Chapter 5
The Semantics of Timbre

Charalampos Saitis and Stefan Weinzierl

Abstract Because humans lack a sensory vocabulary for auditory experiences, timbral qualities of sounds are often conceptualized and communicated through readily available sensory attributes from different modalities (e.g., bright, warm, sweet) but also through the use of onomatopoeic attributes (e.g., ringing, buzzing, shrill) or nonsensory attributes relating to abstract constructs (e.g., rich, complex, harsh). The analysis of the linguistic description of timbre, or timbre semantics, can be considered as one way to study its perceptual representation empirically. In the most commonly adopted approach, timbre is considered as a set of verbally defined perceptual attributes that represent the dimensions of a semantic timbre space. Previous studies have identified three salient semantic dimensions for timbre along with related acoustic properties. Comparisons with similarity-based multidimensional models confirm the strong link between perceiving timbre and talking about it. Still, the cognitive and neural mechanisms of timbre semantics remain largely unknown and underexplored, especially when one looks beyond the case of acoustic musical instruments.

Keywords Auditory roughness · Auditory semantics · Cognitive linguistics · Conceptual metaphor · Crossmodal correspondence · Describing sound · Magnitude estimation · Musical meaning · Qualia · Semantic differential · Sound color · Sound mass · Sound quality · Timbral brightness · Verbal attribute

5.1 Introduction

> After consultations with his teacher and with the great violinist and collector Efrem Zimbalist … Yehudi [Menuhin] played on all three [Stradivari violins] and opted for the "Khevenhüller." (As a test piece he played "The Prayer" from Handel's Dettingen *Te Deum*.). It was to be his principal instrument for over 20 years. He described it as "ample and round, varnished in a deep, glowing red, its grand proportions … matched by a sound

C. Saitis (✉) · S. Weinzierl
Audio Communication Group, Technische Universität Berlin, Berlin, Germany
e-mail: charalampos.saitis@campus.tu-berlin.de; stefan.weinzierl@tu-berlin.de

K. Siedenburg et al. (eds.), *Timbre: Acoustics, Perception, and Cognition*,
Springer Handbook of Auditory Research 69,
https://doi.org/10.1007/978-3-030-14832-4_5

> at once powerful, mellow and sweet." Antonio Stradivarius had made the instrument in 1733, his 90th year, when despite his advancing years he was still at the peak of his powers (Burton 2016, p. 86).

What is a mellow and sweet sound? Imagine yourself listening to a recording of the famous violinist Yehudi Menuhin (1916–1999) performing on his Khevenhüller Strad. How would you describe the *sound* of the violin or the *sound* of Menuhin? What about the *sound quality* of the recording? Musicians, composers, sound artists, listeners, acousticians, musical instrument makers, audio engineers, scholars of sound and music, even sonar technicians, all share a subtle vocabulary of verbal attributes when they need to describe timbral qualities of sounds. These verbalizations are not crucial for processing timbre—listeners can compare (McAdams, Chap. 2), recognize (Agus, Suied, and Pressnitzer, Chap. 3), or memorize and imagine (Siedenburg and Müllensiefen, Chap. 4) timbral qualities without having to name them (Wallmark 2014). However, the way we talk about sensory experiences can disclose significant information about the way we perceive them (Dubois 2000; Thiering 2015). Menuhin's mellow and sweet sound is a particular *concept*, an abstract yet structured idea anchored to and allowing one to make sense of a particular perceptual representation (Wallmark 2014). As such, a relation must exist between the physical properties of a sound that give rise to timbre and its semantic description.

Results of multidimensional scaling of pairwise sound dissimilarity ratings (McAdams, Chap. 2) usually show that timbre may be adequately explained on the basis of just two or three dimensions; a number many times smaller than the plethora of words and phrases used to communicate timbral impressions. On the one hand, this might be due to specific perceptual features of individual sounds (referred to as *specificities*) that are not mapped onto the shared dimensions of the prevailing timbre space. For example, the suppression of even harmonics in clarinet tones, which typically elicits an impression of hollowness, was not accounted for by classic geometric timbre models alone (e.g., McAdams et al. 1995). On the other hand, individual verbalizations can be thought of as representing *microconcepts*—basic elements of semantic knowledge activated by a stimulus object that are not fully meaningful on their own but instead yield meaning when assembled into broader semantic categories (Saitis et al. 2017). Among the diverse timbre vocabulary, therefore, many seemingly unassociated words may share the same meaning and refer to the same perceptual dimension.

Accordingly, the main goals of the research ideas and tools discussed in this chapter are twofold: to identify the few salient semantic substrates of linguistic descriptions of timbral impressions that can yield consistent and differentiating responses to different timbres along with their acoustic correlates and to quantify the relationship between perceptual (similarity-based) and semantic (language-based) representations for timbre. Important questions include the following:

- How similar are semantic timbre spaces between different categories of sound objects, for example, between instrument families and between instruments, voices, and nonmusical sounds?

- Do timbre verbalizations rely explicitly on acoustic cues or are they subject to source-cause categorical influences?
- Are timbre verbalizations a product of cultural dependencies or is timbre semantics cross-cultural?
- What are the neurobiological mechanisms underlying timbral semantic processing?
- In what ways does timbre contribute to larger-scale musical meaning?
- What is the relation between emotion and the semantics of timbre?

Subsequent sections attempt to address these questions. Section 5.2 examines how different communities of listeners verbally negotiate sound qualities and the underlying conceptualizations of timbre. In general, verbal attributes of timbre are predominantly metaphorical in nature, and empirical findings across different types of sounds and analytical approaches converge to a few salient semantic substrates, which are not very different from early theorizations for a low-dimensional semantic space of timbre by Stumpf (1890) and Lichte (1941). These findings are described in Sect. 5.3 and examined further in Sect. 5.4 through psychophysical investigations and interlanguage comparisons.

As with most aspects of timbre, much work on timbre semantics has investigated acoustic musical instruments by means of recorded samples or synthetic emulations. However, talking about instrumental timbre always implicates the acoustic environment in which the instrument is heard. In what ways do the semantics of spaces interact with the semantics of timbre? A preliminary discussion on this important but understudied question is given in Sect. 5.5. Finally in Sect. 5.6, overarching ideas are summarized and new directions for future research are proposed.

Two considerations are necessary before proceeding. First, sound source identification (e.g., this is not a violin) is in itself a type of timbre semantics. The consistent use of onomatopoeia in verbal descriptions of musical and environmental timbres (see Sect. 5.2.1) is one example of identification acting as semantics. In practice, however, timbre semantics is typically defined as *qualia* (this chapter) and sound source perception is studied separately (see McAdams, Chap. 2; Agus, Suied, and Pressnitzer, Chap. 3). Second, in studying timbre semantics as qualia, a distinction will be made between *timbre* as sound quality of complex spectra (this chapter) and *sound quality* as an evaluation of functionality and pleasantness in audio reproduction and industrial sound design contexts (see Lemaitre and Susini, Chap. 9).

5.2 Musical Meaning and the Discourse of Timbre

Listening to a sound (speech, music, environmental events, etc.) involves not only detection-perception of the acoustic signal, but also the interpretation of auditory information (e.g., pitch or the lack thereof, timbre, duration, dynamics). According

to Reybrouck (2013), musical semantics, the processing of meaning emerging from musical auditory information, relies on evolutionarily older mechanisms of meaningfully reacting to nonmusical sound, and

> "… listeners can be conceived as *adaptive* devices, which can build up new semiotic linkages with the sounding world. These linkages can be considered as by-products of both biological and cultural *evolution* and can be helpful in providing coordinative frameworks for achieving diversity of thought, cultural invention, social interaction and optimal coregulation of affect" (pp. 602–603; emphasis added).

Combining previous theoretical accounts of musical semantics with empirical neurobiological evidence, Koelsch (2011) concluded that there are three fundamentally different classes of musical meaning: *extramusical*, *intramusical*, and *musicogenic*. Extramusical meaning arises from the interpretation of musical sound cues through iconic, indexical, and symbolic sign qualities. Iconic qualities resemble qualities of objects and abstract concepts. Indexical meaning emerges from emotion and intention recognition. Symbolic meaning emerges from social and cultural associations. For example, a musical excerpt may sound buzzing, warm, complex, happy, ethnic, patriotic, and so on. Intramusical meaning emerges from the interpretation of structural references between musical units without extramusical associations, such as chord functions during the course of a cadence. Finally, musicogenic refers to meaning that stems from the interpretation of physical, emotional, and self-related responses evoked by musical cues, as opposed to interpreting musical cues per se. A musical performance can thus prompt one to dance, shed tears, or remember a past experience. Within the framework posited by Koelsch (2011), verbal attributes of timbral qualities can generally be thought of as falling into the class of iconic signs (Zacharakis et al. 2014).

5.2.1 Speaking about Sounds: Discourse Strategies

Wake and Asahi (1998) used musical, vocal, and environmental stimuli, and pairs of naïve listeners to study how they describe different types of sounds. Unlike sound experts (i.e., musicians, composers, sound artists, recording engineers, sound and music scholars) the naïve listeners lack a specialized auditory vocabulary. One person in each pair listened to a sound and subsequently described it to their interlocutor, who then had to imagine the described sound and, after listening to the actual stimulus, assess the similarity between the two. The verbalizations used to convey the different sounds were mainly of three types. The first type describes the *perception of the sound itself* using onomatopoeias (i.e., words or vocables considered by convention to phonetically mimic or suggest the sound to which they refer; e.g., chirin-chirin for the sound of a wind bell) or acoustic terminology (e.g., high pitched). The second type describes the *recognition of the sounding situation* using references to the object that made the sound (e.g., a bird) or the action that produced it (e.g., twittering) or other contextual information (e.g., in the morning). The third

type describes the *sound impression* using metaphors and similes (e.g., clear, cool). Wake and Asahi (1998) proposed a model of auditory information processing, according to which recognition and impression are processed either independently (perception then recognition or impression) or sequentially (perception then recognition then impression).

In his empirical ethnographic research on the management of talk about sound between music professionals in the United States, Porcello (2004) identified five strategies that are common to the discourse of timbre among producers and engineers: (1) spoken/sung vocal imitations of timbral characteristics; (2) lexical onomatopoeic metaphors; (3) pure metaphor (i.e., non-onomatopoeic, generally referencing other sensory modalities or abstract concepts); (4) association (citing styles of music, musicians, producers, etc.); (5) evaluation (judgements of aesthetic and emotional value). Thus, a snare drum might sound like /dz:::/ and a muted trombone like wha-wha, a wolf tone on the cello (a persistent beating interaction between string vibrations and sympathetic body resonances) is usually howling and rough or harsh, and a violin tone might sound baroque or like Menuhin or beautiful. In comparison to the taxonomy of Wake and Asahi (1998), Porcello (2004) distinguishes between lexical onomatopoeias and vocal mimicry of nonvocal timbres, including in the latter category nonlexical onomatopoeias, and also considers three types of sound impression descriptions: pure metaphor, association, and evaluation.

