5 Perceptual Evaluation of Sound-Producing Objects

Bruno L. Giordano, Patrick Susini, and Roberto Bresin

The experimental study of sonic interactions can elucidate one of the most important aspects of the design process: "How should the sonic interaction be structured to produce a target perceptual result or to induce a specific motor behavior of the user?" Very similar questions are the object of basic research on the human processing of sensory events (e.g., "What is the perceptual effect of this sound stimulus?"), and have spurred the development of a large number of experimental methods. This chapter is meant as an introductory guide to the behavioral methods for the experimental study of complex sound events and sonic interactions. Table 5.1 reports a list of design questions that can be answered with each of the presented methods. Throughout this chapter, we reference a number of empirical studies of complex and naturalistic sounds based on the described methods. Interested readers can find in these studies more detailed descriptions of the various behavioral paradigms.

Studying sonic interactions in the laboratory implies focusing on conditions where the action or motor behavior of the participant influences the properties of the presented sounds. This generic definition encompasses a large number of everyday events: a sonic interaction can indeed be as simple as the playback of a sound following a button press (e.g., touch tones of a mobile phone). Sonic interactions can be tentatively organized along a continuum of complexity according to the number of sound properties that can be modified by a change in the motor behavior of the user (see figure 5.1). The generation of touch tones lies at one extreme of this continuum because the only property of a sound that can be modified by the user is its presence or absence (the same sound will be played back independently of large variations in the force exerted on the key). Examples of complex sonic interactions, in order of increasing complexity are the turning of a volume knob in a sound amplification system, the striking of an object with a hammer (e.g., Giordano, Avanzini, Wanderley, & McAdams [37]) and the crunching of potato chips (Zampini & Spence [154]), up to the perhaps most complex type of sonic interaction—conducting a symphonic

Table 5.1
Examples of questions answered with the methods described in this chapter

Section and method	Examples of answered questions
5.1: Psychophysical methods	Can the user perceive each of the configurations of a sonic interaction? Can the user differentiate between configurations?
5.2: Identification and categorization	What naturalistic object is recognized in each of the configurations? What emotional category is recognized in a sonic artifact?
5.3: Scaling and rating	How does perceived effort vary between sonic feedbacks for robotic surgery applications? How should the user-controlled gain for sound level vary so as to produce a linear increase in perceived loudness?
5.4: Dissimilarity estimation	Which properties of a complex sonic interaction are most relevant to the user? Do different individuals focus on different attributes of the sensory events?
5.5: Sorting	How many categories of perceived materials can a sound synthesis algorithm reproduce? What is the most typical configuration for each of the material categories?
5.6: Verbalization	Which words capture the semantic correlates of a sonic interaction? What are the individual interactive strategies? Are there problems in the prototype design?
5.7: Semantic differential	Which configuration has the highest aesthetic and functional value? How do preference, perceived sound brightness, and perceived efficiency covary for these particular sonic interactions?
5.8: Preference estimation	Which configuration has the highest aesthetic and functional value? Which configuration is the least annoying?
5.9: Continuous evaluation	Do users' gestures map onto changes in the perceptual attributes of the sonic events? How does the emotional response to a complex sound vary in time?
5.10: Multisensory contexts	What influences most strongly preference for cars? The sound of its doors closing or their felt weight? Do sonic feedbacks significantly shorten the time it takes to park a car?
5.11: Measurement of acoustical information	What sound properties should be manipulated to induce a target perceptual result (e.g., maximize preference)?
5.12: Motion capture	How do we use our body in interaction with a sonic artifact? How do gestures and artifacts mutually influence a sonic interaction?

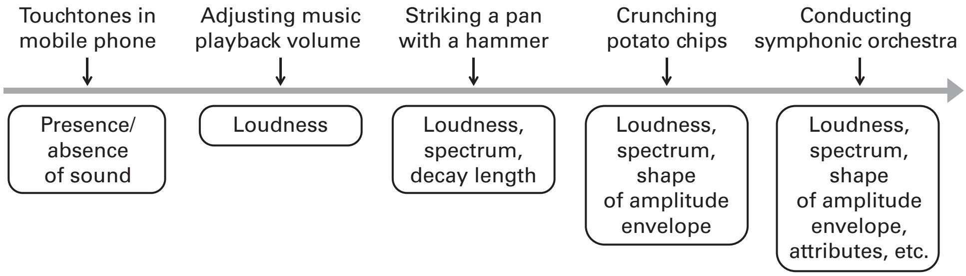

Figure 5.1
Variable complexity of everyday sonic interactions. With more complex interactions, changes in the motor behavior of the user lead to changes in a higher number of properties of the sound signal (from left to right).

orchestra (e.g., Kolesnik & Wanderley [74]). Importantly, a large number of experimental methods can be adopted to study both simple and complex sonic interactions. For example, participants in the study by Giordano, Rocchesso, and McAdams [41] triggered with the click of a mouse the playback of sounds recorded by striking objects of different hardness, whereas participants in the study by Lederman [85] actively generated sounds by scraping a rough surface with their fingers. Nonetheless, both studies adopted the same method, ratings, to measure the perceived properties of the sound-generating objects (hardness and surface roughness, respectively).

Another important distinction between studies of sonic interactions is the type of variable relevant to the experimenter. To make this distinction clear, it is helpful to summarize the various stages involved in the interactive production of sounds and in their perception (see figure 5.2). In general, a sonic interaction begins with a motor behavior or action carried out on a mechanical system, a sound-generating object (e.g., slapping the membrane of a bongo drum with the hand). The action-induced displacements of the components of the mechanical system will ultimately result in the production of a sound, which constitutes a source of acoustical information for the listener. At the same time, the motor behavior itself and the sound-generating object will produce information for additional sensory systems: kinesthetic, tactile, and visual. All these types of sensory information will then trigger various physiological and neural processes, resulting in conscious sensations, perceptions, and cognitions. Eventually, the processing of sensory information will feed back into the planning and control of further sound-generating actions.

The experimental study of sonic interactions can thus focus on four different types of variables: (1) quantitative measures of the motor behavior, as frequently measured

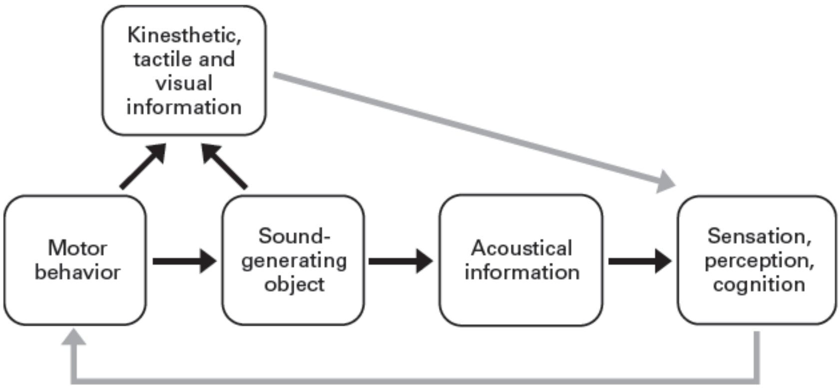

Figure 5.2
Chain of events and processing stages involved in the perception and production of interactive sonic events. Arrows symbolize causal connections.

with motion capture systems (e.g., [73]); (2) measures of the properties of the sound-generating system (e.g., the movement of the hammers and keys of a piano). Section 5.12 reviews a number of studies based on the measurement of the motor behavior and of the behavior of sound-generating systems in interactive contexts. (3) A third type of variable measures the properties of the stimuli impinging on the sensory systems (e.g., properties of the sound signals). Section 5.11 illustrates the main approaches that can be used to establish what properties of the stimulation, specifically of sound stimuli, affect the perceptual responses of experiment participants. Section 5.10 illustrates a number of experimental paradigms for comparing the perceptual effects of information from different sensory modalities. (4) A final category of variables aims at quantifying the sensory, perceptual, and cognitive responses to the sensory stimulation. In contrast with the previously described components of the sonic interaction chain, these variables are not measured directly, but are commonly inferred from the responses of experiment participants in a variety of behavioral tasks (e.g., "rate the hardness of the object you are manipulating," see section 5.3). The majority of the methods presented in this chapter are designed just for this purpose (table 5.1).

