11 Psychological measurement for sound description and evaluation

Patrick Susini,[1] *Guillaume Lemaitre,*[1] *and Stephen McAdams*[2]

[1]Institut de Recherche et de Coordination Acoustique/Musique
Paris, France

[2]CIRMMT, Schulich School of Music, McGill University
Montréal, Québec, Canada

11.1 Introduction

Several domains of application require one to measure quantities that are representative of what a human listener perceives. *Sound quality evaluation*, for instance, studies how users perceive the quality of the sounds of industrial objects (cars, electrical appliances, electronic devices, etc.), and establishes specifications for the design of these sounds. It refers to the fact that the sounds produced by an object or product are not only evaluated in terms of annoyance or pleasantness, but are also important in people's interactions with the object. Practitioners of sound quality evaluation therefore need methods to assess experimentally, or automatic tools to predict, what users perceive and how they evaluate the sounds. There are other applications requiring such measurement: evaluation of the quality of audio algorithms, management (organization, retrieval) of sound databases, and so on. For example, sound-database retrieval systems often require measurements of relevant perceptual qualities; the searching process is performed automatically using similarity metrics based on relevant descriptors stored as metadata with the sounds in the database.

The "perceptual" qualities of the sounds are called the *auditory attributes*, which are defined as percepts that can be ordered on a magnitude scale. Historically, the notion of auditory attribute is grounded in the framework of psychoacoustics. Psychoacoustical research aims to establish quantitative relationships between the physical properties of a sound (i.e., the properties measured by the methods and instruments of the natural sciences) and the perceived properties of the sounds, the auditory attributes. The physical properties of a sound that are related to the auditory attributes can be computed from the sound signal. These values therefore predict the auditory attributes from the sound signal alone and once well understood can be substituted for experimental measurements. They are called *psychoacoustical* *descriptors*. Psychoacoustical research has isolated several auditory attributes: loudness, pitch, duration, and sharpness, among others. Methods have been developed to measure these attributes experimentally, and algorithms have been devised to compute corresponding psychoacoustical descriptors.

Here we use the term "auditory attribute" in a slightly broader sense than the psychoacoustical definition. Indeed, listeners can recover many kinds of information from a sound. Not only do they perceive percepts that can be directly mapped to the physical properties of the sound, but most of the time they also recognize the source that caused the sound and identify its properties. Gaver (1993a, 1993b) initially formalized this idea by introducing the concepts of *musical listening* (focus on the sound itself) and *everyday listening* (focus on the properties of the source). By *measuring auditory attributes*, we therefore mean here "providing quantities representative of what a user perceives."

The purpose of this chapter is to present the measurement of these auditory attributes from an applied perspective. Some of these attributes are easily understood (and have a name) and have been studied in depth. For instance, loudness, pitch, and duration are auditory attributes for which experimental methods, and even mathematical predictive models, are easily accessible. Section 11.1 briefly summarizes some of the results and methods associated with these attributes. Other attributes are less easily specified and often require metaphors from other sensory modalities to be described: brightness (or sharpness), roughness, fluctuation strength, and so on. In Section 11.2, we present more specifically the methods used to explore these attributes. Because they cannot be easily and unequivocally specified to a listener, these attributes require indirect and multidimensional methods that allow exploration of sound perception. Section 11.2 presents several families of methods: semantic scales, similarity judgments and multidimensional scaling, sorting tasks, and cluster analyses. Section 11.3 presents examples of applications in sound quality. Finally, perspectives in the realm of sonic interaction design are briefly introduced.

11.1 Basic knowledge and methods

11.1.1 Peripheral auditory system

We provide here a broad overview of the peripheral auditory system.* For a more complete description, interested readers should refer to Moore (2003).

11.1.1.1 Description

The human peripheral auditory system is composed of three parts: the outer ear, the middle ear, and the inner ear. The *outer ear* is mainly composed of the pinna and the auditory canal between the pinna and the eardrum. The outer ear amplifies the sound level at the eardrum for frequencies around 3 kHz. The *middle ear*, composed of three

* Animations by Prof. Herbert Hudde from Bochum University can be found at the following URL: http://www.ruhr-unibochum.de/ika/ika/forschung/gruppe_hudde/bohear_en.htm

very small ossicles, matches impedance between the air in the auditory canal (outer ear) and the fluids in the cochlea (inner ear). It also improves sound transmission for frequencies in the range of 0.5–4 kHz. From a psychoacoustical point of view, the most important part of the *inner ear* is the basilar membrane (BM) that can be considered as a "frequency analyzer." An incoming sound sets in motion the BM with a maximum displacement at a certain position that differs according to the frequency of the sound; the position of the maximum varies from the beginning (base) of the BM (oval window) for high frequencies to the end (apex) of the BM for low frequencies. The frequency producing a maximum of displacement on the BM is the center frequency of a bandpass filter for that position. Because different fibers of the auditory nerve are connected to different positions along the basilar membrane, the frequency selectivity of the basilar membrane results in a frequency decomposition of the sounds in the auditory nerve. The frequency selectivity of the auditory system has very important consequences for audition. Particularly, the "masking" phenomenon has introduced the concepts of critical bands (CB) and auditory filters and has resulted in a model that is the basis for the computation of psychoacoustical descriptors.

11.1.1.2 Masking, critical bands, and models

Fletcher (1940) introduced the concept of critical bands to account for masking phenomena. For very narrow bands, he showed that the threshold of detection for a pure tone increases as the noise bandwidth increases. After a certain bandwidth, increasing the noise bandwidth no longer changes the tone threshold. Fletcher assumed that only an effective part of the noise masker, close to the frequency of the tone, has the power to mask the tone. The corresponding frequency region is the *critical band*. Further investigations showed that a model consisting of a bank of bandpass filters, the bandwidth of which increases with the center frequency, could account for masking (Zwicker, 1961; Zwicker & Fastl, 1972; Moore & Glasberg, 1983, 1990). The shape of each filter is asymmetric: roll-off is sharp for frequencies below the center frequency (100 dB/octave) and smooth for frequencies above the center frequencies. The steepness of the roll-off decreases as the level of the stimulus increases.

There are several models of these filters. Third-octave bandpass filters can roughly model the auditory filters. Fourth-octave bandpass filters have also been proposed and shown to approximate fairly well the auditory filters except for low frequencies (Hartmann, 1997). A more complex model uses the Gammatone filters (Patterson & Holdsworth, 1991). Finally, based on this concept of critical bands, several scales have been proposed: the *Bark scale* (Zwicker & Terhardt, 1980) and the *Equivalent Rectangular Bandwidth (ERB) scale* (Moore & Glasberg, 1983).

11.1.1.3 Psychoacoustical descriptors

Models of the auditory system based on critical bands are used to compute psychoacoustical descriptors. The classical psychoacoustical descriptors are summarized in Zwicker and Fastl (1999) and Moore (2003).

The descriptor of loudness is widespread. Models have been standardized: ISO 532-A (Stevens' model); ISO 532-B for (Zwicker's model). A BASIC program is also available in Zwicker (1984). ANSI S3.4-2005 is a revision proposed by Moore and Glasberg (1996) and Moore, Glasberg, and Baer (1997). Corrections of this model have also been proposed allowing a better account of impulsive sounds in a background masking noise (Vos, 1998) and of time-varying sounds (Glasberg & Moore, 2002). Another descriptor of loudness (Meunier, Boulet, & Rabau, 2001) has been proposed for environmental and synthesized impulsive sounds. The loudness is well explained by a combination between the logarithm of the release time and the energy.

Psychoacoustical descriptors corresponding to other auditory attributes are also commonly used: spectral centroid and sharpness, roughness (Daniel and Weber, 1997), and so on (see Zwicker & Fastl, 1999, and Fastl, 1997, for summaries). They have also been implemented in several commercial software packages: BAS and ArtemiS by Head Acoustics, dBSonic by 01dB-Metravib, PULSE by Brüel & Kjaer, and LEA by Genesis. The available descriptors that have been implemented are based on experimental results using abstract sounds, thus these psychoacoustical descriptors sometimes need to be adapted for real sounds (see the work by Misdariis et al., 2010, on this question). Only the loudness descriptors have been standardized. They provide reliable results for stationary sounds, but further development is needed for nonstationary sounds.

11.1.2 Classical psychoacoustical methods

The traditional psychoacoustical approach is unidimensional: it aims to establish a quantitative relationship between a single auditory attribute and a physical property of the sound.

11.1.2.1 Indirect methods

11.1.2.1.1 Thresholds. The indirect method is based on the measurement of thresholds. The *absolute threshold* is the minimum detectable level of a sound. For instance, for a pure tone it depends on the frequency of the tone. Under normal conditions, a young listener can hear frequencies between 20 Hz and 20 kHz. For most adults, the threshold increases rapidly above about 15 kHz. The *differential threshold* or *difference limen* (DL) is the smallest change in a sound to produce a *just-noticeable difference* (jnd) in the related auditory attribute.

11.1.2.1.2 Confusion scaling. By varying a physical parameter and measuring the DL for a given auditory attribute, a confusion scale for this attribute can be set. Assuming that all DLs correspond to equal changes of the auditory attribute (jnd), Fechner's law (1860, published in English in 1966) can be determined:

$$\psi = k \log (\phi)$$

where ψ is the magnitude of the auditory attribute, ϕ is the physical parameter, and k is a constant specific to each auditory attribute.