Porcello (2004) further advances a distinction between vocal imitations and onomatopoeias on the one hand (which he calls "sonic iconicity") and the pure iconicity of metaphors originating in nonauditory sensory experiences or abstract concepts on the other hand. These, he observes, are usually "codified, especially among musicians and sound engineers," (Porcello 2004, p. 747). Following their investigation of the relation between verbal description and gestural control of piano timbre, Bernays and Traube (2009, p. 207) similarly concluded that "high level performers … have developed over the years of practice … an acute perceptive sensibility to slight sonic variations. This … results in an extensive vocabulary developed to describe the nuances a performer can detect." Furthermore, as noted by Traube (2004), this vocabulary is traditionally communicated from teacher to student in both the musician and sound engineer communities.

Lemaitre and colleagues (2010) analyzed free sortings of environmental sounds made by expert and nonexpert listeners along with scores of source-cause identification confidence and source-cause verbalizations. For the latter, participants were asked to provide nonmetaphorical nouns and verbs to describe the object and action that produced each sound. Participants were also asked to describe what sound properties they considered in grouping different sounds together. They showed that naïve listeners categorized environmental sounds primarily on the basis of source-cause properties. When these could not be identified, nonexpert listeners turned to the timbral properties of the sound, which they described using metaphors or vocal imitations. In contrast, musicians and other expert listeners relied more on timbral characteristics, verbalizing them using metaphors almost exclusively. This finding may offer support to the auditory information processing model proposed by Wake and Asahi (1998), who assert that timbral impression is processed independently of

or following source recognition. It could also help to explain why Porcello's taxonomy of timbre verbalizations, which is derived from the discourse of sound experts, does not include descriptions of the physical cause of a sound, such as those grouped under "sounding situation" by Wake and Asahi (whose taxonomy is based on verbalizations by nonexpert listeners).

Wallmark (2018) conducted a corpus linguistic analysis of verbal descriptions of instrumental timbre across eleven orchestration treatises. The collected verbalizations were categorized according to: (1) affect (emotion and aesthetics); (2) matter (physical weight, size, shape); (3) crossmodal correspondence (borrowed from other senses); (4) mimesis (sonic resemblance); (5) action (physical action, movement); (6) acoustics (auditory terminology); and (7) onomatopoeia (phonetic resemblance). This scheme is very similar to the one suggested by Porcello (2004), whose notion of "pure" metaphor could be seen as encompassing categories (2) to (6). Whereas onomatopoeic words were prevalent among music producers and engineers in Porcello's study, they accounted for a mere 2% of Wallmark's orchestration corpus, driven primarily by a small number of mostly percussion instruments. In fact, certain instruments and instrument families were found to have a systematic effect on verbal description category. For example, the trombone was described more frequently with affect and mimesis than other brass instruments, while the violin, viola, and cello all shared similar descriptive profiles (cf., Saitis et al. 2017). By means of principal components analysis, the seven categories were further reduced to three latent dimensions of musical timbre conceptualization: *material* (loaded positively onto onomatopoeia and matter), *sensory* (crossmodal and acoustics), and *activity* (action and mimesis).

Notwithstanding the diverse metaphorical timbre lexicon in orchestration books, taxonomies of musical instruments and the kinds of sounds they produce are usually based on the nature of the sound-producing material and mechanism. Koechlin (1954–1959; cited in Chiasson et al. 2017, p. 113–114) proposed instead to organize instrument sounds for orchestration purposes on the basis of volume and intensity. Volume is described as an impression of how much space an instrument sound occupies in the auditory scene ("extensity" is used by Chiasson et al. 2017; see also Rich 1916). Based on an inverse relationship between volume and intensity, Koechlin (cited in Chiasson et al. 2017) further proposed a third attribute of density versus transparency: a musical sound is dense when it is loud but with a small volume, and it is transparent when it has a large volume but low intensity. There is evidence that in the later Middle Ages it was typical to think of musical instruments in terms of volume of sound (Bowles 1954). In orchestras, and for other musical events, instruments with a big, loud sound (*haut* in French) would be grouped together against those with a small, soft sound (*bas*).

Schaeffer (1966) offered a typo-morphology of "sonorous objects" (i.e., sounds experienced by attending to their intrinsic acoustic properties and not to their physical cause) based on sustainment (*facture* in French) and mass. Sustainment refers to the overall envelope of the sound and mass is described as "the quality through which sound installs itself … in the pitch field" (Schaeffer 1966, p. 412), which appears similar to Koechlin's notion of volume. Interestingly, Koechlin and

Schaeffer were both French, shared a composition background, and published their typologies within 10 years of each other. Mass extends the concept of pitch in pure tones (i.e., single frequencies) and tonal sounds (i.e., nonnoisy) to include sounds with fluctuating or indeterminate pitch (e.g., cymbals, white noise). Each mass has a particular timbre associated with it—a set of "secondary" qualities that are either nonexistent (pure tones) or exist at varying degrees from being dissociated (musical notes) to indistinguishable (white noise) from mass. Given the definition of sonorous objects, Schaeffer's timbre is free from any source-cause associations and is thus situated clearly in the realm of quality as opposed to identity (Siedenburg, Saitis, and McAdams, Chap. 1).

In tonal sounds, Schaeffer argues, mass can be low or high (in terms of location in the pitch field) and thick or thin (in terms of extensity in the pitch field); timbre can be dark or light (location), ample or narrow (extensity), and rich or poor (in relation to the intensity of the mass). The latter appears closely related to Koechlin's notion of density as they both describe a mass or volume, respectively, in relation to its intensity. In Smalley's (1997) *Theory of Spectromorphology*, which has its origins in Schaeffer's ideas, pitch field is replaced by "spectral space". The latter is described in terms of emptiness versus plenitude (whether sound occupies the whole space or smaller regions) and of diffuseness versus concentration (whether sound is spread throughout the space or concentrated in smaller regions). Like Koechlin and Schaeffer, Smalley also relies on extra-auditory concepts to serve as discourse for an organization of auditory material that focuses on *intrinsic* features of the sound independently of its source.

5.2.2 Metaphors We Listen With

Wallmark (2014) argues that the metaphorical description of timbre is not simply a matter of linguistic convention, and what Porcello singles out as "pure metaphor" is central to the process of conceptualizing timbre by allowing the listener to communicate subtle acoustic variations in terms of other more commonly shared sensory experiences (nonauditory or auditory-onomatopoeic) and abstract concepts. De Ceuster (2016) points out that timbre has been described with metaphors based on experiences since the presumed birth of the term in the mid-eighteenth century (Dolan 2013). Jean-Jacques Rousseau's "Tymbre" entry in Diderot and D'Alembert's *Encyclopedié* reads:

> A sound's *tymbre* describes its harshness or softness, its dullness or brightness. Soft sounds, like those of a flute, ordinarily have little harshness; bright sounds are often harsh, like those of the *vielle* [medieval ancestor to the modern violin] or the oboe. There are even instruments, such as the harpsichord, which are both dull and harsh at the same time; this is the worst *tymbre*. The beautiful *tymbre* is that which combines softness with brightness of sound; the violin is an example (cited and translated in Dolan 2013, p. 56).

Building on accounts of ecological and embodied cognition, Wallmark (2014) proposes an embodied theory of timbre whereby metaphorical descriptions are indexes of conceptual representations grounded in perception and action. They can be grouped into three categories based on the conceptual metaphors (Lakoff and Johnson 2003): (1) *instruments are voices* (e.g., nasal, howling, open); (2) *sound is material* (e.g., bell-like, metallic, hollow, velvety); and (3) *noise is friction* (e.g., harsh, rough) (cf., Wallmark 2018). The *sound is material* metaphor can be broken down into four subtypes: (2a) naming the source directly (e.g., a bell-like sound); (2b) referencing the physical qualities of the source (e.g., a metallic sounding cymbal); (2c) blending physical and connotative elements of source and sound (e.g., a hollow bassoon); and (2d) referencing physical qualities of unrelated objects (e.g., velvety strings).

Why are instruments voices? Consider phonemes. They can be categorized based on distinctive features associated with the physiology of voice production and articulation that are generally inherent in all languages (Jakobson and Halle 1971). Phonemes can be nasal (coupling between the oral and nasal cavities) or oral (no coupling); compact (spectral dominance of a single central formant when the mouth is wide open) versus diffuse; strident (airstream forced to strike the teeth, high-intensity fricative noise) versus mellow; tense or lax (greater versus lesser deformation of the vocal tract); grave (larger and less compartmented mouth cavity volume, concentration of energy in the lower register) versus acute; flat (smaller lip opening but larger between-lip area, weakening of upper frequencies) or nonflat; and sharp (dilated pharyngeal pass, strengthening of upper frequencies) versus nonsharp. In singing, a low versus high laryngeal position produces a covered versus open vocal timbre or simply a low versus high pitch (Miller 1986). In medicine, hoarse is used to describe the presence of high frequency noise components accompanied by decreased harmonics in the voice due to laryngeal diseases (Isshiki et al. 1969). Attributes such as howling, throaty, hissing, and breathy eventually refer to the associated vocal source or as Sundberg (2013, p. 88) puts it: "The perception of voice seems to be influenced by familiarity with one's own voice production." This observation echoes the motor theory of speech perception, which considers that the latter is based on articulatory motor representations (Liberman and Mattingly 1985) and which Wallmark (2014) extends to a motor theory of all timbre perception in preparation for the *instruments are voices* metaphor.

Albersheim (1939) drew analogies between vowels and colors to propose a geometrical model of *acoustic color* (*Akustischer Farbenkörper* in German) in the form of a cylinder. Its height and radius represented variation in color brightness and saturation, respectively. Changes in color hue were mapped onto a helical line along the surface of the cylinder. Slawson (1985) developed a theory of *sound color*, which he defined as the static spectral envelope of a sound, as opposed to its temporally varied spectrum, based on the distinctive phoneme features of *openness*, *acuteness*, and *laxness*, and their relation to the pitch-invariant formant structure of vowels. The term "openness" was chosen as a perceptually more intuitive depiction of compactness. More open vowels have a higher first formant, while acuteness increases with increasing frequency of the second resonance. Lax vowels have a lower total energy that is less spread out over the spectrum. A fourth dimension was

termed *smallness*: the lower the first and second formants are, the smaller the vowel. Schumann (1929), Reuter (1997), and Lembke and McAdams (2015), among others, have discussed the vowel-like pitch-invariant formant structure of many (but not all) musical instruments and its role in timbre perception.