5.1 Psychophysical Methods: Detection, Discrimination, and Equivalence

The psychophysical methods presented in this section make it possible to answer basic questions concerning the user experience: Is a particular attribute of the sonic interaction perceivable? Are two different settings of a designed sonic interaction perceptually equivalent? Throughout the section, we focus on a simple attribute of the sonic interaction: the loudness of the sound. This example could be easily translated to a variety of sonic interactions such as the perceived loudness of sound effects in a videogaming context.

Psychophysics is the study of the mapping from physical attributes of the stimuli (e.g., sound level), to attributes of the corresponding sensations (e.g., loudness; see Gescheider [35] for an excellent handbook of psychophysics methodology). Classical psychophysical methods are often concerned with the measurement of two sensory quantities: the absolute threshold, which is the smallest or highest detectable value of a stimulus attribute (e.g., the lowest detectable sound level), and the differential threshold, which is the smallest discriminable difference in a stimulus attribute (e.g., the smallest discriminable difference in level).[1]

The *method of constant stimuli* has been widely used for measuring both absolute and differential thresholds. When measuring an absolute threshold (e.g., absolute threshold for sound level), participants are repeatedly presented with a small set of stimuli ranging from hardly perceivable (e.g., very low level) to clearly perceivable. Participants are asked if they detect the stimulus or not. When measuring a differential threshold, on each trial participants are presented with two stimuli: a standard stimulus whose properties remain constant across all trials (e.g., a 60 dB SPL sound) and a comparison stimulus that varies from trial to trial (e.g., one sound from a set ranging from barely weaker to barely louder than the comparison stimulus). Participants indicate for which of the paired stimuli the target attribute has the largest or lowest value (e.g., which of the two stimuli is louder). Figure 5.3 shows the likely outcome of a constant stimuli experiment: the function relating response probabilities to stimulus values is called the *psychometric function*. The absolute threshold can thus be defined as the stimulus intensity perceived 50 percent of the time; the differential threshold can instead be measured as the average of the stimulus values judged greater than the standard 25 percent and 75 percent of the time. More often than not, none of the presented stimuli is associated with the exact response probabilities used to calculate the thresholds. In part for this reason, and in part for the need to integrate experimental data across all the investigated stimuli, a psychometric function is usually fit to the observed response probabilities for all stimuli (e.g., a cumulative normal distribution [152, 153]), and the thresholds are estimated from the parameters of the fitted function.

The method of constant stimuli provides another important measure of sensation: the *point of subjective equality* (PSE). The paradigm used for this purpose is the same as

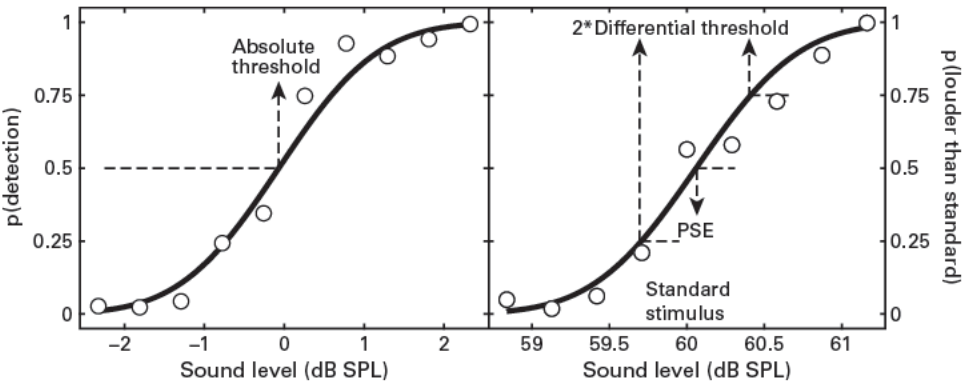

Figure 5.3
Hypothetical outcome of a constant-stimuli experiment for the measurement of the absolute and differential threshold for sound level (left and right panel, respectively). Circles show the response probabilities measured for each of the stimuli; solid lines show the psychometric function fitted to the experimental data and used to calculate the threshold measures. PSE, point of subjective equality. The differential threshold corresponds to half the distance between the stimuli associated to a response probability of 0.25 and 0.75, respectively. Note the slight misalignment of the PSE relative to the standard stimulus (60 dB SPL).

that used for measuring the differential threshold. In the level discrimination experiment described above, the PSE equals the value of the comparison stimulus judged louder than the standard 50 percent of the time (see figure 5.3). It should be noted that when the standard and comparison stimuli differ only in intensity, a PSE that significantly deviates from the intensity of the standard stimulus is a sign of imperfections in the experimental design [35, pp. 52–53]. A more interesting case of PSE measurement is the study by Robinson and Dadson [119]. In this experiment, the comparison stimulus was a 1-kilohertz tone of variable intensity, and the standard stimulus was a fixed-intensity sound of different frequency (e.g., 2 kilohertz). Participants judged whether the comparison or the standard stimulus was louder. The PSE was defined as the intensity of the comparison stimulus judged louder than the standard 50 percent of the time and measured the change in loudness brought by the frequency difference between the standard and comparison stimulus.

One shortcoming of the method of constant stimuli is its susceptibility to response biases. In an experiment designed to measure the absolute threshold for sound level, a very cautious participant might, for example, answer “I do not hear a sound” 80 percent of the time even though level is above threshold for only 50 percent of the trials. The framework of *signal detection theory* (SDT [48]) remedies this problem by computing bias-independent measures of sensitivity. The reader is referred to the works of McNicol [104] and of MacMillan and Creelman [95] for recent handbooks on SDT.

A second potential shortcoming of the method of constant stimuli is a low engagement of the experimental participant, who might quickly grow bored with the repetitive task. The *adjustment method* is a less tedious alternative that measures absolute and differential thresholds and PSEs. Accordingly, participants actively adjust the value of a stimulus property until a desired sensory result is achieved. For example, the participant might use a volume knob to adjust the intensity of a 1-kilohertz comparison stimulus so that it is perceived as loud as a 200-hertz standard stimulus of fixed intensity, thus producing an estimate of the PSE. The price of the adjustment method is an increase in the noise of the experimental data. In the PSE measurement experiment, imperfections in the manual control of the volume knob can, for example, produce a reduced accuracy of the adjustment response.

A final shortcoming of the method of constant stimuli is its low efficiency, which is the fact that many answers are required for each of the stimuli, even for the least informative ones, to yield reliable estimates of the parameters of the psychometric function. For example, when measuring the absolute threshold, the experimenter might be interested in a single point of the psychometric function associated with a response probability of 50 percent. However, the method of constant stimuli would also require collecting many responses for stimuli that are far from the absolute threshold. *Adaptive methods* are a more efficient alternative to the method of constant stimuli because the presented stimuli are concentrated around the point of interest on the psychometric function [34, 88, 143]. This is achieved by determining the level of the stimulus presented at a given trial based on the responses given at the preceding trials. The simplest and earliest example of adaptive method is the staircase or von Békésy tracking method [18, 146]. In an experiment for the measurement of the absolute threshold for sound intensity, the participant is asked to tell whether he hears the presented sound or not. Importantly, if the participant reports a detection at one trial, the intensity at the succeeding trial is decreased, whereas the intensity at the succeeding trial is increased if no detection has been reported (see figure 5.4). This simple rule for determining the value of the presented stimuli allows the experimenter to present sound intensities that are close to the absolute threshold. Alternative methods for determining the stimulus values provide target points on the psychometric function other than the 50 percent [88].

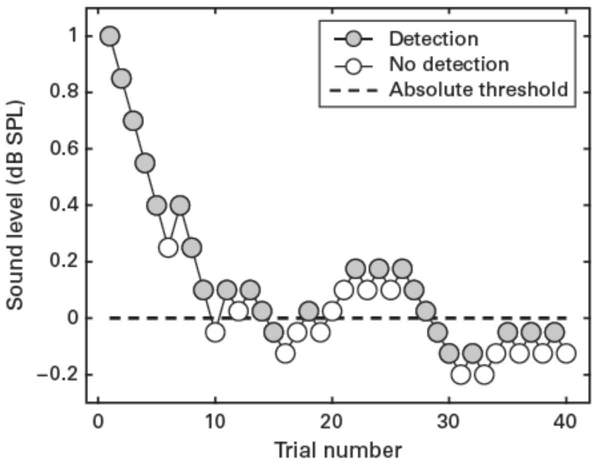

Figure 5.4
Hypothetical data from an experiment for the measurement of the absolute threshold for sound level based on the von Békésy tracking method. The dashed line shows the hypothetical absolute threshold. The level is decreased or increased by a constant step size after a detection or a no-detection response, respectively. Note the slightly larger step size for the initial trials meant to accelerate the convergence around the absolute threshold.

