unwanted psychoacoustic effects for the sound qualities. Since we are accustomed to filtered sounds in our everyday contexts (phone, room acoustics, sound source distances), the effect of the actual filtering may be small and mostly reflect the dampened part of the perceptually salient frequencies (around 3,000 Hz), which gives the impression of being subdued or distant in its character since the high-frequency components lose energy due to air absorption faster than the low-frequency components (Middlebrooks & Greenhaw, 1991). The second alternative explanation is that the instrument templates are stronger than we assumed. In other words, the listeners had no problems recognizing the instruments and used the instrument itself as the main source of their affect ratings and adjusted it according to sound qualities, such as richness, which was in fact reduced by the filtering in comparison to the sounds of Subset 1.

Further work is needed to disentangle these possible explanations. With respect to the main aim of Experiment 3, failing to accurately predict the direction of the manipulations is of minor concern since the purpose was to explore the acoustic correlates of affects in isolated instruments sounds using a more varied set of sounds, and this was achieved with the manipulations. The analyses regarding the articulations also have a caveat: articulations are specific to particular instruments that conventionally produce the sounds. For instance, the string instruments responsible for the staccato articulations tended to receive high Valence and Energy ratings also in Experiment 2, and hence the elevated levels may again relate to the instrument itself, rather than to the articulation. A factorial design would be needed to fully explore the separate contribution of each of these aspects.

Comparison of the affects in Experiments 1 and 3. Experiment 1 had 110 sounds that represented a large variety of real instruments from different eras, music genres, and sound generation types. Experiment 3 relied on a more restricted palette of sounds (35 sounds with three manipulations), representing the most common instruments in the classical orchestra. One could say that Experiment 1 had a larger variety of sounds than Experiment 3. Therefore it is crucial to test whether the models created with linear regression, and thus the features and their weights, are valid for both experiments. To this end, we devised two comparisons: the first was a simple table of correlations between the ratings and features for both experiments (Table 7) and the second is a new cross-validation of the regression models. For most of the acoustic features, correlations appear to be stable across the two experiments: For Valence, which is nearly identical, significant correlations for Ratio of HF-LF energy, Spectral Regularity, and Sub-Band No. 6 Flux are displayed. For Energy, most acoustic features (Ratio of HF-LF energy, Spectral Skewness, Envelope Centroid) also exhibit similar correlations. The most notable difference between the experiments concerned the Attack Slope, in which the difference was not a large one and concerned only the magnitude of the correlation, and not a reversal of a sign and the relationship. From these comparisons, we could infer that the acoustic features selected, and the affect ratings given for sounds in both experiments, were relatively stable and operated consistently across the dimensions.

In Experiments 1 and 3, a five-fold cross-validation was used to assess the degree of overfit of the models

TABLE 7. **Correlations Between Features and Affect Ratings in Experiments 1 and 3.**

	Valence		Energy	
	Expt. 1	Expt. 3	Expt. 1	Expt. 3
Attack Slope	.25**	.10	.04	.25*
Envelope Centroid	-.10	-.04	-.25**	-.44***
HF - LF energy ratio	-.74***	-.56***	.56***	.46***
Spectral Skewness	-.11	-.07	-.23*	-.38***
Spectral Regularity	-.26**	.21*	-.42***	-.18
Spectral Flux	-.15	.09	.12	.22*
Sub-Band No.6 Flux	-.23*	-.26**	.13	-.04

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; $df = 108$ in Experiment 1 and $df = 103$ in Experiment 3.

within the experiments. This new cross-validation was to address the issue of validating regression models between the experiments. In other words, to ensure that we could predict the results of Experiment 3 using the regression model from Experiment 1 and vice versa. This resulted in considerable correlations between the model from Experiment 1 and the ratings of Valence, $r(108) = .43$, $p < .001$, and Energy, $r(108) = .80$, $p < .001$, from Experiment 3. Although we can see the Valence ratings suffered the most in this comparison, it proved to be the most difficult concept to predict and its features across both the experiments differed the most. The high success in the cross-validation for Energy attests to the overall robustness of the observations. When the reverse situation is considered, these cross-validations yield the following correlations between the model from Experiment 3 and prediction of ratings from Experiment 1: Valence, $r(103) = .45$, $p < .001$, and Energy, $r(103) = .77$, $p < .001$. Again, both are statistically significant at $p < .001$ and the pattern is nearly identical to that of the previous comparison. These cross-validations of the regression models further underline the high correspondence of the results obtained in both experiments. In sum, the relationships between the acoustic features and the behavioral ratings in both these experiments seem to share an underlying structure.

