Timbre, Sound Quality, and Sound Design

Guillaume Lemaitre, Patrick Susini

Timbre, Sound Quality, and Sound Design

Guillaume Lemaitre [1]✉

Patrick Susini [1]

Email susini@ircam.fr

[1] STMS-IRCAM-CNRS-UPMC, Paris, France

Abstract

Sound quality evaluation applies the results of timbre research to the assessment of the sound quality of manufactured products (domestic appliances, transportation, etc.). This chapter first provides an overview of one methodology. A number of acoustic descriptors reflecting perceived timbre dimensions are established and used to predict users' preference judgements. Whereas such a methodology has proven very effective, it also has some limitations. In fact, most studies only consider the pleasantness of the sounds and often overlook other potential roles of sounds in products and interfaces. In the second part, the chapter introduces sound design. Whereas sound quality evaluation merely proposes a diagnostic of the timbre of existing products, sound design aims to create or modify the timbre of product sounds to meet specific intentions. These intentions consider the pleasantness, but also several other aspects of product sounds: functionality, identity, and ecology. All these aspects are interdependent and often closely related to the temporal and timbral characteristics of the sound. The chapter continues with a discussion of the roles and practices of sound designers and introduces a set of tools that foster communication about timbre between the different participants of a sound design process. In particular, the focus is on the necessity for these participants to share a common timbre vocabulary, and the potential impact of education about sounds is considered. Finally, an important functional aspect of product sound is discussed: how to design the timbre of sounds to support user interactions with the product.

Keywords

Auditory pleasantness
Product sound
Sonic interaction design
Sound evaluation
Sound function
Sound identity
Timbre descriptors

9.1. Introduction

Picture yourself driving your car. The engine purrs quietly. Then you need to overtake a slower car in front of you. The turn signal emits a satisfying mechanical tick and, at your command, the engine roars energetically; the rising engine sound (its timbre, level, pitch) matches the sporty performance that you expect of your car. As you enter a denser traffic area, passing by other vehicles creates an alternating whoosh that is quite unpleasant.

Similar to cars, most manufactured products make sounds. Some of these sounds are useful, some are annoying, some are funny, some are intrusive, and some are intriguing. Some may enhance the perceived quality of a product (the tick of a luxury watch), whereas some others are so inappropriate that they are deleterious to the overall impression of the product (the irritating hiss of a poorly fitted vacuum cleaner hose). It is therefore important for product makers to evaluate how users perceive the sounds made by the products, in other words, their sound quality. Furthermore, it is extremely useful for product designers to be able to connect the perceived quality of a product sound to measurable quantities.

Predicting the perceived sound quality of a product from quantities measured on the sound (such as timbre features) is the purpose of sound quality evaluation that is discussed in Sect. 9.2. The outcome of the methodology is an algorithm that takes a sound signal at the input and produces a numerical indicator of quality at the output. This methodology has proven very useful for a number of industrial products (cars and transportation in particular). It also has a number of limitations and, in particular, considers only one aspect of the perception of product sounds: whether they are pleasant or unpleasant.

Product sounds are not only pleasant or unpleasant, however. They serve many other purposes: they contribute to the brand image and the coherence of a product, elicit emotional reactions in users, and even have functional aspects in terms of information. As such, product designers not only want to diagnose the quality of a product sound, they also want to design its timbral and temporal characteristics to address different interdependent aspects, such as pleasure, identity, and functionality, as well as taking into account the environment in which it will be heard. As an example, most people in France associate the jingle played before any vocal announcement in French railway stations with the French national railway company (SNCF). The timbral features and temporal properties of the jingle have been specifically designed to attract the attention of users and to communicate the values of the company. In addition, this sound has been designed to be enjoyable in the complex sonic environment of railway stations. Therefore, Sects. 9.3.1 and 9.3.2 of this chapter discuss the process of sound design for the creation of a new sound and how the design process considers different aspects, such as functionality, pleasantness, identity, and ecology, and their relation to timbre. Then the problem of a common vocabulary to communicate about timbral characteristics during a project design process that involves different participants is discussed in Sects. 9.3.3 and 9.3.4. In particular, a lexicon based on previous studies, using semantic descriptions of timbre as described by Saitis and Weinzierl (Chap. 5), is proposed to help participants learn to perceive and to communicate about timbre features and temporal properties. Finally, Sect. 9.3.5 focuses on *sonic interaction design*: using sounds produced by users' interactions with the product to guide or facilitate that interaction.

Coming back to the initial car example, most drivers of manual cars change gear based on how the loudness, pitch, and timbre of the engine sound changes as they step on the accelerator. This is an example of a sonic interaction, resulting from the physical behavior of the product. These sounds could in theory also be designed and engineered by electronic means, as this is already the case in many modern vehicles, to warn pedestrians of otherwise silent electric cars or to promote economic driving.

9.2. Sound Quality Evaluation: A Critical Overview

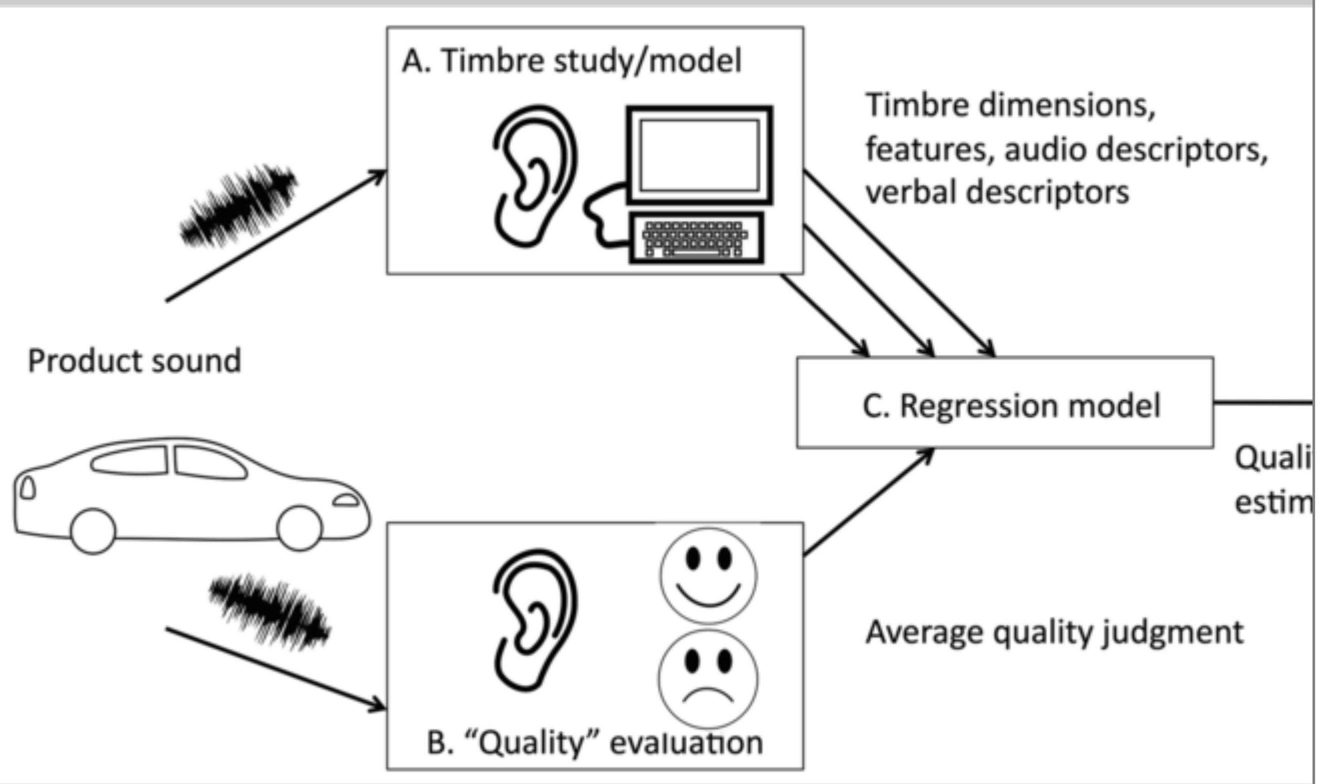
This section describes a methodology that is common to many sound quality studies reported in the literature. Such studies have a practical goal: provide the developer of a product with a tool to measure how users will appraise the sounds of the product or prototype. More specifically, this section focuses on studies that seek to develop a model that can estimate the perceived quality of a product sound (e.g., the sounds of different vacuum cleaners) from the sound signals alone (i.e., without conducting listening tests). In such studies, the quality is defined by a single numerical value, corresponding to the average judgement of a set of typical users listening to and evaluating the sounds. A general overview of this methodology is first provided in Sect. 9.2.1, followed by a more detailed description of its different parts (an even more detailed account is provided by Susini et al. 2011). The method is illustrated in Sect 9.2.3 by one practical example: the sound quality of air-conditioning units. The limitations of such a methodology are discussed in Sect. 9.2.3.

9.2.1. The Classical Methodology of Sound Quality Evaluation: Objectivation

Most sound quality studies available in the literature follow the methodology represented in Fig. 9.1. Starting from a set of recordings of a given product or family of products (e.g., a car engine, a camera, a vacuum cleaner, etc.), the methodology has three main parts. One part (detailed in Sect. 9.2.1.1) involves characterizations of the timbre of the product sounds as a set of *sound descriptors*, that is, numerical quantities calculated from the sound signal (sound descriptors are sometimes called sound features or metrics). For another part of the sound quality study, the researcher collects judgements about the "quality" of each of these sounds using listening tests, as described in Sect. 9.2.1.2. Finally, Sect. 9.2.1.3 details the mathematical models that connect the quality judgements with the sound descriptors. Thus, the final outcome of such a procedure is an algorithm that takes the sound signal as an input and produces a quantitative indicator (a numerical value) at the output, estimating the quality of the sounds. Such an indicator is extremely valuable in an industrial context since it allows the engineers and designers to quickly evaluate how users will appraise the sounds of the product without actually collecting judgements from them with time-consuming listening tests each time a new sound is created. In this context, the whole procedure is sometimes called *objectivation*: it estimates subjective judgements (i.e., resulting from listeners' evaluations) with an objective indicator (i.e., computed from the sound signal). This methodology is rather general, and there are several variations to each part. In the next section, details are provided for each of the parts.