5.2 Identification and Categorization

Within a sonic interaction design context, identification and categorization methods can be adopted to assess the ecological interpretation of the displays: "Does this sonic interaction correspond to the scraping of a metallic surface or of a piece of styrofoam?" For this reason, these methods are among the most frequently adopted in the study of the perception of naturalistic sound events [39, 55, 89, 112, 118, 148]. In general, identification and categorization experiments allow the designer to assess the mapping from a set of display configurations to a set of meaningful verbal labels (e.g., identification of naturalistic events but also mapping of a stimulus set onto emotion-related categories such as "sad" or "happy").

During an identification/categorization experiment, participants are asked to assign each of the stimuli to one among a set of prespecified verbal labels (e.g., "Is this sound a violin, a guitar, or a flute tone?"). Whereas in an identification experiment the number of response categories equals the number of stimuli (e.g., "violin," "guitar" and "flute" in an experiment with three stimuli, one violin, one guitar, and one flute tone), with categorization the response categories are fewer than the number of stimuli (e.g., multiple violin, guitar, and flute tones of different pitch). Whereas in a classical psychophysical experiment stimuli are often highly controlled and vary along a very low number of dimensions (e.g., only sound level in the absolute-threshold example reported in section 5.1), identification and categorization experiments can be carried out with complex stimuli that differ along a large number of properties.

Data analysis can focus either on measures of performance (e.g., "Which among these tones has been identified correctly most often?") or on the raw probabilities of assigning each stimulus to each of the response categories (e.g., "Which among these tones has been most frequently identified as a violin tone?"). A third analysis option is to adopt SDT methods to compute measures of the sensory distance between response categories (e.g., "Is the sensory distance between violin and guitar tones shorter than that between violin and flute tones?") independent of response biases (e.g., in a categorization experiment with an equal number of violin, guitar, and flute tones, the tendency to use the response category "violin" more often than any other). This latter analysis approach was adopted by Giordano et al. [42] in an experiment investigating the effects of multisensory information (auditory, tactile, kinesthetic) on the identification of walked-upon materials. Measures of bias-independent sensory distance among walked-upon materials were computed within the framework of *general recognition theory* (GRT, see figure 5.5 [1, 2]). Unlike classical SDT methods, GRT makes it possible to deal with experiments in which stimuli vary along multiple properties, a frequent case when one is investigating naturalistic stimuli, and in which participants are allowed more than two response categories. Another advantage of GRT is that it considers within a single theoretical and analytical framework data from a variety of methods: identification, categorization but also dissimilarity (section 5.4), and preference (section 5.8).

5.3 Scaling and Rating

Perceptions can be organized in several different ways. In section 5.2 we saw that perceptions can be mapped onto discrete categories, each described by a verbal label. In this section we present a number of methods for measuring ordered relations among perceptions (e.g., the touch sensation of silk is smoother than that of wool). The concept of *sensory continuum* is central to these methods. Sensory continua are the result of a mental computation that allows us to order stimuli relative to a specific attribute and can be conceptualized as directional lines in a cognitive space (e.g., the sound of a flying bee is closer to the origin of the sensory continuum for loudness than the sound of a jet plane). Within a sonic-interaction design context, the methods presented in this section allow the answering of questions such as: "How does

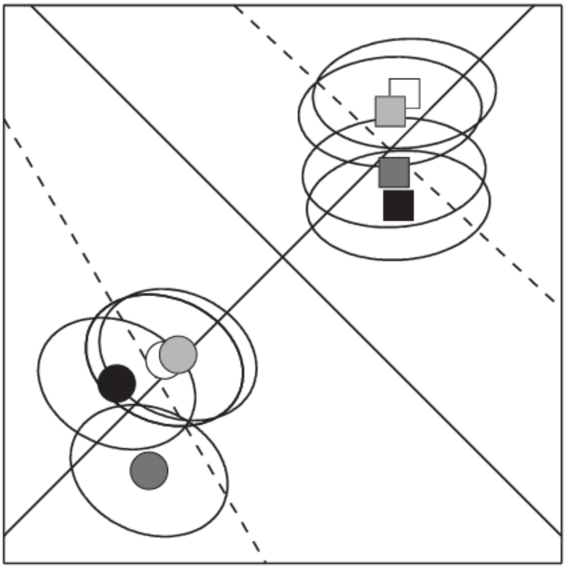

Figure 5.5
General recognition theory analysis of data for the multisensory identification of walked-upon materials [42]. The identification responses modeled in this example were collected after participants walked blindfolded on one of eight different materials: four solids (e.g., ceramic; circle symbols) and four aggregates (e.g., small gravel; square symbols). The sensory representation of each material is modeled as a normal distribution in a two-dimensional space of sensory effects. The average of each normal distribution (filled symbols) denotes the most frequent sensory effects for a particular walking ground; the oval surrounding each filled symbol (0.05 equal-likelihood contour) gives an approximate representation of the extent to which the sensory effects for a given material vary across repeated presentations. Four lines (decision boundaries) divide the two-dimensional space into eight regions: sensory effects that fall within the same region receive the same identification response (e.g., sensory effects above the top-left-to-bottom-right diagonal line are identified as aggregate materials; sensory effects below the same diagonal are identified as solid materials). The measure of the ability to differentiate between two walking grounds independent of response biases is approximated in this figure by the area of overlap between their respective normal distributions.

perceived effort vary between sonic feedbacks for robotic surgery applications?" and "How should the user-controlled gain for sound level vary to produce a linear increase in perceived loudness?"

Scaling methods make it possible to measure the *psychophysical function*, that is, the function that relates physical attributes of the stimuli to sensory continua (e.g., sound level and perceived loudness).[2] We can distinguish between two families of scaling methods: partition and ratio scaling (see Gescheider [35] for a thorough presentation).

In a *partition scaling* experiment, participants are asked to divide a target physical dimension (e.g., sound level) into perceptually equivalent intervals. For example, participants are presented with two stimuli that bracket the range of variation of the physical dimension of interest (e.g., a low- and a high-level sound) and are asked to divide the sensory continuum of interest (e.g., loudness) into a prespecified number of equal sensory intervals (e.g., to adjust the level of this sound so that the difference between its loudness and that of the low-level sound is equal to the difference between its loudness and that of the high-level sound). Similar information can be collected using the *category scaling* method. Accordingly, participants are presented with a set of stimuli that differ along the physical dimension of interest and are asked to assign them to one of a set of prespecified categories, each representing a different level of the sensory quantity and each bracketing an equal range of sensory magnitudes (e.g., to assign each of these fifty differently loud sounds to five classes of progressively increasing loudness, with each class bracketing an equal range of loudness variation).

The methods of *ratio scaling* and *magnitude scaling* produce a mapping from the sensory continuum to a numeric continuum. Both methods have one *estimation* variant and one *production* variant. In a ratio estimation task, participants estimate numerically the ratio between the sensory magnitude of two stimuli (e.g., "What is the ratio between the loudness of sound A and that of sound B?"). In a ratio production task, participants adjust the physical properties of a stimulus so that the ratio of its sensory magnitude to that of a reference stimulus equals a prespecified number (e.g., "Adjust the level of sound A so that its loudness is one-fourth of that of the reference sound B"). With magnitude estimation, participants assign a number to the sensory magnitude of the first presented stimulus and estimate numerically the sensory magnitude of subsequent stimuli based on the number assigned to the first stimulus (e.g., "Given that you estimated the loudness of sound A to equal 100, what number quantifies the loudness of sound B?"). With magnitude production, the sensory magnitude for a reference stimulus is assigned a numerical value, and the participant is asked to manipulate the physical properties of a new stimulus so that its sensory magnitude equals a given number (e.g., "The loudness of this reference sound equals 20; adjust the level of sound B so that its loudness equals 50"). Results from estimation and production scaling methods are known to diverge systematically because participants tend to avoid extreme values along the response continuum (the numerical and physical continuum, for estimation and production methods, respectively). This systematic divergence is termed regression bias [136]. The regression bias can be dealt with by estimating an unbiased psychophysical function defined as the "average" of the functions obtained with estimation and production methods.