The correlations between the acoustic features and the affect ratings show interesting connections with the previous studies. To start with the observations from outside the domain of music, the most preferred environmental wind sounds have a higher concentration of energy in the lower portions of the spectrum and the harsh sounds were associated with a concentration of energy in the higher frequencies (Kidd & Watson, 2003). This compares favorably to our results, where there was also a negative relationship between the Valence ratings and Ratio of high-frequency to low-frequency energy values for the sounds. Also, Kidd and Watson reported that appealing sounds were characterized by a high spectral variation, akin to positive correlations between Valence ratings and Spectral Flux. In the pioneering study by Scherer and Oshinsky (1977), filtration cut-off level was found to relate to Pleasantness ratings for the synthetic stimuli in a negative fashion, and to Activity ratings with a positive coefficient in the regression model, which are both in high agreement with the results of the present study. In general, the pattern of results also resembles the findings from the expressive performance studies, particularly for the acoustic features of Ratio of high-frequency to low-frequency energy and Articulation, which have been linked to systematic differences between the basic emotions expressed in music (e.g., Bresin & Friberg, 2000; Gabrielsson & Lindström, 1995; Juslin, 1997, 2000). These observations, in turn, have been linked with similar patterns of results in expressive speech (e.g., Ilie & Thompson, 2006; Juslin & Laukka, 2003; Laukka, Neiberg, Forsell, Karlsson, & Elenius, 2011).

General Discussion

First, listeners are able to consistently rate brief, isolated instrument samples in terms of affect dimensions. This is something that previous studies within the context of emotions (Bigand et al., 2005; Filipic et al., 2010; Peretz et al., 1998) using short excerpts of actual music passages or production studies (Juslin, 1997, 2000) have hinted at though not studied separately. Recent results (Krumhansl, 2010), which show agreement between the evaluated emotional content in both long and short clips, further endorses the importance of timbre in emotion perception, even if this study did not investigate the specific acoustic cues that contributed to those emotions being communicated. Although the recognition of songs or genres is different from emotional communication, it could be argued that emotional communication in general benefits from such aspects that are immediately perceptible and do not require long preparation time.

Second, affect ratings of such short instrument timbres were moderately well explained using a small set of acoustic features. A compact set of seven acoustic features (spanning spectral, temporal, and spectro-temporal characteristics of the sounds) was used to predict the listeners' ratings across affect dimensions. For most concepts, a dominant part of the variance was explained with two to four acoustic features. These models were then cross-validated across the two experiments and the common underlying features were identified. Many of the successful acoustic features have been identified as salient expressive features in previous studies relating to expressive content of speech and music (Juslin & Laukka, 2003; Scherer & Oshinsky, 1977).

Third, our attempt was to control the timbre without resorting to artificial sound generation schemes. This was achieved by preserving the natural variation within the common instrument timbres and having listeners rate the instruments in isolation. In Experiment 3 we also explored more systematically how articulation and dynamic variation contributed to the perceived affective dimensions. However, further studies of timbre should probably employ radically different timbre spaces as the current exploration emphasized classical music instruments in isolation. For instance, an experiment using synthetic sounds in the style of Grey and Moorer (1977) and McAdams, Beauchamp, and Meneguzzi (1999), or that morphed the existing sets of sounds (Haken, Fitz, & Christensen, 2007) could more clearly disambiguate the learned associations of the instrument characteristics from the acoustic features relating to the emotional expressions.