11.1.2.2 Direct methods

Ratio scaling is a direct method relying on the ability of participants to make numerical judgments of the ratio between the magnitudes of their sensations. The usual methods are *magnitude estimation* and *production*. For magnitude estimation, the participants are required to assign a number proportional to their sensation (e.g., loudness) of the intensity for sounds presented at different levels. For the magnitude production method, the participant is required in this case to adjust the level of a test sound to a specified number proportional to its loudness. The relation between the expressed sensation (e.g., loudness) using such methods and the corresponding acoustical values (e.g., sound pressure level) leads to the well-known psychophysical law, Steven's law:

$$\psi = k\, \phi^{\alpha}$$

where ψ is the magnitude of the auditory attribute, ϕ is the physical parameter, and k and α are constants specific to each auditory attribute. For instance, for the loudness of a 1-kHz tone, the exponent is 0.6: a 10-dB increase leads to a 2-sone increase. For a 3-kHz tone, the exponent is 0.67. Steven's law for loudness has led to the derivation of the sone scale.

The *cross-modal matching* method was proposed by S. S. Stevens (1959). The task consists in matching two sensations (e.g., loudness and muscular force sensation), one of which has been calibrated beforehand by a direct estimation method (Stevens, 1959). The matching function between the sensations is known or experimentally obtained. Then ratings related to the other sensation are directly deduced by way of the matching function. This method can be used to scale the loudness of time-varying sounds (see the next section).

11.1.3 Perspectives: Loudness of time-varying sounds

The classical psychoacoustical methods have been broadly used to study the perception of short and stationary sounds. Everyday sound events and musical pieces, however, are usually nonstationary. The temporal fluctuations and durations (up to 20 minutes) of such nonstationary sounds do not allow the use of classical methods, but require continuous ratings of the sounds. The participant must in this case respond instantaneously to any variation of the sound. The methods and the devices usually proposed can be sorted into five categories.

1. The *method of continuous judgment using categories* was proposed by Kuwano and Namba (1978, 1985) with the aim of studying temporal fluctuations of the level of urban sounds. In this procedure, participants judge

the loudness at each instant using a response box with seven buttons corresponding to seven categories: very, very loud–very loud–loud–medium–soft–very soft–very, very soft. This process is applicable to long-duration stimuli, because the task is not difficult and participants experience little fatigue. The participants modify their judgment as soon as they perceive a change equivalent to the distance between two categories. The main disadvantage of the continuous category judgment method is that it does not allow one to obtain analogical responses as a function of the signal contour.

2. The *audiovisual adjustment method* was developed by Kuwano and Namba (1990). In this method, participants express their judgment by continuously adjusting the length of a line with a cursor so that the length is proportional to the auditory sensation. The main problem with this method comes from the clipping or ceiling effect at the top end of the judgment scale, because the length of the line is limited (computer screen, sheet of paper, etc.). To get around this limitation, Kuwano and Namba (1990) elaborated a device with which the line presented on the terminal screen is projected on a large screen with an overhead projector. In a similar manner, Fastl (1989, 1991) performed an experiment in which the participant judged the instantaneous loudness by associating in real time the displacement of a potentiometer on a mixing table. However, this device provides little feedback (aside from hand/arm position) to the user.
3. The *continuous cross-modal matching method* proposed by Susini, McAdams, and Smith (2002) is based on the cross-modal matching method with a force-feedback device. The participant has to adjust a muscular force sensation to the perceived loudness. This device was used to assess 1-kHz pure tones (Susini, McAdams, & Smith, 2002), urban sound sequences (Susini & Maffiolo, 1999), and sounds of accelerating cars (Susini & McAdams, 2008). The method has proved to be a flexible experimental procedure allowing an individual calibration of the device as a function of each participant's perceptual scale, with the aim of avoiding compression or saturation effects in the responses.
4. The *analog categorical scaling* proposed by Weber (1991) combines the categorical and analogical methods. Participants can slide a cursor continuously along five discrete categories labeled (for example): very loud–loud–medium–soft–very soft. The distance between each category is considered as equivalent. This method has been widely used: for loudness evaluation of variable-amplitude sinusoidal sounds (Susini, McAdams, & Smith, 2002, 2007), for assessing speech quality (Hansen & Kollmeier, 1999; Gros & Chateau, 2001), for assessing the comfort of an urban sequence of a running bus (Parizet, Hamzaoui, Segaud, & Koch, 2003), and for brightness ratings of various sounds (Hedberg & Jansson, 1998).
5. The *semantic scale used in real-time* was introduced by several authors to study more complex auditory attributes than loudness, and more specifically to study real-time emotional response to music. The continuous response digital interface (CRDI) developed by Madsen (1996) allows a

continuous tracking of temporal variations of musical works, as does the two-dimensional emotional space (2DES) proposed by Schubert (1996) with which musically evoked emotions are evaluated in real time in a two-dimensional semantic space. Several authors used continuous rating to measure emotional force in music pieces (Sloboda & Lehmann, 2001; McAdams, Vines, Vieillard, Smith, & Reynolds, 2004).

11.2 Multidimensional and exploratory methods

It is not always easy to specify an auditory attribute a priori. Apart from pitch and loudness, very few words are specific to sound or easily understood by nonspecialists. Therefore, unidimensional techniques such as described above cannot be used to measure auditory attributes not easily communicated to participants, or those that are simply unknown to the experimenter. This section reports methods to explore or measure unspecified auditory attributes and more generally to determine the psychological aspects of sound perceived by listeners.

11.2.1 Judgments on multiple semantic scales

The use of multiple semantic scales is a fruitful technique to assess different psychological aspects of sounds: auditory attributes (e.g., loudness, roughness), appraisal (e.g., preference), emotional response (e.g., beauty, arousal), and connotative dimensions of the sound source (e.g., the power of a sports car).

Semantic scales are category scales defined either by a single semantic descriptor (unipolar scale) or by a pair of antonymic descriptors (bipolar scale). The scales usually have between three and seven categories. It is usually preferred to use an odd number of intervals to include the middle point of the scale.

11.2.1.1 Method and analysis

The most used technique is the semantic differential (SD). Participants are asked to judge each stimulus directly along a set of scales labeled with two opposed semantic descriptors. Usually true antonym labels are used (e.g., good–bad, pure–rich, etc.), but alternatives have been proposed (e.g., good–not good).

The labels of the scales are called *semantic descriptors*. The ratings of a stimulus on the different semantic scales yield a multidimensional representation called the *semantic profile*; an example is presented in Figure 11.1. A factor analysis can combine semantic scales into main factors. A multiple regression analysis can highlight relationships between factors corresponding to cognitive aspects (e.g., preference) and factors corresponding to auditory attributes (e.g., loudness, roughness). The latter factors are interpreted by looking for acoustical or psychoacoustical descriptors that are correlated with them. Each semantic descriptor is hypothesized to be psychologically relevant to the whole set of stimuli under examination. On the other hand, it has to be understood by the participants of the experiment.

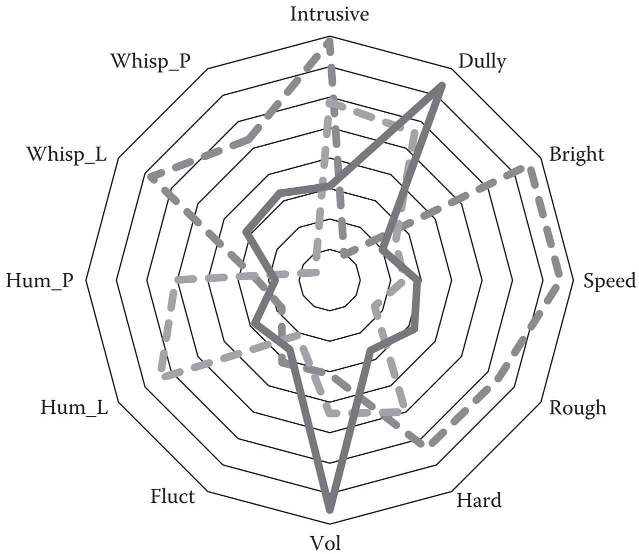


Figure 11.1 (See color insert.) Sensory profiles obtained for three air-conditioning noises from Siekierski, Derquenne, and Martin (2001). Labels of the *semantic descriptors* are Intrusive, Dully, Brightness, Speed, Roughness, Hardness, Voluminous, Fluctuation, Humming, Whispering. The letters L and P correspond to the Level and Pitch, respectively, of the whispering (noise) part and the humming (motor) part.

11.2.1.2 Examples of semantic scales used to describe sounds

Since Solomon (1958) and von Bismarck (1974), the semantic differential technique proposed by Osgood (1952) has been widely used in the realm of sound perception to describe the multidimensional character of the timbre of musical instruments (Wedin & Goude, 1972; Pratt & Doak, 1976; Kendall & Carterette, 1992; Stepánek, 2006), environmental sounds (Björk, 1985; Zeitler & Hellbrück, 2001), and sound products, such as cars, vacuum cleaners, air conditioning noises, or refrigerators (Chouard & Hempel; 1999; Kyncl & Jiricek, 2001; Siekiersky, Derquenne, & Martin, 2001; Jeon, 2006).

Typically, results from the different studies have shown that the set of semantic differentials can be combined into three or four main factors that account for a great deal of the variance in the judgments. For instance, in von Bismarck's study on the timbre of synthetic sounds, results revealed four independent scales: "dull–sharp" (44% of the variance explained), "compact–scattered" (26%), "full–empty" (9%), and "colorful–colorless" (2%). Only the scales referring to sharpness were considered as candidates for a generally usable scale for the measurement of timbre. In Pratt and Doak (1976), three scales ("dull–bright", "pure–rich", "cold–warm") were selected to be the more significant descriptors for instrumental timbres. In summary, results from different studies revealed that descriptors related to sharpness ("sharp," "bright," "metallic," or at the opposite, "dull," "muffled," "round") are appropriate to describe the most salient aspect of timbre.