Fig. 9.1

A common methodology found in many sound quality studies consists of connecting (**A**) timbre descriptors with (**B**) quality judgements through (**C**) a regression model (Original figure)



9.2.1.1. Characterizing the Timbre of Product Sounds

The goal of this section is to characterize each sound as a vector of numerical values that represent the listener's (i.e., the product user's) perception: the sound descriptors. Usually, only a handful of descriptors are considered. As such, sound descriptors can be considered as a low-dimensional representation of the sounds (of much lower dimensionality than the number of sound signals themselves), but they still convey the information that is important to the users in order to assess the quality of the product sounds.

Sound quality studies generally consider descriptors that are representative of a listener's perception: loudness, pitch (for sounds having pitch), duration, and timbre. As described by McAdams (Chap. 2), timbre is here considered as a multidimensional quality and characterized as a set of dimensions, each related to a sound descriptor. Product sound quality studies, therefore, use methods similar to those used to characterize the timbre of musical instruments. One such method, *semantic differentials*, consists of collecting a common vocabulary used by listeners to create scales. Another method, *dissimilarity judgements*, does not use words at all and relies instead on judgements of perceptual distances between sounds. Finally, many studies simply do not conduct any listening tests and rely on common timbre descriptors computed by software packages.

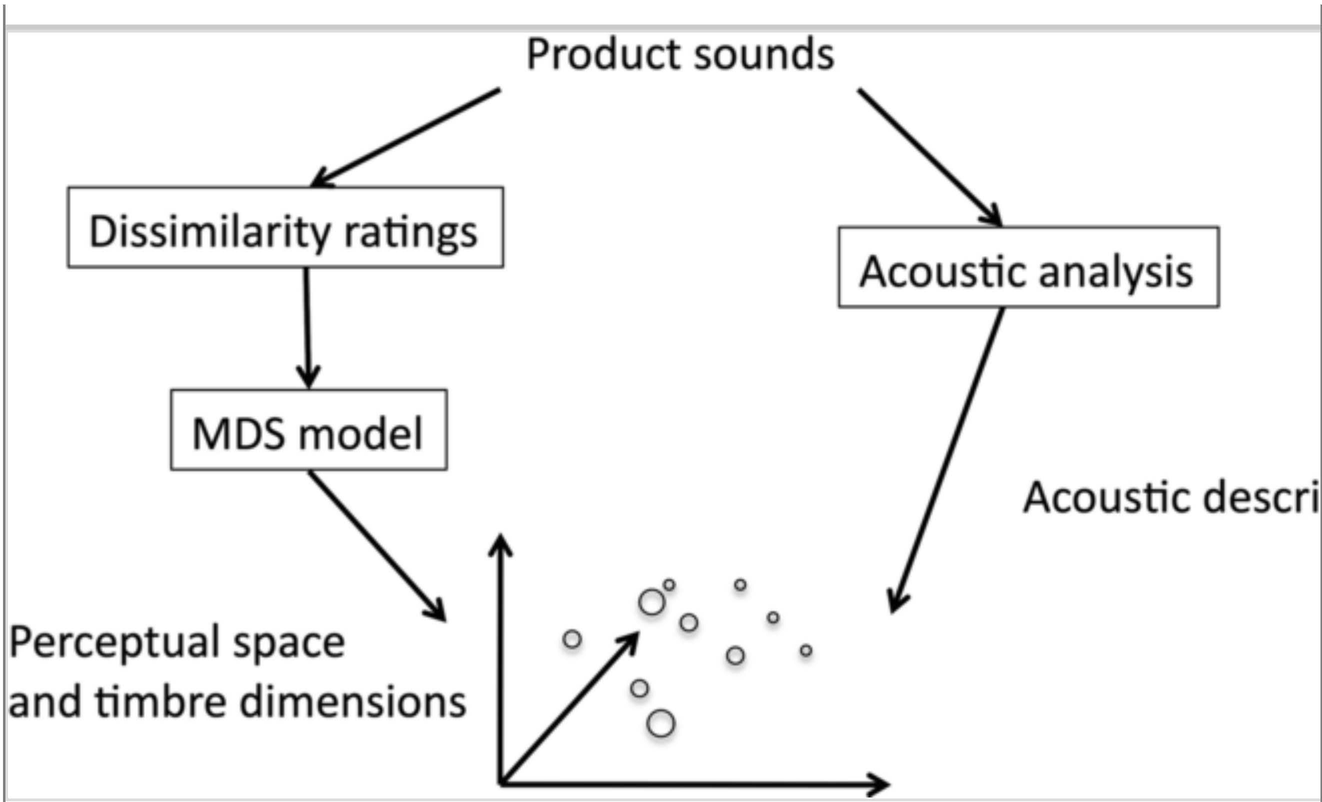
In the semantic differentials approach, the experimenters create scales, usually labeled by two opposed adjectives (e.g., clear/hazy, bright/dull): the semantic differentials (for more details, see Saitis and Weinzierl, Chap. 5). Test participants rate each sound along the set of scales and statistical techniques are used to cluster the scales into main (and often independent) factors. These main factors are interpreted and connected to the acoustic properties of the sounds, usually by listening to the sounds and picking out the best-correlated metrics.

The advantage of this method is that it allows the experimenters to generate a set of semantic descriptors of their product sounds: each main factor corresponds to a dimension of timbre that is actually perceived by the listeners and can be described by labels with a meaning shared by the listeners. In the context of sound quality evaluation, this method was used by Jeon et al. (2007), who studied refrigerator sounds with semantic differentials and identified four main factors: “booming” (and clustering pairs of adjectives such as booming/dry, trembling/flat, irregular/regular, etc.), “metallic” (metallic/deep, sharp/dull, etc.), and “discomforting” (unpleasant/pleasant, discomfort/comfort, etc.). The method, however, restricts the study precisely to what listeners can describe with words, and it is quite possible that some percepts, though perceived by listeners, may not be easily described with words. As such, the outcomes of the method strongly depend on the adjectives selected at the input in terms of relevance for describing a specific set of sounds but also in terms of the meaning of the words. This issue will be discussed in Sect. 9.3.

Instead of using a set of semantic scales, another method based on dissimilarity ratings and multidimensional scaling analysis (MDS) directly uses the ANSI definition of timbre: “the way in which musical sounds differ once they have been equated for pitch, loudness and duration” (American Standard Association 1960; Krumhansl 1989). This method was initially applied to characterize the timbre of musical instruments (Grey 1977; McAdams et al. 1995) and then of product sounds (Susini et al. 1999). In the first step of the method, listeners scale the dissimilarity between each pair of sounds from the set of product sounds under study (see Fig. 9.2).

Fig. 9.2

Dissimilarity ratings and multidimensional scaling (MDS) analysis. Listeners rate the dissimilarity between the two sounds of each possible pair from a set of product sounds. An MDS model is then fit to the data and yields a perceptual space in which the geometrical distance between two sounds corresponds to the perceived dissimilarity between them. The dimensions of the perceptual space are then interpreted by correlating them with acoustic descriptors. The diameter of the circles corresponds to the position along the depth dimension, with larger circles closer to the viewer (Original figure)



In a second step, an MDS algorithm creates a geometrical space in which the geometrical distance between two sounds represents the perceived dissimilarity between them. The dimensions of the space are interpreted as continuous dimensions of the timbre shared by the sounds under study. As for the semantic differential method, correlations between the dimensions and the sounds' features allows for the selection of acoustic descriptors that characterize each semantic dimension (for more detail on these two approaches, see McAdams, Chap. 2; Saitis and Weinzierl, Chap. 5).

The multidimensional scaling framework has the great advantage that it does not impose any predefined rating criteria on the listener. The task is thus simple and completely exploratory: as no dimension is postulated a priori, the outcome of the method (the dimensions and their descriptors) may be completely unexpected by the experimenters. This method was used by Lemaitre et al. (2007), who studied the timbre of car horn sounds and highlighted roughness, sharpness, and a third dimension specific to this set of sounds (spectral deviation), which could not have been specified without this exploratory method.

An alternative to these methods is to not conduct any listening test and simply rely instead on software packages that implement acoustic and psychoacoustic descriptors that have been found in previous timbre studies (McAdams, Chap. 2; Saitis and Weinzierl, Chap. 5). Ircam's and McGill University's "Timbre Toolbox" (Peeters et al. 2011) and Lartillot and Toiviainen's (2007) "MIR Toolbox" are popular sets of Matlab functions designed for Music Information Retrieval that implement many variations of these descriptors (and see Caetano, Saitis, and Siedenburg, Chap. 11). Among commercial packages, Head Acoustics' ArtemiS (https://www.head-acoustics.de, last retrieved on July 4, 2017), Genesis's LEA (http://genesis-acoustics.com, last retrieved on July 4, 2017), and Brüel and Kjær's PULSE have been widely used in industrial contexts. These software routines are incorporated into larger program suites that also do data acquisition, analysis, and reporting and are part of the basic toolkits for many industries.

