Sound Source Mechanics and Musical Timbre Perception: Evidence From Previous Studies

—

Bruno L. Giordano & Stephen McAdams
McGill University, Montréal, Canada

TIMBRE HAS BEEN CONCEIVED OF AS A MULTIDIMENSIONAL sensory attribute and as a carrier of perceptually useful information about the mechanics of the sound source. To date, research on musical timbre has focused on defining its *acoustical* correlates, whereas fragmentary evidence is available on the influence of *mechanical* parameters. We quantified the extent to which mechanical properties of the sound source are associated with structures in the data from published identification and dissimilarity-rating studies. We focus on two macroscopic mechanical properties: the musical instrument family and excitation type. Identification confusions are significantly more frequent for same-family instruments. With dissimilarity ratings, same-family or same-excitation tones are judged more similar and tend to occupy the same region of multidimensional-scaling spaces. As such, significant associations between the perception of musical timbre and the mechanics of the sound source emerge even when not explicitly demanded by the task.

T HE SCIENTIFIC STUDY OF MUSICAL TIMBRE attracts the attention of researchers from a variety of different fields: acoustics (e.g., Hutchins, 1981), machine learning (e.g., Herrera-Boyer, Peeters, & Dubnov, 2003), psychophysics (e.g., Iverson & Krumhansl, 1993), cognition (e.g., Tillmann & McAdams, 2004), neuropsychology (e.g., Samson, Zatorre, & Ramsay, 2002), and psychobiology (e.g., Halpern, Zatorre, Bouffard, & Johnson, 2004). Theories of the perception of musical timbre place similar emphasis on its dependence on the sound structure and on the mechanics of the objects whose interaction produces the sound (Hajda,

Kendall, Carterette, & Harshberger, 1997; Handel, 1995; McAdams, 1993; McAdams & Giordano, 2009). Nonetheless, empirical research has focused on the measurement of the acoustical determinants of timbre (e.g., McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Marozeau, de Cheveigné, McAdams, & Winsberg, 2003), frequently confining considerations of its mechanical determinants to secondary observations at best (see McAdams, Chaigne, & Roussarie, 2004, for a notable exception, and Lakatos, 2000, for a qualitative attempt in this direction). As a consequence, whereas it is widely recognized that timbre is influenced by changes in the mechanics of a musical instrument, it is largely unknown what in the mechanics is differentiated perceptually by listeners. In the speech domain, a similar research bias would reveal, for example, that formant frequencies shape our sensation of human vocalizations, but would not assess their effect on the perception of the gender and age of a speaker (cf. Smith & Patterson, 2005). In order to fill this empirical void, we quantify the extent to which data from 17 published studies on the identification of, and dissimilarity ratings among, musical tones reflect differences in two macroscopic properties of the musical sound source: the type of excitation and the family of a musical instrument.

The term timbre denotes those attributes of auditory sensation that allow a listener to tell that two sounds differ even when they are equated for pitch, loudness, and duration (American National Standards Institute, 1973), and when they have the same spatial location and are produced in environments with the same reverberant properties (Levitin, 1999; McAdams & Giordano, 2009). A popular research paradigm for the study of timbre seeks to quantify its acoustical determinants. To this purpose, the behavioral technique of dissimilarity ratings is combined with the data analysis model of multidimensional scaling (MDS; Borg & Groenen, 1997). Accordingly, participants rate the dissimilarity of sounds presented in pairs (i.e., the perceptual distance between paired sounds). Sounds are often equalized in nontimbral sensory properties such as pitch and loudness in order to focus the listener perceptually on timbre. MDS algorithms

are then used to create a spatial representation of the stimulus set from the input dissimilarities, frequently a Euclidean space with a variable number of dimensions in which more dissimilar sounds are located farther apart. The Euclidean dimensions are finally related statistically to known features of the sound stimuli: those features that are most strongly associated with the dimensions are taken as the acoustical correlates of the sensory dimensions of timbre used by listeners to estimate the dissimilarities. Overall, results from different studies based on this methodology reveal that timbre is a multidimensional attribute of auditory sensation, influenced by a number of different acoustical factors (Handel, 1995). It is worth pointing out that the acoustical properties that influence timbre perception depend, at least in part, on the properties of the investigated stimulus set (Bregman, 1990, p. 124; see Goldstone, Medin, & Halberstadt, 1997, for effects of context on dissimilarity ratings, and McAdams & Giordano, 2006, for the contextual stability of the dissimilarity of musical timbres). As such, it is notable that studies conducted with different stimulus sets frequently find that the same two acoustical properties influence timbre perception: 1) attack time (piano tones have a faster attack than flute tones); and 2) spectral center of gravity (SCG—or the amplitude-weighted average frequency), the main acoustical correlate of the timbral dimension of brightness (e.g., a more sombre bassoon tone has a lower SCG than a bright trumpet tone; McAdams et al., 1995; see Caclin, McAdams, Smith, & Winsberg, 2005, for a confirmatory study; see McAdams, Giordano, Susini, Peeters, & Rioux, 2006, for a meta-analysis of the acoustical determinants of musical timbre). Finally, it is relevant to note that, to date, no strong evidence has emerged concerning an effect of expertise, i.e., of music training, on the structure of the MDS spaces of musical timbres (cf. McAdams et al., 1995; for a discrimination study showing superior performance in musicians, see Chartrand & Belin, 2006).

Beyond the study of its sensory dimensions and acoustical determinants, timbre has been shown to serve a variety of perceptual processes: the formation of auditory streams (Bregman, 1990; McAdams & Bregman, 1979); the dynamics of perceived tension and relaxation within a musical composition (Paraskeva & McAdams, 1997); the integration of tones of different musical instruments into novel timbral blends (Kendall & Carterette, 1991; Sandell, 1995). A function of timbre that has been more neglected by empirical investigations is its role in the recognition of sound-generating objects and events. From a theoretical point of view, earlier reports of accurate abilities to perceive sound source properties in purely auditory contexts (e.g., the hardness of a mallet striking

a metallic pan, Freed, 1990) likely shaped the currently accepted views of timbre (Hajda et al., 1997; Handel, 1995; McAdams, 1993). Accordingly, timbre is at the same time a multidimensional attribute of auditory sensation and the perceptual correlate of the mechanics of the sound source (Handel, 1995). More explicitly, because the mechanics of the sound source structures the acoustics of a sound (e.g., harder mallets generate sounds with a higher SCG, Fletcher & Rossing, 1991), and the features of a sound determine the sensory dimensions of the corresponding auditory object (e.g., the higher the spectral centroid the brighter the sound), the dimensions of auditory sensations, including those determining timbre, are at the same time a reflection of, and the vehicle for, the perception of the mechanics of the sound source.