In other words, timbre can be experienced with reference to the human and non-human voice—a conceptualization already evident in Helmholtz's (1877) choice to synthesize vowel-like sounds for his *Klangfarbe* experiments and in Schilling's definition of the German term as "denoting mostly the accidental properties of a voice" (Schilling 1840, p. 647; cited in Kursell 2013). Timbre can also be experienced as a material object that can be seen, touched, and even tasted. Furthermore, noise-like timbres (e.g., excessive high-frequency content, inharmonicity, flat spectrum) can be understood in terms of frictional material interaction. Very similar metaphorical conceptualizations can be found in verbalizations of other perceptual aspects of sound, such as pitch and loudness (Eitan and Rothschild 2011; Saitis et al. 2017). In general, conceptual metaphors of timbre and auditory semantics may originate in more universal neural processes and structures beyond auditory cognition (cf., Gallese and Lakoff 2005; Walsh 2013).

5.3 Semantic Spaces of Timbre

Scientific interest in timbre semantics started as early as the experimental exploration of timbre itself (Helmholtz 1877; Stumpf 1890). Stumpf (1890) proposed that the various verbal attributes of timbre can be summarized on the basis of semantic proximities by three pairs of opposites: dark–bright (*dunkel–hell* in German), soft–rough (*weich–rauch*), and full–empty (*voll–leer*). Hereafter, these symbols will be used: '–'to indicate antonyms and '/' to indicate synonyms. Discussing a set of psychoacoustic experiments, Lichte (1941) concluded that brightness, roughness, and fullness, as defined by Helmholtz, form independent attributes of sound in addition to pitch and loudness. More systematic efforts to understand the complex multivariate character of timbre semantics were made possible by methodological tools such as factor analysis of ratings on verbal scales that were developed in the 1950s and were first applied to timbre by Solomon (1958) (Sect. 5.3.1). Studies using multidimensional scaling of adjective dissimilarities and psycholinguistic analyses of verbalization tasks have provided additional insight regarding particular aspects that contribute to the semantic description of timbre (Sect. 5.3.2).

5.3.1 Semantic Scales: Methodology and Main Results

Osgood (1952) developed a quantitative method for measuring meaning based on the use of multiple verbal scales. Each scale was defined by pairs of antonymic descriptive adjectives, such as dark–bright and smooth–rough, which he termed *semantic differentials*. The method postulates a semantic space within which the

operational meaning of a given concept can be specified. This "space" is physically thought of as a Euclidean spatial configuration of unknown dimensionality; each semantic differential represents an experiential continuum, a straight line function that passes through the origin of this space. Many different continua are psychologically equivalent and, hence, may be represented by a single latent dimension. The minimum number of such (orthogonal) dimensions can be recovered by means of factor analysis and used to define the semantic space of the concept. Ratings on semantic scales also can be analyzed with principal component analysis or, when appropriately reorganized (e.g., dissimilarity distances, cross-correlations), clustering or multidimensional scaling techniques. The reliability and validity of the semantic differential model depend on a number of methodological considerations (Susini et al. 2012; Saitis et al. 2015). For example, it is important to use verbal scales that are psychologically relevant and commonly interpreted across all raters. And even then, the derived factors are not always easy to interpret with respect to the scales and/or raters.

Solomon (1958) had sonar technicians rate recordings of passive sonar sounds on bipolar scales comprising perceptual attributes (e.g., smooth–rough) that are typically used by experienced sonar operators but also a large number of aesthetic–evaluative adjectives (e.g., beautiful–ugly). A seven-factor solution was obtained, which accounted for only 42% of the total variance in the collected ratings. The first and most salient factor (15%) indicated a "magnitude" dimension, explained by such scales as heavy–light and large–small. The third factor (6%) was identified by such words as clear, definite, and obvious, and labeled as "clarity." The remaining factors were essentially aesthetic–evaluative, probably because many such differentials were used in the design of the study. Generally speaking, such scales are likely to be of little help when one tries to access perceptual representations through language, as affective reactions tend to be less stable across individuals than sensory descriptions.

Jost (1967; cited in Webster et al. 1970, p. 481–483) carried out a semantic differential study of four clarinet notes played at six different loudness levels and found two salient factors of density and volume. However, these appeared to correlate with stimuli variations in pitch and loudness, respectively. Von Bismarck (1974a) sought to address three important issues with applying semantic differentials to the study of timbre semantics: selecting verbal attributes that are perceptually relevant, normalizing sound stimuli for pitch and loudness, and psychophysically explaining the extracted factors. Sound stimuli comprised synthetic steady-state signals of two types: vowel-like and instrument-like harmonic complexes, and consonant-like noises. These had spectral envelopes varying systematically along three parameters: frequency location of overall energy concentration, slope of the envelope, and frequency location of energy concentrations within the spectrum. All sounds were normalized in loudness by means of perceptual adjustment to a given reference. The harmonic complexes were further equalized in fundamental frequency at 200 Hz. Sixty-nine differential scales were initially rated for suitability to describe timbre on a scale from "very unsuitable" to "highly suitable". From thirty-five scales with the highest mean suitability ratings, seven scales deemed

synonymous were further discarded. The scales soft–loud and low–high were included to test the effectiveness of loudness and pitch normalization, respectively.

Factor analysis of ratings by a group of musicians and another group of nonmusicians yielded similar, although not identical, four-factor solutions that explained more than 80% of the variance in the data. The four factors were defined by the differentials dull–sharp, compact–scattered, full–empty, and colorful–colorless. Although participants were instructed to ignore pitch and loudness as much as possible, ratings on the soft–loud and low–high scales were highly correlated with those on dull–sharp and dark–bright, respectively. This illustrates how the same word can have different connotations in different contexts. Even when sounds were equalized in loudness and pitch, listeners still used related attributes to describe other impressions. In agreement with the view that verbal attributes of timbre are "codified" among musically trained listeners (see Sect. 5.2.1), ratings from nonmusicians were more scattered than those of musicians. Prompted by the finding that the dull–sharp factor explained almost half of the total variance in the data, von Bismarck (1974b) confirmed in subsequent psychoacoustic experiments that a dull–sharp scale had desirable measurement properties (e.g., doubling, halving) and concluded that sharpness may represent an attribute of sounds distinguishable from pitch and loudness.

Von Bismarck's is arguably the first comprehensive investigation of timbre semantics, markedly improving upon the earlier studies, but certain aspects have been questioned. For example, aesthetic-evaluative and affective scales were still used. In addition, the preliminary assessment of whether or not a scale was suitable for describing timbre was carried out in an undefined context, without presentation of the timbres to be described, while further discarding of scales was based on an arbitrary judgement of word synonymy. Perhaps more importantly, a *semantic* issue with the semantic differentials is the assumption of bipolarity that underlies the model (Heise 1969; Susini et al. 2012). Are soft–loud and dark–bright always true semantic contrasts? Is sharp the true semantic opposite of dull when talking about timbre?

One way to address potential biases associated with prescribing antonymic relationships between adjectives is to use adjective checklists. These were used extensively in musical affect research up until the late 1950s (for a review, see Radocy and Boyle 2012) but have largely been replaced by semantic scales. Similarly to von Bismarck (1974a), Pratt and Doak (1976) attempted to first find verbal scales suitable for describing timbre. An initial list of 19 "commonly used" adjectives was reduced to seven items by means of a checklist task. By (arbitrarily) discarding synonyms and "not very useful" words, the list was further reduced to the attributes brilliant, rich, and warm; dull, pure, and cold, respectively, were (arbitrarily) chosen as opposites to form semantic differentials. From ratings of different synthesized harmonic spectra on the three scales, it was found that the former were most consistently discriminated by the brilliant–dull scale.

In a separate study (Abeles 1979), each of twenty-four recorded isolated clarinet notes was presented three times, each time with five adjectives randomly selected from a list of forty words. Three independent groups of clarinetists, nonclarinetist

musicians, and nonmusicians were asked to check as many adjectives as they thought best described the timbre of each note. Factor analysis of the combined data across the three listener groups (no individual group analyses were reported) yielded a three-factor solution of shape (round/centered–pinched/thin), density (clear/brilliant–fuzzy/airy), and depth (resonant/rich/projecting; no negatively loaded adjectives were reported). Edwards (1978) and Pratt and Bowsher (1978) found a very similar set of semantic dimensions for the trombone (see Sect. 5.3.2), which is also a wind instrument.

Kendall and Carterette (1993a, b) attempted a systematic use of verbal scales bounded by an attribute (e.g., bright) and its negation (e.g., not bright), which was termed the verbal attribute magnitude estimation (VAME) method because the task for the rater is to assess how much of a single attribute is possessed by a stimulus. Unipolar scales offer a way of dealing with polysemy and nonexact antonymy within the semantic differential framework. Accordingly, antonymic or synonymic relationships can be assessed a posteriori through negative or positive correlations between ratings on different unipolar scales.

A first pair of experiments (Kendall and Carterette 1993a) sought to explore the extent to which von Bismarck's (1974a) semantic space, which had resulted from synthetic vowel-like sounds, is relevant in describing the timbre of natural (recorded) instrument sounds. The stimuli comprised dyads of wind instrument notes produced in unison, and they were rated by nonmusicians on eight VAME scales that loaded high on the first (hard, sharp, loud, complex), second (compact, pure), and fourth (dim, heavy) von Bismarck factors. Analyses converged to a two-dimensional solution accounting for almost 98% of the variance; however, it mapped weakly onto a two-dimensional similarity space of the same dyads, prompting the authors to conclude that von Bismarck's scales were less relevant in rating natural versus synthetic timbres. In subsequent experiments (Kendall and Carterette 1993b), the same stimuli were rated by musicians on twenty-one VAME scales induced from adjectives describing instrumental timbre in an orchestration book. Similar analyses resulted in a two-dimensional semantic space of nasal–rich and brilliant–reedy adjectives, which explained 96% of the data variance and corresponded more strongly with similarity ratings.

The work of Kendall and Carterette constitutes the first systematic effort to combine semantic ratings with similarity judgements to directly examine the relationship between the perception of timbre and its verbal communication. In this context, these results illustrate that the validity of a semantic space as a perceptual construct depends on a number of issues such as the type of sounds tested, the type of verbal scales used, and the musical background of raters. Especially when considering differences in how musically experienced versus naïve listeners conceptualize timbral qualities (see Sect. 5.2.1), it is plausible that the better results obtained in the second set of experiments (Kendall and Carterette 1993b) were not only a result of selecting more relevant semantic scales but also of recruiting musically trained listeners. Von Bismarck (1974a) and Abeles (1979) both found that in rating the same sounds on the same semantic scales musicians were generally more consistent than nonmusicians.