Fourth, the most consistently clear affect structure throughout the experiments was best represented in the two affect dimensions of Valence and Energy arousal. These two dimensions provided the least redundant descriptions of the affects when they were directly rated. Furthermore, they were retrieved by means of a multidimensional scaling of data obtained from an emotional similarity task. The affect ratings in Experiments 1 and 3 were initially gathered using the three dimensional model, due to its specific advantages in terms of capturing emotions in music (Ilie & Thompson, 2006) and because it was good for particular self-reported affects (Schimmack & Grob, 2000). However, this model failed to produce distinct results for each of the three dimensions, as was found in a previous study by Eerola and Vuoskoski (2011), and so it was eventually dropped. Nevertheless, it would be premature to conclude that the affect structure provided by instrument sounds is always two dimensional, since the coverage of possible sounds was not exhaustive. These concerns can all be easily addressed in future studies.

The results of the present study do raise a number of questions related to the timbral characteristics of emotional expression. The fact was that a number of essential music parameters such as alterations in dynamics, pitch, and harmony were predominantly missing from the stimulus materials, and yet the affective nuances of the instrument sounds were nevertheless communicated to the participants in a consistent manner. This attests to the importance of timbral features in conveying affects.

Future research could look at the effect of orchestration on affect by making empirical evaluations of certain classical instrument combinations. These could be, for example, a brass quintet or a string quartet, or they could be whole sections from an orchestra (e.g., strings, woodwind, or brass). Interesting hypotheses for the presumed affect dimensions could be obtained from orchestration manuals (Piston, 1955; Read, 2004), which Kendall and Carterette (1993) already used for evaluating the verbal attributes of simultaneous instrument dyads.

A stronger argument about the importance of timbre and particularly specific timbral characteristics, can be expressed by drawing attention to the underlying physiological mechanisms that reflect the body-states related to affective experiences. These so-called *push effects* (Scherer, Johnstone, & Klasmeyer, 2003) have been documented widely. Pleasant affective states are reflected in faucal and pharyngeal expansion that manifest in relatively more low- than high-frequency energy (Scherer, 1986), whereas the high-arousal emotions (anger and joy) are related to an increase in high-frequency energy (Banse & Scherer, 1996; Johnstone & Scherer, 2000). Even the xylophone can be interpreted as being unable to mimic the speech cues used to convey sadness and depression due to its physical characteristics (Schutz et al., 2008). In our opinion, this argument is more consistent with the results observed in the present three experiments since the patterns are easier to connect to findings in the vocal expression of emotion (Scherer et al., 2003; Juslin & Laukka, 2003), and even to infant-directed singing, than to the conventions of music. For example, infant-directed singing has relatively more energy at lower frequencies (Trainor, Clark, Huntley, & Adams, 1997), rendering it more pleasant to the listener (infant and adult), which is consistent with the observations of the present study concerning the Valence dimension. Also, the few studies dealing with aspects of timbre in communicating emotions in music (Goydke et al., 2004; Juslin, 2000; Scherer & Oshinsky, 1977) have drawn mutually consistent observations about the role of high-frequency energy content and articulation type relating to different emotional expressions.

The timbral cues available in short, isolated instrument sounds may partly capitalize common cues of emotional expression in addition to being subject to the conventions of culture. Although further research is required to obtain answers to these fundamental issues, it is clear that timbral cues to affects in music are relatively strong, and they follow a distinct pattern that resemble findings made in other domains. More importantly, the results provide a number of tantalizing new prospects for studying the pivotal role of timbre in the perception of emotional expression in music.

Author Note

The work was funded by the BrainTuning FP6-2004-NEST-PATH-028570 (first author) and by the Finnish Centre of Excellence in Interdisciplinary Music Research (second and third authors). We would like to thank Anthony Prechtl and Ezekiel Oak Brockmann for their helpful assistance in this project. An early part of the study was presented in ICMPC9 in Sapporo, Japan.

Correspondence concerning this article should be addressed to Tuomas Eerola, Department of Music, University of Jyväskylä, Finland. E-MAIL: tuomas.eerola@jyu.fi

References

AL-KANDARI, N., & JOLLIFFE, I. (2001). Variable selection and interpretation of covariance principal components. *Communications in Statistics-Simulation and Computation, 30*, 339–354.