These packages include at their core a few descriptors whose calculation was formalized by Zwicker and Fastl (1990): sharpness, roughness, and fluctuation strength. *Sharpness* corresponds to a percept whereby sounds can be ordered on a scale ranging from dull to sharp or bright. It is correlated with the spectral balance of energy: sounds with more energy in low frequencies are perceived as dull whereas sounds with more energy in high frequencies are perceived as bright or sharp. *Fluctuation strength* and *roughness* both correspond to the perception of amplitude modulations in the signal, each corresponding to a different range of modulation frequencies. When modulations are slow (around 4 Hz), the sounds are perceived as fluctuating (wobbling): this is the percept of fluctuation strength. Faster modulations (around 70 Hz) are perceived as rough (harsh): this is the percept of roughness. In addition to Zwicker and Fastl's descriptors, *tonalness* (also called pitch strength or pitch salience) also plays an important role. Tonalness refers to the magnitude of the sensation of pitch in a sound (from a weak to a strong sensation of pitch) (see Hansen et al. 2011). Usually, it is estimated as the ratio of manually identified tonal components over noisy components (tone-to-noise ratio, prominence ratio) (Terhardt et al. 1982), but the perceived tonalness of sounds with multiple tonal components is still under study (also see Saitis and Weinzierl, Chap. 5).

9.2.1.2. Measuring Quality

The other important part of any sound quality study consists of collecting "quality" judgements about the product sounds. The term quality is used here in a very broad sense, as it may actually correspond to several slightly different ideas: pleasantness, unpleasantness, annoyance, merit, preference, and others. In any case, quality judgements are always obtained through a listening test and are averaged over participants into a single numerical value for each sound. Eventually, the outcome of the whole sound quality study will be an algorithm that estimates these judgements with the aim of replacing listening tests. There are two main classes of methods: sound-wise scaling procedures and paired comparisons.

In the case of *sound-wise scaling*, listeners rate each sound on a scale or a set of scales. In its simpler form, subjects rate each sound on a single scale: The annoyance of washing machine noises (Jeong et al. 2015) and the amenity (i.e., pleasantness in this context) of refrigerator noises (Sato et al. 2007) are examples of such scales. Producing absolute judgements of quality can sometimes be difficult for listeners without a context or a reference. Therefore, one variation of the method uses reference sounds to anchor the judgements. For example, Lemaitre et al. (2015a) have adapted the MUSHRA procedure (MUltiple Stimuli with Hidden Reference and Anchor, International Telecom Union 2001–2003) for collecting quality judgements of unpleasantness for wind buffeting noises. This procedure allowed the listeners to compare different sounds and then rate each sound on a scale ranging from "the least unpleasant" to "the most unpleasant." For each subset of sounds, the least and the most unpleasant sounds were systematically included to anchor the listeners' judgements.
AQ1

Another difficulty of the method is to choose the label of the scale. "Quality," "amenity," "annoyance," "unpleasantness," and "noisiness" have been used, but they are also sometimes difficult to interpret for the listeners. Therefore, sets of semantic differentials (and dimensionality reduction techniques) are also used at times, just as occurs for characterizing the timbres of the sounds (see Sect. 9.2.1.1). In fact, most studies use only one listening test with semantic differential scales corresponding to both timbre dimensions and quality judgements. In one such example, the participants of a study by Hoffmann et al. (2016) rated a set of road tire noises using the following: pleasant, sharp, loud, rough, stressful, activating, and the pitch.

Another class of methods uses *paired comparisons*. The advantage is that listeners find it much easier to produce a comparison than an absolute judgement. In this case, listeners hear all possible combinations of two sounds in the set and make comparison judgements, which can be binary or on a scale. For example, they may have to select which one of the two environmental sounds is the most unpleasant (Ellermeier et al. 2004) or rate two diesel engine sounds, A and B, on a scale ranging from "I prefer A a lot" to "I prefer B a lot" (Parizet et al. 2004). The judgements for each pair of sounds are then transformed into quality scores for each sound. The simplest method consists of averaging the preference judgements for one sound across all the other sounds to which it has been compared. More complex methods rely on statistical models connecting individual quality scores to the probability of choosing sound A over B, and statistical methods are used to fit the parameters of the model. In the Bradley-Terry-Luce model (BTL), the probability of preferring sound A over B is proportional to the ratio of the quality of sound A over the sum of the quality of sounds A and B (Bradley and Terry 1952; Ellermeier et al. 2004).

Other models are derived from Thurstone's case V model, whereby the preference probability is proportional to the cumulative normal distribution of the difference of the two quality scores (Thurstone 1927; Susini et al. 2004). One limitation of both approaches is that quality judgements are averaged across listeners and thus ignore potential individual differences in preference; indeed, two listeners can perceive the same difference between two sounds because they are different in roughness, but their preferences according to roughness can be opposed. An alternative that avoids this potential pitfall is discussed in Sect. 9.2.3. Another limitation, from a practical point of view, is that the number of pairs grows rapidly with the number of sounds. Thus, this method is limited to a rather small number of sounds and requires lengthy listening tests (see Sect. 9.2.3) unless the stimulus pairs are partitioned across subjects (Elliott et al. 2013).

9.2.1.3. Connecting Timbre with Quality

The final step of the method is to connect quality judgements with sound descriptors. Such a connection is made through a model that takes the sound descriptors as input vectors and produces estimates of quality judgements at the output. The most commonly used model is the *multivariate linear model*, whereby the quality judgement of a given sound is estimated from a linear combination of a set of descriptors. The coefficients of the model (i.e., the contribution of each descriptor to the quality of the sounds) are determined by fitting the model to the data (sound descriptors and quality judgements).

One of the difficulties of this approach is to select the descriptors that enter the model. When the timbre of the product sound has been characterized with a listening test (see Sect. 9.2.1.1), the most straightforward solution is to use the results of this initial phase as inputs to the model. When the initial timbre characterization is missing, hundreds or thousands of descriptors are often available to investigators from various software packages. Since the quality judgements result from listening tests, the set usually consists of 10–100 examples (it is very difficult for listeners to provide more judgements in a test). Using all available descriptors in the model is simply not possible because it would overfit the data; therefore, experimenters have to select a subset of them. One simple method requires the investigators to listen carefully to the sounds, consider the listeners' comments, and manually pick out the descriptors that are the most likely candidates. More systematic methods (e.g., stepwise regression, Monte Carlo) test different versions of the model to select the best subset of descriptors (Lemaitre et al. 2015a).

Linear models rely on the assumption that the contribution of each descriptor to the quality of the product sound is linear and therefore monotonic. This assumption is reasonable for loudness, as users usually prefer quieter over louder product sounds. But this is not necessarily the case for every descriptor. For example, Pietila and Lim (2015) found that listeners disliked the sounds of golf clubs hitting a ball that were too low or too high in pitch. In such cases, nonlinear models are required. One solution is to use polynomial regression or create nonlinear transformations of the descriptors within a linear regression model. These methods, however, require specifying the exact form of the nonlinearity.

All of these methods usually consider quality judgements averaged across listeners. However, different listeners may have different preferences, as illustrated by the example below.

9.2.2. Example: The Sound Quality of Air-Conditioning Units

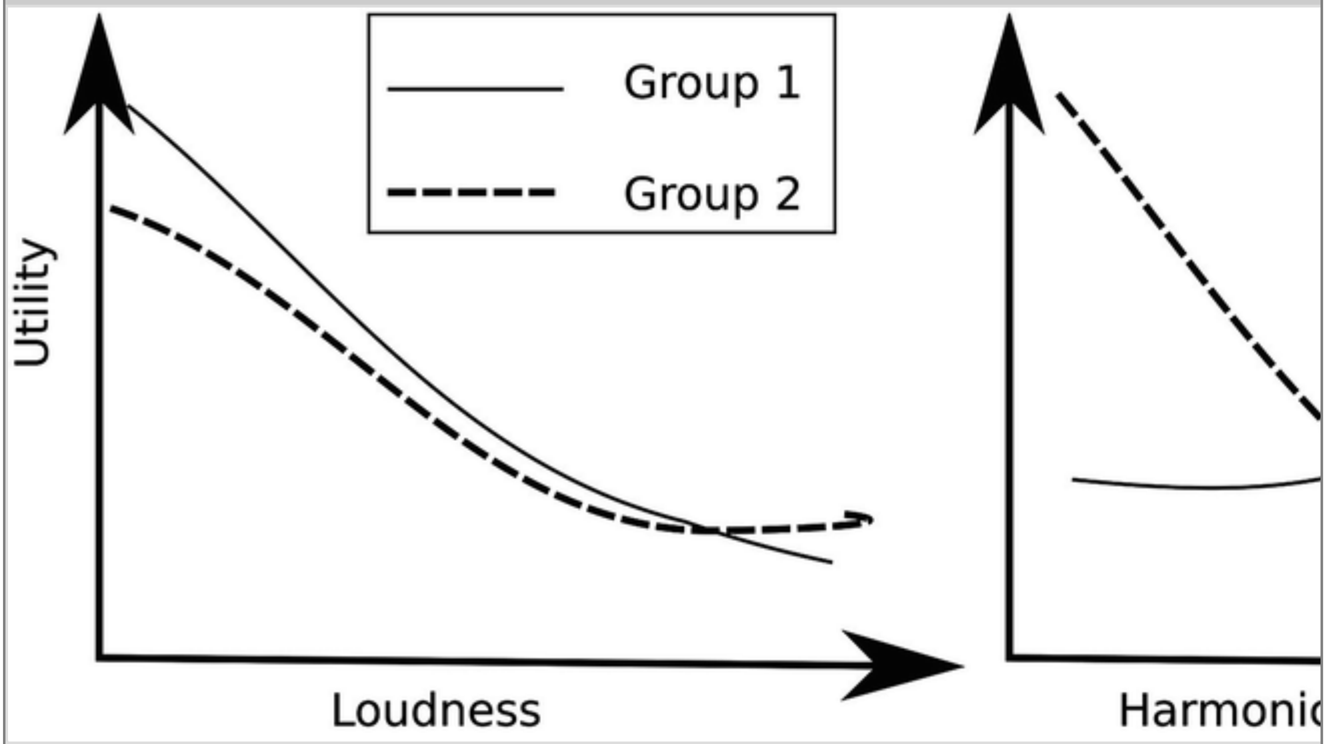
Air conditioning units in homes, offices, and vehicles are sources of noises that can sometimes be extremely tiresome and unpleasant. Aerodynamic turbulences are created by air being blown out of a duct through a vent, and sometimes a grid, resulting in wideband noises and hisses. As a consequence, many studies have sought to quantify which timbral characteristics of these sounds are unpleasant and thus focus the efforts of engineers on reducing these unpleasant characteristics.