Kuwano and Namba (2001) report three main factors ("powerful," "metallic," and "pleasant") that had consistently been extracted in most of their former studies of sound quality. Furthermore, the semantic descriptor "powerful" was usually well correlated with computed loudness (Zwicker's loudness level based on ISO 532B), and the semantic descriptor "metallic" was well correlated with computed sharpness. The "pleasant" factor was related to cognitive and cultural factors as well as to physical properties of sounds. In Zeitler and Hellbrück's study (2001) on environmental sounds, four factors were linked, respectively, to a hedonic aspect ("ugly"–"beautiful"), timbre ("dark–light"), power ("weak–strong"), and rapid temporal variations ("unstable–stable"). The three latter factors were well correlated with three calculated descriptors, sharpness, loudness, and roughness, respectively. Results from other studies on sounds (speech or sonar sounds) are quite similar: the most important factors were generally interpreted as representing loudness, timbre (sharpness) or pitch, and an overall subjective impression.

11.2.1.3 Prerequisites to use semantic scales

11.2.1.3.1 Controlling loudness, pitch, and duration. First, acoustical parameters such as loudness and pitch, as well as variations over time, strongly affect the perception of timbre. To study auditory attributes independently of those obvious parameters, it is therefore recommended to control them and to use steady-state sounds, equalized in loudness, pitch[*] and duration. This statement is in agreement with the current ANSI definition and summarized by Krumhansl (1989, p. 44): timbre is "the way in which musical sounds differ once they have been equated for pitch, loudness and duration." Otherwise, it is recommended to ask participants to ignore these parameters, following the proposal by Pratt and Doak (1976), who define timbre as "that attribute of auditory sensation whereby a listener can judge that two sounds are dissimilar using any criteria other than pitch, loudness or duration."

11.2.1.3.2 Selecting an appropriate number of semantic scales. Second, a restricted number of semantic pairs suitable for describing timbre have to be selected. Indeed, the preselection of semantic descriptors by the experimenter may strongly affect the results, for these descriptors may not necessarily conform with those a participant would use spontaneously. For instance, Pratt and Doak (1976) investigated (by a questionnaire) what were the most appropriate adjectives for describing timbre of instruments among a list of 19 commonly used terms. Seven words emerged as favorites: rich, mellow, colorful, brilliant, penetrating, bright, and warm. Similarly,

[*] For pitch, this can be done as for loudness by an adjustment procedure using the real-time SuperVP software program based on the phase vocoder technique; it is then possible to transpose, stretch, or shorten sounds in real-time.

in von Bismark (1974), participants were asked to rate each of 69 semantic scales in terms of their suitability for describing timbre. Finally, 28 scales were considered as representative. However, it should be noted that in von Bismarck's study, the 69 scales were rated independently of the sound selected for the study and thus may not be relevant for describing the perceptual dimensions of these sounds.

11.2.1.3.3 Selecting relevant descriptors. The third prerequisite consists in asking participants to judge the relevance of the semantic descriptors concerning the sounds used in the study. Faure (2010) gathered a set of semantic descriptors from a free verbalization experiment on 12 musical sounds. They were used to build semantic scales. The relevance of these scales was then judged for the same set of sounds. Comparison between the relevance judgments of the scales and the vocabulary produced spontaneously showed that several semantic descriptors that were spontaneously produced (such as "strong", "loud", etc.) were not considered as relevant when presented with the scales, even by the participants who produced them. Inversely, several semantic descriptors that were rarely used spontaneously were judged to be globally relevant by the majority of participants (e.g., "soft," "muffled/dull sounding," "metallic," "nasal"). In another study (Kyncl & Jiricek, 2001), participants freely described six vacuum cleaner sounds. Among the 33 pairs of semantic oppositions obtained from the vocabulary spontaneously produced, only 5 were consistently judged as relevant for describing the sounds ("fuzziness," "atypicality," "inefficiency," "loudness," and "pleasantness"). These studies highlight the importance of judging the relevance of descriptors used in SD-scales for a specific corpus of sounds.

11.2.1.3.4 Defining the meaning of the scales. The fourth prerequisite concerns the definition of the scales. Indeed, it is crucial that the participants correctly understand the meaning of the labels. For instance, in Faure (2010), the stimuli were equalized in loudness. Surprisingly, the participants spontaneously used the word "loud" to describe the sounds. Actually, the participants' comments revealed that they used "loud" to describe different perceptions: "strong in sonic presence, the attack," "evokes the power and persistence of the sound." Similarly, in von Bismarck (1974), although the sounds were equalized in loudness, participants used the scale "soft–loud" to describe attributes other than loudness, such as "unpleasant." Therefore, the experiment must clearly define the meaning of the semantic scales to eliminate any risk of semantic ambiguity. Presenting them in a sentence can help define the meaning of the descriptors. For instance Parizet and Nosulenko (1999) showed that ratings of internal noises of vehicles were more reliable when the semantic descriptors were presented in a sentence than when presented in isolation. Susini, Houix, Misdariis, Smith, and Langlois (2009) introduced the semantic descriptor "loud" by the sentence "The TV is too loud, we can't have a discussion." This sentence aimed at clearly indicating that "loud" referred to the sound level and not the unpleasantness.

In addition to the several prerequisites presented above, other recommendations should be taken into consideration when using the SD-scale technique to rate a corpus of sounds.

- Several studies have shown that subjects feel uncertain in giving ratings unless they can refer them to the whole sample of sounds. Thus the entire range of sounds has to be presented before the main experiment, and participants must be instructed to use the full range of the scale. In addition, it is recommended that the range of sensitivity corresponding to each semantic descriptor of the selected set of sounds be broad enough.
- Many studies on timbre have used the traditional semantic differential paradigm (e.g., dull–sharp). Bipolar adjective pairs raise the question of presenting the right antonym labels (is *dull* the opposite of *sharp* when used to describe sounds?). In Chouard and Hempel (1999), clear antonyms were found in about 23% of the cases for a list of 242 adjectives produced by the participants to describe interior car sounds. Thus an important problem in the use of bipolar opposites is that the "opposite" is sometimes unobtainable or not always a good antipode. To solve this problem, Kendall and Carterette (1992) proposed using a scale bounded by an attribute and its negation (e.g., sharp–not sharp) to rate the timbre of musical sounds. The authors termed this method *verbal attribute magnitude estimation* (VAME), because the task for the participant is to rate the degree to which an attribute is possessed by a stimulus.
- Finally, we recommend presenting the whole set of sounds for each semantic descriptor instead of the classical way consisting in presenting one sound to the participant, who has to evaluate it on the whole set of semantic descriptors. In a study by Parizet and colleagues (1999, 2005), the comparison of the two methods showed that the former proved to be more accurate and with a shorter duration than the classical one, because listeners were focused on one semantic descriptor at a time while hearing a new stimulus. In addition, to measure subject reliability or accuracy, random presentation of the stimuli can be repeated. Cross-correlation coefficients are calculated between the data from both presentations of the repeated stimuli to compute subject reliability.

11.2.2 Dissimilarity judgments and multidimensional scaling technique

Semantic scales compare stimuli along dimensions directly described semantically. It is therefore possible to assess various psychological aspects of a corpus of sounds, ranging from elementary auditory attributes to cognitive and emotional aspects. The disadvantage is that the number of scales is often too high and, with the exception of a few studies mentioned in the previous section, some of the selected semantic descriptors are not perceptually relevant to the corpus studied and are sometimes redundant in relation to each other. However, this approach is appropriate to study the perception of various environmental sounds, as long as several prerequisites are taken into account.

In contrast, the multidimensional scaling technique (MDS) is based on dissimilarity ratings and thus does not require a priori assumptions concerning the number of perceptual dimensions or their nature, unlike the methods that use ratings along specified dimensions.

11.2.2.1 MDS and auditory perception

The multidimensional scaling technique is a fruitful tool for studying perceptual relations among stimuli and for analyzing the underlying auditory attributes used by the participants to rate the perceived similarity between two sounds. MDS represents the perceived similarities in a low-dimensional Euclidean space (so-called *perceptual space*), so that the distances among the stimuli reflect the perceived similarities (see McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995, for a review of the different MDS algorithms). Each dimension of the space (so-called *perceptual dimension*) is assumed to correspond to a perceptual continuum that is common to the whole set of sounds. It is also assumed that each dimension can be well explained by an acoustic parameter or a psychoacoustical descriptor. In other words, the MDS technique is appropriate for describing sounds that are comparable along continuous auditory attributes, which means that it is appropriate for studying homogeneous corpora of sounds, that is, those made of sounds produced by the same type of source.

11.2.2.2 Method and analysis

Participants rate the perceived dissimilarity between each pair of sounds under consideration, that is, $N(N-1)/2$ ratings for N stimuli, on a continuous scale labeled “Very Similar” at the left end and “Very Dissimilar” at the right end. Then, the dissimilarities are modeled as distances in a Euclidean space of R dimensions expected to be the most relevant perceptual dimensions shared by the sounds. In the perceptual space, a large dissimilarity is represented by a large distance. The final and the most difficult part of this approach lies in matching perceptual dimensions to acoustical or psychoacoustical descriptors.

11.2.2.3 Example of MDS studies to describe timbre of musical sounds

Many researchers have applied the MDS technique to characterize the perceptual dimensions of sounds, since the seminal studies by Peters (1960) and Plomp (1970). Peters (1960) started to apply the MDS technique to a corpus of sounds with a known dimensionality (16 pure tones composed of 4 frequencies at 4 sound pressure levels: the acoustical dimensionality is therefore 2). The analysis of the dissimilarity judgments from 39 participants successfully highlighted the two expected auditory attributes: pitch and loudness. He therefore concluded that the MDS technique might be useful to explore sets of sounds the auditory attributes of which would be unknown. To test this idea, he applied the technique to other corpora of sounds, for which the salient auditory attributes were unknown (synthetic complex sounds and speech sounds). The results were less easily interpretable (he found between three and six dimensions for the complex sounds). But compared to what he obtained with more traditional approaches (free verbal description, partition scaling, magnitude estimation), he concluded that "the most promising approach for the isolation and definition of perceptual dimensions of complex sounds was the MDS model" (p. 52). Plomp (1970) applied MDS to sets of musical sounds, which yielded three orthogonal dimensions.