The nasal–rich dimension of Kendall and Carterette (1993b) summarizes descriptions of nasal/edgy/brittle/weak/light versus rich/round/strong/full. It thus appears to correspond to the shape factor found by Abeles (1979) for clarinet sounds. Abele's density factor seems to be closer to Kendall and Carterette's brilliant–reedy dimension, which relates to impressions of brilliant/crisp/pure versus reedy/fused/warm/complex. In some agreement with these two studies, Nykänen et al. (2009) found four semantic dimensions for a set of saxophone notes, namely, warm/soft, back vowel-like sounding, sharp/rough, and front vowel-like sounding. Considering that back versus front vowels tend to be perceived as dark/round versus bright/thin (Jakobson and Halle 1971), two broader dimensions alluding to shape (warm/soft–sharp/rough) and density (dark/round–bright/thin) may be hypothesized. It therefore appears that most wind instrument timbres can be positioned within a common semantic space. How does this space adapt when sounds from other instrument families are included? Kendall et al. (1999) found that adding a violin note did not affect the semantic space; however, its mapping onto the corresponding perceptual space was less robust.

Using fifteen VAME scales, Disley et al. (2006) obtained a four-dimensional semantic space for twelve orchestral instrument notes of same pitch: bright/thin/harsh/clear–dull/warm/gentle/rich, pure/percussive/ringing–nasal, metallic–wooden, and evolving. Ratings remained fairly consistent across multiple repetitions. Several listeners noted that they used metallic and wooden to describe the recognized material of the instrument rather than a timbral quality, which would explain the loading of these scales on a separate component (one could expect metallic to correlate with bright/harsh and wooden with warm/rich). Similarly, the presence of a fourth dimension solely defined by evolving is likely due to reported listener difficulties in understanding what it meant, although the moderate loading of rich on the same component might indicate a spectral flux type of dimension (see Sect. 5.4.1).

Using a more diverse set of stimuli (twenty-three isolated notes from acoustic, electromechanical, and electronic instruments, with different pitches), twice as many VAME scales, and analyses that accounted for nonlinear relationships between the semantic variables, Zacharakis et al. (2014) arrived at a three-dimensional space summarized as luminance (brilliant/sharp–deep), texture (soft/rounded/warm–rough/harsh), and mass (dense/rich/full/thick–light). This space was largely similar across two independent groups of native English and Greek-speaking listeners (musically experienced). Two different groups of English and Greek listeners provided dissimilarity ratings of the same set of sounds and the respective three-dimensional spaces derived from multidimensional scaling (MDS) were also found to be highly similar. Comparisons between the semantic and perceptual spaces illustrated strong correlations of luminance and texture, on the one hand, and texture with two of the three MDS dimensions on the other, independent of native language. Texture appeared to contribute to all three MDS dimensions. Results for mass were less conclusive. Moderately similar results have been obtained for an even larger set of musical sounds (forty-two sustained orchestral instrument notes of the same pitch) using bipolar scales and different methods of analysis (Elliott et al. 2013). A

strong, but not one-to-one, correspondence between semantic and perceptual dimensions of timbre had previously been shown by Samoylenko et al. (1996) and Faure (2000), who collected free verbalizations during dissimilarity ratings.

5.3.2 *Further Findings from Verbalization and Verbal Dissimilarity Tasks*

Verbalization tasks, where participants are asked to describe timbral impressions in their own words, offer an alternative means of exploring the semantics of timbre. They can be used as a standalone method (Traube 2004; Saitis et al. 2017), or to complement a preceding task (e.g., describe timbral differences during pairwise sound comparisons: Samoylenko et al. 1996; Faure 2000), or to help design further experiments (e.g., extract relevant adjectives for anchoring semantic scales: Rioux and Västfjäll 2001; Grill 2012). Verbalization can be *free*, in the sense that very general open-ended questions are asked and no restriction is imposed on the format of the response, or *constrained*, where questions are more structured and responses must conform to a certain format. A qualitative method of deriving semantic proximities from verbalization data relies on theoretical assumptions about cognitive categories and their relation to natural language (Dubois 2000). From what is being said and how it is being said, relevant inferences can be derived about how people conceptualize sensory experiences (semantic level) and can be further correlated with physical parameters (perceptual level).

Traube (2004) asked classical guitar players to freely describe the timbre of their instrument in relation to how it is produced. The ten most commonly used adjectives were dry, nasal, thin, metallic, bright, round, warm, thick, velvety, dark. By combining linguistic analysis and acoustic measurements, a strong correspondence was found between the plucking position along the string, the frequency location of the generated comb filter formants, and the use of adjectives describing vowel-like timbre for similarly located vocal tract formants, which echoes the *instruments are voices* metaphor (Sect. 5.2.2). As an example, adding the nasal and oral cavities (nasal voice) causes a broadening of all vocal tract formant bandwidths and a flattening of spectral peaks in the range 300–2500 Hz (Jakobson and Halle 1971; Mores 2011). Traube found that guitars sound more nasal/bright/dry when plucked closer to the bridge because of analogous spectral effects. Conversely, plucking between the sound hole and the fingerboard produces spectra similar to nonnasal vowels and is perceived as more velvety/dark/round.

Rioux and Västfjäll (2001) and Saitis et al. (2017) have provided further evidence that, while perceived variations in how an instrument sounds rely on variations in style and the expertise of different musicians (Saitis et al. 2012), the broader semantic categories emerging from verbal descriptions remain common across diverse musical profiles, thus reflecting a shared perception of acoustic information patterns. Importantly, the verbal data revealed that vibrations from the violin body and

the bowed string (via the bow) are used as extra-auditory cues that not only help to better control the played sound but also contribute to its perceived qualities. For example, recent research on the evaluation of piano and violin quality has revealed that an increase in the vibrations felt at the fingertips of pianists and the left hand of violinists can lead to an increase in perceived sound loudness and richness (Saitis et al. 2018). Also, impressions like bright and rich mostly refer to the sustained part of a note, while words like soft tend to describe qualities of transients (cf., Brent 2010; Bell 2015).

An example of constrained verbalization is the *repertory grid technique*. Listeners form bipolar constructs (i.e., antonymic pairs of adjectives) by articulating the difference between two sounds taken from a larger pool that is relevant to the aims of the task at hand (referred to as elements). Alternatively, three sounds are presented and listeners are first invited to select the least similar one and subsequently to verbally explain their grouping. Finally, listeners are asked to rate all elements on each new construct. The resulting grid of constructs and elements, essentially semantic differential ratings, can then be evaluated with factor analytical, clustering, or multidimensional scaling techniques. Using this method, Grill (2012) found an expanded semantic space for electroacoustic "textures", which combined dimensions pertinent mostly to such sounds (ordered–chaotic or coherent–erratic, homogeneous–heterogeneous or uniform–differentiated) with dimensions commonly found for voices and instruments (high–low or bright–dull, smooth–coarse or soft–raspy, tonal–noisy).

A semantic space can also be derived quantitatively through MDS of pairwise distances in a list of adjectives. Moravec and Štěpánek (2003) initially asked conductors, composers, engineers, teachers, and musicians (three groups of bowed-string, wind, and keyboard performers) to provide words they typically use to describe the timbre of any musical instrument. The four most frequently mentioned words across all respondents (sharp, gloomy, soft, clear) were also among the four most frequently used in each of the three musician groups. Still, some within-group preferences were observed. Bowed-string players used sweet and warm more frequently than both keyboard and wind performers. Similarly, narrow was much more popular with wind musicians. The thirty most frequently reported adjectives were subjected to dissimilarity ratings (Moravec and Štěpánek 2005) and MDS identified three dimensions closely matching luminance, texture, and mass (Zacharakis et al. 2014), namely, gloomy/dark–clear/bright, harsh/rough–delicate, and full/wide–narrow, respectively.

Edwards (1978) collected a corpus of free verbalizations of trombone sound quality through interviews and a postal survey of over 300 trombone performers. A subset of the verbal data was arranged in terms of semantic similarity by the author himself on the basis of proximities identified in the corpus. This kind of dissimilarity matrix was subsequently subjected to MDS. With respect to timbre, two dimensions of small–wide and dull/round–clear/square emerged. A different subset of the verbalizations indicated a third timbral aspect referring to "amount" and "carrying" or "penetrating" properties of sound. These seem to generally agree with the findings of Abeles (1979), Kendall and Carterette (1993b), and Nykänen et al. (2009).

In another trombone study, Pratt and Bowsher (1978) selected the scales compact–scattered, dull–bright, and not penetrating–penetrating to correspond to Edwards' three dimensions. It was found that the second and third scales were good discriminators of trombone timbres but compact–scattered was not. Indeed, the latter may not indicate size, which is the label Edwards gave to his first dimension, but may indicate density (see Sect. 5.2.1).

Fritz et al. (2012) had violinists arrange sixty-one adjectives for violin timbre on a two-dimensional grid (EXCEL), so that words with similar meanings lay close together and those with different meanings lay far apart. The collected grids were converted into dissimilarity matrices using a custom distance metric between two cells (see p. 793 in Fritz et al. 2012) and MDS yielded three dimensions: warm/rich/mellow versus metallic/cold/harsh (richness; texture), bright/responsive/lively versus muted/dull/dead (resonance; projection), and even/soft/light versus brash/rough/raspy (texture; clarity). The parenthetical terms potentially correspond to semantic categories from the cognitive model proposed by Saitis et al. (2017). In both studies, violinists used words like lively, responsive, ringing, and even bright to describe the "amount of sound" perceived "under the ear" (resonance) and in relation to spatial attributes (projection). Differences between the labels of the found semantic dimensions for trombone (wind) and violin (bowed string) timbre seem to generally agree with those observed by Moravec and Štěpánek (2003).

In the piano study of Bernays and Traube (2011), fourteen adjectives extracted from spontaneous verbalizations yielded a four-dimensional MDS space. Based on the first two dimensions (78% of the total variance explained) and additional hierarchical clustering, five adjectives were proposed to best represent a semantic space for piano timbre: bright, dry, dark, round, and velvety. Lavoie (2013) performed MDS on dissimilarities between adjectives describing classical guitar timbre. In agreement with Traube (2004), a dimension of velvety/dark–bright/dry was obtained, related to whether the string is plucked between the sound hole and the fingerboard versus closer to the bridge (like nasal); a dimension of round/bright–dull/thin was associated with sound resonance and projection. It is worth noting the highly similar labels of the reported semantic spaces across the two instruments. To a certain extent, this may reflect shared conceptualization structures between musicians whose primary instrument produces impulsive string sounds. On the other hand, given that all three studies were conducted with musicians from the Montreal region, it may be that these results mirror a verbal tradition specific to that geographic location, possibly due to a strong influence by one or more particular teachers in the area (cf., Saitis et al. 2017).