A study on the sound quality of indoor air-conditioning units illustrates the different steps of the methodology outlined in Sect. 9.2.1 (Susini et al. 2004). In a first step, the study characterized the timbre with dissimilarity ratings and MDS. This analysis yielded three dimensions and showed that the first dimension corresponded to the relative balance of the harmonic (motor) and noise (ventilator) components (harmonic-to-noise ratio). The second dimension corresponded to the spectral centroid of the sounds (a descriptor similar in spirit to sharpness; see Siedenburg, Saitis, and McAdams, Chap. 1), and the third dimension corresponded to the loudness of the sounds. The experimenters then collected paired preference judgements: listeners indicated which sound they preferred in each pair of sounds. The preference probabilities were transformed into quality values for each sound with a model based on Thurstone's case V (for details, see de Soete and Winsberg 1993). Finally, the study used a statistical model to relate the quality of the sounds to a nonlinear utility function of each sound descriptor (de Soete and Winsberg 1993).

Interestingly, the analysis found different utility functions for two latent groups of listeners (latent here means that the groups were not predetermined but resulted from the analysis itself). Figure 9.3 represents these utility functions for three descriptors (harmonic-to-noise ratio, spectral centroid, and loudness) for the two groups of listeners.

Fig. 9.3

Utility functions for the sounds of air conditioning units, three descriptors, and two latent groups of listeners (*solid line*, *dashed line*). The utility functions represent how preference judgements change with the sound descriptors. (Adapted from Susini et al. 2004; used with permission from Elsevier)



Unsurprisingly, the results showed that the quality of the air-conditioning unit is related to loudness with the same decreasing monotonic function for the two groups of listeners: listeners unanimously preferred quieter sounds. More unexpectedly, the relationship between quality and the harmonic-to-noise ratio is completely different for the two groups of listeners. Whereas listeners in the first group (solid line) preferred sounds with a strong harmonic component, listeners in the second group (dashed line) preferred sounds with a strong noisy component. The situation is somewhat similar for the spectral centroid: listeners in one group preferring sounds with a lower spectral centroid and listeners in the other group providing a rather flat response. Overall, listeners in the second group focused mainly on loudness to judge their preferences. These results clearly show that, whereas listeners perceive differences of timbre more or less equivalently, each individual may prefer different timbral characteristics. Thus, as discussed by McAdams (Chap. 2), it is very important to consider individual differences and to consider nonlinear relationships between sound quality and timbre dimensions.

9.2.3. Limitations and Issues of Classical Methodology

Over the years, the classical methodology of sound quality evaluation has been applied to a variety of industrial products and has resulted in a number of quality indicators that have been used successfully by industrial practitioners. This methodology has two main advantages. First, it characterizes the timbre of the products under consideration. As such, it provides the product makers with a good understanding of the sounds of their products and, in particular, the features of the sounds that may be deleterious to the overall quality of the products. Second, the nature of the results of such studies (a piece of software takes sound signals as an input and calculates a single numerical estimation of quality at the output) makes them very easy to integrate into a measurement chain. In fact, the results of most published sound quality studies are remarkably consistent. So one may wonder if a universal estimator of sound quality could be designed that is valid for any product sound. This impression may be the result of the limitations of the methodology. The next sections discuss these limitations.

9.2.3.1. Does Sound Quality Evaluation Still Require Experimental Characterization?

Many sound quality studies have been published over the years and some results are remarkably consistent. As a matter of fact, almost all studies have found that listeners prefer quieter over louder sounds. Another common result is that quality judgements are negatively correlated with roughness (rougher sounds are evaluated as more unpleasant than smoother ones), tone-to-noise ratio and related metrics (sounds with prominent tonal components tend to be judged as unpleasant hisses), and fluctuation strength (sounds with too much fluctuation are judged as unpleasant). Convex (i.e., u-shaped) functions that relate quality and sharpness or the spectral gravity center also have been found. Listeners tend to find sounds with prominent high frequencies (shrilling) or low frequencies (rumbling) unpleasant. Of course, specific products may deviate from general tends: typically, rough engine sounds may be appropriate for sport cars or motorcycles, and some listeners in the air-conditioning study reported in Sect. 9.2.3 preferred sounds with greater high-frequency energy.

Overall, these tendencies appear to be quite strong. The logical consequence should be that one could design a universal descriptor of sound quality, valid for any kind of sound, once and for all. In fact, Zwicker and Fastl (1990) proposed such a universal indicator: *sensory pleasantness*. They mathematically defined this indicator of sensory pleasantness as a combination of loudness, sharpness, roughness, and tonality.

There are, however, a number of issues with such an indicator. First, and as illustrated by previous examples, preference may vary from one individual to another. Second, Zwicker and Fastl's indicator uses monotonic and separable functions to relate sensory pleasantness to different timbre descriptors, whereas previous examples have hinted that these relationships may be nonmonotonic and nonlinear in some cases. These are, however, technical difficulties that could be solved in principle. Such a universal indicator would make experimental characterization of timbre and sound quality unnecessary and thus would be a great help for product developers. But the limitations of such potential universal indicators are more conceptual and result from the concepts behind the methodology themselves. The promise of a universal indicator of sound quality based on simple timbre descriptors may in fact result from an overly restrictive perspective on sound quality and timbre. The following sections take another walk through the three parts of the methodology to discuss these concepts.

9.2.3.2. Timbre and Descriptors

Most of the discussions about the timbre of musical instruments carried out in this book (see McAdams, Chap. 2) also apply to product sound quality studies. In particular, one important limitation of the experimental methods used to characterize the timbre of product sounds is that the set of sounds under study must be homogeneous: sounds should be perceived as produced by the same source and should continuously and densely span a common acoustic space (Susini et al. 2011). When different sounds are perceived as produced by completely different sources (e.g., cats, motorcycles, and sea waves) or possess too many idiosyncratic features, listeners may become unable to provide continuous ratings of dissimilarity (in an MDS procedure), and the concept of continuous timbre dimensions is no longer relevant. In fact, sound categorization and source identification are strong cognitive processes that are not always compatible with an MDS procedure (although see McAdams, Chap. 2). To address this issue in a product sound quality study, sounds are carefully selected, and listeners are required to focus on the timbral features irrespective of the sound sources. It is also not uncommon to add synthesized sounds by homogeneously varying a few synthesis parameters or to select sounds that are closely distributed in terms of timbre characteristics.

This creates two potential issues. First, there is a risk that the sound selection may not be representative of the variability of sounds emitted by the products (ecological validity). This issue can be handled by preliminary studies seeking to explore the variability of product sounds and select samples representative of this variability (Susini et al. 2004; Parizet et al. 2008).

The second issue is circularity. Because of the established tradition of sound quality indicators and timbre descriptors, there is a tendency to select sounds that are homogeneously sampled across "classical" descriptors (e.g., sharpness, roughness, tonality, etc.). Unsurprisingly, the results of experimental characterization often yield the same timbre dimensions and descriptors used to select the sounds. In addition, the experimental characterization of timbre requires a limited number of sounds (typically about twenty sounds), and a dense and homogeneous sound selection can only span a few dimensions. This may explain why many experimental studies of product sound timbre systematically yield more or less the same three to five timbre dimensions.

One deleterious consequence is that this leads to the false belief that the timbre of any sound set can be characterized by the same three to five descriptors, conveniently implemented in software packages. However, this book beautifully illustrates a number of other phenomena contributing to the timbre of sounds. First, the timbre of a product is closely related to its identity, which cannot be easily ascribed to the combination of a few descriptors. In particular, a given product generates a large variety of different sounds created by a number of different physical sources. As discussed by McAdams (Chap. 2), the identity of a musical instrument, or for that matter of an industrial product, is precisely characterized by the variability of the sound events produced by the product or products under different modes of operation. Agus, Suied, and Pressnitzer (Chap 3) also show that the identity of a sound source cannot be completely specified by a few systematic common dimensions and descriptors. In particular, idiosyncratic features of sounds play an important role in sound recognition and, by definition, are not generalizable to other sounds.

The temporal evolution of product sounds is also very important for their identity (McAdams, Chap. 2; Caetano, Saitis, and Siedenburg, Chap. 11); yet classical descriptors usually do not take time into account. More generally, whether timbre perception can be reduced to a few common continuous and simple dimensions is an extremely important and still open question in timbre research (see the discussions in Aucouturier and Bigand 2012; Siedenburg et al. 2016), and the different chapters of this book offer different perspectives on this question. Alluri and Kadiri (Chap. 6) show that different auditory dimensions (pitch, loudness, spectral centroid, harmonicity) are specifically encoded in different cortical areas. Agus, Suied, and Pressnitzer (Chap. 3) discuss very sparse spectrotemporal representations, whereas Elhilali (Chap. 12) introduces very rich and redundant representations of sounds based on modulation encoding in the brain. Therefore, characterizing the timbre of product sounds should definitely not be considered as a problem that can be solved with standard methods dating from the early 1990s. Product sound quality as an engineering practice requires standards, but these standards should evolve with timbre research. The methods whereby sound descriptors are actually discovered blindly by deep-learning algorithms (without using the experimenters' intuitions or pre-existing knowledge) are very promising.