Since then, several psychoacoustical studies using MDS have shown clearly that musical timbre is a multidimensional attribute. Grey (1977) identified three salient dimensions shared by a corpus of musical sounds. Using a refinement of the classical MDS technique (EXSCAL, developed by Winsberg & Carroll, 1989), Krumhansl (1989) also found a space with three dimensions shared by a corpus of synthesized musical sounds (winds, bowed string, plucked strings, mallet percussion). The same set of sounds was analyzed by McAdams, Winsberg, Donnadieu, Soete, and Krimphoff (1995), who also found a 3-D space. The first dimension of the perceptual space was correlated with the centroid of the amplitude spectrum. It has generally been reported to correspond to the semantic descriptors "metallic," "sharp," or "brilliant." The second dimension was correlated with the logarithm of the attack time of the amplitude envelope, and corresponds to the semantic descriptors "fast-slow attack," "resonant," or "dry." The third dimension was correlated with the spectral irregularity (logarithm of the spectral deviation of component amplitudes from a global spectral envelope derived from a running mean of the amplitudes of three adjacent harmonics) or the spectral flux (average of the correlations between amplitude spectra in adjacent time windows).

11.2.2.4 Prerequisites for using MDS to study auditory perception

11.2.2.4.1 Controlling loudness, pitch, and duration. It is important to emphasize that the musical sounds used in the studies previously mentioned were equalized in pitch, subjective duration, and loudness, so that ratings would only concern the differences in timbre. Indeed, certain auditory attributes, such as loudness, might dominate and overpower less salient ones, as mentioned in Section 11.2.1.3 for semantic scales. Two sounds that differ mainly in terms of loudness will be judged obviously different according to this dimension, with little contribution from other dimensions of variation being taken into account.

11.2.2.4.2 Selecting a homogeneous corpus of sounds. As mentioned earlier, MDS is hypothesized to represent a corpus of sounds by a limited number of continuous auditory dimensions that are common to all the sounds. That means the corpus has to be composed of homogeneous sound objects (sounds produced by the same type of object or stimuli that sound rather similar, e.g., a class of car sounds) in order to avoid a perceptual structure that is strongly categorical for which the MDS approach is not adapted (see next section). A cluster analysis on the similarity ratings can reveal the degree of homogeneity of the sound corpus. If the tree structure obtained reveals a strong categorization of the corpus, it is advisable to determine which categories best represent the objectives of the study in order to obtain appropriate stimuli.

11.2.2.4.3 Limiting the number of sounds. As participants may become fatigued or lose motivation over time, application of the MDS technique is restricted to a rather small number of sounds (more or less 20 well-chosen sounds), because the number of pairs ($N(N - 1)/2$) grows rapidly with the number of sounds (N). Thus a preliminary categorization experiment may be advisable in order to select the most representative sounds (see Susini, McAdams, Winsberg, Perry, Vieillard, & Rodet, 2004). Another possibility to avoid being confined to a small number of stimuli is to use sorting tasks. Indeed, the validity of using sorting tasks for sounds instead of paired comparisons has been tested and shown to be effective with two different sets of auditory stimuli (Bonebright, 1996). However, further tests have to be performed in order to confirm the validity for collecting data using sorting tasks.

11.2.2.4.4 Collecting information from participants. Once the perceptual configuration is obtained, it is important to identify the perceptual meaning of each dimension or even to label the dimensions using semantic descriptors, and also, to give a physical interpretation by establishing systematic relations between the stimulus characteristics and their locations in the space. Knowledge and familiarity with the sound corpus and perceptually relevant acoustic parameters are thus necessary in order to characterize the dimensions of the space objectively. Another option is to directly ask the participants to describe which sensation they attended to while judging the dissimilarities.

11.2.3 Sorting tasks

The MDS technique is not appropriate for sets of sounds caused by very different and obviously identified sources. For instance, Susini, Misdariis, McAdams, & Winsberg (1998) applied an MDS analysis to an extremely heterogeneous set of environmental sounds (trains, cars, and planes). The analysis yielded a strongly categorical perceptual structure: listeners identified the sound sources rather than comparing them along continuous dimensions. Therefore, this predominant cognitive factor—recognition, classification, and identification of the sound source (see McAdams, 1993)—violated the assumption of underlying continuous dimensions required by the MDS technique. In this case, other experimental approaches are needed and, particularly, the sorting tasks.

11.2.3.1 Sorting task, categorization, and auditory cognition

Sorting tasks are very commonly used in cognitive psychology to address the questions of identification and categorization of sound sources. These questions are tightly bound: identifying the source of sound can be viewed as connecting auditory perception to concepts, and concepts to language, in a bidirectional relationship (McAdams, 1993; Goldstone & Kersten, 2003). Several approaches to the organization and processing of concepts and categories have been developed (see Goldstone & Kersten, 2003, or Komatsu, 1992 for a review). Before presenting the technical procedure of sorting tasks, we briefly recall the general principles of the prototypical approach to categorization developed by Rosch (1978), which is very often used as an underlying framework in sorting tasks. This approach is based on the notion of similarity and is therefore well adapted to account for perceptual concepts such as those used to describe sounds.

Rosch's approach to categorization relies on two principles. First categorization is based on the *cognitive economy* principle: categories allow organisms to handle the infinite number of stimuli by treating them as equivalent when the differentiation is irrelevant for the purpose at hand. The second principle is that *the world has structure*. Categorization of the world is thus not arbitrary, but relies on its perceived structure (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

Two concepts are often borrowed from Rosch's work: first, Rosch and her colleagues have experimentally identified three *levels* in taxonomies of objects:

- The base level: items in these categories share many elements in common.
- The superordinate level: this level is more inclusive than the base level, but items in the categories at this level share fewer elements in common.
- The subordinate level: items in the categories at this level share many elements in common, but the classes are less inclusive.

Second, Rosch has introduced the notion of analog category membership: categories are internally structured into a *prototype* and nonprototype members. For these latter members, there is a gradient of *category membership* (Rosch et al., 1978).

11.2.3.2 Method and analysis

In a *sorting task*, listeners are required to sort a set of sounds and to group them into classes. When the experimenter does not specify any specific criteria that the listeners have to use, the task is called a *free-sorting task*. Usually, the listeners are also required to indicate the meaning of each class.[*] Sometimes, the listeners also have to select a prototype in each category (the most representative member).

Technically, because personal computers are widespread in the lab,[†] the procedure amounts to providing the listeners with an interface allowing them to listen to the sounds by clicking on icons and moving the icons so as to form groups.

To analyze the results, the partition of the sounds created by each listener is coded in an *incidence matrix* (in the matrix, 0 indicates that two sounds were in separate groups and 1 that they were in the same group). A *co-occurrence* matrix is then obtained by summing the incidence matrices, which can be interpreted as a proximity matrix (Kruskal & Wish, 1978). Therefore, as with dissimilarity ratings, sorting tasks result in estimating similarities between the sounds. However, the structure of these data might be different depending on the procedure. For instance, Aldrich,

[*] A *classification* of the sounds is the result of a sorting task. "Categories are equivalence classes of different (i.e., discriminable) entities and categorization is the ability to form such categories and treat discriminable entities as members of an equivalence class" (Sloutsky, 2003, p. 246).

[†] Things were rather more complicated without computers; see Vanderveer (1979).

Hellier, and Edworthy (2009) showed that dissimilarity ratings encouraged participants to use acoustical information, whereas a free-sorting procedure emphasized categorical information. Different techniques are available to visualize the proximity data. When the data follow the triangular inequality, but not the ultrametric inequality,* they are best represented in a low-dimensional geometrical space (e.g., by using MDS). When they also follow the ultrametric inequality, they are best represented in a tree representation (Legendre & Legendre, 1998). Cluster analyses create such representations. The most popular tree representation is the *dendrogram*. It consists in representing the data in a hierarchical tree. In such a tree, the leaves represent the sounds, and the height of the node that links two leaves represents the distance between the two sounds. The representation is hierarchical, well suited to represent class inclusion, and therefore fits well with Rosch's framework.

11.2.3.3 Examples of urban soundscape categorization

Sorting tasks have been largely used to study the categorization of everyday sounds and soundscapes† (Guyot, 1996; Guyot, Castellengo, & Fabre, 1997; Vogel, 1999; Maffiolo, Dubois, David, Castellengo, & Polack, 1998; Guastavino, 2007; see Schulte-Fortkamp & Dubois, 2006, for a review of recent advances).