5.4 Semantic Spaces of Timbre Revisited

Despite important methodological differences, the findings described in the previous section show remarkable similarities when certain classes of timbres (e.g., individual instrument families) and mixed sets across distinct classes (e.g., various

orchestral instruments) are rated on verbal scales, but similarities are also evident when verbal descriptions are collected in the absence of sound examples (e.g., verbalization tasks, adjective dissimilarity ratings). The most salient dimensions can be interpreted broadly in terms of brightness/sharpness (or luminance), roughness/harshness (or texture), and fullness/richness (or mass). The boundaries between these dimensions are sometimes blurred, while different types of timbres or scenarios of timbre perception evoke semantic dimensions that are specific to each case (e.g., nasality, resonance/projection, tonalness–noisiness, compact–scattered). Generally, no striking differences between expert and naïve listeners are observed in terms of semantic dimensions, although the former tend to be more consistent in their perceptions than the latter. In this section, the identified semantic dimensions of timbre are examined further through looking at their acoustic correlates (Sect. 5.4.1) and comparisons between different languages and cultures (Sect. 5.4.2).

5.4.1 Acoustic Correlates

Impressions of brightness in timbre perception are typically found correlated with the spectral centroid, a scalar descriptor defined as the amplitude-weighted mean frequency of the spectrum (Siedenburg, Saitis, and McAdams, Chap. 1; Caetano, Saitis, and Siedenburg, Chap. 11), which indicates the midpoint of the spectral energy distribution (cf., Lichte 1941). In other words, frequency shifts in spectral envelope maxima are systematically perceived as changes in brightness. The spectral centroid is typically found correlated with one of the dimensions (usually of three) that describe timbre dissimilarity spaces. A higher proportion of high-frequency energy also characterizes brightness in timbral mixtures arising from multitrack recorded music, although the absence of high pitch in such stimuli rendered them as less bright (Alluri and Toiviainen 2010). This is because frequency shifts in pitch, too, are systematically perceived as changes in brightness (Cousineau et al. 2014; Walker 2016). The sharpness factor in von Bismark's (1974a) study (dull–sharp, soft–hard, dark–bright) was also strongly related to the frequency position of the overall energy concentration of the spectrum, with sharper/harder/brighter sounds having more energy in higher frequency bands. Similarly, Bloothooft and Plomp (1988) observed that verbal attributes of stationary sung vowels related to sharpness (including sharp–dull, shrill–deep, metallic–velvety, angular–round, and cold–warm) referred primarily to differences in spectral slope between the vowels. Acute (i.e., sharp) phonemes are also characterized by a concentration of energy in the higher frequencies of the spectrum (Jakobson and Halle 1971; Slawson 1985).

A model for estimating sharpness, originally proposed by von Bismarck (1974b), calculates the midpoint of the weighted specific loudness values in critical bands (Fastl and Zwicker 2007). Critical bands correspond to equal distances along the basilar membrane and represent the frequency bands into which the acoustic signal is divided by the cochlea. Grill (2012) found a strong correlation between bright–dull electroacoustic textural sounds and the sharpness model, which is consistent

with the origin of the latter in psychoacoustic experiments with wideband noise spectra. However, Almeida et al. (2017) showed that the sharpness model insufficiently predicted brightness scaling data for tonal sounds. Marozeau and de Cheveigné (2007) proposed a spectral centroid formula based on the same concept of weighted partial loudness in critical bands, which better modeled the brightness dimension of dissimilarity ratings and was less sensitive to pitch variation compared to the classic spectral centroid descriptor.

Yet another verbal attribute that has been associated with spectral energy distribution is nasality. Etymologically, nasality describes the kind of vocal sound that results from coupling the oral and nasal cavities (Sects. 5.2.2 and 5.3.2). However, it is sometimes used to describe the reinforcement of energy in higher frequencies at the expense of lower partials (Garnier et al. 2007; Mores 2011). In violin acoustics, nasality is generally associated with a strong frequency response in the vicinity of 1.5 kHz (Fritz et al. 2012). Kendall and Carterette (1993b) found that nasal versus rich wind instrument sounds had more energy versus less energy, respectively, in the upper harmonics, with rich timbres combining a low spectral centroid with increased variations of the spectrum over time. Sounds with a high versus a low spectral centroid and spectral variation were perceived as reedy versus brilliant, respectively. Adding a violin note in a set of wind instrument timbres confirmed a strong link between nasality and the spectral centroid, but rich and brilliant were correlated only with spectral variation and only to some modest degree (Kendall et al. 1999). Helmholtz (1877) had originally associated the nasality percept specifically with increased energy in odd numbered upper harmonics, but this hypothesis remains unexplored.

Are timbral brightness and sharpness the same percept? Both of them relate to spectral distribution of energy, and most of the related studies seem to suggest at least partial similarities, but there is still no definite answer to this question. Štěpánek (2006) suggested that a sharp timbre is one that is both bright and rough. However, semantic studies of percussive timbre reveal two independent dimensions of brightness and sharpness/hardness (Brent 2010; Bell 2015). Brighter percussive timbres appear associated with higher spectral centroid values during attack time, while sharp/hard relates to attack time itself (i.e., sharper/harder percussive sounds feature shorter attacks). Attack time refers to the time needed by spectral components to stabilize into nearly periodic oscillations, and it is known to perceptually distinguish impulsive from sustained sounds (McAdams, Chap. 2). Furthermore, concerning brightness, there seems to exist a certain amount of interdependency with fullness. Sounds that are described as thick, dense, or rich are also described as deep or less bright and brilliant, while nasality combines high-frequency energy with low spectral spread and variability. The acoustic analyses of Marozeau and de Cheveigné (2007) and Zacharakis et al. (2015) suggest that brightness may not only relate to spectral energy distribution but also to spectral detail.

To further complicate things, a number of studies based on verbalizations that were collected either directly from musicians or through books and magazines of music revealed a semantic dimension of timbre associated with a resonant and ringing but also bright and brilliant sound that can project (Sect. 5.3.2). This suggests an

aspect of timbre that is primarily relevant to playing an instrument and is associated with assessing how well its sound is transmitted across the performance space. It also suggests an interaction between perceived sound strength and timbral brightness. Based on sound power measurements and audio content analysis of single notes recorded at pianissimo and fortissimo across a large set of standard orchestral instruments (including some of their baroque and classical precursors), Weinzierl et al. (2018b) were able to show that the intended dynamic strength of an instrument can be identified as reliably by sound power as by combining several dimensions of timbral information. Indeed, the most important timbral cue in this context was found to be spectral skewness (Caetano, Saitis, and Siedenburg, Chap. 11) with a left-skewed spectral shape (i.e., a shift of the peak energy distribution toward higher frequencies) indicating high dynamic strength.

Helmholtz (1877) claimed that the sensation of roughness arises from the increasingly dissonant (unpleasant) sounding intervals formed between higher adjacent partials above the sixth harmonic. Empirical data from Lichte (1941) and later Schneider (1997) support this view, which has also lent itself to theories of musical tension (McAdams, Chap. 8). However, Stumpf (1898) disagreed with Helmholtz and provided examples of dissonant chords that were judged as not rough, highlighting a difference between *musical* dissonance and *sensory* dissonance. More recent evidence also suggests that roughness (expressing sensory dissonance) and musical dissonance may constitute distinct percepts (McDermott et al. 2010; Bowling et al. 2018). Physiologically, impressions of roughness and/or sensory dissonance can be linked to the inability of the cochlea to resolve frequency pair inputs whose interval is smaller than the critical band, causing a periodic "tickling" of the basilar membrane (Helmholtz 1877; Vassilakis and Kendall 2010).

Further psychophysical experiments have linked roughness to envelope fluctuations within a critical band produced by amplitude-modulation frequencies in the region of about 15–300 Hz (Fastl and Zwicker 2007; Vassilakis and Kendall 2010). For a given amplitude spectrum and a given modulation depth, modulations with an abrupt rise and a slow decay have been shown to produce more roughness than modulations with a slow rise and an abrupt decay (Pressnitzer and McAdams 1999). For electroacoustic sounds, the effect of sudden changes in loudness over broad frequency ranges is described as coarse and raspy (Grill 2012). Existing psychoacoustic models estimate roughness using excitation envelopes (Daniel and Weber 1997) or excitation-level differences (Fastl and Zwicker 2007) produced by amplitude modulation in critical bands. Nykänen et al. (2009) found that both models of sharpness (von Bismarck 1974b) and roughness (Daniel and Weber 1997) contributed to predictions of roughness of saxophone sound, but sharpness was a much more important contributor. However, and as noted already, these models were originally designed based on experiments with wideband noise spectra and thus may not be applicable for more natural and tonal sounds like those made by a saxophone (or any musical instrument for that matter).

Sounds perceived as rough are also described as harsh—ratings on the latter are typically found correlated with ratings on the former. However, acoustic analyses tend to associate harshness mainly with too much high-frequency energy (i.e.,

unpleasant). This is also evident in psycholinguistic studies of violin timbre (Fritz et al. 2012; Saitis et al. 2013) and voice quality (Garnier et al. 2007). Such descriptions include strident, shrill, piercing, harsh, and even nasal. Note that an implicit connection of roughness to energy in higher frequencies is also claimed by Helmholtz's hypothesis. Zacharakis et al. (2014, 2015) found that sounds with stronger high partials were described as rough or harsh and the opposite as rounded or soft and, to a lesser extent, as bright or sharp. They went on to suggest that spectral energy distribution is manifested primarily in descriptions of texture and not of brightness, which also relied on spectral detail. Rozé et al. (2017) showed that inappropriate bowing and posture coordination in cello performances resulted in energy transfer toward higher frequency harmonics, a decrease in attack time, and an increase in amplitude fluctuation of individual harmonics; this kind of timbre was perceived as harsh and shrill. Under optimal playing conditions, cello sounds were described as round.

A concept related to sensory dissonance, but distinct from roughness, is that of noisiness versus tonalness. The latter signifies the perception of strong stationary and near-periodic spectral components. As such, it has a close relation to pitch patterns. In this case, the timbre tends to be described as pure, clear or clean, and even bright. When random transients dominate the spectrum, the timbre tends to be described as noisy or blurry and messy. A dimension of tonal–noisy has been found for different types of timbres, including electroacoustic sounds (Sect. 5.3). However, specifically in bowed-string instruments, audible noise can still be present even when a clear and steady tonal component is established (Štěpánek 2006; Saitis et al. 2017). One source of such noise, sometimes described as rustle, is the self-excitation of subfundamental harmonics, particularly in the upper register (Štěpánek and Otcěnásek 1999). Another source is the differential slipping of bow hairs in contact with the string (McIntyre et al. 1981). In fact, adding such audible noise to synthesis models for instrumental sounds is known to enhance their perceived naturalness (Serra 1997).

Helmholtz (1877) and Lichte (1941) found that the predominance of odd harmonics in a spectrum (such as clarinet notes) elicits an impression of hollowness or thinness compared to sounds with more balanced spectral envelopes (such as bowed strings) that are perceived as full. Despite explicitly synthesizing odd and even harmonic spectra to test the thin–full hypothesis, von Bismarck (1974a) did not report any relation between those stimuli and his fullness factor. Hollowness has also been found connected to the amount of *phantom partials* (nonlinearly generated frequencies due to string tension modulation) in piano sounds (Bensa et al. 2005). A small number of phantom partials produces a hollow timbre; gradually increasing the presence of such partials gives a rounder timbre, but sounds with a very large number of phantom partials (i.e., more such partials in the upper register) can appear metallic and aggressive.