9.2.3.3. Connecting Quality with Timbre Descriptors

The techniques used to connect quality judgements with timbre descriptors also have a number of limitations. First, the most used method is multivariate linear regression, which, by definition, assumes a linear relationship between quality judgements and timbre descriptors. However, many examples reported above show that such an assumption is not true in general. Linear regression with nonlinear transformation of descriptors, polynomial or spline regressions, can handle nonlinearity, but the experimenters have to define the exact shape of this nonlinearity (e.g., order of the polynomials, nodes, and order of the spline functions, etc.).

Machine-learning methods can address such issues because they can learn nonlinear functions empirically, directly from the data, without the need for the experimenter to specify the nature of the nonlinearities. Pietila and Lim (2015) used a nested artificial neural network to connect preference judgements of golf club sounds to three descriptors (pitch, loudness, and timbre). The results showed complex nonlinear functions connecting preference judgements to each descriptor, which could not have been found by regression-based techniques without a priori assumptions about these functions. Another advantage of this technique is that it directly connects preference judgements to sound descriptors without the specification of an intermediate model connecting preference to quality (e.g., BTL, Thurstone's case V model).

Another issue with the connection of quality judgements to timbre descriptors is the selection of descriptors that enter the model. In most reported cases, experimenters manually select the descriptors on the basis of their own listening experience. Even when the timbre has been experimentally characterized with listening tests, there is no guarantee that listeners will use the same sound features to assess both preference and the dissimilarity between these sounds. In theory, listeners may consider that one aspect of the sounds that does not contribute much to their dissimilarity (e.g., a tiny buzz) is in fact deleterious to the quality of the sounds. Here again, machine-learning techniques can address such issues. Caetano, Saitis, and Siedenburg (Chap. 11) review deep-learning techniques that can learn sound features from the sounds themselves. For example, Dai et al. (2003) used a three-layer neural network to estimate annoyance judgements of brake squealing noises. Instead of using loudness, pitch, and timbre descriptors, the input to their neural network was a vector containing the amplitude of the spectral peaks between 2 and 18 kHz (80 values). The neural network thus estimated the annoyance judgements directly from the spectrum.

More generally, using machine-learning techniques has a huge potential for sound quality evaluation (for a review, see Pietila and Lim 2012). It should be noted, however, that the classical sound quality methodology (derived from psychoacoustics) and machine learning techniques have different philosophies as regards the generalizability of the results. On the one hand, psychoacoustics is based on inferential statistics: The quality judgements collected during a listening test are assumed to be randomly sampled for a population distribution, and the result of the procedure is the probability that the quality estimation corresponds to the average judgements in the population. On the other hand, the machine learning approach is empirical. Some empirical data (i.e., quality judgements) are used to train the model, and other empirical data are used to estimate the prediction power of the model. The upside of this approach is that it is assumption-free, and the generalizability of the model is tested on real data. The downside is that, as a consequence, it requires a large quantity of empirical data to be reliable. This is a very important issue in practice, since human listeners can only provide a few quality judgements in a reasonable amount of time. An additional downside is that the resulting network structure is difficult to interpret in behavioral and/or neuropsychological terms.

9.2.3.4. Quality Evaluation

Finally, probably the biggest issue with the classical methodology is that the notion of quality itself is extremely limited. Despite variations in the practicalities, the different methods all consist of letting listeners simply rate the quality of sounds, considered as a single notion, or indicate which of two sounds they prefer.

There are several issues with this approach. First, sounds are usually played out of context. Listeners are seated in a laboratory setting, listen to sounds over headphones or loudspeakers, and indicate their ratings on some kind of interface. This is not a particularly ecological setting. In reality, product sounds are made by a visible and tangible product, manipulated by a user aiming to do something with it, in a given situation, and in a given multisensory environment. Users have expectations about the sounds of a product, especially in relation to their previous knowledge and the other sensory aspects of the product. Playing sounds to listeners in a laboratory setting eliminates most of these aspects, which do have an important contribution to the perceptual quality of a product. As a matter of fact, many studies report that a large proportion of variance in annoyance judgements of community noise (e.g., transportation noise) is related to nonacoustic factors such as personal or socio-economic factors (Paté et al. 2017). Therefore, it should be stressed that the methodology for sound quality studies deals only with the acoustic determinants of sound quality.

Even with this qualification, a second important aspect should be considered. The “quality” of a product sound is not something that can be ordered along a single dimension. Typically, listeners dislike louder sounds when required to judge sounds heard in a laboratory setting. But take the example of a vacuum cleaner: a completely silent vacuum cleaner would be unusable because people use the sucking noise to monitor how the vacuum cleaner is operating. Furthermore, the loudness of the vacuum cleaner is often associated with power. It is difficult for users (and buyers) to consider that a very quiet vacuum cleaner may still be powerful and deserve a high price tag. So, pleasantness, usability, and power perception may actually be examples of factors that contribute to the quality of a product and are somewhat independent.

Finally, this approach advocates for the idea that product sound quality should be addressed by considering what the sounds are used for in a given product. The legal function of car horns, for example, is to warn road users of a potential danger; accordingly, designers must create new sounds that are still recognized as car horns (Lemaitre et al. 2007). Carmakers provide electrical vehicles with exterior sounds. Such sounds must be heard in the acoustic situation of normal traffic. Accordingly, studying the quality of such sounds must consider the detectability of sounds in such an environment (Parizet et al. 2014). More generally, Sect. 9.3 discusses sound design, which considers sounds in terms of different aspects that go beyond the unique notion of quality but are still related to timbre.

9.3. From Sound Quality to Sound Design

As discussed in Sect. 9.2, most studies consider only listener’s overall preference and do not consider other roles that sounds could serve. In fact, many products are strongly associated with their sounds: the sound of the Harley Davidson motorcycle is a classic example. In 2013, the Italian composer Andrea Cera designed the sound of ZOE—the electric car produced by the French carmaker Renault—to inform pedestrians of its movement on the street; its timbre is now emblematic of the car’s identity and is nicely integrated into the urban sound environment. For products or devices, there are also several examples that reveal how sounds can improve useful information: monitoring in intensive care units or indicating performance for physical rehabilitation or sports activities. Thus, extending the sound quality approach, the *sound design* approach embraces the fine articulation of functionality, pleasantness, identity, and ecology of product sounds and their environment. Furthermore, sound design not only considers the diagnostic of existing product sounds; it is a process to design, engineer, or modify the dynamic and interactive timbral characteristics of product sounds to meet specific intentions or requirements defined during the development of products.

9.3.1. Sound Design: Make the World Sound Better

40 years ago, in *The Tuning of the World* (1977), R. Murray Schafer wrote about a new soundscape in which natural sounds are increasingly replaced by artificial sounds (p. 91), and “warned music educators that they would now have to be as concerned about the prevention of sounds as about their creation” (p. 98). Today sounds are widely used in a variety of products, ranging from desktop computers to mobile phone applications and from safety warnings (e.g., for hospitals, aircraft) to electric cars. These new artificial sounds are functional sounds added into our environment for specific purposes—not as decorations and not as pieces of art. The aim is to produce a sound to communicate efficient information to a user. Acoustic features such as rhythm, pitch, and loudness make an alarm audible and urgent (Stanton and Edworthy 1999). Using sounds to warn of a danger, to confirm actions, or to guide someone toward a specific direction is one kind of functional sound feedback, but sounds can also be designed to improve users’ performances in terms of learning and control of a device (see Sect. 9.3.5 for a discussion on sonic interactions).

Fine-tuning acoustic features to provide efficient information using a sound, such as an alarm, could be done by an engineer based on ergonomic inputs (Stanton and Edworthy 1999) or psychoacoustic tests. In this way, the alarm sound is created based on functional recommendations. In the specific case of electric vehicles, which may be dangerous for pedestrians in urban areas because they are too quiet, the addition of sounds quickly appeared necessary to allow pedestrians to not only detect the presence of the vehicle but also its position, distance, and speed. Engineers can apply those functional recommendations in terms of intensity and fundamental frequency for the sound but, in addition, a sound designer would also shape the timbre to achieve a “harmonious solution.”

From an ecological perspective, a harmonious solution is obtained if the creation of the functional sound feedback is conceived in terms of its integration into the environment in which it will be heard as if it had always been part of it, thereby producing a feeling of pleasure or satisfaction by the intrinsic characteristics of the sound. From an industrial perspective, a harmonious solution is obtained by taking into account the brand values in order to make audible the identity of the brand (or the car) through functional sound feedback. The sound designer considers the global coherence of a new functional sound and evaluates the timbre characteristics in relation to a user’s pleasure, brand identity, and the environment of use.

9.3.2. Timbre Is a Key Element for Sound Design

Susini et al. (2014) have proposed a general definition of sound design: A sound design approach is implemented to create new sounds in order to make intentions audible in a given context of use. The sounds created are referred to as *intentional sounds*. There are different types of intentions: the first intention is to efficiently reach a goal through the sound (functionality), and the second intention is to produce a harmonious solution in the environment (ecology) that combines satisfaction of the user (pleasantness) and coherence with the product (identity). Successful sound design should be the articulation of the different intentions in order to produce new interactions through sound. Thus, a designed sound must simultaneously satisfy different aspects; the consequence is that its timbre is shaped by multiple nonindependent recommendations. Indeed, the formal aspect of a sound is already largely shaped by acoustic features that are related to functional recommendations; an alarm sound has strong spectrotemporal characteristics corresponding to its warning function, and these characteristics tend to limit the possibility to shape the sound in coherence with the environment. This is the tricky part of the process undertaken by the sound designer. Functional constraints, usually related to temporal variations in intensity or pitch, must be combined with other aspects such as pleasantness, identity, and ecology, which are more related to timbre.