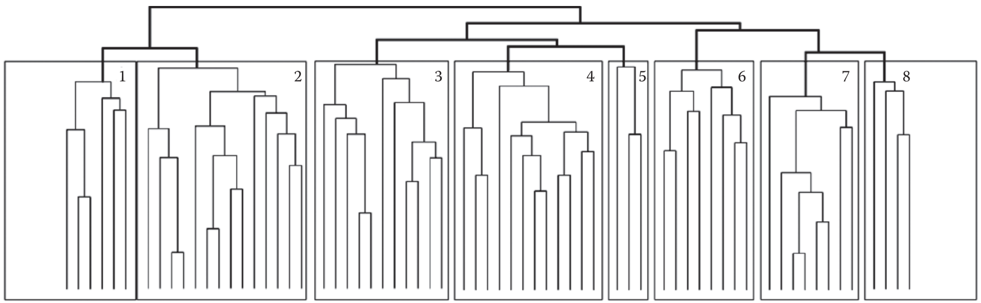
More recently, Tardieu, Susini, Poisson, Lazareff, and McAdams (2008) conducted an experiment that aimed to highlight the different types of auditory information that are perceived in the soundscapes of train stations. The goal was also to determine the information that participants used in the recognition of the space typology. Sixty-six soundscape samples were presented to participants in a free-categorization task with verbalization. The results showed that the listeners grouped together the samples into eight global categories. Further analysis aimed to explain the categories on the basis of the free verbalizations. Each verbalization was reduced to the words that contained a descriptive meaning. For example, the text "I have grouped here the sequences that took place in a ticket office. We clearly hear people talking about price and ticket" is reduced to the words "ticket office, clearly hear, people, talking about price." This reduction was made with the help of the software LEXICO (2003), which automatically counts every word in a text. Then, words are grouped into semantic fields that are deduced from the verbal descriptions. Five semantic fields were deduced (Figure 11.2): sound sources (e.g., trains, departure boards, ticket-punching machines, whistle, etc.), human activities (e.g., conversations, steps,

---

* The triangular inequality states that for any three points A, B, and C, $d(A,C) \geq d(A,B) + d(B,C)$, where d is the distance between the two points. In a Euclidean space, the length of any side of a triangle cannot be greater than the sum of the other two sides. In an ultrametric space, this inequality is replaced by $d(A,C) \leq \max\{d(A, B), d(B, C)\}$. In this kind of space, any given side must be less than or equal to the longer of the other two sides. Note that this is less constraining than the Euclidean case. The ultrametric inequality is to most forms of hierarchical clustering what the triangle inequality is to two-way multidimensional scaling.

† The term "soundscape" was introduced in the late 1970s by the Canadian composer R. Murray Schafer (1977), who defined soundscape as the auditory equivalent to landscape. Beside Schafer's project, the term soundscape perception is used in a scientific context to characterize how inhabitants perceive, experience, and appraise their sonic environment.

Participants' Categories:



Participants' Verbalizations:

Semantic Fields:	1	2	3	4	5	6	7	8
Sound sources	Ticket dispenser	–	Announcement music	Wheeled suitcases	Departure board	Trains	Jingles	Music, trains
Human activities	Conversation, waiting, steps	Voice, people	Waiting conversations	Steps, voice		Arrivals, departures	Voice, arrivals, departures	Voices, conversations
Room effect	Small space	Closed space, small	Reverberation, confused	Echoes, resonances		External	Resonances, reverbs.	Large reverberation
Type of space	Waiting room, ticket office	Bar, waiting room	Bar, hall	Corridor, hall		Platforms	Hall	Hall
Judgments	Quiet space	Quiet, intimate	Unpleasant	Noisy, rhythmic		–	–	Noisy, unpleasant

Figure 11.2 *(See color insert.)* Dendrogram representing the categories of train station soundscapes found in Tardieu et al. (2008). The descriptions of the categories provided by the participants are reported on the table below, grouped into five semantic fields. The most cited verbalizations are shown in a larger size font.

transaction, departure, etc.), room effect (e.g., reverberation, confined, exterior/interior, etc.), type of space (e.g., waiting room, platforms, halls, etc.), and personal judgment (e.g., annoying, pleasant, beautiful, musical, etc.).

11.2.3.4 Prerequisites for using sorting tasks

The sorting task is very intuitive for the listeners, and, in the case of the free-sorting task, has the great advantage of leaving the listeners free to arrange the sounds as they wish. Contrary to dissimilarity ratings, a large number of the sounds can be handled by the listeners in a session.

11.2.3.4.1 Considering a large number of existing sounds. It is possible with sorting tasks to test many existing sounds that are representative of the variety of sounds under consideration. For instance 74 environmental sounds were presented in Bonebright's study (2001), 150 recorded sound effects in Scavone, Lakatos, and Harbke's study (2002), 48 alarm sounds in Susini, Gaudibert, Deruty, and Dandrel's study (2003), and 66 train station soundscapes in Tardieu et al.'s study (2008).

11.2.3.4.2 Collecting information on the type of similarities used for each category. From a practical point of view, contrary to the MDS approach, categorization tasks are well adapted to describe perceptually heterogeneous corpora of sounds and to reveal different levels of similarities between the sounds. However, great care has to be taken when analyzing the categories because the type of similarities used by the participants may vary from one category to another, depending on the difficulty in identifying the sounds and on the expertise of the participants (more or less skill with sound evaluation). Indeed, three types of similarities have been identified (Lemaitre et al., 2010), based on acoustical properties (loudness, roughness, intensity fluctuations, etc.), identified physical interactions causing the sound (impact sound on glass, rattle sound on metal, sound effect, etc.) and meanings associated with the identified sound sources (sounds of breakfast, sounds that remind one of childhood, etc.).

11.2.3.4.3 Selecting the type of similarities. Semantic analyses of the verbal descriptions of the categories provide rich insights that reveal, on the one hand, the strategy used by the participants to form the categories, and on the other hand, the type of information used. However, semantic analyses are often time consuming and have to be done rigorously by experts. Lemaitre et al. (2010) proposed an alternative, which consists of asking the participants to rate for each category which type of similarity (acoustical, causal, semantic) they had used. The results may help the experimenter to understand the level of the perceptual structures underlying each category.

11.3 Application: Sound quality of environmental sounds

The quality of the acoustic environment is currently an important issue. Efforts are being made to account for the annoyance caused by noises (Guski, 1997). At the same time, designers are seeking to improve the sound quality of industrial products.

The idea of sound quality has emerged relatively recently. It refers to the fact that the sounds produced by an object or product are not only annoying or unpleasant, but are also a way for people to interact with an object. In the case of industrial products, it is therefore of major importance to design sounds to meet consumer expectations.

Since the beginning of the 1990s, sound quality has been conceived of mainly in the paradigm of psychoacoustics. This has led to the design of experimental methods and auditory descriptors relevant to sound quality. For instance, Zwicker and Fastl (1999) asked participants to rate *pleasantness* on a unidimensional scale (e.g., ratio scale). Then the pleasantness scores were correlated with psychoacoustical descriptors. Ellermeier, Mader, and Daniel (2004) gathered preference judgments of environmental sounds using a 2AFC (two alternative forced choice) procedure and analyzed them using the BTL technique (Bradley–Terry–Luce). This technique represented the perceived unpleasantness on a ratio scale. The unpleasantness scores were then predicted by a linear combination of psychoacoustic descriptors (roughness and sharpness). The semantic differential technique is also used to evaluate sound quality. It has been largely used for cars (Bisping, 1997; Chouard & Hemepl, 1999), vacuum cleaners (Ih et al., 2002), and refrigerators (Jeon, 2006). However, as noted in Section 11.2.3.1, defining the appropriate semantic descriptors of the scales must be done carefully.

Most of the studies use psychoacoustical descriptors (loudness, roughness, etc.) to explain unpleasantness scores or semantic ratings. These descriptors are currently included in most sound quality software packages, yet they are not always adapted to describing all kinds of everyday sounds. Indeed, it appears that relevant perceptual dimensions are different from one study to another according to the corpus of sounds under consideration. Therefore, there are no "universal" acoustical or psychoacoustical descriptors that can be used to measure relevant auditory attributes for all categories of environmental sounds, and which would thus provide the same effect on the sound quality of any product.

11.3.1 Application of the MDS technique to describe environmental sounds

A crucial aspect for the research in sound quality is to determine the relevant auditory attributes related to a specific family of environmental sounds. The MDS technique has been shown to be a fruitful tool for revealing and characterizing the unknown perceptual dimensions underlying the timbre of musical sounds. During the last decade, the MDS technique has been successfully applied to different kinds of environmental sounds: everyday sounds (Bonebright, 2001), interior car sounds (Susini, McAdams, and Smith, 1997), air-conditioning noises (Susini et al., 2004), car door closing sounds (Parizet, Guyader, and Nosulenko, 2006), and car horn sounds (Lemaitre, Susini, Winsberg, McAdams, and Letinturier, 2007). For all the mentioned studies, MDS analyses led to 3-D perceptual spaces (Figure 11.3 presents the 3-D space obtained for car sounds) and all the dimensions except one were described by different acoustical parameters.

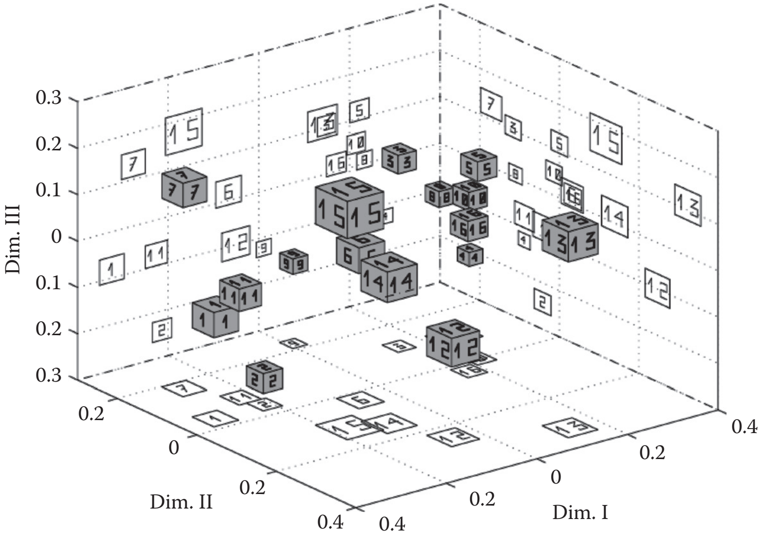


Figure 11.3 (See color insert.) Three-dimensional space for car sounds: dimension I is explained by the energy ratio between the harmonic and noisy parts, dimension II by the spectral centroid, and dimension III by the decrease in the spectral envelope.