ALLURI, V., & TOIVIAINEN, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception, 27*, 223–242.

BALKWILL, L.-L., & THOMPSON, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*, 43–64.

BANSE, R., & SCHERER, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614-636.

BARALDI, F. B., DE POLI, G., & RODA, A. (2006). Communicating expressive intentions with a single piano note. *Journal of New Music Research, 35*, 197-210.

BARRETT, L., & FOSSUM, T. (2001). Mental representations of affect knowledge. *Cognition and Emotion, 15*, 333–363.

BELIN, P., FILLION-BILODEAU, S., & GOSSELIN, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40,* 531–539.

BIGAND, E., VIEILLARD, S., MADURELL, F., MAROZEAU, J., & DACQUET, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion, 19*, 1113–1139.

BOX, G., & COX, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*, 211–252.

BRADLEY, M. M. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37*, 204–215.

BRESIN, R., & FRIBERG, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal, 24*, 44–63.

BURGOYNE, J., & MCADAMS, S. (2008). A meta-analysis of timbre perception using nonlinear extensions to CLASCAL. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer music modeling and retrieval: Sense of sounds* (Volume 4969, pp. 181–202). Berlin: Springer.

CACLIN, A., BRATTICO, E., TERVANIEMI, M., NÄÄTÄNEN, R., MORLET, D., GIARD, M., ET AL. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience, 18,* 1959–1972.

CACLIN, A., MCADAMS, S., SMITH, B., & GIARD, M. (2008). Interactive processing of timbre dimensions: An exploration with event-related potentials. *Journal of Cognitive Neuroscience, 20,* 49–64.

CACLIN, A., MCADAMS, S., SMITH, B. K., & WINSBERG, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America, 118,* 471–482.

COX, T. F., & COX, M. A. A. (2001). *Multidimensional scaling.* New York: Chapman and Hall.

DE LEEUW, J., & MAIR, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software, 31*, 1–30.

EEROLA, T., & FERRER, R. (2008). Instrument library (MUMS) revised. *Music Perception, 25*, 253–255.

EEROLA, T., & VUOSKOSKI, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*, 18–49.

FILIPIC, S., TILLMANN, B., & BIGAND, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin and Review, 17,* 335–341.

FLETCHER, N. H., & TARNOPOLSKY, A. (1999). Blowing pressure, power, and spectrum in trumpet playing. *Journal of the Acoustical Society of America, 105,* 874–881.

GABRIELSSON, A., & LINDSTRÖM, E. (1995). Emotional expression in synthesizer and sentograph performance. *Psychomusicology, 14,* 94–116.

GABRIELSSON, A., & LINDSTRÖM, E. (2010). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, application* (pp. 367-400). New York: Oxford University Press.

GJERDINGEN, R. O., & PERROTT, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research, 37*, 93–100.

GOYDKE, K., ALTENMÜLLER, E., MÖLLER, J., & MÜNTE, T. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research, 21,* 351-359.

GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61*, 1270–1277.

Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluation of synthesized musical instrument tones. *Journal of the Acoustical Society of America, 62,* 454–462.

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate data analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harschberger, M. L. (1997). Methodological issues in timbre research. In I. Deliege & J. A. Sloboda (Eds.), *Perception and cognition of music* (pp. 253–307). Hove, UK: Psychology Press.

Haken, L., Fitz, K., & Christensen, P. (2007). Beyond traditional sampling synthesis: Real-time timbre morphing using additive synthesis. In J. Beauchamp (Ed.), *Analysis, synthesis, and perception of musical sounds: The sound of music* (pp. 122–144). New York: Springer.

Hamann, S., & Adolphs, R. (1999). Normal recognition of emotional similarity between facial expressions following bilateral amygdala damage. *Neuropsychologia, 37,* 1135–1141.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception, 19,* 1–64.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23,* 319–329.

Jensen, K. (1999). *Timbre models of musical sounds.* Unpublished doctoral dissertation, University of Copenhagen, Copenhagen, Denmark.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of emotions* (Vol. 2, pp. 220–235). New York: Guilford Press.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14,* 383–418.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1797–1813.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814.