Most sound quality studies seek to improve the quality of everyday existing sounds resulting from natural physical/mechanical/electrical phenomena called *nonintentional sounds*: the sound of a door lock, a car engine, an electric razor, etc. Usually, this work is done by professionals in acoustics, psychoacoustics, electronics, or mechanics. In the best-case scenario, recommendations are based on a perceptual analysis of the timbre differences and similarities between a set of sounds collected from several devices covering the full range of a product, such as the example of the air-conditioning units presented in Sect. 9.2.

But what happens when the product does not exist yet, or when the device was silent but to which it has become necessary to add sounds? Typically, an electric car is quieter than a car with an internal combustion engine. However, it is widely accepted that it is necessary to add intentional sounds to alert pedestrians and also to inform the driver about the car's state of functioning (e.g., its speed). New sounds must be imagined and created that satisfy functional constraints (e.g., detectability) as well as constraints in terms of pleasantness, identity, and ecology. The fruitful approach providing relations between users' preferences and timbre attributes based on an analysis of a collection of existing sounds (i.e., the sound quality methodology) is not useful in this case. Fortunately, practice in sound design is led by a strong creative process based on different sources of inspiration; in addition to their technical skills, sound designers are characterized by creative abilities to make sound sketches composed of different timbres, which can make all the difference in producing a successful articulation between functionality, pleasantness, and identity of a new product sound with respect to the sound environment. As has been done for science fiction movies, sound designers have to imagine and create new intentional sounds for our everyday environments.

Research on sound perception and cognition can be very informative for designing sounds. Indeed, the choice of one sound rather than another can be done arbitrarily by a sound designer; however, that does not mean that any choice would do equally well. The best choice must be an intelligent fit to human perception. Imagine a case in which one has to make a choice: is it advisable to use abstract sounds[1] (artificial tones such as beeps) rather than ecologically produced sounds (everyday sounds related to a source or an action)? Compared to abstract sounds, ecological sounds are often identifiable and thus more rapidly used in line with their function: the sound of a door lock may be used as a depiction for a "closing file" action on a computer because it will be quickly and easily understandable by users. However, it has been shown that abstract sounds work just as well after users have been exposed to them for a while and have learned to associate the sound with the function (see Agus, Suied, and Pressnitzer, Chap. 3). This finding encourages designers to propose new sounds and then to drop the old clichés (e.g., the use of a reflex camera sound for a digital camera). Such knowledge is very helpful in making a decisive choice in a sound design process.

In sound design, knowledge of timbre perception is especially important. The timbre study of air-conditioning units presented in Sect. 9.2.3 is a useful example about the relevant timbre dimensions that can be used to shape a product sound with respect to a listener's preferences. Knowledge of the relation between sound identification and timbre for different physical characteristics of everyday sounds is also fundamental in sound design in order to manipulate sound descriptors to achieve specific target sounds, for example, in terms of material (e.g., a digital sound that evokes wood or metal) or form (e.g., a sound that has the same resonance as a large plate).

Finally, knowledge of timbre dimensions with respect to the values of a brand is also very useful for the sound designer working on the sound identity of the brand. For example, during co-design sessions (involving marketing professionals, sound designers, ergonomists, etc.), three of the main brand values of the French railway company (SNCF)—benevolent, simple, and efficient—were associated with several words related to timbre features (e.g., dull, round, warm, and dynamic). Those timbre features were then used as recommendations by the sound designers to propose sounds for different railway station equipment (e.g., departure boards or ticket dispensers). Verbal descriptions of timbre are fundamental in initiating the creative work of the sound designer in relation to the brand identity. This last example raises several questions related to the verbal description of timbre features in a sound design process. How do different participants involved in the sound design process communicate about timbre? Is it possible to share a common language? Are there enough specific words to describe timbre and translate them directly in terms of recommendations?

9.3.3. Communication About Sounds and Timbre: Different Strategies

Saitis and Weinzierl (Chap. 5) describe different methods or tasks to investigate verbal descriptions of sounds. In the free verbalization task, participants usually produce verbal descriptions related to their expertise and to their ability to recognize sounds. As Chion (1994) has argued, "we hear as we speak."

For everyday sounds, the most common strategy is to describe the source of the sounds ("this is the sound of a hairdryer," "it is a vacuum cleaner," "this is a trumpet") or the action that produced them ("someone is hitting a glass," "this is the sound of a string being pinched," "she is pushing a switch"). This is the *causal strategy*, which is the most intuitive way to speak about sounds for nonexperts. Descriptions are sometimes solely related to a specific meaning in a specific context or location: alarm sounds in intensive care units have a specific meaning only for the staff. This is the *contextual strategy*: Verbal descriptions are not specific to a sound's features but are more context-dependent. Finally, it seems that descriptions are seldom based on the sound itself in terms of acoustic characteristics and timbre features. This is the *reduced listening strategy*: Descriptions are directly related to the features of a sound independently of the meaning, the process that produced the sound, or its location.

For abstract sounds—artificial tones or beeps that differ in fundamental frequency, harmonic series, amplitude envelope shapes, rhythms and modulations—or sounds whose source is not easily identifiable, several strategies have been highlighted. Participants use vocal imitations (i.e., they vocally reproduce spectromorphological characteristics), onomatopoeic sounds (bleep, buzz, pop), illustrative analogies (like a fog horn, siren), and verbal descriptions of timbre (soft, high, short, tonal, dull, strident, warm). The type of strategy is mostly related to a listener's expertise in the field of sound. Indeed, sound experts will be more inclined to provide verbal descriptions related to the sound itself, such as acoustic characteristics and timbre features, than illustrative analogies. Schaeffer (1966) was perhaps the first to focus on descriptions related to a sound per se; based on a phenomenological approach, he provided several sound examples that illustrated different typo-morphological concepts such as mass, grain, and melodic profile. Although original and very detailed, Schaeffer's description remains quite complex and very few people use it today in its fullest form. However, the original idea of a lexicon illustrated with sound examples is still inspiring and challenging as a way to describe timbre (see Sect. 9.3.4).

In the sound design process, especially when it involves participants with diverse levels of expertise (from clients to sound designers), this diversity of strategies can be a serious obstacle for communication. Sound designers usually need information in terms of acoustic characteristics related to timbre features or temporal properties, but initial intentions are often expressed by the client with terms related to a context or a meaning, a function or an image (cf. contextual strategy). For example, the intention for an alarm sound in the context of a hospital could be described as “alerting but kind” rather than sound features such as long, smooth, continuous, high-pitched, loud enough, which are directly informative for the sound designers. Unfortunately, there is no common practice for talking about acoustic characteristics and timbre features, and sound designers often complain about the lack of tools to communicate about sounds in a sound design process.

Recently, in order to overcome this lack of a common language, an academic review of a large number of works dealing with verbal descriptions of timbre was performed for different kinds of sounds, from abstract to everyday sounds. Then, a lexicon of thiry-five relevant timbre descriptors (e.g., dry, bright, rough, warm, round, nasal, complex, strident) was proposed as an extension of Schaeffer’s (1966) fundamental concept (Carron et al. 2017).

9.3.4. Learning to Talk About Timbre in a Sound Design Process

A standardized lexicon of specific terms to describe relevant sound features is a very promising tool for the field of sound design from a practical point of view, for example, to assist in the training of the different participants involved in a sound design project to perceive and use relevant timbre features for the design process. The lexicon also would be useful to teach pupils in a sound design or post-production course who are learning to listen to timbre features and could then describe those features with a common vocabulary. In the lexicon proposed by Carron et al. (2017), each term is presented on a computer interface, defined and illustrated by musical, vocal, environmental, abstract, and effect sounds, in order to provide a large diversity of examples. This tool has been tested and approved in different case studies in which industrial partners were involved (Carron et al. 2015). From a training perspective, a set of audio tests also has been developed to evaluate participants’ understanding of the lexicon; it is a complementary and indispensable element of applying the lexicon. The tests assess whether using the lexicon may improve listeners’ perception of a specific feature as well as their ability to describe sounds with only the terms of the lexicon.

One set of tests is called “from word to sound”; participants are asked to choose from among five sounds the most typical sound related to a specific term. Another set of tests is called “from sound to words”; participants have to choose three words among the list of terms provided to describe the prominent auditory dimensions of a specific sound (e.g., a continuous and warm sound with a slow attack). Results of the tests are compared with previous results of twenty well-trained participants.

During a training session, individual and collective explorations of the lexicon are alternated with the different tests. After each test, terms are discussed collectively to ensure a common understanding, and eventually there is a refinement of the sound examples provided for the lexicon. This global training ensures that participants involved in the same project have a rich and varied vocabulary that is adapted to describe a large number of timbre features and temporal properties appropriate for an important variety of sounds. This procedure is an alternative to sensory evaluation often used to reveal a list of words specific to the timbre of a set of sounds in relation to consumer preferences. The sensory evaluation requires several steps of discussion, training, and testing with a panel of experts, a process which is often very long (several weeks) and specific to a set of sounds.

9.3.5. Sounds to Support User Interaction

As discussed before, designing sounds has the potential to address the different roles of sounds in products, interfaces, places, or brands. The following sections focus on fostering user interaction. In fact, sounds are particularly well-suited to guide and facilitate interactions between users and a product. Because sounds are dynamic stimuli, they can react instantly and continuously to users' actions and gestures and thus provide users with real-time feedback of their actions. Furthermore, the tight bond between audition and motor actions makes it possible to use sounds to continuously guide and facilitate gestural interactions.