The spectral centroid* is the acoustical descriptor shared by all the perceptual spaces related to environmental sounds. Therefore, this descriptor appears to describe musical sounds as well as environmental sounds and is related to the semantic descriptors “metallic,” “sharp,” or “brilliant.” Aside from the spectral centroid, no universal auditory attributes exist to characterize the timbre of any sound, and an inventory of the different salient auditory attributes to describe the different family of sounds is needed. A meta-analysis of 10 published timbre spaces conducted by McAdams, Giordano, Susini, Peeters, and Rioux (2006) using multidimensional scaling analyses (CLASCAL) of dissimilarity ratings on recorded, resynthesized or synthesized musical instrument tones, revealed four primary classes of descriptors: spectral centroid, spectral spread, spectral deviation, and temporal envelope (effective duration/attack time).

* The spectral centroid is the weighted mean frequency of the spectrum of the signal; each partial tone is weighted by its corresponding amplitude. The calculation of this feature can be more or less complex (see the work by Misdariis et al., 2010), but the basic expression is:

$$SC = \frac{\sum_i A_i \times f_i}{\sum_i A_i}$$

where A_i and f_i are the amplitude and frequency of the corresponding partial.

11.3.2 A general framework for sound quality

In a more general framework, the MDS technique may be combined with another approach based on a semantic study of the corpus of sounds under consideration, in order to map preference judgments onto both relevant objective descriptors and appropriate semantic descriptors. Figure 11.4 presents the framework of the different stages of these related approaches. This general framework was applied using air-conditioning noises as an example in a three-part study by Susini, Perry, Winsberg, Vieillard, McAdams, and Winsberg (2001), Siekierski et al. (2001), and Junker, Susini, and Cellard (2001).

The first step consists in determining the perceptual space using the MDS technique. Then, in a second step, the acoustical descriptors that are correlated with the positions of the sounds along the perceptual dimensions are determined. In a parallel third step, the sounds are verbally described through a descriptive analysis that involves a small number of trained listeners. This step provides a list of selected semantic descriptors—which will be used to define relevant semantic scales—and a verbal description of the auditory cues used by the participants to compare the sounds in order to guide the research of the objective descriptors correlated with the auditory dimensions obtained in the previous stage. In the last step, participants rate their preference (or annoyance) of the sounds. The degree of preference (or, inversely, annoyance) associated with each sound is related to a function of the significant objective descriptors on the one hand, and the semantic descriptors on the other. The advantage of this global approach is that it does not limit the exploration and characterization of the components of sound quality to acoustical and semantic descriptors that are already known. It provides a method for finding new objective

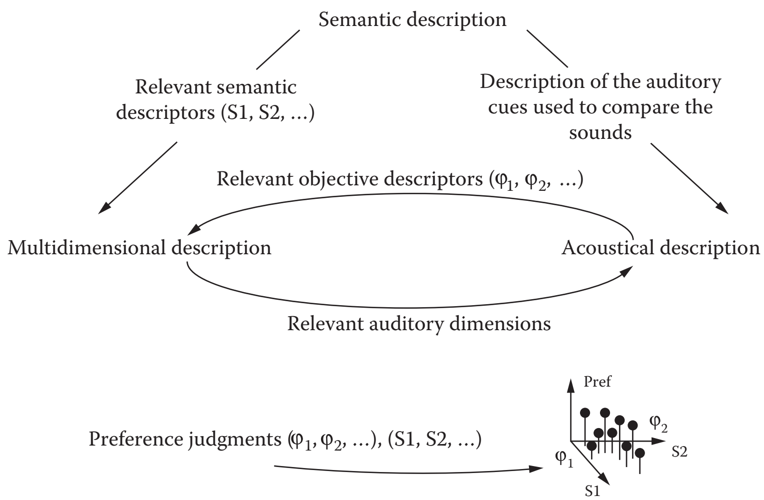


Figure 11.4 Framework for a global sound quality approach, involving multidimensional, acoustical, and semantic descriptions combined with preference judgments, based on Susini et al. (2001) and Siekiersky et al. (2001).

and semantic descriptors that are perceptually relevant for describing and evaluating a sound object in the design process.

11.4 Perspectives: Sounds in continuous interactions

The methods reported in this chapter all address the measurement of quantities representative of what a human listener perceives. The evaluation of the perceived *sound quality* of industrial objects is a very important domain in which these methods are applied. Traditionally, the paradigm of sound quality evaluation considers a listener passively receiving information from the sounds of the product. Such evaluations would, for instance, study the acoustical properties of a car engine roar that a user prefers (aesthetics) and that are representative of a sports car (functionality).

New technologies for sensing and embedded computation, however, have made it possible for designers to consider sonic augmentations of a much wider array of everyday objects that incorporate electronic sensing and computational capabilities. Where continuous auditory feedback is concerned, the sound is no longer produced in a static or isolated way, but is rather coupled to human action in real time. This new domain of applications is called *sonic interaction design*.

From the standpoint of perception, the level of dynamical interactivity embodied by such artifacts is very different from the situation of passive listening in which most of the methods reported are carried out. In sonic interactions, participants are not listening to sequences of static sounds selected by an experimenter, but instead dynamically explore the sounds of an interactive object. This context may be thought to be more closely allied with enactive views of perception (e.g., Bruner, 1966) than with some of the more traditional approaches found in experimental auditory psychology.

The study of sonic interaction entails an understanding of perceptual–motor behavior, because these processes underlie any form of human interaction. New methods may therefore be required. Such methods experimentally study how users perform when required to do a task involving sonic interaction. An interesting example is provided in work by Rath (Rath & Rocchesso, 2005; Rath, 2006, 2007; Rath & Schleicher, 2008). They describe the Ballancer, a tangible interface consisting of a wooden plank that may be tilted by its user in order to drive a virtual ball rolling along the plank. The authors used this interface to study participants' abilities to use this auditory feedback in a task involving guiding the ball to a target region along the length of the plank, depending on the kind of sound used. Lemaitre et al. (2009) used another tangible interface (the Spinotron) implementing the metaphor of a child's spinning top to study how continuous sonic interactions guide the user in making a precise gesture.

References

Aldrich, K. M., Hellier, E. J., & Edworthy, J. (2009), What determines auditory similarity? The effect of stimulus group and methodology, *Quarterly Journal of Experimental Psychology, 62(*1), 63–83.

Bisping, R. (1997). Car interior sound quality: Experimental analysis by synthesis, *Acta Acustica United with Acustica, 83,* 813–818.

Björk, E. A. (1985). The perceived quality of natural sounds, *Acustica, 57,* 185–188.

Bonebright, T. L. (1996). *Vocal affect expression: A comparison of multidimensional scaling solutions for paired comparisons and computer sorting tasks using perceptual and acoustic measures*, Doctoral dissertation, University of Nebraska.

Bonebright, T. L. (2001). Perceptual structure of everyday sounds: A multidimensional scaling approach. In *ICAD*, Helsinki, Finland.

Brunner, J. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Chouard, N., & Hempel, T. (1999). A semantic differential design especially developed for the evaluation of interior car sounds. In *Joint Meeting: ASA/EAA/DEGA* (*JASA* (105), Berlin), p. 1280.

Daniel, P., & Weber, R. (1997). Psychoacoustical roughness: Implementation of an optimized model," *Acta Acustica United with Acustica, 83*, 113–123.

Ellermeier, W., Mader, M., & Daniel, P. (2004). Scaling the unpleasantness of sounds according to the BTL model: Ratio-scale representation and psychoacoustical analysis, *Acta Acustica United with Acustica, 90,* 101–107.

Fastl, H. (1989). Average loudness of road traffic noise. In *InterNoise 89*, Newport Beach, CA.

Fastl, H. (1991). Evaluation and measurement of perceived average loudness. In A. Schick, J. Hellbrück, and R. Weber (Eds.), *Fifth Oldenburg Symposium on Psychological Acoustics* (BIS, Oldenburg), 205–216.

Fastl, H. (1997). The psychoacoustics of sound-quality evaluation, *Acta Acustica United with Acustica, 83,* 754–764.

Faure, A. (2010). Des sons aux mots, comment parle-t-on du timbre musical? Editions Edilivre—APARIS, Paris.

Fechner, G. T. (1966). *Elements of psychophysics*. New York: Holt, Rinehart and Winston (Original 1860).

Fletcher, H. (1940). Auditory patterns, *Review of Modern Physics*, *12,* 47–65.

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

Gaver, W. W. (1993b). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, *5,* 285–313.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds, *Journal of Audio Engineering Society, 50,* 331–342.

Goldstone, R. L., & Kersten, A. (2003). Concepts and categorization. In A. F. Healy & R. W. Proctor (Eds.), *Comprehensive handbook of psychology, Vol. 4: Experimental psychology*, chapter 22, pp. 599–621. New York: Wiley.

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

Gros, L., & Chateau, N. (2001). Instantaneous and overall judgments for time-varying speech quality: Assessments and relationships, *Acta Acustica United with Acustica, 87,* 367–377.

Guastavino, C. (2007). Catagorisation of environmental sounds, *Canadian Journal of Experimental Psychology, 61,* 54–63.

Guski, R. (1997). Psychophysical methods for evaluating sound quality and assessing acoustic information, *Acta Acustica United with Acustica, 83,* 765–774.

Guyot, F. (1996). Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative: Une approche par la catégorisation. Doctoral dissertation, Université du Maine.

Guyot, F., Castellengo, M., & Fabre, B. (1997). Etude de la catégorisation d'un corpus de bruits domestiques. In D. Dubois (Ed.), *Catégorisation et cognition,* Paris: Kimé.

Hansen, M., & Kollmeier, B. (1999). Continuous assessment of time-varying speech quality. *Journal of the Acoustical Society of America 106,* 2888–2899.

Hartmann, W. H. (1997). *Signals, sound, and sensation*. New York: AIP Press.