A model of sound production influential on the theory of timbre perception is the source-filter model by Handel (1989, 1995, 2006; see Fant, 1960, for the speech-production antecedent of this theory). Accordingly, a musical tone is the product of the interaction between a source (e.g., a vibrating object such as the string in a violin) and a filter (e.g., the violin body) that selectively amplifies and suppresses different frequencies of the mechanical vibration of the source and radiates the sound. Because a source has to be set into vibration with some process of excitation, such as the act of bowing or plucking violin strings, the source-filter model is better termed the excitation-source-filter (ESF) model.

Musical instruments are traditionally classified in terms of the source component of the ESF model. Accordingly, von Hornbostel and Sachs (1914) distinguish four families of nonelectrical musical instruments based on the nature of the primary vibrating element: a stretched string in the family of chordophones or string instruments (e.g., viola, guitar, harp, and piano); a stretched membrane in membranophones (e.g., snare drum, tympani, and kazoo); the body of the instrument itself in idiophones, usually a stiff material that can vibrate in absence of tension (e.g., metal in cymbals and tubular bells; wood in xylophones); the air itself in aerophones or wind instruments (e.g., flute). Further distinctions can be made within the family of the aerophones, depending on whether the sound results from the blowing of an air jet (e.g., flute), from the buzzing of a single reed (e.g., clarinet, saxophone) or a double reed (e.g., oboe, bassoon), or, as with instruments like the trumpet or the trombone, through the buzzing of "lip reeds" (Fletcher & Rossing, 1991). It should be noted that instruments belonging to the same family frequently share general properties of the filter (e.g., with many chordophones the filter is a wooden box, whereas with many aerophones it is a tube; with many idiophones there is no filter).

A second mechanical distinction focuses on the excitation component of the ESF model (e.g., Hajda et al., 1997): 1) impulsive excitations are very limited in time (e.g., fractions of a millisecond, as in the strike of a mallet on a xylophone bar; Fletcher & Rossing, 1991) and are located at sound onset (e.g., drums, marimba, piano, guitar, pizzicato violin); 2) continuous excitation occurs when energy is introduced into the vibrating system throughout most of the duration of a tone (e.g., bowed violin, blown flute); 3) multiple impacts occur when multiple objects are struck within a short time period (e.g., bamboo chimes; Lakatos, 2000) or the same object is struck repeatedly (e.g., a tremolo played on a xylophone; Hajda, 1995).

The relationship between source mechanics and auditory perception has received the attention of a growing body of literature that focuses on nonmusical sounds (see Lutfi, 2007, for an up-to-date review). In contrast, very few studies have quantified the role of source mechanics in the perception of musical timbre: bar length and density on the dissimilarity ratings of xylophone sounds in McAdams et al. (2004); mouthpiece depth on the perceived brightness of trumpet sounds in Poirson, Petiot, and Gilbert (2005) (see also Aramaki, Baillères, Brancheriau, Kronland-Martinet, & Ystad, 2007, and Buksnowitz, Teischinger, Müller, Pahler, & Evans, 2007, for the preference of wood properties in violin and xylophone sounds). Consistently with current theories of timbre, both of these studies, carried out with tones from the same instrument, show that listeners' perceptions reflect the mechanics of the sound source, due to the mediation of various acoustical factors (e.g., SCG and measures of the rate of change in the amplitude envelope in McAdams et al., 2004). It should be noted, however, that in both studies stimuli were generated by manipulating only the target mechanical properties. For this reason, significant mechanical-perceptual associations were highly likely.

Mechanical distinctions between families of musical instruments and between excitation types have been at least implicitly considered in the vast majority of studies on musical timbre. Together, these factors are likely to be among the most important sources of variability in the domain of acoustic music. From the acoustical point of view, very different musical instruments, such as a plucked violin and a blown flute, likely offer the perceptual system a large number of acoustical parameters that discriminate reliably between excitation types and instrument families. For example, several properties of the time-varying loudness would distinguish an impulsive piano tone from a continuant flute tone (Hajda et al., 1997). In spite of this strong acoustical differentiation, the perceptual ability to differentiate families of musical instruments and types of excitation shouldn't be taken for granted. First, it is unknown whether sound properties that reliably distinguish different-family and different-excitation tones are accurately processed by the auditory system. For example, SCG might not always differentiate between a string instrument and a reed instrument for particular choices in the reed and string materials, and the attack time of a flute tone might tend toward that of a piano tone for particular playing styles such as the contemporary "slap-tongue" technique. From a methodological standpoint, a democratic task such as dissimilarity rating doesn't constrain judgments along specific dimensions of the stimulus, but it also doesn't guarantee that participants will focus on acoustical properties that are diagnostic of differences in instrument family or excitation type. Similarly, even though listeners can identify a single instrument at better-than-chance levels, they might in principle confound equally often tones generated by similar source-filter mechanical systems (e.g., two air jet aerophones such as a flute and a recorder), and tones generated by different source-filter mechanical systems excited in a similar way (e.g., a struck idiophone such as the marimba and a struck membranophone such as the tympani). With such outcomes, it would be difficult to conclude that perception reflects the mechanics of the sound source, because the patterns of between-instrument confusions would not be indicative of similarities in the sound source mechanics.

The published empirical evidence on the perceptual relevance of changes in musical instrument family and in excitation type is fragmentary at best and is frequently limited to observations void of any statistical verification. Some relevant studies have investigated the identification of musical instruments mainly using continuant tones (in an identification experiment, participants are asked to assign one experimenter-defined label to each sound: "Is this tone produced by a guitar or a violin?"). These studies frequently report better-than-chance identifications of individual instruments (Clark, Luce, Abrams, Schlossberg, & Rome, 1963). Interestingly, instrument families are reported to be identified more accurately than individual instruments (Martin, 1999; Srinivasan, Sullivan, & Fujinaga, 2002). Importantly, this difference might not be indicative of a genuine perceptual ability to better discriminate instrument families but rather might be the result of the statistical computations: performance in the identification of instrument families is computed by considering both the correct identifications of single instruments and the incorrect confusions between instruments from the same family. This potential problem is obviated by comparing the probability of

confusing instruments from different families with that of same-family confusions (referred to from now on as between- and within-family confusions, respectively). Van Dinther and Patterson (2006) indeed observed higher within- than between-family confusions. However, because participants in their study were given trial-by-trial feedback on the correct identification of the instrument family, results are better interpreted as showing that listeners can learn how to discriminate families, rather than as proof of a spontaneous perceptual focus on this mechanical distinction. Interestingly, pitch variations larger than one octave can induce between-family confusions in nonmusicians (Handel & Erickson, 2001, 2004; Steele & Williams, 2006), whereas with musicians between-family confusions increase drastically only for pitch variations larger than two octaves (Steele & Williams, 2006). As such, nontimbral factors appear to have a stronger role in instrument identification with nonmusicians than with musicians. However, we still do not know whether within-family confusions are significantly less frequent than between-family confusions in the absence of trial-by-trial feedback.