The idea of using sounds to support user interactions is relatively new in the field of sound design and has been labeled as *sonic interaction design* (Serafin et al. 2011). Sonic interaction design also introduces new research questions. One new research area is determining which timbre features best support interaction. Intuitively, ecologically produced sounds (e.g., the sound of rubbing a finger against a rough surface) seem the best candidates to support an interaction (e.g., interacting with a touch screen), because most people would know how timbre features change with changing interaction parameters (e.g., a faster rubbing gesture producing higher frequencies). In the next subsections, evidence is presented that, in fact, this first intuition may be too simplistic and that further work is needed to fully understand how gestures and timbre become associated in memory.

9.3.5.1. Sonic Interaction Design: A Rationale

Designing sonic interactions consists of using and creating sounds to design, help, or augment how users interact with a product or a machine. As such, sonic interaction design fits under the larger umbrella of *interaction design*: designing the ways users may interact with systems and computer interfaces in particular (Crampton-Smith 2007).

There is a rich history of sounds in human-computer interfaces. Computer interfaces emit a variety of different sound signals, each aiming to communicate a different message (e.g., computer starting up, different types of errors). One key question for designers, therefore, is how to convey a specific message with nonspeech sounds? One common strategy consists of using artificial tones (beeps or sequences of beeps forming a melody) and relying on an arbitrary code mapping the sounds' features (pitch, loudness, duration, timbre) to the message. The messages conveyed by artificial beeps are called “earcons” (Brewster 2009). The main disadvantage is that users must learn the mapping between sound features and meaning, in other words, the code connecting the different melodies to their meaning. William Gaver, an influential perception researcher and interaction designer, proposed another strategy that does not rely on an arbitrary mapping but is instead based on the spontaneous identification of sound sources that are called “auditory icons”. The most famous auditory icon created by Gaver is probably the sound of a crumpled sheet of paper thrown into a garbage can, used as feedback to indicate file deletion, which was developed for Apple computers. In this example, the meaning of the sound results from the spontaneous identification of the sound source and a metaphor: deleting an electronic document being equivalent to physically crumpling and discarding a sheet of paper (Gaver 1986, 1989).

Similar to this example, many computer sounds rely on a desktop metaphor; however, ubiquity and mobility have drastically changed how users interact with mobile phones, tablets, and connected watches. Screen-and-keyboard interactions are simply not possible in many tiny devices, and vocal and gestural interactions thus have become more and more important and sophisticated (think of the different gestures used on mobile touch screens).

The use of sounds, however, has evolved more slowly, and most electronic devices still use simple and discreet beeps. There is, nonetheless, a great potential for sounds to guide, foster, facilitate, or augment gestural interactions. Sounds are dynamic stimuli. As such, they can very easily be made to react continuously and in real time to dynamic data, such as movement data, through some sort of model or mapping (i.e., data sonification). This idea is illustrated by Tajadura-Jiménez et al. (2014), who showed that changing the timbre of the sounds made by fingers rubbing the surface of an object changed the velocity and the pressure of the fingers, as well as the tactile perception of the surface.

In fact, audition and motor actions are tightly bound, and continuous sounds created by gestures may influence the gestures themselves. Lemaitre et al. (2009) have shown in particular that a continuous sonic feedback, reacting in real time to the users' gestures, can help them learn a fine gesture more rapidly than does visual feedback. Lemaitre et al. (2015b) also have shown that playing the sound of an action (e.g., tapping, scraping) can facilitate (when the action is congruent with the sound) or hinder (when incongruent) the subsequent execution of another action.

This idea of a continuous coupling of action and sound is exactly what happens when a person learns how to play a musical instrument. To produce a good tone, a violinist bows a string, and (particularly during training) adjusts bowing action continuously by listening to the sound that is produced. This sonic feedback guides the player's control, modifying bow speed, pressure, angle, and so forth. In fact, users also use continuous sounds when interacting with other products: most people use the change of pitch in the sound of their car engine while accelerating to decide when to manually change gears. The study conducted by Jérémy Danna and colleagues illustrates another such designed sonic interaction (Danna et al. 2013, 2015). They investigated the real-time continuous sonification of handwriting gestures to facilitate graphomotor learning. They devised a system whereby the movements of a pen on a sheet of paper changed the timbre of the resulting sound. When the pen movements were too fast, the timbre became squeaky and unpleasant. Jerky movements also resulted in unpleasant crackly sounds. Overall, the results showed that this designed sonic interaction improved the kinematics of the handwriting movements. There was, however, a lack of a long-term effect, thus raising the question of the persistence of the timbre-gesture associations in memory.

9.3.5.2. New Research Questions

The key issue of sonic interaction design is to design the timbre of sounds that can effectively support an interaction with a physical object or an interface (see also Ystad, Aramaki, and Kronland-Martinet, Chap. 13). "Supporting" an interaction may in fact correspond to several aspects. It may, for example, contribute to the aesthetics of the product, in other words, make it appealing or intriguing. Such ideas are, in fact, very close to the concepts of sound quality and sound design previously discussed. But more interestingly, supporting the interaction can also mean that the sounds produced by the users' gestures effectively help or guide the interaction. As such, evaluating the effectiveness of the sound design cannot be conducted only with the method of sound quality evaluation described in Sect. 9.2, but evaluation should mainly rely on measuring users' performances at performing a task (e.g., accuracy, speed at achieving a goal, etc.). As suggested by the above example of handwriting, another very important aspect of evaluating such designs is to assess the long-term benefits and, for example, whether gesture facilitation persists after the sonic feedback has been removed.

Because sonic interactions deal specifically with continuous sounds, static timbre descriptors, such as those previously discussed, are no longer relevant. Instead, the dynamic aspects of timbre become crucial. Coming back to the example of changing gears while driving a car, it is the change of timbre during acceleration that makes a car driver change the gear.

Sonic interaction design also poses specific challenges to the relationship between sound design and timbre dimensions. Whereas sound quality evaluation consists of mapping a user's judgements to timbre features or dimensions, designing sonic interactions consists of designing sounds that can effectively guide an action. A first intuition is that ecological sounds (i.e., sounds produced by real-world phenomena or connected to gestures via models of physical phenomena) should work best (Rath and Schleicher 2008; Lemaitre et al. 2009). However, empirical results have not confirmed this intuition so far. In particular, Lemaitre et al. (2015a, b) have shown that playing the sound of an action can prime that action. When participants had to respond to a vocal cue by physically tapping or scraping on a response interface (*ecological sound-gesture mapping*), playing, tapping, or scraping sounds before the cue could facilitate or hinder the gesture. But the same priming effect also occurred when the tapping or scraping gestures produced simple tones at different frequencies (*arbitrary sound-gesture mapping*). No advantage was observed for ecological mappings when compared to arbitrary mappings.

AQ2

These results (as well as those of Danna et al. 2015) suggest that the pitch, loudness, and timbre of sounds and motor programs can be coupled in memory, and that this association can be so strong and bidirectional that simply playing a sound that has been produced by a gesture can influence the later execution of that gesture, no matter whether the coupling has existed for a long time (such as in an ecological association) or has just been created (such as in an arbitrary sound-gesture mapping). The question thus becomes that of how are sounds memorized together with other sensory stimuli or motor programs? Similarly, the question of whether the benefits of a sonically augmented device can persist for a long time is in fact the question of the persistence in memory of sensory-motor couplings. Memory for timbre is thus a very timely and crucial research question (see Siedenbug and Müllensiefen, Chap. 4).

9.4. Summary

Most manufactured products, objects, and electronic devices that make up our environment produce sounds at some point. Some of these sounds are useful, some are annoying, some are funny, some are intrusive, and some are intriguing. Such sounds may thus enhance the perceived quality of a product, whereas others may be so inappropriate that they are deleterious to the overall impression of a brand. It is therefore utterly important to be able to assess how product sounds are perceived and to design them with intentions. These are the purposes of sound quality evaluation and sound design, two areas of applied research and professional activity that rely heavily on research on timbre perception and cognition.

The first part of this chapter described and discussed a general methodology used to evaluate sound quality of products. This methodology has three main elements: describing the timbre of the product sounds (using experimental methods or software packages), collecting listeners' quality or preference judgements, and connecting timbre features with preference judgements through various sorts of regression techniques. This method has been used extensively in a variety of applied studies and has proved to be a valuable tool for engineers, as it eventually produces an algorithm taking a sound signal at the input and producing a single indicator of quality at the output. Such quality indicators are powerful, simple to use, and do not require costly user testing.

There are, however, a number of limitations to this set of methods. First of all, they are based on a definition of timbre that may be oversimplified: it considers only homogeneous sets of sounds, a limited number of timbre dimensions, and it does not consider the sound source identities, sound idiosyncrasies, or individual differences among listeners. Furthermore, this methodology relies on a very narrow conception of the quality of a set of sounds. Most methods only consider listeners' preferences or one-dimensional ratings of quality of the sounds without taking into account the context of use.

The second part of this chapter discussed the sound design approach. In contrast to sound quality evaluation, sound design embraces a wider range of aspects of intentional sounds: functionality, pleasantness (listeners' satisfaction), identity (coherence between sound and product), and ecology (integration with the context of use).