Juslin, P. N., & Sloboda, J. A. (2010). *Handbook of music and emotion: Theory, research, applications.* New York: Oxford University Press.

Kendall, R., & Carterette, E. (1993). Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's "Orchestration." *Music Perception, 10,* 469–501.

Kidd, G. R., & Watson, C. S. (2003). The perceptual dimensionality of environmental sounds. *Noise Control Engineering Journal, 51,* 216–231.

Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2008). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music, 36,* 101–126.

Krumhansl, C. L. (2010). Plink: "Thin slices" of music. *Music Perception, 27,* 337–354.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception and Psychophysics, 62,* 1426–1439.

Lartillot, O., & Toiviainen, P. (2007). MIR in Matlab (II): A toolbox for musical feature extraction from audio. In S. Dixon, D. Bainbridge, & R. Typke (Eds.), *Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 237–244). Vienna, Austria: Österreichische Computer Gesellschaft.

Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion, 19,* 633–653.

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2011). Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Computer Speech and Language, 25,* 84–104.

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

Maffiolo, V., Castellengo, M., & Dubois, D. (1999). Is pleasantness for soundscapes dimensional or categorical? *Journal of the Acoustical Society of America, 105,* 943–943.

McAdams, S., Beauchamp, J., & Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *Journal of the Acoustical Society of America, 105,* 882.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes. *Psychological Research, 58,* 177–192.

McGraw, K., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

Menon, V., Levitin, D., Smith, B., Lembke, A., Krasnow, B., Glazer, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage, 17,* 1742–1754.

Middlebrooks, J., & Greenhaw, D. (1991). Sound localization by human listeners. *Annual Review of Psychology, 42,* 135–159.

Ohta, K., Kuwano, S., & Namba, S. (1999). Sound quality of impulsive sounds in relation to their physical properties. *Technology reports of the Osaka University, 49*, 189-199.

Opolko, F., & Wapnick, J. (2006). The McGill University Master Samples Collection on DVD (3 DVDs). Quebec, Canada: McGill University.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68,* 111–141.

Piston, W. (1955). *Orchestration*. New York: Norton.

Pitt, M., & Crowder, R. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 728–738.

RAWLINGS, D., & LEOW, S. H. (2008). Investigating the role of psychoticism and sensation seeking in predicting emotional reactions to music. *Psychology of Music, 36,* 269–287.

READ, G. (2004). *Orchestral combinations: The science and art of instrumental tone-color.* Lanham, MD: Scarecrow.

REDONDO, J., FRAGA, I., PADRÓN, I., & PIÑEIRO, A. (2008). Affective ratings of sound stimuli. *Behavior Research Methods, 40,* 784–790.

ROCCHESSO, D. (2004). Introduction to sound processing. University of Verona, Italy.

RUSSELL, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.

SCHELLENBERG, E., IVERSON, P., & MCKINNON, M. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin and Review, 6,* 641–646.

SCHERER, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143–165.

SCHERER, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice, 9*, 235–248.

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

SCHERER, K. R., & OSHINSKY, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1,* 331-346.

SCHIMMACK, U., & GROB, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. European *Journal of Personality, 14,* 325–345.

SCHIMMACK, U., & REISENZEIN, R. (2002). Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion, 2,* 412–417.

SCHUTZ, M., HURON, D., KEETON, K., & LOEWER, G. (2008). The happy xylophone: Acoustics affordances restrict an emotional palate. *Empirical Musicology Review, 3,* 126–135.

SETHARES, W. (1998*). Tuning, timbre, spectrum, scale*. Berlin, Germany: Springer-Verlag.

SLOBODA, J. A., & JUSLIN, P. N. (2010). At the interface between the inner and outer world: Psychological perspectives. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 73–97). New York: Oxford University Press.

STREET, J. O., CARROLL, R. J., & RUPPERT, R. J. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician, 42*, 152–154.

STRONG, W., & CLARK, M. (1976). Synthesis of wind-instrument tones. *Journal of the Acoustical Society of America, 41,* 39–52.