The *rating method* can be conceived as a variant of the magnitude estimation method. Accordingly, participants estimate the sensory magnitude by choosing an integer number within a prespecified range (e.g., "Rate the hardness of a hammer used to generate this sound by using the integer numbers from 1 to 7" [33]). Before any experimental data are collected, it is good practice to allow participants to establish a mapping between the range of variation of the target sensory property within the set of stimuli and the response scale. To illustrate the need of this additional step, we can hypothesize that participants are asked to rate loudness on a scale from 1 to 10, and that participants are not familiarized with the experimental set of stimuli before rating each of them. A participant at the beginning of this experiment might, for example, rate the loudness of one stimulus using the highest allowed response (10), and find out that subsequent stimuli are louder than the previously heard ones. In such a case, the ratings of the participant will not accurately measure the perceived loudness. A variant of the rating task uses a nonnumerical continuous response scale. For example, in an experiment on the estimation of the hardness of struck sounding objects [41], participants rated hardness by moving a slider along a continuous scale marked "very soft" and "very hard" at the endpoints. This approach circumvents eventual response biases originating from the tendency to use particular numbers more frequently than others (e.g., multiples of 5 in a 1 to 100 scale) and maximizes the amount of experimental information (e.g., the number of different rating answers that a participant can give is usually much larger with a slider than with a numerical scale including only integer numbers).

The method of *cross-modality matching* finally relies on the comparison of sensory magnitudes from different modalities. With this method, participants adjust the properties of a comparison stimulus so that the target sensory magnitude it evokes matches the magnitude of a target sensory attribute of the reference stimulus presented in a different modality. This method was, for example, used by Grassi [47] to investigate the estimation of the size of a ball from the sound it makes when bouncing on a plate.

On each trial, the reference stimulus was a bouncing sound. Participants were asked to estimate the size of the bouncing ball by manipulating the diameter of a circle presented on a computer screen. Focusing on the design of sonic interactions, cross-modality matching could, for example, be adopted to calibrate the sensory properties of simultaneous auditory and tactile displays.

5.4 Dissimilarity Ratings

Many naturalistic sonic interactions generate rich sensory signals from various modalities (e.g., walking on various gravels or grass; crumpling a sheet of paper). Similarly, sonic interactions designed in the laboratory can involve many different parameters that control the sensory information delivered to the user. Within such rich domains it is often unclear what sensory properties dominate perceptions (e.g., which sensory properties of a naturalistic event; which synthesis parameters of a designed sonic interaction). The method of *dissimilarity ratings*, also known as paired comparisons method, sheds light on this issue and, in combination with particular data-analysis methods, makes it possible to answer questions such as: "Which properties of a complex sonic interaction are most relevant for the user?" and "Do different individuals focus on different properties of the sensory event?" Dissimilarity ratings have been frequently adopted to characterize the perception of complex auditory stimuli such as musical sounds [51, 75, 102] or environmental sounds [57, 105].

The goal of a dissimilarity ratings experiment is to measure the perceptual distance between stimuli: very similar/dissimilar stimuli are separated by a short/large perceptual distance. The structure of a dissimilarity ratings experiment is very similar to that of a standard ratings experiment. On each, participants are presented with two stimuli and are asked to rate how dissimilar they are on a "very similar" to "very dissimilar" scale. Throughout the experiment, participants rate the dissimilarity of each possible pair of stimuli.[3] so that each of the participants yield a matrix of between-stimulus perceptual distances.

The most common strategy for the analysis of dissimilarity ratings relies on the mathematical model of multidimensional scaling (MDS [9, 20])[4]. In general, MDS represents the dissimilarity ratings as the distance between the stimuli in a Euclidean space with a given number of dimensions (see figure 5.6).[5,6] Pairs of stimuli rated as very dissimilar are also far apart in the MDS space; stimuli rated as very similar are very close in the MDS space.

A notable limitation of classic MDS arises from the fact that it yields *rotationally invariant* solutions. A rotationally invariant solution can be rotated arbitrarily without

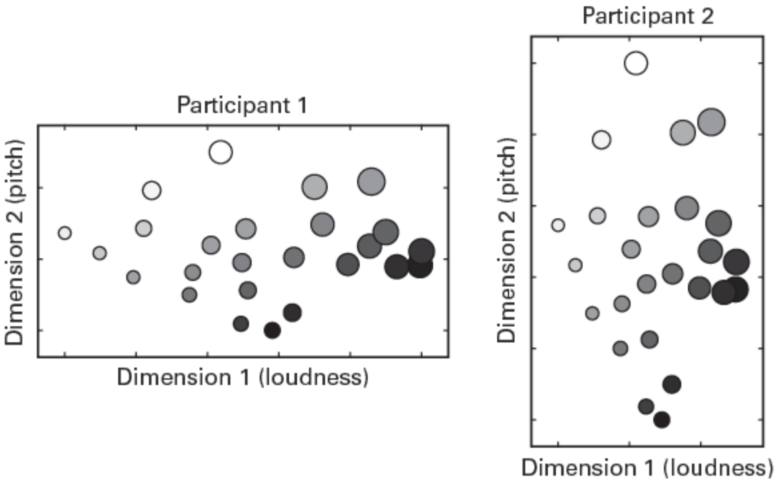

Figure 5.6
Weighted Euclidean model for a hypothetical dissimilarity ratings experiment carried out with 25 sounds differing in loudness (larger symbols denote louder sounds) and pitch (brighter symbols denote higher-pitched sounds). The MDS model in this example has two dimensions: dimension 1 is strongly correlated with the loudness of the stimuli; dimension 2 is strongly correlated with pitch. Participant 1 (left) carries out the task by focusing on loudness differences, hence the wider spread of the stimuli along the loudness-related dimension. Participant 2 (right) focuses on pitch differences, hence the larger spread of the stimuli along the pitch-related dimension.

affecting the extent to which the MDS distances accurately reproduce the input dissimilarities (e.g., a geographic map can be arbitrarily rotated without affecting the extent to which it accurately represents the distances between cities). In this sense, the dimensions of classic MDS spaces, that is, the location of the stimuli along each of the dimensions, do not provide meaningful information about the perceptual structure of the stimuli. A second limitation of classic MDS algorithms is that they accept in input only one matrix of dissimilarities and are thus not appropriate for modeling interindividual differences. The weighted Euclidean MDS model (wMDS) is free of these limitations (e.g., INDSCAL [141]): it is capable of analyzing simultaneously dissimilarity matrices from multiple experiment participants, and yields psychologically meaningful dimensions.

As in classical MDS, wMDS represents dissimilarities as the distance between stimuli in a Euclidean space. Notably, wMDS assumes that each of the participants weights differently the dimensions of a common group space. The individual weights are multiplicative terms specific to each of the participants that modify the spread of the stimuli along the dimensions of the group space, and allow her to reconstruct a space of mental distances specific to the participant that better accounts for her perceptions (see figure 5.6). Importantly, the wMDS model is not rotationally invariant. For this reason, the dimensions of the group space can be taken as a model of the response criteria followed to carry out the dissimilarity ratings task. For example, stimulus features that strongly correlate with the location of the stimuli along a specific dimension of the wMDS model are likely to have been used by participants to estimate the between-stimulus dissimilarity (note that the correlation between a dimension of the wMDS model and a stimulus feature is not affected by the multiplication of the dimension by a participant-specific weight). Further, within the wMDS model the range of variation of a given dimension for each of the participants can be taken as a measure of the perceptual relevance of the dimension to the participant herself, that is, if for one participant the first dimension has a larger range of variation than the second dimension, she likely carried out the dissimilarity ratings task by focusing more on the feature that correlates with the first dimension than on the feature that correlates with the second dimension (see figure 5.6).

A very important aspect of the dissimilarity ratings method is that it does not constrain participants to focus on a specific property of the stimuli. Indeed, such a generic task as “rate the dissimilarity between these two stimuli” leaves the participant free to decide which stimulus properties most strongly affect dissimilarity. In this sense, dissimilarity rating is the most liberal among the experimental methods presented up to this point. Dissimilarity rating thus frees the experimenter from assumptions about the perceptual structure of the stimulus set: asking participants to scale a particular property of the stimuli (e.g., the size of a bouncing ball from its sound [47]) indeed comes with the hidden assumption that perceptions are organized along the rated property. Such an assumption can prove hard to be tested empirically (e.g., What is the object of perception when hearing the sound of a bouncing ball? Size, weight, or the acceleration at the moment of impact with a sound-generating object?). The liberal nature of dissimilarity ratings thus makes it a particularly good method for exploratory studies of complex perceptual domains (e.g., environmental sounds [57]), for which it might just not be possible to formulate a working hypothesis on what stimulus attributes are perceptually relevant (e.g., “Which stimulus attributes capture the perceptual difference between the sound of an airplane engine and that of a dog howling?”). For all these reasons, the quantification of perceptual distances can be considered the method of choice for the exploration of novel, complex perceptual domains (see also Borg & Groenen [9, pp. 9–11]).