Decision-bound theories of categorization (Ashby & Lee, 1991) describe the relationship between identification and dissimilarity judgments (see Smits et al., 2006, for the superior ability of decision-bound models to account for the categorization of synthetic sound stimuli). Accordingly, in a dissimilarity-rating task, participants estimate the distance of stimuli within a space of perceptual features. In an identification task, to the contrary, participants partition the same space of perceptual features into independent regions using, for example, linear or quadratic decision bounds (cf. Ashby & Maddox, 1998) and assign identification responses depending on the regions of the space in which the perceptual effects of the stimuli fall (see Figure 1 for a visual example). Within this framework, if mechanically equivalent tones such as same-family or same-excitation tones are frequently confused in an identification task, one or both of two outcomes should be expected with the dissimilarity ratings of the same tones. First, mechanically similar stimuli should be perceived as more similar and thus cluster together in the derived MDS spaces (clustering outcome). Second, mechanically different stimuli should lie in different regions of the MDS spaces, regions that in an identification task are separated by a decisional bound (regional separation outcome). Both outcomes would be perhaps more compelling evidence for the perceptual relevance of mechanical features, because in contrast to an identification task, dissimilarity raters are free to focus on any acoustical property of the sound stimuli, including those that are not diagnostic of differences in
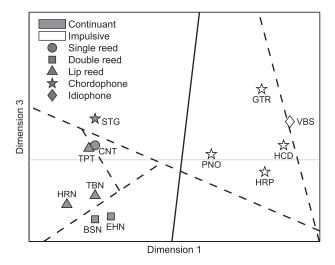

FIGURE 1. An example of regional analysis applied to the MDS configuration published in McAdams et al. (1995, Euclidean dimensions 1 and 3). Continuous and dashed lines separate different-excitation and different-family tones, respectively. Note that the lines can represent linear decision bounds in a hypothetical identification task. BSN = bassoon; CNT = clarinet; EHN = English horn; GTR = guitar; HCD = harpsichord; HRN = French horn; HRP = harp; PNO = piano; STG = string; TBN = trombone; TPT = trumpet; VBS = vibraphone.

the mechanics of the sound source. Note that both the clustering and regional separation outcomes are not completely independent, i.e., a regional separation can be observed even in the absence of a clustering of mechanically similar stimuli. For example, if different-family tones lie on two very close parallel lines, they can be separated by a middle linear decision bound even though mechanically similar stimuli might be as far from each other as mechanically different stimuli.

Previous dissimilarity-rating studies have frequently observed that impulsive and continuant tones occupy disjoint regions of the MDS spaces (Iverson & Krumhansl, 1993; McAdams et al., 1995). On the contrary, it still remains a controversial issue whether tones from the same family of musical instruments are regionally separated and/or cluster together in MDS spaces (Hajda et al., 1997, pp. 265–266). The most mechanically heterogeneous sets of musical sounds were investigated by Lakatos (2000) in a study on the dissimilarity ratings of wind/string and percussive tones (e.g., violin and bamboo chimes, respectively). The perceptual relevance of mechanical factors was assessed through the qualitative inspection of overlapping cluster structures fitted to the behavioral data (Corter & Tversky, 1986). He concluded that mechanical similarities contributed to the clustering of percussive timbres, whereas wind/string tones clustered based on acoustical rather than mechanical similarities. When related to the results of identification

studies, which focused exclusively on wind/string tones, these results would suggest that in the absence of trial-by-trial feedback, listeners will confuse equally often same-family and different-family instruments. Note that the analyses carried out by Lakatos (2000) were qualitative and not quantitative. As such, it is still largely unclear to what extent the perceptual dissimilarity of largely different musical timbres reflects the mechanics of sound generation.

In this study, we investigate the extent to which identification and dissimilarity ratings of musical instrument tones reflect differences in the mechanics of the sound-generation process. To this purpose, we reanalyzed the data from 17 studies on the identification and dissimilarity rating of musical tones. We focused on two macroscopic mechanical variables of the musical sound source: the type of excitation and the family of musical instruments. With the exception of the study of Grey (1977), included here for historical reasons, all of the other considered studies investigated tones of real or simulated musical instruments with minimal editing of the sound samples (e.g., no reverse playback, etc.). This choice was meant to focus on the experiments in which the acoustical information about the properties of the mechanics of the sound-generating event was maximized.

Identification of Musical Instruments

We measured the extent to which identification performance and confusion probabilities in previously published studies on musical instrument identification changed as a function of the similarities in the family of the musical instruments. Based on the literature, we expected higher-than-chance identification performance with individual instruments and even better performance in the identification of musical instrument families. In the case of a significant association between musical instrument families and perceptual identifications, we also expected a higher rate of within-family confusions than between-family confusions. Because the large majority of previous identification studies were conducted exclusively with continuant tones, the perceptual relevance of the distinction between excitation types could not be assessed.

METHOD

Data sets. We considered identification studies conducted with tones of real or simulated musical instruments and, in particular, publications that reported the matrices of identification confusions (an important study that did not report confusion matrices was Saldanha & Corso, 1964). Unpublished confusion matrices could be obtained from the authors of a published study in only one case (27 tones condition in Srinivasan et al., 2002). Only in Grey's (1977) study, included here for its historical relevance, did participants receive trial-by-trial feedback on identification performance, although only for the identification of individual instruments and not for that of families. Table 1 reports the classification of musical instruments for these studies along with a number of methodological variables such as constant vs. variable pitch. Note that among all the investigated stimuli, only the piano tone in Gfeller, Knutson, Woodworth, Witt, and DeBus (1998) is impulsive. In some of the studies, decay transients, and part of the stationary portions of the tones, could be missing for

TABLE 1. Datasets Considered for the Meta-analysis of Studies on the Identification of Musical Instruments.