The mass dimension of Zacharakis et al. (2014) exhibited three strong correlations in the English listeners' group (results for the Greek group were less conclusive). Thickness and density increased with inharmonicity and with fluctuation of the spectral centroid over time and decreased with fundamental frequency. Similar

to the first correlation, Bensa et al. (2005) observed that synthetic piano sounds with the least high-frequency inharmonic partials were perceived as poor, whereas increasing their number resulted in richer timbres. The second correlation appears to be in agreement with the connection between richness and high spectral variation reported for wind instruments by Kendall and Carterette (1993b) and for sustained instruments by Elliott et al. (2013) and may relate, at least partially, to multiple-source sounds with higher spectral flux values below 200 Hz that are perceived as fuller (Alluri and Toiviainen 2010).

The correlation between thickness/density and fundamental frequency found by Zacharakis et al. (2014) emerged largely due to the presentation of stimuli with different pitches. This acoustic interpretation of thickness/density alludes to an attribute of pure tones described by Stumpf (1890) as volume (*Tongröße* in German), which aligns inversely with pitch in that lower/higher pitches are larger/smaller. Together, the three attributes of volume, pitch, and loudness determine what Stumpf termed *tone color* (*Tonfarbe*). Rich (1916) provided empirical evidence that volume (he used the word extensity) can be distinct from pitch in pure tones. Terrace and Stevens (1962) showed that volume can also be perceived in more complex tonal stimuli, specifically, quarter-octave bands of pitched noise, and that it increases with loudness but decreases with pitch. Stevens (1934) observed that pure and complex tones further possess an attribute of density, which changes with loudness and pitch in a manner similar to perceptions of brightness: the brighter the tone, the louder and the less dense it is (Boring and Stevens 1936; cf., Zacharakis et al. 2014). Empirical observations of volume and density perceptions for pure tones have cast doubt on Schaeffer's (1966) claim that these have no timbre (Sect. 5.2.1).

Further experiments by Stevens et al. (1965) provided empirical support to Koechlin's claim that density is proportional to loudness and inversely proportional to volume (Sect. 5.2.1). An inverse relation between spectral centroid and volume was observed, which has been confirmed by Chiasson et al. (2017). They found that high energy concentrated in low frequencies tends to increase perceived volume, whereas low energy more spread out in higher frequencies tends to decrease it. Similarly, Saitis et al. (2015) showed that violin notes characterized as rich tended to have a low spectral centroid or stronger second, third, and fourth harmonics, or a predominant fundamental. Given that in harmonic sounds the fundamental is the lowest frequency, these findings generally agree with Helmholtz's (1877) claim that the stronger versus weaker the fundamental is relative to the upper partials, the richer versus poorer the sound is perceived.

5.4.2 Influence of Language and Culture

In the interlanguage study of Zacharakis et al. (2014, 2015), the overall configurational and dimensional similarity between semantic and perceptual spaces in both the English and Greek groups illustrates that the way timbre is conceptualized and communicated can indeed capture some aspects of the perceptual structure within

a set of timbres, and that native language has very little effect on the perceptual and semantic processing involved, at least for the two languages tested. There also seems to be some agreement regarding the number and labeling of dimensions with studies in German (von Bismarck 1974a; Štěpánek 2006), Czech (Moravec and Štěpánek 2005; Štěpánek 2006), Swedish (Nykänen et al. 2009), and French (Faure 2000; Lavoie 2013). Chiasson et al. (2017) found no effect of native language (French versus English) on perceptions of timbral volume. All these studies were conducted with groups of Western listeners and with sounds from Western musical instruments. Further evidence of whether language (but also culture) influences timbre semantics comes from research involving non-Western listeners and non-Western timbres.

Giragama et al. (2003) asked native speakers of English, Japanese, Bengali (Bangladesh), and Sinhala (Sri Lanka) to provide dissimilarity and semantic ratings of six electroacoustic sounds (one processed guitar, six effects). Multidimensional analyses yielded a two-dimensional MDS space shared across the four groups and two semantic factors (sharp/clear and diffuse/weak) whose order and scores varied moderately between languages and related differently to the MDS space. For Bengali and Sinhala, both Indo-Aryan languages, the similarity between the respective semantic spaces was much stronger, and they correlated better with the MDS space than for any other language pair, including between the Indo-European English and Indo-Aryan relatives. Furthermore, the sharp/clear and diffuse/weak factors closely matched the semantic space of electroacoustic textures found by Grill (2012), whose study was conducted with native German speakers.

Alluri and Toiviainen (2010) found a three-dimensional semantic timbre space of activity (strong–weak, soft–hard), brightness (dark–bright, colorless–colorful), and fullness (empty–full) for Indian pop music excerpts rated by Western listeners who had low familiarity with the genre. Here timbre refers to timbral mixtures arising from multiple-source sounds. Both the number and nature of these dimensions are in good agreement with Zacharakis et al. (2014). Furthermore, similar semantic spaces were obtained across two groups of Indian and Western listeners and two sets of Indian and Western pop music excerpts (Alluri and Toiviainen 2012). Acoustic analyses also gave comparable results between the two cultural groups and between the two studies. Intrinsic dimensionality estimation revealed a higher number of semantic dimensions for music from one's own culture compared to a culture that one is less familiar with, suggesting an effect of enculturation. Furthermore, Iwamiya and Zhan (1997) found common dimensions of sharpness (sharp–dull, bright–dark, distinct–vague, soft–hard), cleanness (clear–muddy, fine–rough), and spaciousness (rich–poor, extended–narrow) for music excerpts rated separately by Japanese and Chinese native speakers (type of music used was not reported). These dimensions appear to modestly match those found by Alluri and Toiviainen (2010) and by Zacharakis et al. (2014).

Taken as a whole, these (limited) results suggest that conceptualization and communication of timbral nuances is largely language independent, but some culture-driven linguistic divergence can occur. As an example, Zacharakis et al. (2014) found that, whereas sharp loaded highest on the luminance factor in English, its

Greek equivalent *οξύς* (oxýs) loaded higher on the texture dimension of the respective semantic space. Greek listeners also associated *παχύς* (pakhús), the Greek equivalent of thick, with luminance rather than mass. Furthermore, a well-known discrepancy exists between German and English concerning the words *Schärfe* and sharpness, respectively (see Kendall and Carterette 1993a, p. 456). Whereas *Schärfe* refers to timbre, its English counterpart pertains to pitch. On the one hand, such differences between languages may not imply different mental (nonlinguistic) representations of timbre but rather reflect the complex nature of meaning.

On the other hand, there exists evidence that language and culture can play a causal role in shaping nonlinguistic representations of sensory percepts, for example, auditory pitch (Dolscheid et al. 2013). This raises a crucial question concerning the use of verbal attributes by timbre experts such as instrument musicians: To what extent does experience with language influence mental representations of timbre? Based on their findings, Zacharakis et al. (2015) hypothesized that "there may exist a substantial latent influence of timbre semantics on pairwise dissimilarity judgements" (p. 408). This seems to be supported from comparisons between general dissimilarity, brightness dissimilarity, and brightness scaling data by Saitis and Siedenburg (in preparation), but more research is needed to better understand the relationship between linguistic and nonlinguistic representations of timbre. Nevertheless, semantic attributes, such as brightness, roughness, and fullness, appear generally unable to capture the salient perceptual dimension of timbre responsible for discriminating between sustained and impulsive sounds (Zacharakis et al. 2015).

5.5 Timbre Semantics and Room Acoustics: Ambiguity in Figure-Ground Separation

Imagine yourself listening to recordings of the famous violinist Yehudi Menuhin (1916–1999) performing on his Khevenhüller Strad in different concert halls. Does your impression of the sound of the violin or the sound of Menuhin change from one recording or hall to another? The answer would be almost certainly yes. The perceived timbre of a sound is not only a result of the physical characteristics of its source: It is always influenced by the properties of the acoustic environment that connects the sound source and the listener. Putting it differently, in evaluating the timbre of a sound, listeners invariably evaluate timbral characteristics of the presentation space too. The influence of the latter on the spectral shape of a sound, as illustrated by the room acoustic transfer function (Weinzierl and Vorländer 2015), is manifested in a characteristic amplification or attenuation of certain frequencies, superimposed by an increasing attenuation of the spectral envelope toward higher frequencies due to air absorption. The extent of these effects can vary substantially from one space to another, depending on the geometry and materials of the room.

When listeners try to perceptually separate the properties of the sound source from the properties of the room, they face a situation that has been described as

figure-ground organization in Gestalt psychology. Although its origins lie in visual scene analysis, organizing a perceptual stream into foreground (figure) and background (ground) elements has been shown to apply also in the auditory realm (Bregman 1990). Listeners can group foreground sounds across the spectral or temporal array and separate them from a background of concurrent sounds. When timbre acts as a contributor to sound source identity (Siedenburg, Saitis, and McAdams, Chap. 1), figure-ground segregation is generally unambiguous. A violin note will always be recognized as such, categorically as well as relative to concurrent notes from other instruments, regardless of the performance venue—excluding deliberate attempts to blend instrumental timbres (Lembke and McAdams 2015). However, figure-ground separation becomes more complicated when one looks beyond sound source recognition.

During language socialization of musicians or music listeners, where timbre functions as qualia (Siedenburg, Saitis, and McAdams, Chap. 1), there is not a single moment when a musical instrument is heard without a room acoustic contribution (except under anechoic room conditions). Even if the specific characteristics of the respective performance spaces are different, it can be assumed that common properties of *any* room acoustic environment (e.g., high-frequency spectral attenuation and prolongation by reverberation) will, to a certain degree, become part of the mental representation of an instrument's sound. It can be shown, for instance, that the early part of a room's reverberation tail tends to merge with the direct sound perceptually, increasing the perceived loudness of the sound rather than being attributed to the response of the room (Haas 1972). In addition, many musical instruments have their own decay phase, and with decay times of up to 3 s for violins on the open string (Meyer 2009), it becomes difficult to predict the extent to which listeners can successfully segregate the source and room streams when communicating timbral qualities.

The role of timbre in the characterization of room acoustic qualities has traditionally received little attention. In the current standard on room acoustic measurements of musical performance venues, there is not a single parameter dedicated to the timbral properties of the hall (ISO 3382-1:2009). However, recent studies have highlighted timbre as a central aspect of room acoustic qualities (Lokki et al. 2016), with brilliance, brightness, boominess, roughness, comb-filter-like coloration, warmth, and metallic tone color considered as the most important timbral attributes of a specific performance venue (Weinzierl et al. 2018a, b). The ways in which the semantics of spaces *interact* with the semantics of timbre and the extent to which figure-ground separation is reflected in the language of space and source are objects for future research.