We have seen that dissimilarity ratings, in conjunction with wMDS modeling, makes it possible to (1) discover perceptually relevant attributes of the stimuli; (2) establish hierarchies of perceptual relevance for stimulus attributes (e.g., loudness is more perceptually relevant than pitch); and (3) analyze interindividual differences for the perceptual relevance of stimulus attributes. From the applied perspective, the designer of sonic interactions might focus the modeling efforts on those properties of the interactive events that dominate the perceptions of the majority of the individuals.

5.5 Sorting

One of the goals for a designer of sonic interactions might be to establish a perceptually meaningful palette of presets for the parameters of complex audio-haptic synthesis algorithms. Palettes of presets are, for example, common in the control of digital color spaces (e.g., the user is able to select from various color presets rather than having to specify the RGB value corresponding to the desired color). Within the context of sonic interactions, the designer might, for example, plan to create a palette of presets for the control of the perceived material of virtual audio-haptic objects. Faced with this problem the designer needs to decide how many presets should be included in the palette and what parameter values should be assigned to each of them: "How many categories of perceived materials can the synthesis algorithm reproduce, and what is the most typical configuration for each of them?" Sorting methods make it possible to answer these questions and, in general, can be adopted to define meaningful categories of events within a perceptual domain. Further, sorting methods can be adopted as a more efficient although less reliable and accurate alternative to dissimilarity ratings (see section 5.4) [38].

With sorting experiments, participants are quite simply asked to create groups of stimuli. In an image-sorting experiment, each image might be printed on a card, and participants might be asked to create groups of cards. In a sound-sorting experiment, each sound might be represented by an icon on a computer screen, and participants might be asked to group the icons. One of the first decisions to be made when designing a sorting experiment how many groups of stimuli the participant should create. In the *free sorting* variant [121], the experimenter does not specify the number of groups and leaves the decision up to the participant. A free sorting task could, for example, be used to answer the materials-palette problem described at the beginning of this section (e.g., "group together stimuli generated with the same material; create as many groups you think are necessary"). With *constrained sorting* the number of groups is instead specified by the experimenter. Gingras, Lagrandeur-Ponce, Giordano, and McAdams [36] used constrained sorting to test the recognition of the individuality of music performers. In this experiment participants were presented with excerpts of organ-performance recordings from six different performers and were asked to create six different groups of excerpts played by the same individual.

A second important decision to be made when designing a sorting experiment concerns the criteria that participants should follow when creating the groups of stimuli. In the above examples participants receive specific instructions about the criteria they should follow to create the groups. When this is the case, sorting methods bear a strong resemblance to the categorization task described in section 5.2. The only major differences are indeed that categories are not labeled verbally by the experimenter (free and constrained sorting) and that the number of categories may not be specified (free sorting). Most often, however, participants are given the more generic instructions to create groups based on the similarity of the stimuli (e.g., for studies carried out with complex naturalistic sounds [8, 38, 53, 57]). When this is the case, the sorting task is the "categorization analogue" of dissimilarity ratings and shares with this method the ability to uncover the structure of perceptual domains within an assumption-free framework (see section 5.4). Similarity-based free sorting further allows one to measure the so-called basic level of categorization in a stimulus domain [120]. Sorting data, coding for the group membership of each of the stimuli, are frequently converted into co-occurrence data (see Coxon [21] for various data-analytic strategies for sorting data). For each possible pair of stimuli, the square co-occurrence matrix uses a binary variable to code whether stimuli have been assigned to the same group or not. Co-occurrences collected with a similarity-focused sorting are often considered as binary measures of dissimilarity: dissimilar stimuli do not belong to the same group (co-occurrence = 0), whereas similar stimuli do (co-occurrence = 1).

Sorting makes it possible to measure a full dissimilarity matrix in a shorter time than dissimilarity ratings [4, 38, 121]. For this reason sorting is often adopted to measure the dissimilarity of large sets of stimuli, or the dissimilarity of stimuli that might easily produce adaptation effects (e.g., taste stimuli [84]). The higher efficiency of sorting comes, however, with a price: dissimilarity data are less accurate, that is, less likely to reflect stimulus properties, and less reliable, that is, less likely to be replicated with a different group of participants [38]. Another drawback of sorting methods concerns the data-modeling aspect, in particular the extent to which MDS algorithms accurately model sorting dissimilarities. As explained above, each of the participants in a sorting experiment yields a binary dissimilarity matrix, the co-occurrence matrix.

Notably, MDS models of binary dissimilarities are known to present strong artifacts that prevent an accurate account of the input data [44]. As such, interindividual-differences MDS models of sorting data (e.g., wMDS, see section 5.4) are prone to significant errors. For this reason sorting data are less than ideal for measuring and modeling interindividual differences in perceived dissimilarity. It is thus advisable to carry out MDS analyses of sorting data by focusing on the co-occurrence matrix pooled across participants.

The *hierarchical sorting method* finally yields richer participant-specific data than free sorting (e.g., Rao & Katz [117]; see Giordano et al. [38] for various comparisons between sorting methods and dissimilarity ratings). In the agglomerative variant of this method, participants start from a condition where each of the stimuli is in a different group and, at each stage of the procedure, merge together the two most similar stimuli or groups of stimuli. The procedure is iterated until all stimuli are merged in the same group. Between-stimulus dissimilarity is estimated by the step of the merging procedure when two stimuli are first grouped (e.g., dissimilarity = 1 and 10 if stimuli A and B have been merged together at the first or tenth stage of the sorting procedure). With this method, each of the participants thus yields a square dissimilarity matrix with as many different values as the number of merging steps. Figure 5.7 shows the hierarchical sorting for one of the participants in an experiment carried out with environmental sounds [40] (also see other studies of sound stimuli based on the hierarchical sorting method [38, 61]). Note that in the first step of this hierarchical sorting experiment participants created 15 groups of similar stimuli. In other words, the hierarchical sorting started with a constrained sorting step (*truncated hierarchical sorting*; see Giordano et al. [38]).

5.6 Verbalization

Verbalization tasks can be used to explore qualitatively the perceptually relevant attributes of sound-interactive events: “Do users focus on the characteristic of the sound signal (e.g., this sound is very bright), on the characteristic of the source (e.g., this is the sound of a breaking glass), or on more abstract symbolic contents associated with the sound event (e.g., harmfulness for pieces of shattered glass)”? Focusing on interactive events, verbalizations might be collected during an interview where participants are shown the video recording of their own interaction with a sonic prototype (auto-confrontation interview). Analysis of the verbalizations will, for example, provide useful feedback on individual interactive strategies or on problems in the design of the prototype.

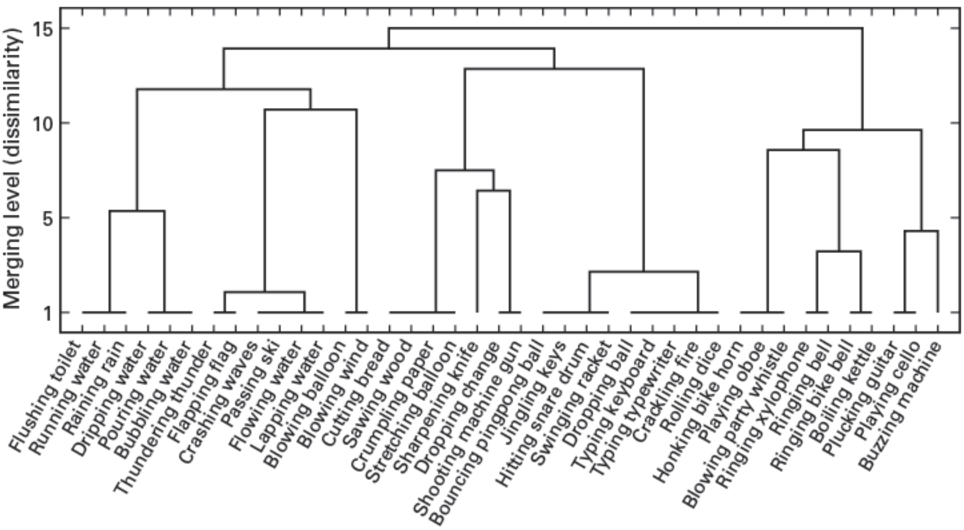

Figure 5.7
Truncated hierarchical sorting data for nonliving environmental sounds (data for one participant in the unbiased condition in Giordano et al. [40]). Participants initially created fifteen groups of similar sounds. At each subsequent step, participants merged together the two most similar sounds or groups of sounds. Similar sounds are assumed to be merged earlier than dissimilar sounds: the sorting step where two sounds are first merged yields an estimate of their dissimilarity.