Dataset	Pitch	Participants	Tones	Aerophones				Chordophones
				Single reed	Double reed	Lip reed	Air jet	
A	N/A	Mus.	Rec.		3	3	1	4
B	Con.	Mus.	Rec.	3	1	5	1	
C	Var.	Mus.	Rec.	1	3	4	1	
D	Con.	Mus.	Rec.	2	2	2	1	2
E	Var.	Mus.	Sim.	5	4	3	1	3
F	Mel.	NMus.	Rec.	1		1		2
G	Var.	Mus.	Rec.	1	3	4	2	4
H	Mel.	Mus.	Rec.	4	3	6	2	4
I	Var.	Mus.	Rec.	2	2	2	1	2
J	Var.	Mus.	Rec.	9	4	6	4	4

Note. The five rightmost columns indicate the number of instruments in each family present in the stimulus set. A = Clark et al. (1963), normal tones condition; B = Berger (1964), unaltered-tone condition; C = Strong and Clark (1967), natural tones; D = Elliott (1975), unaltered-tone condition; E = Grey (1977); F = Gfeller et al. (1998), normal-hearing listeners; G = Martin (1999), isolated-tone condition; H = Martin (1999), tone-sequence condition; I = Srinivasan et al. (2002), nine-tone pre-training condition; J = Srinivasan et al. (2002), 27-tone pre-training condition. Con. = constant; Var. = variable; Mel. = melody; Mus. = musicians; NMus. = nonmusicians; Sim. = simulated/synthetic; Rec. = recorded.

the purpose of limiting the duration of the stimuli. Decay transients are known to play only a secondary role in the identification of nonimpulsive tones (Clark et al., 1963; Saldanha & Corso, 1964).

RESULTS

We measured the extent to which identifications and confusions of musical instrument tones varied as a function of differences in the family of the musical instrument, including the additional distinction between subclasses of aerophones. Analyses were carried out on the raw confusion matrices, without accounting for eventual response biases. Given the multidimensionality of the musical stimuli investigated in the various studies, the rigorous computation of bias-corrected confusion measures is not a trivial problem and goes beyond the goals of our study (see Ashby, 1992, for multidimensional signal-detection theory methods).

We computed four different statistics: the identification performance for individual musical instruments and for musical instrument families, and the rate of confusions between same- and different-family tones. Note that whereas the majority of the datasets allowed one response category for each of the musical instruments in the stimulus set, Gfeller et al. (1998) and Martin (1999) also included as response categories instruments not included in the stimulus set. Further, because participants were asked to identify musical instruments, and not families of musical instruments, in all experiments family identification performance is derived from the data by pooling all responses given to instruments

belonging to the same family, i.e., correct identifications of individual instruments and within-family confusions. All of the statistics were normalized with respect to chance performance in order to relativize their value in comparison to guessing behavior. The guessing participant was assumed to choose each of the responses with the same probability, thus producing a confusion matrix with the same number of responses in each of the cells. Identification performance scores were thus normalized so that values of 0 and 1 corresponded to chance and perfect performance, respectively; confusion rates were normalized so that scores of 0 and 1 corresponded to no confusions and to the guessing rate, respectively. Note that for six out of ten studies, the adopted data normalization produced rates of within-family confusions higher than 1 (see Table 2). These values are obtained when same-family stimuli are confused with each other more frequently than would be expected if participants used each of the different musical-instrument responses with the same probability. In a hypothetical experiment with one flute, one guitar and one violin tone, a guessing participant who is presented a guitar tone will give the answer "flute," on one third of the trials and "violin" on another third of the trials. On the other hand, a participant who always confuses tones from the same family (i.e., the chordophones guitar and violin), but never confuses tones from different families (e.g., guitar and flute), will never answer "flute" when presented a guitar tone and will answer "violin" on half of the guitar trials. Assuming that the violin tone is never confused with the other instruments, in this hypothetical experiment the

TABLE 2. Normalized Identification Performance (0 = Chance; 1 = Perfect) and Confusion Rates (0 = No Confusions; 1 = Guessing Error-Rate) in Musical Instrument Identification Studies.

Dataset	Identification performance			Confusions		
	Instrument	Family	Difference	Within-family	Between-family	Difference
A	.68***	1.00***	.32	1.46	0.00***	1.46***
B	.56***	.77***	.21	0.95	0.23***	0.72***
C	.84***	.93***	.10	0.45***	0.07***	0.38***
D	.72***	.84***	.12	1.22	0.16***	1.05***
E	.74***	.93***	.19	1.12	0.07***	1.05***
F	.79***	.97**	.18*	0.91	0.03***	0.89***
G	.43***	.88***	.45***	2.64	0.12***	2.52***
H	.66***	.96***	.31**	1.78	0.04***	1.75***
I	.90***	.93***	.03	0.34***	0.07***	0.27*
J	.57***	.87***	.30***	1.72	0.13***	1.60***
Mean	.69	.91	.22	1.26	0.09	1.17
SE	.05	.02	.04	0.23	0.02	0.23

Note. The fourth and seventh columns show the difference between the scores in the third and second columns, and between those in the fifth and sixth columns, respectively. *$p < .05$, **$p < .01$, ***$p < .001$. For the source of datasets, see Note in Table 1. Refer to text for an explanation of confusion measures higher than 1.

rate of within-family confusions normalized for guessing behavior will be 3/2.

For each of the studies, we first tested whether chance-normalized identification performance scores were significantly higher than 0 (i.e., better than chance) and whether chance-normalized confusion rates were significantly lower than 1 (i.e., lower than the guessing rate). Second, we tested whether identification of families of musical instruments was significantly better than that of individual musical instruments, and whether within-family confusions were significantly more likely than between-family confusions. Hypothesis testing was carried out using the bootstrap technique (Efron & Tibishirani, 1993) based on the resampling of the cells of the confusion matrix (number of bootstrap replicates = 10,000). Significance levels were computed on the basis of BC_a (bias corrected and accelerated) confidence intervals, a refinement of the confidence intervals based on the percentile method that takes into account asymmetries and biases in the distribution of bootstrap replicates. The results of this analysis are summarized in Table 2.