5.6 Summary

Timbre is one of the most fundamental aspects of acoustic communication and yet it remains one of the most poorly understood. Despite being an intuitive concept, timbre covers a very complex set of auditory attributes that are not accounted for by

frequency, intensity, duration, spatial location, and the acoustic environment (Siedenburg, Saitis, and McAdams, Chap. 1), and the description of timbre lacks a specific sensory vocabulary. Instead, sound qualities are conceptualized and communicated primarily through readily available sensory attributes from different modalities (e.g., bright, warm, sweet) but also through onomatopoeic attributes (e.g., ringing, buzzing, shrill) or through nonsensory attributes relating to abstract constructs (e.g., rich, complex, harsh). These metaphorical descriptions embody conceptual representations, allowing listeners to talk about subtle acoustic variations through other, more commonly shared corporeal experiences (Wallmark 2014): with reference to the human and nonhuman voice (*instruments are voices*), as a tangible object (*sound is material*), and in terms of friction (*noise is friction*). Semantic ratings and factor analysis techniques provide a powerful tool to empirically study the relation between timbre perception (psychophysical dimensions), its linguistic descriptions (conceptual-metaphorical dimensions), and their meaning (semantic dimensions).

Common semantic dimensions have been summarized as brightness/sharpness (or luminance), roughness/harshness (or texture), and fullness/richness (or mass) and correspond strongly, but not one-to-one, with the three psychophysical dimensions along which listeners are known to perceive timbre similarity. In some cases, the dimensions are relatively stable across different languages and cultures, although more systematic explorations would be necessary to establish a cross-cultural and language-invariant semantic framework for timbre. A recent study with cochlear implant listeners indicated a dimension of brightness and one of roughness in relation to variations in electrode position and/or pulse rate (Marozeau and Lamping, Chap. 10). Furthermore, notions of timbral extensity and density have been central to spectromorphological models of listening and sound organization (Sect. 5.2.1) and to theories of sound mass music (Douglas et al. 2017). More generally, timbre is implicated in size recognition across a range of natural (e.g., speech, animals; see Mathias and von Kriegstein, Chap. 7) and possibly even abstract sound sources (Chiasson et al. 2017).

Long-term familiarity with and knowledge about sound source categories influence the perception of timbre as manifested in dissimilarity ratings (McAdams, Chap. 2). An interesting question that has not been fully addressed yet is whether source categories further exert an effect on the semantic description of timbre, given the strong link between linguistic and perceptual representations. In this direction, Saitis and Siedenburg (in preparation) compared ratings of dissimilarity based on brightness with ratings of general dissimilarity and found that the former relied primarily on (continuously varying) acoustic properties. Could the mass dimension be more prone to categorical effects due to its connection with source size recognition? Closely related to this question is the need to specify the role of affective mediation in timbre semantics. For example, bright timbres tend to be associated with happiness, dull with sadness, sharp with anger, and soft with both fear and tenderness (Juslin and Laukka 2004). McAdams (Chap. 8) discusses the effect of timbral brightness on emotional valence in orchestration contexts.

Nonauditory sensory attributes of timbre exemplify a particular aspect of semantic processing in human cognition: People systematically make many crossmodal mappings between sensory experiences presented in different modalities (Simner et al. 2010) or within the same modality (Melara and Marks 1990). The notion of sound color, timbre's alter ego, is exemplified in terms such as the German *Klangfarbe* (*Klang* + *Farbe* = sound + color) and the Greek *ηχόχρωμα* [ichóchroma] (*ήχος* [íchos] + *χρώμα* [chróma] = sound + color) and is itself a crossmodal blend. In viewing timbre semantics through the lens of crossmodal correspondences, questions about the perceptual and neural basis of the former can thus be reconsidered. What timbral properties of sound evoke the analogous impression as touching a smooth surface or viewing a rounded form? Are perceptual attributes of different sensory experiences (e.g., a smooth surface and a rounded form) mapped to similar or distinct timbres? Are crossmodal attributes (e.g., smooth, rounded) a result of supramodal representations (Walsh 2013) or of direct communication between modalities (Wallmark 2014)? Addressing these questions requires a comprehensive examination of auditory-nonauditory correspondences, including the collection of behavioral and neuroimaging data from appropriate tasks that extend beyond the semantic differential paradigm.

Acknowledgements Charalampos Saitis wishes to thank the Alexander von Humboldt Foundation for support through a Humboldt Research Fellowship.

Compliance with Ethics Requirements Charalampos Saitis declares that he has no conflict of interest.

Stefan Weinzierl declares that he has no conflict of interest.

References

Abeles H (1979) Verbal timbre descriptors of isolated clarinet tones. Bull Council Res Music Educ 59:1–7

Albersheim G (1939) Zur Psychologie der Toneigenschaften (On the psychology of sound properties). Heltz, Strassburg

Alluri V, Toiviainen P (2010) Exploring perceptual and acoustical correlates of polyphonic timbre. Music Percept 27:223–242

Alluri V, Toiviainen P (2012) Effect of enculturation on the semantic and acoustic correlates of polyphonic timbre. Music Percept 29:297–310

Almeida A, Schubert E, Smith J, Wolfe J (2017) Brightness scaling of periodic tones. Atten Percept Psychophys 79(7):1892–1896

Bell R (2015) PAL: the percussive audio lexicon. An approach to describing the features of percussion instruments and the sounds they produce. Dissertation, Swinburne University of Technology

Bensa J, Dubois D, Kronland-Martinet R, Ystad S (2005) Perceptive and cognitive evaluation of a piano synthesis model. In: Wiil UK (ed) Computer music modelling and retrieval. 2nd international symposium, Esbjerg, May 2004. Springer, Heidelberg, pp 232–245

Bernays M, Traube C (2009) Expression of piano timbre: verbal description and gestural control. In: Castellengo M, Genevois H (eds) La musique et ses instruments (Music and its instruments). Delatour, Paris, pp 205–222

Bernays M, Traube C (2011) Verbal expression of piano timbre: multidimensional semantic space of adjectival descriptors. In: Williamon A, Edwards D, Bartel L (eds) Proceedings of the international symposium on performance science 2011. European Association of Conservatoires, Utrecht, pp 299–304
Bloothooft G, Plomp R (1988) The timbre of sung vowels. J Acoust Soc Am 84:847–860
Boring EG, Stevens SS (1936) The nature of tonal brightness. Proc Natl Acad Sci 22:514–521
Bowles EA (1954) Haut and bas: the grouping of musical instruments in the middle ages. Music Discip 8:115–140
Bowling DL, Purves D, Gill KZ (2018) Vocal similarity predicts the relative attraction of musical chords. Proc Natl Acad Sci 115:216–221
Bregman AS (1990) Auditory scene analysis. The perceptual organization of sound. MIT Press, Cambridge
Brent W (2010) Physical and perceptual aspects of percussive timbre. Dissertation, University of California
Burton H (2016) Menuhin: a life, revised edn. Faber & Faber, London
Chiasson F, Traube C, Lagarrigue C, McAdams S (2017) Koechlin's volume: perception of sound extensity among instrument timbres from different families. Music Sci 21:113–131
Cousineau M, Carcagno S, Demany L, Pressnitzer D (2014) What is a melody? On the relationship between pitch and brightness of timbre. Front Syst Neurosci 7:127
Daniel P, Weber R (1997) Psychoacoustical roughness: implementation of an optimized model. Acta Acust united Ac 83:113–123
Douglas C, Noble J, McAdams S (2017) Auditory scene analysis and the perception of sound mass in Ligeti's continuum. Music Percept 33:287–305
de Ceuster D (2016) The phenomenological space of timbre. Dissertation, Utrecht University
Disley AC, Howard DM, Hunt AD (2006) Timbral description of musical instruments. In: Baroni M, Addessi AR, Caterina R, Costa M (eds) Proceedings of the 9th international conference on music perception and cognition, Bologna, 2006
Dolan EI (2013) The orchestral revolution: Haydn and the technologies of timbre. Cambridge University Press, Cambridge
Dolscheid S, Shayan S, Majid A, Casasanto D (2013) The thickness of musical pitch: psychophysical evidence for linguistic relativity. Psychol Sci 24:613–621
Dubois D (2000) Categories as acts of meaning: the case of categories in olfaction and audition. Cogn Sci Q 1:35–68
Edwards RM (1978) The perception of trombones. J Sound Vib 58:407–424
Eitan Z, Rothschild I (2011) How music touches: musical parameters and listeners' audio-tactile metaphorical mappings. Psychol Music 39:449–467
Elliott TM, Hamilton LS, Theunissen FE (2013) Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. J Acoust Soc Am 133:389–404
Fastl H, Zwicker E (2007) Psychoacoustics: facts and models, 3rd edn. Springer, Heidelberg
Faure A (2000) Des sons aux mots, comment parle-t-on du timbre musical? (From sounds to words, how do we speak of musical timbre?). Dissertation, Ecoles des hautes etudes en sciences sociales
Fritz C, Blackwell AF, Cross I et al (2012) Exploring violin sound quality: investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties. J Acoust Soc Am 131:783–794
Gallese V, Lakoff G (2005) The brain's concepts: the role of the sensory-motor system in conceptual knowledge. Cogn Neuropsychol 22:455–479
Garnier M, Henrich N, Castellengo M et al (2007) Characterisation of voice quality in Western lyrical singing: from teachers' judgements to acoustic descriptions. J Interdiscipl Music Stud 1:62–91
Giragama CNW, Martens WL, Herath S et al (2003) Relating multilingual semantic scales to a common timbre space – part II. Paper presented at the 115th audio engineering society convention, New York, 10–13 October 2003