Free verbalization is the simplest method for the collection of verbal data. Accordingly, participants are asked to describe in words the presented stimuli, with no restriction being imposed on the format of the response (e.g., "describe the sound you heard"). The high discovery potential of a free verbalization experiment comes at the price of a high complexity of data that can differ greatly across experiment participants. Simpler verbal data that require less sophisticated analysis strategies can be obtained using *constrained verbalization* methods that impose limits on the format of the verbal responses (e.g., "use one verb and one/two nouns to describe the sound-generating event" [40]). Verbalization methods can be used either to characterize stimuli presented in the laboratory [145] or as a method to sample the environment (verbalization of sounds as they are experienced in outside the laboratory [3]). Verbalization can be used as a method of its own or as a complement to a preceding task (e.g., free verbal description of the groups of sounds created in a sorting task, section 5.5) or as a strategy for designing further experiments (e.g., verbalization data can be used to design semantic-differential experiments, section 5.7).

Various software routines are available for the analysis of verbalization data. The software LEXICO [16] can be used to tabulate the terms used by participants according to their frequency of occurrence in the verbalization responses. The software STONE [115] can be used to carry out a lexical analysis that organizes the semantic fields among the different verbal descriptions. Methods of natural language processing can be used to carry out various analyses of the semantic content of the verbal descriptions. For example, latent semantic analysis techniques [83] can be used to compute measures of the semantic similarity of the verbal descriptions used to identify a sound-generating event [40]. Specific to research on complex sounds, methods for the scoring of the accuracy of verbal identifications of a sound-generating event have been detailed [3, 40, 98]. Further methods for the analysis of verbal response data have been developed by Nosulenko and Samoylenko [108]. Notably, the approach developed by Nosulenko and colleagues makes it possible to organize and encode verbal units at different levels: logical, perceptual, and semantic.

Verbalization tasks have been frequently used in the study of complex sounds. Peters [114] investigated the verbalization of synthetic sounds (e.g., pure tones, white noise), speech, music, and environmental sounds. Overall, terms describing the sensory properties of the sounds (e.g., high, low, soft, loud) were used more often than words related to objects, actions, and events involved in the generation of a sound. Nonetheless, only one-third of the participants used terms describing the sensory quality of the stimuli for real sounds. Similarly, in a study on the free identification of environmental sounds, Vanderveer [145] observed that listeners most often described the sound-generating event and, more specifically, the actions and objects involved, , and the context where the sound was generated. Faure [32] asked participants to freely describe pairs of musical sounds presented during a dissimilarity estimation experiment. Verbal responses belonged to one of three semantic categories: descriptors of the sound source (e.g., material, action), descriptors of the temporal evolution of a sound (e.g., attack, progression, resonance), and descriptors of sensory aspects of a sound (e.g., sharp, light, bright). Kyncl and Jiricek [82] investigated the free verbalization of vacuum cleaner sounds. Thirty-three pairs of semantic opposites were derived from the verbal responses, five of which were consistently judged as relevant to describe the vacuum cleaner sounds: fuzziness, atypicality, inefficiency, loudness, and pleasantness.

5.7 Semantic Differential

The *semantic differential* method can be conceived as a multidimensional extension of the ratings method (section 5.3). The main difference between the two methods indeed stands in the number of psychological or perceptual attributes simultaneously evaluated by the experiment participant: one for the latter, more than one for the former. The semantic differential method can, for example, be adopted to optimize a sound-interactive system according to multiple psychological attributes (e.g., "Which configuration has the highest aesthetic and functional attributes?") and to analyze the interdependence among multiple properties of sonic interactions (e.g., "How do preference, perceived sound brightness, and perceived efficiency covary?"). Various studies adopted the semantic differential to assess the multidimensional character of complex sound stimuli (e.g., Solomon [133] and von Bismarck [147] for early applications; and various sources for studies on musical timbre [71, 116, 134, 151]; for studies on environmental sounds [6, 72, 81, 155]; for studies of sounds generated with various human-made objects such as cars, vacuum cleaners, air conditioning systems, and refrigerators [5, 15, 65, 67, 70, 82, 131, 139]).

In the most popular variant of the semantic differential method [109, 110], participants rate each stimulus along several bipolar scales defined by opposing semantic descriptors (e.g., "soft" and "loud" or "pure" and "rich"). Each bipolar scale is usually divided into an odd number of intervals (e.g., seven). The task of the participant is thus to choose which of these intervals most appropriately describes the location of the stimulus along the continuum defined by the opposing semantic descriptors (e.g., this sound has a loudness of 5 along a soft-to-loud scale with seven intervals). As described for the ratings method, the rating scale does not need to be clearly divided into a predefined number of intervals (e.g., ratings along each bipolar scale can be collected using on-screen sliders or by marking with a pen a position along a line connecting the two opposing descriptors). The method of *verbal attribute magnitude estimation* (VAME) is a variant of the semantic differential method that might improve the interpretability of the results [59, p. 259]. With the VAME method, bipolar scales are not defined by opposing semantic descriptors but by one semantic descriptor and its negation (e.g., "not loud" and "loud").

Semantic differential studies are not restricted to particular attributes: they can be adopted to evaluate various properties of the stimuli such as sensory attributes (e.g., sound roughness), higher-level psychological attributes (e.g., pleasantness), or emotional attributes (e.g., dominance). As a result, this method can be a good choice for the study of previously unexplored perceptual domains. Two considerations are important concerning the design of a semantic differential experiment. First, in the absence of previous literature on the investigated stimuli, the choice of what semantic attributes should be considered can be based on a preliminary verbalization experiment (see section 5.6). For example, participants in the preliminary experiment might be presented with each of the stimuli and asked to describe verbally their most salient attributes. The semantic differential can thus be designed by considering those semantic attributes used by the majority of the participants in the verbalization experiments. A second design consideration concerns the number of attributes included in the semantic differential. It is likely that as the number of semantic attributes grows, participants are more likely to answer using response strategies meant to reduce the difficulty of the task. In particular, with an excessive number of semantic attributes, participants might start using response scales in a correlated manner (e.g., sounds that are rated as brighter are also rated as more pleasant) not because of a genuine association between the semantic attributes within the stimulus set (e.g., all feminine voices have a high pitch) but simply in order to minimize fatigue. To minimize these effects it is advisable to limit the number of semantic attributes to the minimum necessary. Independently of the origin of the correlation between different response scales, statistical methods such as factor analysis or principal components analysis can be adopted to reduce the raw semantic differential data into a set of independent dimensions of evaluation.

5.8 Preference

A sound designer often aims to improve the overall quality of sonic interaction experiences. To this purpose she might seek an answer to questions such as "Which configuration of a sonic feedback system do users prefer?" or, almost equivalently, "Which configuration is the least annoying?" Preference judgments have been often adopted in the applied sector to improve the design of a variety of products [54] (see Ellermeier & Daniel [27, 28] for studies of tire sounds and environmental sounds; Susini and colleagues, [139, 140] for studies of car sounds and air-conditioning noises; Lemaitre, Susini, Winsberg, McAdams, & Letinturier [87] for a study of car horns).

We can distinguish between two types of preference data: revealed and stated preferences [7]. Revealed preferences are usually derived from choice data collected in ecological conditions (e.g., product sales). One of the main disadvantages of revealed preferences is that the complete set of alternative choices is often unknown (e.g., data on the sales of Gibson Les Paul and of Fender Stratocaster guitars lack information about all of the brands and makes of electric guitars considered by the costumers of a musical instrument shop). In the following, we focus on stated preferences which are directly elicited from the participants and are not characterized by this important drawback of revealed preferences.