Hedberg, D., & Jansson, C. (1998). *Continuous rating of sound quality*. Stockholm: Karolinska Institutet.

Ih, J.-G., Lim D.-H., Jeong H., & Shin, S.-H. (2002) Investigation on the correlation between sound quality and spectral composition vacuum cleaner sound by using the orthogonal array. In *Sound Quality Symposium*, Dearborn, MI.

Jeon, J. Y. (2006). Sound radiation and sound quality characteristics of refrigerator noise in real living environments, *Applied Acoustics, 68,* 1118–1134.

Junker, F., Susini, P., & Cellard, P. (2001). Sensory evaluation of air-conditioning noise: Comparative analysis of two methods. In *ICA,* Rome, Italy, p. 342.

Kendall, R. A., & Carterette, E. C. (1992). Verbal attributes of simultaneous wind instrument timbres. I. von Bismarck's adjectives. *Music Perception, 4,* 185–214.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112*(3), 500–526.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielsen, and O. Olsson (Eds.), *Structure and perception of electrostatic sound and music* (pp. 43–53). Amsterdam: Elsevier.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

Kuwano, S., & Namba, S. (1978). On the loudness of road traffic noise of longer duration (20 min) in relation to instantaneous judgment. In *ASA & ASJ* joint meeting. *(Journal of the Acoustical Society of America),* Berlin, *64,* 127–128.

Kuwano, S., & Namba, S. (1985). Continuous judgement of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness. *Psychology Research, 47,* 27–37.

Kuwano, S., & Namba, S. (1990). Continuous judgement of loudness and annoyance. In F. Müller (Ed.), *Fechner Day, Proceedings of the 6th Annual Meeting of the International Society for Psychophysics*, Würzburg, Germany, pp. 129–134.

Kuwano, S., & Namba, S. (2001). Dimensions of sound quality and their measurement. In *ICA,* Rome, Italy, p. 56.

Kyncl, L., & Jiricek, O. (2001). Psychoacoustic product sound quality evaluation. In *ICA,* Rome, Italy, p. 90.

Lamalle, C., Martinez, W., Fleury, S., Salem, A., Fracchiolla, B., Kuncova, A., & Maisondieu, A. (2003). *LEXICO 3, Outils de statistique textuelle—Manuel d'utilisation*, SYLED-CLA2T (Paris 3).

Legendre, P., & Legendre, L. (1998). *Numerical ecology*. Amsterdam: Elsevier.

Lemaitre, G., Houix, O., Misdariis, M., & Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds, *Journal of Experimental Psychology: Applied, 16,* 16–32.

Lemaitre, G., Houix, O., Visell, Y., Franinovic, K., Misdariis, N., & Susini, P. (2009) Toward the design and evaluation of continuous sound in tangible interfaces: The spinotron. *International Journal of Human Computer Studies*, *67*, 976–993.

Lemaitre, G., Susini, P., Winsberg, S., McAdams, S., & Leteinturier, B. (2007). The sound quality of car horns: A psychoacoustical study of timbre, *Acta Acustica United with Acustica, 93,* 457–468.

Madsen, C. K. (1996). Empirical investigation of the "aesthetic response" to music: Musicians and nonmusicians. In *4th International Conference on Music Perception and Cognition,* Montreal, Canada, pp. 103–110.

Maffiolo, V., Dubois, D., David, S., Castellengo, M., & Polack, J. D. (1998). Loudness and pleasantness in structuration of urban soundscapes. In *InterNoise,* Christchurch, New Zealand.

McAdams, S., & Bigand, E. (Eds.). (1993). *Thinking in sound—The cognitive psychology of human audition,* Oxford: University Press.

McAdams, S., Giordano, B., Susini, P., Peeters, G., & Rioux, V. (2006). A meta-analysis of acoustic correlates of timbre dimensions. In *ASA,* Hawaii.

McAdams, S., Vines, B. W., Vieillard, S., Smith, B., & Reynolds, S. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception, 22,* 297–350.

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

Meunier, S., Boullet, I., & Rabau, G. (2001). Loudness of impulsive environmental sounds. In *ICA,* Rome, Italy, p. 91.

Misdariis, N., Minard, A., Susini, P., Lemaitre, G., McAdams, S., & Parizet, E., (2010). Environmental sound perception: Metadescription and modeling based on independent primary studies. *Journal on Audio, Speech, and Music Processing.* doi:10.1155/2010/362013.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). New York: Academic Press.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, *74 (3)*, 750–753.

Moore, B. C. J., & Glasberg, B. R. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47,* 103–138.

Moore, B. C. J., & Glasberg, B. R. (1996). A revision of Zwicker's loudness model. *Acta Acustica United with Acustica, 82,* 335–345.

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness and partial loudness. *Journal of the Audio Engineering Society, 45,* 224–240.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin, 49*, 197–237.

Parizet, E. (1999). Comparison of two semantic-differential test methods. In *EAA & ASA joint meeting*. *Journal of the Acoustical Society of America*, Berlin, p. 1189.

Parizet, E., & Nosulenko, V. N. (1999). Multi-dimensional listening test: Selection of sound descriptors and design of the experiment. *Noise Control Engineering Journal, 47,* 1–6.

Parizet, E., Guyader, E., & Nosulenko, V. (2006). Analysis of car door closing sound quality. *Applied Acoustics, 69*, 12–22.

Parizet, E., Hamzaoui, N., and Sabatié, G. (2005). Comparison of some listening test methods: A case study. *Acta Acustica United with Acustica, 91,* 356–364.

Parizet, E., Hamzaoui, N., Segaud, L., & Koch J.-R. (2003). Continuous evaluation of noise discomfort in a bus. *Acta Acustica United with Acustica*, *89*, 900–907.

Patterson, R. D., & Holdsworth, J. (1991). A functional model of neural activity patterns and auditory images. *Advances in speech, hearing and language processing, Vol. 3.* London: JAI Press.

Peters, R. W. (1960). *Research on psychological parameters of sound*, (pp. 60–249). Hattiesburg: Mississippi Southern College.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. P. Smoorenburg (Eds.), *Frequency anaysis and periodicity detection in hearing* (pp. 397–414). Leiden: A. W. Sijthoff.

Pratt, R. L., & Doak, P. E. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration, 45,* 317–328.

Rath, M. (2006). On the relevance of auditory feedback for subjective and objective quality of control in a balancing task. In *Proceedings of the 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems*, Berlin, pp. 85–88.

Rath, M. (2007). Auditory velocity information in a balancing task. In *ICAD*, Montréal, Canada, pp. 372–379.

Rath, M., & Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE MultiMedia, 12*, 60–69.

Rath, M., & Schleicher, D. R. (2008). On the relevance of auditory feedback for quality of control in a balancing task. *Acta Acustica United with Acustica, 94*, 12–20.

Robinson, D. W., & Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics, 7,* 166–181.

Rosch, E. (1978). Principles of categorisation. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization,* (pp. 27–47). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., Mervis, C. B., Gray, W. D, Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–443.

Scavone, G. P., Lakatos, S., & Harbke, C. R. (2002). The sonic mapper: An interactive program for obtaining similarity ratings with auditory stimuli. In *ICAD,* Kyoto.

Schafer, R. M. (1977). *The tuning of the world*. New York: Random House.

Schubert, E. (1996). Continuous response to music using the two dimensional emotion space. In *ICMPC,* Montreal, Canada, pp. 263–268.

Schulte-Fortkamp, B., & Dubois, D. (2006). Recent advances in soundscape research. *Acta Acustica United with Acustica, 92*(6), 5–6.

Siekierski, E., Derquenne, C., & Martin, N. (2001). Sensory evaluation of air-conditioning noise: Sensory profiles and hedonic tests. In *ICA,* Rome, Italy, 326.

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception, 19,* 87–120.

Sloutsky, V. M. (2003) The role of similarity in the development of categorization. *Trends in Cognitive Science, 7,* 246–251.

Solomon, L. N. (1958), Semantic approach to the perception of complex sounds. *Journal of the Acoustical Society of America, 30*(5), 421–425.

Stepánek, J. (2006). Musical sound timbre: Verbal description and dimensions. In *DAFx-06,* Montréal.

Stevens, J. C., & Tulving, E. (1957). Estimations of loudness by a group of untrained observers, *American Journal of Psychology, 70,* 600–605.

Stevens, S. S. (1959). Cross-modality validations of subjective scales for loudness, vibrations, and electric shock. *Journal of Experimental Psychology, 57,* 201–209.

Susini, P., Gaudibert, P., Deruty, E., & Dandrel, L. (2003). Perceptive study and recommendation for sonification categories. In *ICAD*, Boston.

Susini, P., Houix, O., Misdariis, N., Smith, B., & Langlois, S. (2009.) Instruction's effect on semantic scale ratings of interior car sounds. *Applied Acoustics, 70*(3), 389–403.

Susini, P., & Maffiolo, V. (1999). Loudness evaluation of urban soundscapes by a cross-modal matching method. In *EAA & ASA joint meeting, Journal of the Acoustical Society of America*, Berlin, p. 943.

Susini, P., & McAdams, S. (1998). Global and continuous judgements of sounds with time-varying intensity: Cross-modal matching with a proprioceptive input device. In *ICA & ASA joint meeting*, *Journal of the Acoustical Society of America,* Seattle, p. 2812.

Susini, P., & McAdams, S. (2008). Loudness asymmetry ratings between accelerating and decelerating car sounds, in *Acoustics'08, ASA-EAA Joint Conference,* Paris.

Susini, P., McAdams, S., & Smith, B. (2002). Global and continuous loudness estimation of time-varying levels. *Acta Acustica United with Acustica, 88,* 536–548.

Susini, P., McAdams, S., & Smith, B. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica United with Acustica, 93,* 623–631.