THAYER, R. E. (1989). *The biopsychology of mood and arousal.* Oxford, UK: Oxford University Press.

TRAINOR, L., CLARK, E., HUNTLEY, A., & ADAMS, B. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development, 20,* 383–396.

TSANG, C. D., & TRAINOR, L. J. (2002). Spectral scope discrimination in infancy: Sensitivity to socially important timbres. *Infant Behavior and Development, 25,* 183–194.

TZANETAKIS, G., & COOK, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10,* 293–302.

VIEILLARD, S., PERETZ, I., GOSSELIN, N., KHALFA, S., GAGNON, L., & BOUCHARD, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion, 22,* 720–752.

VON BISMARCK, G. (1974). Sharpness as an attribute of the timbre of steady sounds. *Acustica, 30,* 159–172.

ZENTNER, M. R., & EEROLA, T. (2010). Self-report measures and models of musical emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 187–221). New York: Oxford University Press.

APPENDIX A: **Selected Sounds from the McGill University Master Samples (MUMS).**

N	Disc	Folder	Subfolder 1	Subfolder 2	Subfolder 3	Filename
1	1	Accordion	Accordion Treble Notes			Accordion Treble
2	1	Brass	Cornet			Cor
3	1	Brass	French Horn	French Horn		Fhrn
4	1	Brass	French Horn	French Horn Muted		Fhrn
5	1	Brass	Trombones	Trombone Tenor		TTbn
6*	1	Brass	Trombones	Trombone Tenor Muted		TTbn
7	1	Brass	Trumpets	Trumpet in C		CTpt
8	1	Brass	Trumpets	Trumpet in C Harmon Mute		CTpt
9	1	Brass	Trumpets	Trumpet Bach		BachTpt
10	1	Brass	Trumpets	Trumpet Bucket Loud		BbMuTpt
11	1	Brass	Trumpets	Trumpet Bucket Soft		BbMuTpt
12*	1	Brass	Trumpets	Trumpet Cup Loud		BbMuTpt
13	1	Brass	Trumpets	Trumpet Hard Attack		BbTpt
14	1	Brass	Tuba			Tba
15	1	Guitars	Acoustic Guitar	Harmonics		Acoustic Gtr Harmonics
16*	1	Guitars	Acoustic Guitar	Normal		Ac Gtr Normal
17	1	Guitars	Acoustic Guitar	Pizzicato		Gtr Pizz
18	1	Guitars	Acoustic Guitar	Sul Ponticello		Acoustic guit sulpont
19	1	Guitars	Acoustic Guitar	Sul Tasto		Acoustic Gtr Sul Tasto
20	1	Guitars	Electric Guitar	Electric Guitar		Electric Guitar
21	1	Guitars	Electric Guitar	Electric Guitar Harmonics		Egtr Harm
22*	1	Guitars	Electric Guitar	Electric Guitar Stereo Chorus		Egtr Stereo
23	1	Harp	Harp Harmonics			Harp Harmonics
24	1	Harp	Harp Single Notes			Harp
25*	2	Keyboards	Harpsichord	Harpsichord 8 Stop		Harpsichord 8 Stop
26	2	Keyboards	Harpsichord	Harpsichord 8/4 Stops		Harpsichord
27	2	Keyboards	Harpsichord	Harpsichord Buff Stop		Harpsichord Buff Stop
28*	2	Keyboards	Organ	Baroque Plenum		Baroque Plenum-
29	2	Keyboards	Organ	Crumhorn		Crumhorn
30	2	Keyboards	Organ	Gemshorn		Gemshorn
31	2	Keyboards	Organ	Solo Trumpet		Trumpet
32	2	Keyboards	Organ	SymPlenum _56		Symphonic Plenum
33	2	Keyboards	Pianos	Concert Hall Steinway Soft		Steinway Grand Soft
34	2	Keyboards	Pianos	Hamburg Steinway Loud		Hamburg Grand Loud
35	2	Keyboards	Pianos	Piano Harmonics		Piano Harmonics
36	2	Keyboards	Pianos	Piano Mpp Loud		Mpp Piano Loud
37*	2	Keyboards	Pianos	Piano Mpp Medium		Mpp Piano Medium
38	2	Keyboards	Pianos	Piano Mpp Soft		Mpp PianoSoft

(continued)

APPENDIX A: Continued.