Grill T (2012) Perceptually informed organization of textural sounds. Dissertation, University of Music and Performing Arts Graz
Haas H (1972) The influence of a single echo on the audibility of speech. J Audio Eng Soc 20:146–159
Heise DR (1969) Some methodological issues in semantic differential research. Psychol Bull 72:406–422
Helmholtz H (1877) Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik, 4th edn. F. Vieweg und Sohn, Braunschweig. English edition: Helmholtz H (1954) On the sensations of tone as a physiological basis for the theory of music (trans: Ellis AJ), 2nd edn. Dover, New York
Isshiki N, Okamura H, Tanabe M, Morimoto M (1969) Differential diagnosis of hoarseness. Folia Phoniatr 21:9–19
Iwamiya S, Zhan M (1997) A comparison between Japanese and Chinese adjectives which express auditory impressions. J Acoust Soc Jpn 18:319–323
Jakobson R, Halle M (1971) Fundamentals of language, 2nd edn. Mouton, The Hague
Juslin PN, Laukka P (2004) Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. J New Music Res 33:217–238
Kendall RA, Carterette EC (1993a) Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives. Music Percept 10:445–468
Kendall RA, Carterette EC (1993b) Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's orchestration. Music Percept 10:469–501
Kendall RA, Carterette EC, Hajda JM (1999) Perceptual and acoustical features of natural and synthetic orchestral instrument tones. Music Percept 16:327–363
Koelsch S (2011) Toward a neural basis of processing musical semantics. Phys Life Rev 8:89–105
Kursell J (2013) Experiments on tone color in music and acoustics: Helmholtz, Schoenberg, and Klangfarbenmelodie. Osiris 28:191–211
Lakoff G, Johnson M (2003) Metaphors we live by. University of Chicago Press, Chicago
Lavoie M (2013) Conceptualisation et communication des nuances de timbre à la guitare classique (Conceptualization and communication of classical guitar timbral nuances). Dissertation, Université de Montréal
Lemaitre G, Houix O, Misdariis N, Susini P (2010) Listener expertise and sound identification influence the categorization of environmental sounds. J Exp Psychol Appl 16:16–32
Lembke S-A, McAdams S (2015) The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. Acta Acust United Ac 101:1039–1051
Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cogn 21:1–36
Lichte WH (1941) Attributes of complex tones. J Exp Psychol 28:455–480
Lokki T, Pätynen J, Kuusinen A, Tervo S (2016) Concert hall acoustics: repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences. J Acoust Soc Am 140:551–562
Marozeau J, de Cheveigné A (2007) The effect of fundamental frequency on the brightness dimension of timbre. J Acoust Soc Am 121(1):383–387
McAdams S, Winsberg S, Donnadieu S et al (1995) Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychol Res 58:177–192
McDermott JM, Lehr AJ, Oxenham AJ (2010) Individual differences reveal the basis of consonance. Curr Biol 20:1035–1041
McIntyre ME, Schumacher RT, Woodhouse J (1981) Aperiodicity in bowed-string motion. Acustica 49:13–32
Melara RD, Marks LE (1990) Interaction among auditory dimensions: timbre, pitch and loudness. Percept Psychophys 48:169–178
Meyer J (2009) Acoustics and the performance of music. Springer, Berlin
Miller R (1986) The structure of singing: system and art of vocal technique. Schirmer Books, New York

Moravec O, Štěpánek J (2003) Verbal description of musical sound timbre in Czech language. In: Bresin R (ed) Proceedings of the Stockholm Music Acoustics Conference 2003. KTH, Stockholm, p 643–646

Moravec O, Štěpánek J (2005) Relations among verbal attributes describing musical sound timbre in Czech language. In: Proceedings of Forum Acusticum Budapest 2005: the 4th European congress on acoustics. Hirzel, Stuttgart, p 1601–1606

Mores R (2011) Nasality in musical sounds – a few intermediate results. In: Schneider A, von Ruschkowski A (eds) Systematic musicology: empirical and theoretical studies. Peter Lang, Frankfurt am Main, pp 127–136

Nykänen A, Johansson Ö, Lundberg J, Berg J (2009) Modelling perceptual dimensions of saxophone sounds. Acta Acust United Ac 95:539–549

Osgood CE (1952) The nature and measurement of meaning. Psychol Bull 49:197–237

Porcello T (2004) Speaking of sound: language and the professionalization of sound-recording engineers. Soc Studies Sci 34:733–758

Pratt RL, Bowsher JM (1978) The subjective assessment of trombone quality. J Sound Vib 57:425–435

Pratt RL, Doak PE (1976) A subjective rating scale for timbre. J Sound Vib 45:317–328

Pressnitzer D, McAdams S (1999) Two phase effects in roughness perception. J Acoust Soc Am 105:2773–2782

Radocy RE, Boyle JD (eds) (2012) Psychological foundations of musical behavior, 5th edn. Thomas Books, Springfield

Reuter C (1997) Karl Erich Schumann's principles of timbre as a helpful tool in stream segregation research. In: Leman M (ed) Music, gestalt, and computing. Studies in cognitive and systematic musicology. Springer, Heidelberg, pp 362–372

Reybrouck M (2013) From sound to music: an evolutionary approach to musical semantics. Biosemiotics 6:585–606

Rich GJ (1916) A preliminary study of tonal volume. J Exp Psychol 1:13–22

Rioux V, Västfjäll D (2001) Analyses of verbal descriptions of the sound quality of a flue organ pipe. Music Sci 5:55–82

Rozé J, Aramaki M, Kronland-Martinet R, Ystad S (2017) Exploring the perceived harshness of cello sounds by morphing and synthesis techniques. J Acoust Soc Am 141:2121–2136

Saitis C, Fritz C, Guastavino C, Giordano BL, Scavone GP (2012) Investigating consistency in verbal descriptions of violin preference by experienced players. In: Cambouropoulos E, Tsougras C, Mavromatis P, Pastiadis K (eds) Proceedings of the 12th international conference on music perception and cognition and 8th triennial conference of the European Society for the Cognitive Sciences of Music, Thessaloniki

Saitis C, Fritz C, Guastavino C, Scavone GP (2013) Conceptualization of violin quality by experienced performers. In: Bresin R, Askenfelt A (eds) Proceedings of the Stockholm music acoustics conference 2013. Logos, Berlin, p 123–128

Saitis C, Fritz C, Scavone GP et al (2017) Perceptual evaluation of violins: a psycholinguistic analysis of preference verbal descriptions by experienced musicians. J Acoust Soc Am 141:2746–2757

Saitis C, Järveläinen H, Fritz C (2018) The role of haptic cues in musical instrument quality perception. In: Papetti S, Saitis C (eds) Musical Haptics. Springer, Cham, pp 73–93

Saitis C, Scavone GP, Fritz C, Giordano BL (2015) Effect of task constraints on the perceptual evaluation of violins. Acta Acust United Ac 101:382–393

Samoylenko E, McAdams S, Nosulenko V (1996) Systematic analysis of verbalizations produced in comparing musical timbres. Int J Psychol 31:255–278

Schaeffer P (1966) Traité des objets musicaux: essai interdisciplines. Editions du Seuil, Paris. English edition: Schaeffer P (2017) Treatise on musical objects: an essay across disciplines (trans: North C, Dack J). University of California Press, Oakland

Schneider A (1997) "Verschmelzung", tonal fusion, and consonance: carl stumpf revisited. In: Leman M (ed) Music, gestalt, and computing. Springer, Berlin, pp 117–143

Schumann KE (1929) Physik der Klangfarben (Physics of timbres). Habilitation, Universität Berlin
Serra X (1997) Musical sound modelling with sinusoids plus noise. In: Roads C, Pope S, Piccialli A, de Poli G (eds) Musical signal processing. Swets Zeitlinger, Lisse, pp 91–122
Simner J, Cuskley C, Kirby S (2010) What sound does that taste? Cross-modal mappings across gustation and audition. Perception 39:553–569
Slawson W (1985) Sound color. University of California Press, Berkeley
Smalley D (1997) Spectromorphology: explaining sound-shapes. Organised Sound 2:107–126
Solomon LN (1958) Semantic approach to the perception of complex sounds. J Acoust Soc Am 30:421–425
Štěpánek J (2006) Musical sound timbre: verbal description and dimensions. In: Proceedings of the 9th international conference on digital audio effects. McGill University, Montreal, p 121–126
Štěpánek J, Otcěnásek Z (1999) Rustle as an attribute of timbre of stationary violin tones. Catgut Acoust Soc J (Series II) 3:32–38
Stevens SS (1934) Tonal density. J Exp Psychol 17:585–592
Stevens SS, Guirao M, Slawson AW (1965) Loudness, a product of volume times density. J Exp Psychol 69:503–510
Stumpf C (1890) Tonpsychologie (Psychology of sound), vol 2. Hirzel, Leipzig
Stumpf C (1898) Konsonanz und Dissonanz (Consonance and dissonance). Barth, Leipzig
Sundberg J (2013) Perception of singing. In: Deutsch D (ed) The psychology of music, 3rd edn. Academic, London, pp 69–105
Susini P, Lemaitre G, McAdams S (2012) Psychological measurement for sound description and evaluation. In: Berglund B, Rossi GB, Townsend JT, Pendrill LR (eds) Measurement with persons: theory, methods, and implementation areas. Psychology Press, New York, pp 227–253
Terrace HS, Stevens SS (1962) The quantification of tonal volume. Am J Psychol 75:596–604
Traube C (2004) An interdisciplinary study of the timbre of the classical guitar. Dissertation, McGill University
Thiering M (2015) Spatial semiotics and spatial mental models. Figure-ground asymmetries in language. De Gruyter Mouton, Berlin
Vassilakis PN, Kendall RA (2010) Psychoacoustic and cognitive aspects of auditory roughness: definitions, models, and applications. In: Rogowitz BE, Pappas TN (eds) Human vision and electronic imaging XV. SPIE/IS&T, Bellingham/Springfield, p 75270
von Bismarck G (1974a) Timbre of steady tones: a factorial investigation of its verbal attributes. Acustica 30:146–159
von Bismarck G (1974b) Sharpness as an attribute of the timbre of steady sounds. Acustica 30:159–172
Wake S, Asahi T (1998) Sound retrieval with intuitive verbal expressions. Paper presented at the 5th international conference on auditory display, University of Glasgow, 1–4 November 1998
Walker P (2016) Cross-sensory correspondences: a theoretical framework and their relevance to music. Psychomusicology 26:103–116
Wallmark Z (2014) Appraising timbre: embodiment and affect at the threshold of music and noise. Dissertation, University of California
Wallmark Z (2018) A corpus analysis of timbre semantics in orchestration treatises. Psychol Music. https://doi.org/10.1177/0305735618768102
Walsh V (2013) Magnitudes, metaphors, and modalities: a theory of magnitude revisited. In: Simner J, Hubbard E (eds) Oxford handbook of synesthesia. Oxford University Press, Oxford, pp 837–852
Webster J, Woodhead M, Carpenter A (1970) Perceptual constancy in complex sound identification. Br J Psychol 61:481–489
Weinzierl S, Lepa S, Ackermann D (2018a) A measuring instrument for the auditory perception of rooms: the Room Acoustical Quality Inventory (RAQI). J Acoust Soc Am 144:1245–1257
Weinzierl S, Lepa S, Schultz F et al (2018b) Sound power and timbre as cues for the dynamic strength of orchestral instruments. J Acoust Soc Am 144:1347–1355

Weinzierl S, Vorländer M (2015) Room acoustical parameters as predictors of room acoustical impression: what do we know and what would we like to know? Acoust Aust 43:41–48

Zacharakis A, Pastiadis K, Reiss JD (2014) An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. Music Percept 31:339–358

Zacharakis A, Pastiadis K, Reiss JD (2015) An interlanguage unification of musical timbre: bridging semantic, perceptual, and acoustic dimensions. Music Percept 32:394–412