Susini, P., McAdams, S., and Winsberg, S. (1997). Perceptual characterisation of vehicle noises. In *EEA Symposium: Psychoacoustic, in industry and universities,* Eindhoven, The Netherlands.

Susini, P., McAdams, S., Winsberg, S., Perry, I., Vieillard, S., & Rodet, X. (2004). Characterizing the sound quality of air-conditioning noise. *Applied Acoustics, 65*(8), 763–790.

Susini, P., Misdariis, N., McAdams, S., & Winsberg, S. (1998). Caractérisation perceptive de bruits. *Acoustique et Techniques*, *13*, 11–15.

Susini, P., Perry, I., Winsberg, S., Vieillard, S., McAdams, S., & Winsberg, S. (2001). Sensory evaluation of air-conditioning noise: Sound design and psychoacoutic evaluation. In *ICA,* Rome, Italy, p. 342.

Tardieu, J., Susini, P., Poisson, F., Lazareff, P., & McAdams, S. (2008). Perceptual study of soundscapes in train stations, *Applied Acoustics, 69,* 1224–1239.

Vanderveer, N. J. (1979) *Ecological acoustics: Human perception of environmental sounds.* Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Vogel, C. (1999). *Etude sémiotique et acoustique de l'identification des signaux sonores d'avertissement en contexte urbain*. Doctoral dissertation, Université Paris 6.

von Bismarck, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes, *Acustica, 30,* 146–158.

Vos, J. (1998). The loudness of impulse and road-traffic sounds in the presence of masking background noise, *Acta Acustica United with Acustica, 84,* 1119–1130.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of timbres. *Scandinavian Journal of Psychology, 13,* 228–240.

Weber, R. (1991). The continuous loudness judgement of temporally variable sounds with an "analog" category procedure. In A. Schick, J. Hellbrück, and R. Weber (Eds.), *Fifth Oldenburg Symposium on Psychological Acoustics,* (pp. 267–294) Oldenburg: BIS.

Winsberg, S., & Carroll, D. (1989). A quasi non-metric method for multidimal scaling via an extended Euclidian model. *Psychometrika, 54*, 217–229.

Zeitler, A., & Hellbrück, J. (2001). Semantic attributes of environmental sounds and their correlations with psychoacoustic magnitude. In *ICA,* Rome, Italy, p. 114.

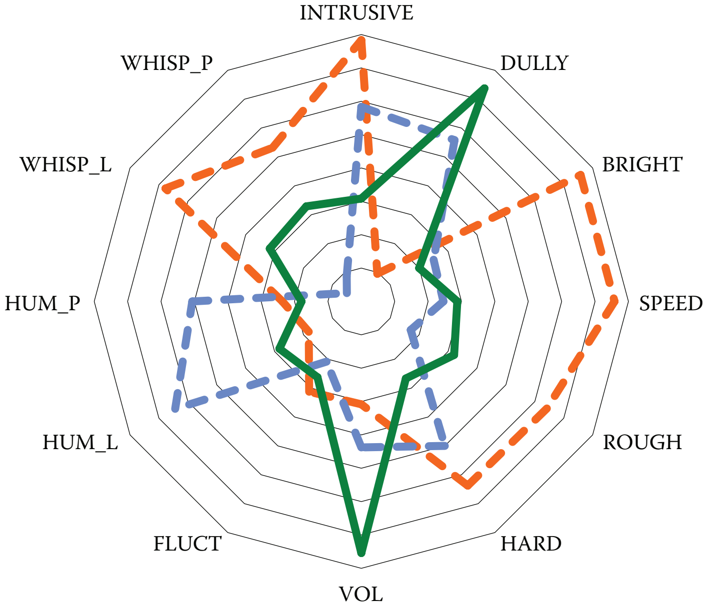
Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands, *Journal of the Acoustical Society of America, 33,* 248.

Zwicker, E. (1984). BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532-B, *Acustica, 55,* 63–67.

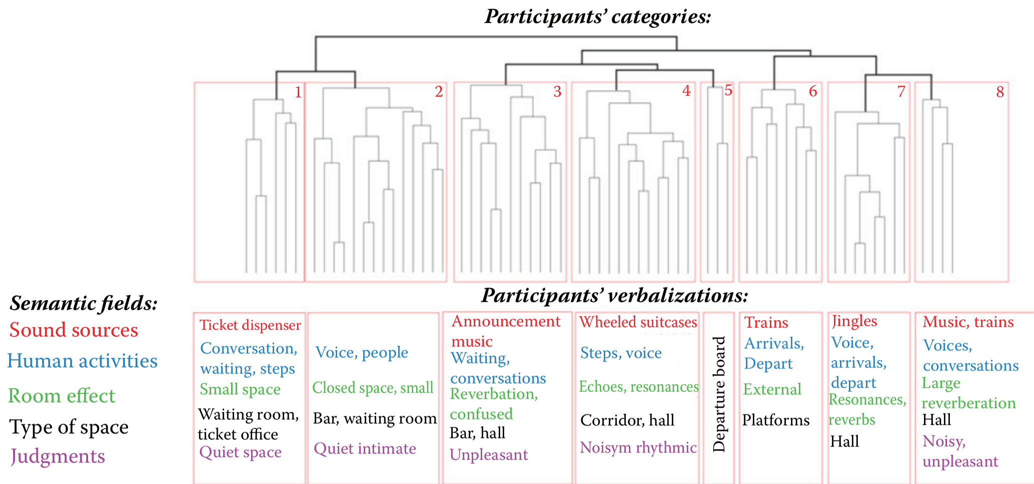
Zwicker, E., & Fastl, H. (1972). On the development of the critical band, *Journal of the Acoustical Society of America, 52*, 699–702.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2nd ed.). Heidelberg: Spinger-Verlag.

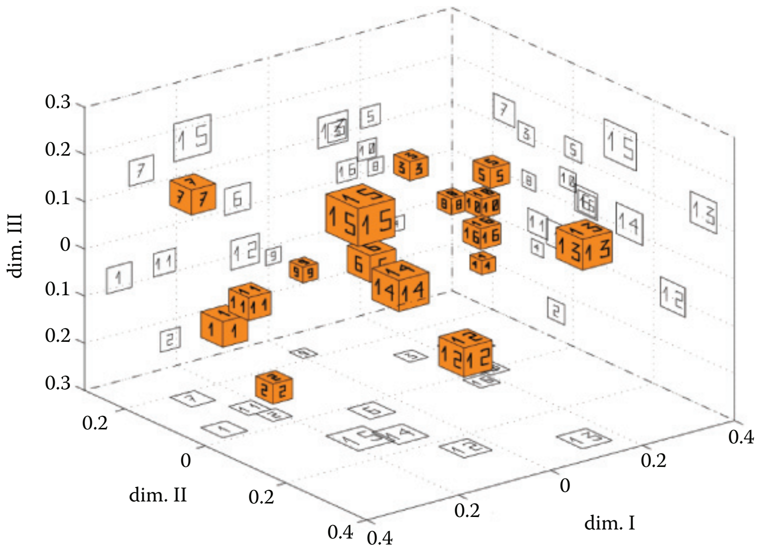
Zwicker, R., and Terhardt, E. (1980). Analytical expressions for critical band rate and critical bandwidth as a function of frequency, *Journal of the Acoustical Society of America, 68,* 1523–1525.



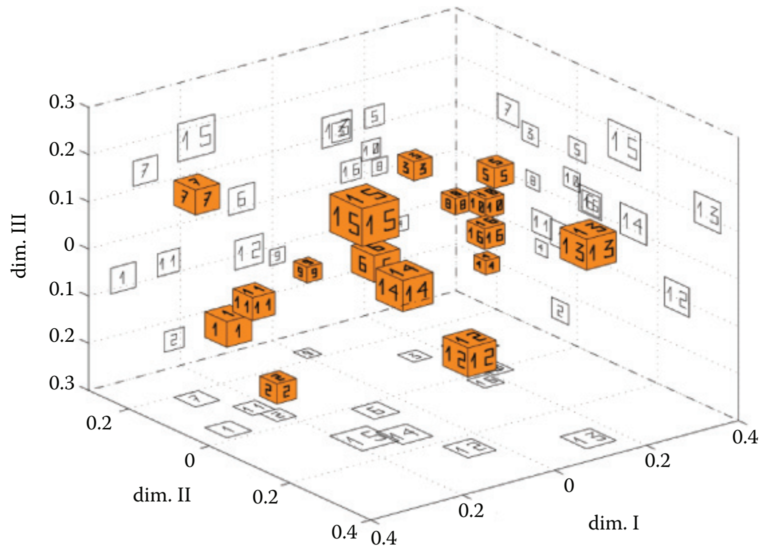
Color Figure 11.1 Sensory profiles obtained for three air-conditioning noises from Siekierski, Derquenne, and Martin (2001). Labels of the *semantic descriptors* are Intrusive, Dully, Brightness, Speed, Roughness, Hardness, Voluminous, Fluctuation, Humming, Whispering. The letters L and P correspond to the Level and Pitch, respectively, of the whispering (noise) part and the humming (motor) part.



Color Figure 11.2 Dendrogram representing the categories of train station soundscapes found in Tardieu et al. (2008). The descriptions of the categories provided by the participants are reported on the table below, grouped into five semantic fields. The most cited verbalizations are shown in a larger size font.



Color Figure 11.3 Three-dimensional space for car sounds: dimension I is explained by the energy ratio between the harmonic and noisy parts, dimension II by the spectral centroid, and dimension III by the decrease in the spectral envelope.



Color Figure 11.3 Three-dimensional space for car sounds: dimension I is explained by the energy ratio between the harmonic and noisy parts, dimension II by the spectral centroid, and dimension III by the decrease in the spectral envelope.