N	Disc	Folder	Subfolder 1	Subfolder 2	Subfolder 3	Filename
39	2	Keyboards	Pianos	Piano Rigth pedal Vol9		Piano Right Pedal
40	2	Keyboards	Pianos	Steinway Piano Plucked		Steinway Pno Plucked
41	2	Percussion	Marimba	Grand Symphonic Marimba		Grand Symphonic Marimba
42	2	Percussion	Marimba	Marimba Crescendo Roll		Marimba Roll
43	2	Percussion	Marimba	Marimba Soft Mallet		Marimba Soft Mallet
44	2	Percussion	Steel Drum	Steel Drum Loud		Steel Drum Loud
45	2	Percussion	Steel Drum	Steel Drum Rolls		Steel Drum Rolls
46	2	Percussion	Steel Drum	Steel Drum Soft		Steel Drum
47	2	Percussion	Tubular Bells			Tubular Bells
48	2	Percussion	Vibraphone	Vibraphone Bowed		Vibraphone Bowed
49	2	Percussion	Vibraphone	Vibraphone Hard Mallet		Vibe Hard Mallet
50	2	Percussion	Vibraphone	Vibraphone Soft Mallet		Vibe Soft Mallet
51*	3	Strings	Bass	Acoustic Bass	Acoustic Bass Amplified	Dbs
52	3	Strings	Bass	Acoustic Bass	Acoustic Bass Bowed Vibrato	Dbs
53	3	Strings	Bass	Acoustic Bass	Acoustic Bass Muted	Dbs
54	3	Strings	Bass	Acoustic Bass	Acoustic Bass Pizz	Dbs
55	3	Strings	Bass	Electric Bass Bright		Electric Bass Bright
56	3	Strings	Bass	Electric Bass Deep		Electric Bass Deep
57	3	Strings	Bass	Electric Bass Harmonics		Electric Bass Harmonics
58	3	Strings	Cellos	Cello		Cel
59	3	Strings	Cellos	Cello Martelé		Cel
60*	3	Strings	Cellos	Cello Muted Vibrato		Cel
61	3	Strings	Cellos	Cello Non-Vibrato		Cel
62	3	Strings	Cellos	Cello Pizzicato		Cel
63*	3	Strings	Lutes	Archilute		Archilute
64*	3	Strings	Lutes	Renaissance 8 Course Lute		Lute
65	3	Strings	Violas	Viola Martelé		Vla
66	3	Strings	Violas	Viola Muted		Vla
67*	3	Strings	Violas	Viola Non-Vibrato		Vla
68	3	Strings	Violas	Viola Pizzicato		Vla
69	3	Strings	Violas	Viola Vibrato		Vla
70	3	Strings	Violins	Violin 1 Non Vibrato		Vln
71	3	Strings	Violins	Violin 2 Non Vibrato		Vln
72	3	Strings	Violins	Violin 3 Non Vibrato		Vln
73	3	Strings	Violins	Violin Martelé		Vln
74	3	Strings	Violins	Violin Muted Vibrato		Vln
75	3	Strings	Violins	Violin Pizzicato		Vln

(continued)

APPENDIX A: Continued.