Preference data can be collected with various experimental methods [50], some of which have been described in previous sections. In a *paired preference comparison* task, participants are presented with all the possible pairs of stimuli, one at a time, and are asked to choose which of the two they prefer. In a *preference ranking* task, participants are presented with all of the stimuli at once and are asked to arrange them from the least to the most preferred. With *preference ratings*, participants rate their preference for each of the stimuli on a categorical or continuous scale (see section 5.3).

The most peculiar aspect of the study of preference is not the behavioral method used to collect the data, but strategy adopted for their analysis. Here, we briefly describe several models of preference judgments and list various references for the interested reader.

The law of comparative judgment by Thurstone [142] can be considered as one of the earliest models developed for the analysis of paired preference comparison data. Within this framework, the probabilities of preferring one stimulus over the other are used to determine the position of each of the stimuli along a preference continuum. A similar representation can be computed based on the Bradley-Terry-Luce (BTL) model [10, 90]. More complex preference models locate stimuli in a preference space with a given number of dimensions. We can distinguish between two classes of such models (see figure 5.8): ideal point and ideal vector models. Ideal point models represent experiment participants as points in the same preference space where stimuli are positioned. The location of the participant within the space models the hypothetical stimulus he prefers the most, that is, the ideal point, and his preference for each of the experimental stimuli is represented by their distance from the ideal point (stimuli farther from the ideal point are less preferred). Ideal point models can, for example, be computed by using unfolding algorithms [9, 17, 20, 24, 25]. With ideal vector models, stimuli are also represented as points in a space with a given number of dimensions, whereas participants are represented as vectors oriented toward a maximum preference point located at infinity. The preference for the experimental stimuli is thus modeled as the location of their projections onto the ideal vector, with preferred stimuli located further along the direction of the ideal vector. Ideal vector models can be fitted using the MDPREF algorithm [14].

The dimensions of various multidimensional models of preference can, in general, be interpreted similarly as the dimensions of the weighted Euclidean model (wMDS). For example, with the MDPREF model the dimensions can be interpreted as corresponding to the stimulus attributes that most strongly influence preference in the population of participants. Further, within the same model the length of the projection of a participant-specific ideal vector onto a dimension measures the relevance of that dimension, that is, stimulus attribute for the participant.

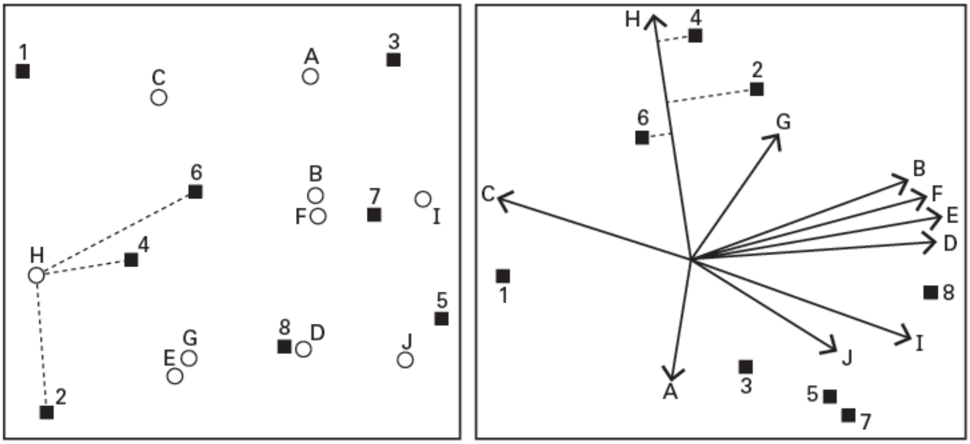

Figure 5.8
Multidimensional models of preference (hypothetical data). In both panels stimuli are represented as numbers, and participants are represented as letters. In the ideal point model (left), preferred stimuli are located closer to the ideal point (one for each of the participants). For example, the three most preferred stimuli for participant H are stimuli numbers 4, 2, and 6. In the ideal vector model (right), participants are represented as directional vectors of preference. The preferences of a given participant can be visualized by projecting the stimuli onto his ideal vector (dashed lines): preferred stimuli are located farther along the direction of the ideal vector.

All of the abovementioned models of preference operate on behavioral data and do not require a knowledge of the stimulus properties (e.g., price of a guitar). For this reason the study of what properties of the stimuli influence preference is usually carried out after the preference model has been created (see section 5.11). The quantification and prediction of choice and preference based on measured properties of the stimuli are instead an integral part of a family of experimental and analysis methods that flourished within the field of marketing research: conjoint measurement [49]. The reader interested in conjoint measurement has several sources [13, 50, 54].

In summary, the mathematical modeling of preference can be used by the designer to discover those configurations of a sonic feedback system that are most preferred (or least annoying) for the majority of the participants. In the case of meaningful dimension of a multidimensional preference space, the designer might thus decide to focus the modeling efforts on those stimulus properties/synthesis parameters that most strongly correlate with the most relevant dimensions of the preference space (see also section 5.11).

5.9 Continuous Evaluation

Sonic interactions are time-varying events: posture, sound, touch, and visual properties all change in time as a consequence of the motor behavior of the user. Importantly, the temporal variation in sensory information is likely to result in a temporal variation of the experienced properties of the sonic interaction itself. The behavioral methods presented in the previous sections are not able to account for this level of complexity of a sonic interaction because, in general, they produce one single data point for each of the presented stimuli. Continuous evaluation methods instead permit the measurement of the temporal dynamics of the perceptions of a user (e.g., "How does the emotional response to a complex sound vary in time?") because they continuously sample perceptions throughout the entire duration of the stimulus.

Participants in a continuous evaluation experiment are asked to repeatedly judge some perceptual/cognitive attribute as the stimulus unfolds in time (e.g., a musical composition or a soundscape). No limitation is imposed on the type of task carried out by the listener: we can thus have a continuous ratings task (e.g., rate continuously along a pleasant/unpleasant scale), a continuous categorization task (e.g., "What emotions are you currently feeling? Sadness, happiness, fear, or anger?"), or a continuous preference estimation task (e.g., "Rate your current preference for the sound material on a least preferred to most preferred scale"). Table 5.2 provides a classification of the type of judgments and stimuli investigated in a number of continuous evaluation studies.

Table 5.2
Summary of behavioral methods used in previous continuous evaluation studies

Behavioral method	Sound stimuli	Study
Ratings (categorical scale)	Road traffic	[78, 79]
	Trains	[106]
	Helicopters	[80]
	Car accelerations	[77]
Ratings (continuous scale)	Road traffic	[30, 31, 45]
	Music	[107, 97, 125, 101, 144]
	Speech	[19]
Ratings (continuous and categorical scales)	Road traffic	[150, 64, 111]
	Synthetic sounds	[62, 137, 138]
	Speech	[60, 52]
Cross-modality matching	Pure tones	[137]

The design of a continuous evaluation experiment should take into account the maximum temporal resolution of the responses that an experiment participant can reliably give. The temporal resolution of the response is indeed limited by various factors such as the constants of temporal integration governing a particular perceptual or sensory process (e.g., loudness integration) or the speed of the motor responses of a participant (e.g., how rapidly a participant can move a slider). For example, whereas it is unlikely that a listener can rate the time-varying loudness of a stimulus with a temporal resolution higher than 100 milliseconds (ten ratings per second), it is plausible that a reliable rating can be collected at slower temporal rates such as two ratings per second. Because of these limitations, continuous evaluation methods are better suited for quantifying the perceptual dynamics for relatively long stimuli (e.g., several minutes) rather than for very short ones (e.g., sounds shorter than 1 second).

In regard to sonic-interaction contexts, logistics make it hard if not impossible to ask a participant to interactively generate a sound while evaluating simultaneously his time-varying perceptions. For this reason, it might be advisable to carry out continuous evaluations on offline recordings of sonic interactions (e.g., videos, sound recordings). Despite this limitation, this method will permits to address important sonic-interaction issues, such as the temporal correlation between the gestures of a user and the perceptions of a passive listener. The implementation of continuous evaluation must finally meet a number of requirements: judgments must be made easily, rapidly, and without discontinuity. Several methods combined with specific devices for response collection have been proposed. The interested reader is referred to the work of Schubert [126, 127] for a description of the methodological issues associated with the continuous evaluation of complex musical materials.