With all of the datasets, performance in the identification of both individual instruments and musical instrument families was significantly better than chance. Although identification performance for families was better overall than that for individual instruments, this difference was significant for a minority of the studies (4/10). This result might be due simply to the low power of the statistical test resulting from the low number of cells of the confusion matrix used to test the difference (e.g., eight cells for the individual musical instruments in an experiment with eight instruments and eight response categories). Most importantly, for the majority of the studies (8/10), within-family confusions were not significantly lower than those expected under guessing behavior. It should be noted that lower than chance within-family confusions were observed only for the study of Clark et al. (1963) and in the nine-instrument condition of Srinivasan et al. (2002). The origin of these latter results is unclear. Finally, with all of the considered studies, between-family confusions were significantly less frequent than would be expected under guessing behavior, and within-family confusions were significantly more frequent than between-family confusions.

DISCUSSION

We quantified identification performance and confusions in previously published studies on the identification of musical instruments. In all of the datasets, identification of individual instruments was higher than chance, although far from perfect. Not surprisingly, identification of families of musical instruments was also

better than chance for all of the studies and was close to perfect with several of them. Although identification was better overall for families than for individual musical instruments, this difference was significant in the minority of the studies (four out of ten). More relevant results emerged from the analysis of identification confusions. In most studies, the probability of confusing an instrument with another one from the same family was not significantly lower than one would expect if the participant were guessing. Further, the probability of confusing instruments belonging to different families was lower than the guessing rate and close to zero for all of the studies. Indeed, within-family confusions were significantly more frequent than between-family confusions.

Along with the identification performance results, the analysis of confusion errors clearly shows that the cause of the less-than-perfect performance in the identification of individual musical instruments can be traced back to a relative inability to perceptually distinguish instruments belonging to the same family. It should be remembered that in the majority of the studies considered, multiple pitches were presented for each of the musical instruments. As such, at least two explanations can be advanced for the inability to discriminate instruments from the same family. First, acoustical information is not present that differentiates between musical instruments from the same family in a reliable way across pitch variations. Second, acoustical information is present, but it is less perceptually salient than pitch variations. In both of these cases, listeners would likely identify instruments based on their knowledge of their typical pitch (e.g., viola tones have a lower pitch than violin tones), thus producing the observed patterns of within-family confusions. Whatever the correct explanation, the analysis of timbre identification data shows that large differences in mechanical properties of the sound source are associated with significantly better identification ability.

Dissimilarity Ratings of Musical Tones

We tested the extent to which the geometry of the MDS spaces published in previous studies on the dissimilarity ratings of musical tones reflected similarities and differences in the family of a musical instrument or in its excitation type. To this purpose, we measured whether mechanically-similar stimuli were clustered together and/or occupied the same region of the MDS space (clustering and regional separation outcomes, respectively). Note that both a clustering and a regional separation based on musical instrument family would be consistent with the infrequent between-family identification confusions.

Most importantly, both outcomes would demonstrate significant associations between the mechanics of the sound source and judgments on musical timbre, even when the task does not explicitly demand a perceptual distinction between different mechanical systems.

METHOD

Data sets. Analyses focused on published MDS representations derived from the dissimilarity ratings of tones from multiple musical instruments. For the study by Wedin and Goude (1972), we considered the MDS model computed by Hajda et al. (1997). Part of the synthetic tones investigated by Krumhansl (1989) and McAdams et al. (1995) were hybrids of two different musical instruments (e.g., the *trumpar* tone was generated with a model that combined characteristics of both trumpet and guitar). The hybrid tones were excluded from the analyses, because the properties of the simulated sound source could not be established unambiguously. Kendall, Carterette, and Hajda (1999) investigated the identification and dissimilarity ratings of recorded and synthetic tones. An identification experiment revealed significantly lower identification performance for synthetic tones, as compared to real ones, thus casting a doubt on the perceptual validity of the synthesis models. For this reason, we focused our analyses on the MDS configuration computed for the judgments of natural tones averaged across a variety of experimental conditions (Kendall et al., 1999, p. 343). Among the various datasets collected by Hajda (1995), we considered only those conducted on recorded tones. In particular, we considered the MDS space computed on the matrix of judgments for natural tones averaged across a variety of different experimental conditions (Hajda, 1995, p. 79). For all of the studies, eventual edits of the real and synthetic sounds were kept to a minimum (e.g., removing an offset transient). The selected datasets are summarized in Table 3. Note that a *martelé* playing style for the violin (Lakatos, 2000) was considered as an impulsive excitation, given the extremely short duration of the contact between the bow and the string.

TABLE 3. Datasets Considered for the Meta-analysis of Studies on the Dissimilarity Ratings of Musical Tones.

					Aerophones						
Dataset	Pitch	Participants	Tones	Excitation	Single reed	Double reed	Lip reed	Air jet	Chordophones	Membranophones	Idiophones
K	Con.	Mix.	Rec.	Cont.	1	2	3		2		
L	Con.	N/A	Rec.	Cont.	1	2	3		3		
E	Con.	Mus.	Sim.	Cont.	5	4	3	1	3		
M	Con.	Mus.	Sim.	Cont.	1	3	3		1		
				Imp.					5		1
N	Con.	Mus.	Rec.	Cont.	2	3	5	1	2		
				Imp.					1		2
O	Con.	Mix.	Rec.	Imp.					3		3
				M. Imp.							2
P	Con.	Mix.	Sim.	Cont.	1	2	3		1		
				Imp.					4		1
Q	Con.	Mix.	Rec.	Cont.	4	3	2	1	1		
R	Con.	Mix.	Rec.	Cont.	3	2	3	4	1		
				Imp.					4		
S	Con.	Mix.	Rec.	Cont.						1	2
				Imp.						4	9
				M. Imp.						1	1
T	Con.	Mix.	Rec.	Cont.	2		2	2		1	2
				Imp.					4	3	4
U	Var.	Mix.	Rec.	Cont.	2	2	2		2		
				Imp.					6		
V	Var.	Mix.	Rec.	Cont.	2	2	2		2		
				Imp.					6		

Note. Datasets: K = Wedin and Goude (1972); L = Wessel (1973); E = Grey (1977); M = Krumhansl (1989); N = Iverson and Krumhansl (1993), entire tones condition; O = Hajda (1995), recorded tones condition; P = McAdams et al. (1995); Q = Kendall et al. (1999); R = Lakatos (2000), harmonic set condition; S = Lakatos (2000), percussive set condition; T = Lakatos (2000), combined set condition; U = Marozeau et al. (2003), two-semitone difference condition; V = Marozeau et al. (2003), 11-semitone difference condition. Con. = constant; Var. = variable; Mel. = melody; Mus. = musically trained; Mix. = mixed musicians/nonmusicians; Sim. = simulated/synthetic; Rec. = recorded; Cont. = continuant; Imp. = impulsive; M. Imp. = multiple impacts. Note that some percussive stimuli in Lakatos (2000) did not have a clear pitch.