Because of its wider scope, sound design faces a number of methodological issues. One such issue is that sound designers have to interact with many different practitioners in a company, with varying levels of sound expertise, and communication clearly is an issue. These issues are also research challenges. Chief among them is the question of how to standardize verbal descriptions of sounds. This chapter thus discussed the results of research that has studied how people talk about sounds. Most people describe the sources of the sounds, the actions that produce the sounds, the context and the location, or they produce vocal imitations, but rarely do respondents talk about timbre! The chapter then described a tool that has been designed to teach and train stakeholders of sound design projects to understand, share, and use technical descriptions of timbre (such as dry, bright, rough, warm, round, nasal, complex, strident), but also to describe the temporal characteristics (such as constant/fluctuating, ascending/descending, discontinuous, etc.) and the general qualities (such as soft/loud, low/high, short/long, etc.).

Finally, the chapter focused on one particular aspect of sound functionality: sonic interaction. In fact, recent research has shown a tight bond between audition and motor behaviors. Sonic interaction design seeks to exploit this close connection by designing sounds to help, guide, foster, or augment how users interact with a product or a machine. This approach is very promising, especially for motor rehabilitation applications, but the field of sonic interaction faces important research challenges. In particular, the dynamic aspects of timbre, timbre memory, and the nature of the interactions of sound representations with other sensory or motor modalities are important areas that contemporary research on timbre and other chapters of this book explore.

9.5. Compliance with Ethics

Guillaume Lemaitre declares that he has no conflict of interest.

Patrick Susini declares that he has no conflict of interest.

References

AQ3

American Standard Association (1960) USA acoustical terminology S1.1–160. American Standard Association

Aucouturier JJ, Bigand E (2012). Mel Cepstrum & Ann Ova: The difficult dialog between MIR and music cognition. In: Proceedings of the international conference on music information retrieval (ISMIR), Porto, Portugal, 2012

Bradley RA, Terry ME (1952) Rank analysis of incomplete block designs: I. the method of paired comparisons. Biometrika 39(3/4):324–345

Brewster S (2009) Non-speech auditory outputs. In: Sears A, Lacko JA (eds) Human-computer interaction: fundamentals. CRC Press, Boca Raton, p 213

Carron M, Dubois F, Misdariis N, Susini P (2015) Définir une identité sonore de marque: méthodologie et outils (Defining a brand's sonic identity: methods and Tools). Acoustique et Techniques 81:20

Carron M, Rotureau T, Dubois F, Misdariis N, Susini P (2017) Speaking about sounds: a tool for communication on sound features. J Design Res 15(2):85–109

Chion M (1993) Le promeneur écoutant: essais d'acoulogie (The listening walker: essays in acoulogy). Editions Plume, Paris

Crampton-Smith G (2007) Foreword: What is interaction design. In: Moggridge B (ed) Designing interactions. The MIT Press, Cambridge, MA, pp 7–19

Dai Y, Lim TC, Karr CL (2003) Subjective response simulation of brake squeal noise applying neural network approach. Noise Control Eng J 51(1):50–59

Danna J, Velay JL, Paz-Villagran V, Capel A, Petroz C, Gondre C, Thoret E, Aramaki M, Ystad Y, Kronland-Martinet R (2013) Handwriting movement sonification for rehabilitation of dysgraphia. In: Proceedings of the 10th Symposium on Computer Music Multidisciplinary Research, Marseille, 2013

Danna J, Fontaine M, Paz-Villagran V, Gondre C, Thoret E, Aramaki M, Kronland-Martinet R, Ystad S, Velay JL (2015) The effect of real-time auditory feedback on learning new characters. Hum Mov Sci 43:216–228

de Soete G, Winsberg S (1993) A thurstonian pairwise choice model with univariate and multivariate spline transformations. Psychometrika 58(2):233–256

Ellermeier W, Mader M, Daniel P (2004) Scaling the unpleasantness of sounds according to the BTL model: ratio-scales representation and psychoacoustical analysis. Acta Acust united Ac 90:101–107

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(1):389–404

Gaver WW (1986) Auditory icons: Using sound in computer interfaces. Hum Comput Interact 2(2):167–177

Gaver WW (1989) The sonic finder: An interface that uses auditory icons. Hum Comput Interact 4:67–94

Grey JM (1977) Multidimensional perceptual scaling of musical timbres. J Acoust Soc Am 61(5):1270–1277

Hansen H, Verhey JL, Weber R (2011) The magnitude of tonal content, a review. Acta Acust united Ac 97(3):355–363

Hoffmann A, Bergman P, Kropp W (2016) Perception of tyre noise: can tyre noise be differentiated and characterized by the perception of a listener outside the car? Acta Acust united Ac 102(6):992–998

International Telecom Union (2001–2003) Recommendation ITU-R BS.1534–1: method for the subjective assessment of intermediate quality level of coding systems. International Telecom Union, Geneva

Jeon JY, You J, Chang HY (2007) Sound radiation and sound quality characteristics of refrigerator noise in real living environments. Appl Acoust 68:1118–1134

Jeong U-C, Kim J-S, Jeong J-E, Yang I-H, Oh J-E (2015) Development of a sound quality index for the wash cycle process of front-loading washing machines considering the impacts of individual noise sources. Appl Acoust 87:183–189

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

Lartillot O, Toiviainen P (2007) A Matlab toolbox for musical feature extraction from audio. In: Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07), University of Bordeaux 1, France

Lemaitre G, Susini P, Winsberg S, Letinturier B, McAdams S (2007) The sound quality of car horns: a psychoacoustical study of timbre. Acta Acust united Ac 93(3):457–468

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. Int J Hum Comput Stud 67:976–993

Lemaitre G, Vartanian C, Lambourg C, Boussard P (2015a) A psychoacoustical study of wind buffeting noise. Appl Acoust 95:1–12

Lemaitre G, Heller LM, Navolio N, Zuñiga-Peñaranda N (2015b) Priming gestures with sounds. PLoS One 10(11):e0141791

McAdams S, Winsberg S, Donnadieu S, Soete GD, Krimphoff J (1995) Perceptual scaling of synthesized musical timbres: common dimensions, specificities and latent subject classes. Psychol Res 58:177–192

Parizet E, Brocard J, Piquet B (2004) Influence of noise and vibration to comfort in diesel engine cars running at idle. Acta Acust united Ac 90:987–993

Parizet E, Guyader E, Nosulenko V (2008) Analysis of car door closing sound quality. Appl Acoust 69:12–22

Parizet E, Ellermeier W, Robart R (2014) Auditory warnings for electric vehicles: detectability in normal-vision and visually-impaired listeners. Appl Acoust 86:50–58

Paté A, Lavandier C, Minard A, Le Griffon I (2017) Perceived unpleasantness of aircraft flyover noise: influence of temporal parameters. Acta Acust united Ac 103(1):34–47

Peeters G, Giordano BL, Susini P, Misdariis N, McAdams S (2011) The timbre toolbox: extracting audio descriptors from musical signals. J Acoust Soc Am 130(5):2902–29016

Pietila G, Lim TC (2012) Intelligent systems approaches to product sound quality evaluations – A review. Appl Acoust 73(10):987–1002

Pietila G, Lim TC (2015) Sound quality preference modeling using a nested artificial neural network architecture. Noise Control Eng J 63(2):138–151

Rath M, Schleicher R (2008) On the relevance of auditory feedback for quality of control in a balancing task. Acta Acust united Ac 94(1):12–20

Sato S, You J, Jeon J (2007) Sound quality characteristics of refrigerator noise in real living environments with relation to psychoacoustical and autocorrelation function parameters. J Acoust Soc Am 122(1):314–325

Schaeffer P (1966) Traité des objets musicaux. Editions du Seuil, Paris

Schafer RM (1977) The tuning of the world. Alfred A. Knopf, New York

Serafin S, Franinovic K, Hermann T, Lemaitre G, Rinott M, Rocchesso D (2011) Sonic interaction design. In: Hermann T, Hunt A, Neuhoff JG (eds) Sonification handbook. Logos Verlag, Berlin, pp 87–110

Siedenburg K, Fujinaga I, McAdams S (2016) A comparison of approaches to timbre descriptors in music information retrieval and music psychology. J New Music Res 45(1):27–41

Stanton NA, Edworthy J (1999) Auditory warnings and displays: an overview. In: Stanton NA, Edworthy J (eds) Human factors in auditory warnings. Ashgate, Aldershot

Susini P, McAdams S, Winsberg S (1999) A multidimensional technique for sound quality assessment. Acta Acust united Ac 85:650–656

Susini P, McAdams S, Winsberg S, Perry S, Vieillard S, Rodet X (2004) Characterizing the sound quality of air-conditioning noise. Appl Acoust 65(8):763–790

Susini P, Lemaitre G, McAdams S (2011) Psychological measurement for sound description and evaluation. In: Berglund B, Rossi GB, Townsend JT, Pendrill LR (eds) Measurement with persons–Theory, methods and implementation area. Psychology Press, Taylor and Francis, New York/London

Susini P, Misdariis N, Lemaitre G, Houix O (2012) Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. J Multimodal User Interfaces 5(3–4):175–186

Susini P, Houix O, Misdariis N (2014) Sound design: an applied, experimental framework to study the perception of everyday sounds. The New Soundtrack 4(2):103–121

Tajadura-Jiménez A, Liu B, Bianchi-Berthouze N, Bevilacqua F (2014) Using sound in multi-touch interfaces to change materiality and touch behavior. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. Helsinki, Finland, pp 199–202

Terhardt E, Stoll G, Seewann M (1982) Algorithm for extraction of pitch and pitch salience from complex tonal signals. J Acoust Soc Am 71(3):679–688

Thurstone LL (1927) A law of comparative judgment. Psycho Rev 34(4):273–286

Zwicker E, Fastl H (1990) Psychoacoustics acts and models. Springer Verlag, Berlin

[1] The difference between artificial and everyday sounds is defined in Sect. 9.3.5.1 (cf. earcons versus auditory icons).