N	Disc	Folder	Subfolder 1	Subfolder 2	Subfolder 3	Filename
76	3	Strings	Violins	Violin Vibrato		Vln
77	3	Strings	Violins	Vln Ensemble Dry-Bright		Vln
78	3	Strings	Violins	Vln Ensemble Soft Attack-Wet		Vln
79	3	Strings	Viols	Bass Viol		Bass Viol
80	3	Strings	Viols	Tenor Viol		Tenor Viol
81	3	Strings	Viols	Treble Viol		Treb Viol
82*	3	Woodwinds	Bassoons	Bassoon		Bsn
83	3	Woodwinds	Clarinets	Clarinet Bflat		BbCla
84	3	Woodwinds	Clarinets	Clarinet EFlat		EbCla
85	3	Woodwinds	Flutes and Piccolo	Flute Alto Non Vib		Aflt
86*	3	Woodwinds	Flutes and Piccolo	Flute Alto Vib		Aflt
87	3	Woodwinds	Flutes and Piccolo	Flute Bass Flutter		BFlt
88	3	Woodwinds	Flutes and Piccolo	Flute Bass Vib		BFlt
89	3	Woodwinds	Flutes and Piccolo	Flute Flutter		Flt
90	3	Woodwinds	Flutes and Piccolo	Flute Non Vib		Flt
91	3	Woodwinds	Flutes and Piccolo	Flute Vibrato		Flt
92*	3	Woodwinds	Historical Wind Instr.	Alto Shawm		Alto Shawm
93	3	Woodwinds	Historical Wind Instr.	Crumhorns	Tenor Crumhorn	Tenor Crumhorn
94	3	Woodwinds	Historical Wind Instr.	Oboes	Baroque Oboe (in C)	Baroque Oboe (in C)
95	3	Woodwinds	Historical Wind Instr.	Oboes	Baroque Oboe (with flatterment)	Baroque Oboe (flattermnt)
96*	3	Woodwinds	Historical Wind Instr.	Oboes	Classical Oboe	Classical Oboe
97	3	Woodwinds	Historical Wind Instr.	Oboes	Oboe d'Amore	Oboe d'Amore
98*	3	Woodwinds	Historical Wind Instr.	Recorders	Baroque Tenor Recorder	Baroque Tenor Rec
99	3	Woodwinds	Historical Wind Instr.	Recorders	Renaissance Bassinet Recorder	Renaiss Bassinet Rec
100	3	Woodwinds	Historical Wind Instr.	Recorders	Renaissance Quart Recorder	Renaiss Quart Rec
101	3	Woodwinds	Historical Wind Instr.	Recorders	Renaissance Tenor Recorder	Renaiss Tenor Rec
102*	3	Woodwinds	Historical Wind Instr.	Treble Cornett		Treble Cornett
103	3	Woodwinds	Oboe and English Horn	English Horn		Ehrn
104	3	Woodwinds	Oboe and English Horn	Oboe		Obo
105	3	Woodwinds	Saxophones	Alto Sax Sounds	Alto Sax	Asax
106	3	Woodwinds	Saxophones	Alto Sax Sounds	Alto Sax Growls	Asax
107	3	Woodwinds	Saxophones	Baritone Sax		BarSax
108	3	Woodwinds	Saxophones	Soprano Sax		SSax
109	3	Woodwinds	Saxophones	Tenor Sax Sounds	Tenor Sax	TSax
110	3	Woodwinds	Saxophones	Tenor Sax Sounds	Tenor Sax Growls	TSax

APPENDIX B: Selected Sounds from the Vienna Symphonic Library (VSL), Subset 1

Nro	Instrument	Articulation
1	Bassoon	Staccato
2	Oboe	Vibrato
3	Bassoon	Legato
4	Bassoon	Sforzato
5	Cello	Legato
6	Cello	Marcato
7	Cello	Pizzicato
8	Cello	Sforzato
9	Clarinet	Legato
10	Clarinet	Sforzato
11	Clarinet	Staccato
12	Clarinet	Plain
13	Flute	Legato
14	Flute	Sforzato
15	Flute	Vibrato
16	Horn	Legato
17	Horn	Sforzato
18	Horn	Staccato
19	Horn	Vibrato
20	Marimba	Plain
21	Oboe	Legato
22	Oboe	Sforzato
23	Oboe	Staccato
24	Trombone	Sforzato
25	Trombone	Staccato
26	Trombone	Vibrato
27	Trumpet	Sforzato
28	Trumpet	Staccato
29	Trumpet	Vibrato
30	Vibraphone	Plain
31	Vibraphone	Vibrato
32	Violin	Legato
33	Violin	Marcato
34	Violin	Pizzicato
35	Violin	Sforzato