RESULTS

Analyses were performed using two complementary approaches. In separate analyses, we first tested whether same-family tones or same-excitation tones occupied disjoint regions of the MDS spaces, i.e., were regionally separated. This approach to the analysis of MDS configurations has been called regional analysis (Borg & Groenen, 1997). A second distance-based analysis tested whether same-family or same-excitation tones were significantly more similar, i.e., closer than different-family or different-excitation tones, respectively.

Regional separations were tested within the Euclidean portion of the MDS models. Note that in some of the studies perceptual distances were modeled as the between-stimuli distances in a common Euclidean space, plus the location of each of the stimuli along specific dimensions not shared with all the other stimuli (Krumhansl, 1989, McAdams et al., 1995, and the MDS spaces for the harmonic and combined sets in Lakatos, 2000; see McAdams et al., 1995, for further details on MDS models with specificities as applied to timbre perception data). We used linear logistic classifiers to measure whether tones that shared a mechanical feature (e.g., instruments from the same family) occupied disjoint regions of the Euclidean space, as defined by linear functions of the MDS coordinates (see an example in Figure 1). In practice, the logistic regression model predicted either the instrument family or the excitation type based on the dimensions of the Euclidean spaces (e.g., Agresti, 1996). A perfect prediction was obtained when mechanically similar stimuli could be perfectly separated by means of linear boundaries. In order to provide a measure of the linear separability of the mechanically similar stimuli, we quantified the agreement between the actual classification of the stimuli (e.g., lip reeds, membranophones, etc.) and the classification predicted by the logistic regression model. We used the adjusted Rand index (Hubert & Arabie, 1985) for this purpose, equal to one in case of perfect agreement and to zero for chance-level agreements between actual and predicted classification. We used bootstrap hypothesis testing based on BCa confidence intervals to assess whether the adjusted Rand index was significantly higher than zero. Each of the bootstrap significance tests was carried out by resampling the stimuli (number of bootstrap replicates = 10,000). A significant test was taken as evidence for the regional separation of different groups of mechanically similar stimuli within the MDS space.

The distance-based analysis used as data the distances of stimuli within the MDS models. These were calculated considering the Euclidean portion of the models

TABLE 4. Regional and Distance-Based Analyses of the MDS Models Derived from Dissimilarity Ratings of Musical Tones (Two Leftmost and Rightmost Columns, Respectively).

Dataset	Adjusted Rand index		Within/between class dissimilarity	
	Excitation	Family	Excitation	Family
K		1.00***		.97
L		1.00***		.56***
E		.73***		.87*
M	1.00***	.62***	.50***	.61***
N	1.00***	1.00***	.40***	.40***
O	1.00***	1.00***	98	.66**
P	1.00***	1.00***	.55***	.68***
Q		1.00***		.71**
R	1.00***	.55***	.44***	.53***
S	1.00***	.24*	.73***	.96
T	1.00***	.31***	.54***	.72**
U	1.00***	1.00***	88	.99
V	1.00***	1.00***	.57***	.67***
Mean	1.00	.80	.62	.72
SE	.00	.08	.05	.05

Note. *p < .05, **p < .01, ***p < .001. For the source of datasets, see Note in Table 3.

and, when available, the specificities (cf. McAdams et al., 1995). For each dataset we tested whether stimuli that were mechanically similar in terms of family or excitation were significantly more similar to each other than mechanically different stimuli. To this purpose, we used bootstrap hypothesis testing based on BC$_a$ confidence intervals and on stimulus resampling (number of bootstrap replicates = 10,000). The results of both the regional and distance-based analyses are summarized in Table 4.

Both analyses revealed significant associations between the structure of MDS models and mechanical properties of the stimuli. All of the studies revealed perfect regional separations based on excitation type and, in the majority of the studies (8/13), on the musical instrument family as well. With the majority of the studies, mechanically similar stimuli were more similar than mechanically different stimuli (7/9 for excitation, 10/13 for instrument family). Note that the number of tests for which similarity-based clusterings were significant (17/22) was lower than the number of significant tests for regional separation (22/22). This result might be due in part to a difference in power between the statistical tests used to measure significant clusterings and regional separations. Nonetheless, of the five analyses that revealed nonsignificant clusterings of either instrument families or excitation types, four of the related regional analyses showed perfect regional separation of mechanically similar stimuli.

DISCUSSION

We measured the extent to which differences in the perceived dissimilarity of musical tones were associated with differences in musical instrument family or in excitation type. In particular, we tested whether mechanically similar stimuli were clustered in MDS spaces and whether they defined separated regions in those MDS spaces. Based on the analysis of identification data, we expected significant associations between perceptual dissimilarities and differences in instrument family. With the majority of the analyzed datasets, we observed that tones produced by instruments of the same family consistently clustered together and occupied the same region of the MDS space. Based on observations previously published in the timbre perception literature, we also expected significant associations between perceptual dissimilarities and differences in excitation type. Significant regional separations and stimulus clusterings were consistently associated with this mechanical variable in all of the analyzed datasets. In summary, the analysis of dissimilarity-rating data showed significant associations between the mechanics of sound production and perceptual judgments.

For the majority of analyzed datasets, pitch was either held constant within the stimulus set (from 262 Hz in Iverson & Krumhansl, 1993, to 466 Hz in Kendall et al., 1999) or varied within the stimulus set (from 246 to 466 Hz in the 11-semitone condition of Marozeau et al., 2003). As such, both the regional separation and the clustering results emerged both for stimulus sets with different pitches and for sets that included pitch variation. No empirical data are available to test whether significant mechanical-perceptual associations emerge when dissimilarities are collected on pitches outside this range. Studies on the identification of musical instruments are of little help here. For example, even though between-family discrimination performance drops for nonmusicians and musicians with pitch differences larger than one and two octaves, respectively (Steele & Williams, 2006), it is not possible to make sensible predictions about whether this result is caused by pitch-based perceptual clustering of the stimuli from different families or by a nonoptimal location of the decision bounds that separate different families of instruments. Furthermore, it might be reasonable to expect that some between-family distinctions are more robust to pitch variations than others, and that the perceptual differentiation of the excitation types shows a different degree of pitch-related vulnerability than between-family distinctions. Again, to our knowledge, no empirical evidence is available on this point.

No strong difference emerged between the level of structuring in the dissimilarities associated with excitation types and with families of musical instruments. Indeed, the same proportion of datasets showed significant clusterings based on either excitation types or on instrument family, and significant regional separations of excitation types and instrument families were observed in all studies. The only difference observed between the two mechanical factors was indeed in the proportion of studies that showed perfect regional separation: although stimuli generated with different types of excitation (e.g., impulsive vs. continuant) lie in perfectly disjoint regions of the MDS spaces in all studies, different-family tones (e.g., aerophones vs. chordophones) lie in perfectly disjoint MDS regions for only eight of the thirteen datasets. Based on these results, empirical evidence appears to support stronger perceptual effects of differences in excitation type than in instrument family.

General Discussion

To date, musical timbre has been conceptualized as a multidimensional attribute of auditory sensation and, at the same time, as a perceptual correlate of the mechanics of the sound source (Handel, 1995, p. 426). Within this framework, research has nonetheless emphasized the quantification of the acoustical correlates of the sensory dimensions of musical timbre (McAdams et al., 1995), whereas little effort has been made to understand what in the mechanics of the sound source is actually differentiated in timbre perception.

We analyzed 23 datasets from 17 previously published studies on the identification and dissimilarity ratings of musical timbres. In general terms, we measured the extent to which similarities and differences in the mechanics of the sound source are associated with structures in the behavioral responses. We focused on two mechanical factors that are likely to be among the major sources of acoustical variability in music: the musical instrument family and the excitation type. Consistent with previously published studies on timbre identification, we observed better than chance, although seldom perfect, identifications of both individual instruments and instrument families. Although the chance-corrected identification performance for instrument families was better than for individual instruments, the difference was significant only for a minority of the studies. We interpreted this result in terms of the low power of our statistical tests. More importantly, with the majority of the datasets, instruments from the same family were confused with each other significantly more frequently than were instruments from different families. Notably, with the majority of the studies within-family confusions were as frequent as would be expected if participants were guessing. As such, similarities

in the mechanics of the sound source were associated with increased identification confusions.

Results consistent with the literature on timbre identification emerged from the analysis of dissimilarity ratings. We separately tested whether mechanically similar stimuli clustered together in the MDS spaces, i.e., were more similar than mechanically different stimuli, and whether mechanically similar stimuli occupied the same region of an MDS spaces defined by hypothetical linear decision bounds (Ashby & Maddox, 1998). Overall, tones generated by the same type of excitation or tones generated by instruments from the same family were more similar to each other and occupied independent regions of the MDS spaces. As such, significant mechanical-perceptual associations emerged even when the task did not explicitly require participants to focus on acoustical differences that are diagnostic of mechanical differences in the sound source.

Overall, the analyses reported here are consistent with current theories of timbre, according to which this multidimensional attribute of auditory sensation is indeed the perceptual correlate of the mechanics of the sound source. In contrast to previous music-focused studies relevant to this issue (McAdams et al., 2004; Poirson et al., 2005), we investigated sets of sounds where the mechanical factors of interest—musical instrument family and excitation type—were not the only sources of acoustical variability. Above all, significant mechanical-perceptual associations emerged across within- and between-set variations in pitch. This approach consequently strengthens our test of the mechanical-perceptual association.

Several points were not addressed by our study. At least three of them deserve further investigation. First, the perceptual roles of instrument family and excitation type were tested separately. Given that these factors are to a certain extent correlated within the musical domain (e.g., most of the idiophones are struck), further research is needed to disentangle the perceptual relevance of these two factors. A likely empirical route to this goal might be to investigate stimulus sets that orthogonally combine excitation types and instrument families (e.g., bowed idiophones as in Lakatos, 2000, or wind tones generated with unusually short pulses of air as in the "slap-tongue" technique in flutes and brass instruments). For example, one possible related empirical question is whether excitation type influences the identification of instrument families and vice versa.

Second, the perceptual ability to differentiate between mechanical systems was quantified independently of the acoustical correlates of these abilities. As such, it is unclear what acoustical information listeners use to distinguish between families of musical instruments. Given the relevance of SCG to the dissimilarity ratings of timbres (cf. McAdams et al., 1995), it would be plausible to expect that this acoustical feature plays a role in the discrimination of instrument families. Associated with the hypothesis that listeners make use of whatever acoustical features maximize performance (Handel, 1995, p. 433), one might expect that the pervasiveness of a perceptual focus on SCG serves the task of optimizing the ability to differentiate mechanical systems of sound generation (see Giordano, Rocchesso, & McAdams, 2010, for a quantitative version of this hypothesis in the domain of nonmusical impact sounds). It should be noted that our distinction between families of musical instruments groups two different factors from the ESF model of sound production (Handel, 1989): the source (e.g., a vibrating string) and the filter (e.g., the body of a violin). As we noted, it is common for systems that share similar sources (e.g., strings) to also share mechanically similar filters (e.g., wooden boxes of similar shape in string instruments). Disentangling the perceptual role of these two mechanical factors will require either building novel musical instruments, or, probably more feasibly, carrying out research with stimuli synthesized according to a physical model of the musical sound source.

Finally, the analyses did not consider the role of cognitive factors in the mechanical-perceptual associations. Indeed, a popular hypothesis is that listening to music activates a processing mode that focuses on sensory attributes, independently of how sensation might inform the processes of sound source recognition (Gaver, 1993). One might speculate therefore that when listening to complex musical materials, as opposed to the isolated tones investigated by the majority of the considered studies on timbre, the perceptions of listeners will be less influenced by acoustical features that are diagnostic of differences in the mechanics of sound production. Further empirical investigations will be necessary to shed light on these issues.

Author Note

Correspondence concerning this article should be addressed to Bruno L. Giordano, Schulich School of Music, McGill University, 555 Sherbrooke Street West, H3A1E3, Montréal, Québec, Canada. E-MAIL: bruno. giordano@music.mcgill.ca

References

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.

American National Standards Institute. (1973). *Psychoacoustical terminology*. ANSI S3.20–1973 (p. 56). New York: American National Standards Institute.

Aramaki, M., Baillères, H., Brancheriau, L., Kronland-Martinet, R., & Ystad, S. (2007). Sound quality assessment of wood for xylophone bars. *Journal of the Acoustical Society of America*, *121*, 2407–2420.

Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*, 150–172.

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 251–301). San Diego, CA: Academic Press.

Berger, K. W. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America*, *36*, 1888–1891.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling*. New York: Springer-Verlag.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT press.

Buksnowitz, C., Teischinger, A., Müller, U., Pahler, A., & Evans, R. (2007). Resonance wood [Picea abies (L.) Karst.]—Evaluation and prediction of violin maker' quality-grading. *Journal of the Acoustical Society of America*, *121*, 2384–2395.

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.

Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*, 164–167.

Clark, M., Luce, D., Abrams, R., Schlossberg, H., & Rome, J. (1963). Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones. *Journal of the Audio Engineering Society*, *11*, 45–54.

Corter, J. E., & Tversky, A. (1986). Extended similarity trees. *Psychometrika*, *51*, 429–451.

Efron, B., & Tibishirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Elliott, C. (1975). Attacks and releases as factors in instrument identification. *Journal of Research in Music Education*, *23*, 35–40.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton & Co.

Fletcher, N. H., & Rossing, T. D. (1991). *The physics of musical instruments*. New York: Springer-Verlag.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *Journal of the Acoustical Society of America*, *87*, 311–322.

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, *5*, 1–29.

Gfeller, K., Knutson, J., Woodworth, G., Witt, S., & DeBus, B. (1998). Timbral recognition and appraisal by adult cochlear implant users and normal–hearing adults. *Journal of the American Academy of Audiology*, *9*, 1–19.

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 462–476.

Goldstone, R. L., Medin, D., & Halberstadt, J. (1997). Similarity in context. *Memory and Cognition*, *25*, 237–255.

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

Hajda, J. M. (1995). *The relationship between perceptual and acoustical analyses of natural and synthetic impulse signals*. Unpublished master's thesis, University of California, Los Angeles.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliege & J. A. Sloboda (Eds.), *The perception and cognition of music* (pp. 253–306). London: L. Erlbaum.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*, 1281–1292.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing* (pp. 425–461). San Diego, CA: Academic Press.

HANDEL, S. (2006). *Perceptual coherence: Hearing and seeing.* New York: Oxford University Press.

HANDEL, S., & ERICKSON, M. L. (2001). A rule of thumb: The bandwidth for timbre invariance is one octave. *Music Perception*, *19*, 121–126.

HANDEL, S., & ERICKSON, M. L. (2004). Sound source identification: The possible role of timbre transformations. *Music Perception*, *21*, 587–610.

HERRERA-BOYER, P., PEETERS, G., & DUBNOV, S. (2003). Automatic classification of musical instrument sounds. *Journal of New Music Research*, *32*, 3–21.

HUBERT, L. J., & ARABIE, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

HUTCHINS, C. M. (1981). The acoustics of violin plates. *Scientific American*, *245*, 170–186.

IVERSON, P., & KRUMHANSL, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, *94*, 2595–2603.

KENDALL, R. A., & CARTERETTE, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, *8*, 369–404.

KENDALL, R. A., CARTERETTE, E. C., & HAJDA, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, *16*, 327–364.

KRUMHANSL, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43–51). Amsterdam, Netherlands: Excerpta Medica.

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

LEVITIN, D. J. (1999). Memory for musical attributes. In P. R. Cook (Ed.), *Music, cognition, and computerized sound: An introduction to psychoacoustics* (pp. 209–227). Cambridge, MA: MIT Press.

LUTFI, R. A. (2007). Human sound source identification. In W. A. Yost, R. R. Fay, & A. N. Popper (Eds.), *Auditory perception of sound sources* (pp. 13–42). New York: Springer.

MAROZEAU, J., DE CHEVEIGNÉ, A., McADAMS, S., & WINSBERG, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, *114*, 2946–2957.

MARTIN, K. D. (1999). *Sound-source recognition: A theory and a computational model.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

McADAMS, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford, UK: Oxford University Press.

McADAMS, S., & BREGMAN, A. S. (1979). Hearing musical streams. *Computer Music Journal*, *3*, 26–43.

McADAMS, S., CHAIGNE, A., & ROUSSARIE, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America*, *115*, 1306–1320.

McADAMS, S., & GIORDANO, B. L. (2006). Generalizing timbre space data across stimulus contexts: The meta-analytic approach. *Journal of the Acoustical Society of America*, *119*, 3395.

McADAMS, S., & GIORDANO, B. L. (2009). The perception of musical timbre. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 72–80). Oxford, UK: Oxford University Press.

McADAMS, S., GIORDANO, B. L., SUSINI, P., PEETERS, G., & RIOUX, V. (2006). A meta-analysis of acoustic correlates of timbre dimensions. *Journal of the Acoustical Society of America*, *120*, 3275–3276.

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.

PARASKEVA, S., & McADAMS, S. (1997). Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of tension/relaxation schemas of musical phrases. In *Proceedings of the 1997 International Computer Music Conference* (pp. 438–441). Thessaloniki, Greece: International Computer Music Association.

POIRSON, E., PETIOT, J. F., & GILBERT, J. (2005). Study of the brightness of trumpet tones. *Journal of the Acoustical Society of America*, *118*, 2656–2666.

SALDANHA, E. L., & CORSO, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, *36*, 2021–2026.

SAMSON, S., ZATORRE, R. J., & RAMSAY, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, *125*, 511–523.

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

SMITH, D. R. R., & PATTERSON, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgments of speaker size, sex, and age. *Journal of the Acoustical Society of America*, *118*, 3177–3186.

SMITS, R., SERENO, J., & JONGMAN, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 733–754.

SRINIVASAN, A., SULLIVAN, D., & FUJINAGA, I. (2002). Recognition of isolated instrument tones by conservatory students. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & E. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition, Sidney* (pp. 720–723). Adelaide, Australia: Casual Productions.

STEELE, K. M., & WILLIAMS, A. K. (2006). Is the bandwidth for timbre invariance only one octave? *Music Perception*, *23*, 215–220.

STRONG, W., & CLARK, M. (1967). Synthesis of wind–instrument tones. *Journal of the Acoustical Society of America*, *41*, 39–52.

TILLMANN, B., & MCADAMS, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1131–1142.

VAN DINTHER, R., & PATTERSON, R. D. (2006). Perception of acoustic scale and size in musical instrument sounds. *Journal of the Acoustical Society of America*, *120*, 2158–2176.

VON HORNBOSTEL, E. M., & SACHS, C. (1914). Systematik der Musikinstrumente. Ein Versuch [An attempt at a systematic classification of musical instruments]. *Zeitschrift für Ethnologie*, *4–5*, 553–590.

WEDIN, L., & GOUDE, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology*, *13*, 228–240.

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