Continuous evaluation methods have been used for the study of speech quality [52, 60], for the quantification of the effects of running bus sounds on comfort [111], to estimate auditory brightness [62], and to measure emotional responses to music [76, 101, 132, 144]. Various studies have assessed the relationship between the instantaneous judgment of a given perceptual property of the time-varying stimulus and the global judgment of the overall level of the same property for the entire stimulus. Kuwano and Namba [79] and Fastl [31], for example, showed that judgments of the global loudness of a complex sound are strongly related to the peaks of the time-varying loudness (see also [45, 63, 137]).

5.10 Multisensory Contexts

By definition a sonic interaction involves multiple sensory modalities: audition, touch and kinesthesia (haptics), and possibly vision.[7] Within the context of multisensory interactions, a relevant issue might, for example, be whether each of the involved sensory modalities affects a behavioral variable of interest (e.g., "Does the sound of car doors influence preference?") or which of the sensory modalities has the strongest behavioral effect (e.g., "Which has the stronger impact on preference: the sound of a car's doors closing or the felt weight of the doors?"). Alternatively, the goal of the designer might be to improve some aspect of the performance of the user based on multiple sensory inputs: "Do sonic feedbacks shorten the time required for parking a car?" Answering most of these questions does not usually require experimental tasks different from those described in previous sections. As detailed in this section, the study of multisensory contexts might nonetheless pose specific challenges and require specialized experimental paradigms.

The study of multimodal perception and performance typically involves a comparison of data collected when only one modality is stimulated with data collected in multisensory contexts (e.g., when only touch is available vs. when both sound and touch are available). Designing an experimental condition where information from one single modality is presented might not always be trivial, especially when participants interact with naturalistic objects.

A recent study by Giordano et al. [42] illustrates this issue. The goals of this study were to measure the effects of various sensory modalities and of their combination on the identification and discrimination of walked-on grounds: audition, kinesthesia (e.g., perception of limbs movement), and touch. Participants carried out a simple identification task (see section 5.2). The ideal experiment would have included seven experimental conditions: three conditions for each of the sensory modalities in isolation, three conditions where only one of the three modalities was suppressed, and one condition that combined information from all modalities. The general approach adopted to suppress information from one modality was based on the phenomenon of masking: a signal cannot be perceived if presented together with a sufficiently intense masking stimulus (e.g., random signal or noise). Auditory information was thus suppressed by presenting walking participants with an intense noise over headphones, and leaving both touch and kinesthetic information intact. Similarly, touch information was disrupted by presenting walking participants with a random mechanical vibration at the foot. The acoustical and tactile noises were presented simultaneously in a condition where only kinesthetic information was left intact. Notably, the mechanical actuator used to vibrate the foot also generated an audible noise (the range of frequencies that can be perceived with touch partially overlaps with the range of auditory frequencies). For this reason, an auditory-kinesthetic condition could not be investigated. Further limitations in the experimental design arouse from the impossibility of suppressing kinesthetic information while participants walked. Apart from the logistic challenge, such an experimental condition would have likely produced abnormal, and most importantly, unstable locomotion. For this reason the auditory-tactile condition could not be investigated, and the auditory-only condition could be carried out only with nonwalking participants who were presented with recordings of the sound they had generated while walking on the grounds. The study by Giordano et al. [42] thus exemplifies the compromises and challenges faced in the study of multimodal naturalistic sonic interactions and shows the usefulness of noise-based disruption of sensory information in both the auditory and tactile domains.

A frequent goal of studies on multisensory contexts is to establish modality dominance hierarchies, that is, to assess which modality affects most strongly the perception of multimodal events. This question is often answered by investigating *multisensory conflicts*. Conflicting multisensory events combine contrasting information from different modalities, that is, modality-specific stimuli that induce different perceptual results when presented in isolation (e.g., a light located slightly to the left of the point of fixation and a sound presented slightly to the right of the same point). The experimental paradigm relies on the comparison of perceptions for the unimodal stimuli with perceptions of multisensory conflicts (e.g., "Provided that this light is perceived to be located on the left, and that this sound is perceived to be located on the right, what will be the perceived location of the light-sound event?"). In general, the perceptions for a multisensory conflict will be most similar to those for the dominant-modality stimulus presented in isolation. This basic paradigm can be used to reveal the dominance of vision on audition in the estimation of the spatial position of an event (ventriloquist effect [66]), the effect of seeing the movements of the lips of a speaker on the speech perception [103], or the influence of the number of sound beeps on the perceived numerosity of visual flashes ([129]; see Shams, Kamitani, & Shimojo [130] for an overview of the effects of auditory information on visual perception and Lederman & Klatzky [86] for an overview of research on the multisensory perception of audio, haptic, and visual textures). From a design perspective, it would be reasonable to focus modeling efforts on the modalities that dominate the perceptual responses of a user.

The work of Ernst and Banks [29] combines the study of multisensory conflicts with an interesting experimental manipulation of sensory noise. In simple words, the ML model predicts that in a multimodal context modality-specific information that is better discriminated when presented in isolation affects more strongly the multisensory perceptions. To test this prediction, Ernst and Banks investigated the perception of the height of virtual visual-haptic ridges. The method of constant stimuli was used to measure the perceived height of ridges in visual-haptic conflicting stimuli where haptic and visual height differed. The main goal of the experiment was to measure how the conflicting visual and haptic heights were combined to yield an estimate of visual-haptic height. The experiment included various visual noise conditions where the visual ridge stimulus was perturbed randomly with a noise of variable intensity. Notably, increased noise levels produced a more pronounced impairment in the ability to discriminate ridge height based on visual information alone. In these conditions the ML model predicted that the perceived height of the visual-haptic ridge progressively approached the haptic height as the level of the visual noise increased. The experimental results were in strong agreement with this prediction. From the methodological point of view, the work by Ernst and Banks shows how noise can be used not only to occlude completely the information from one sensory modality but also to modulate its discriminability. The possibility of predicting multisensory dominance based on measures of the discrimination of unimodal information might also be particularly useful to the design process. Indeed, with this paradigm the designer will be able to estimate which modality will dominate a multisensory interaction based on unimodal displays even before further efforts are spent on combining each of the modalities in a single multimodal display.

The paradigm developed by Giordano et al. [37] adopts a different approach for the study of multisensory dominance in interactive contexts, based on measures of motor activity rather than based on non-action responses (e.g., "Is this light presented to your right or left?"). In this study participants are asked to strike an audio-haptic virtual object with a target velocity. Each trial is divided in three phases (see figure 5.9). During a first training phase, participants receive feedback on whether their striking velocity is within the target range. During a subsequent adaptation phase, feedback is eliminated, and the virtual object is kept unchanged. Participants are asked

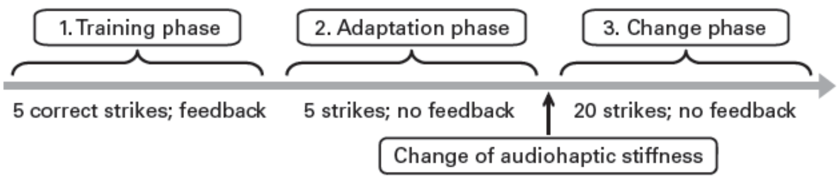

Figure 5.9
Structure of experiment trials in the study by Giordano et al. (audio-haptic condition [37]). The training phase ended after five consecutive correct strikes within the target velocity range. The adaptation and change phases ended after five and twenty strikes, respectively, independently of whether they were within the target velocity range or not. Participants received correctness feedback only during the training phase.

to continue striking the object with the same velocity. In a final change phase, the properties of the virtual object are changed (auditory or haptic stiffness). Multisensory dominance is inferred from the strength of the effects of modality-specific changes on striking velocity, where a change in the dominant modality (haptics, in this experiment) leads to larger changes in striking velocity. The assessment of multisensory perceptions based on the motor activity of the user might be particularly useful for those contexts where a continuous interaction of the user with the display precludes the possibility of easily collecting non-action judgments such as verbally communicated responses.

A further topic of interest in the study of multisensory contexts focuses on the effects of the congruence of different modalities on perceptions [68]. For example, Giordano et al. [37] measured the extent to which the congruence of the auditory and haptic stiffness of the virtual object influenced velocity-tracking abilities. On each audio-haptic trial the change in audio-haptic stiffness could be congruent (e.g., both the auditory and haptic stiffnesses increase above the initial level) or incongruent (e.g., an increase in auditory stiffness and a decrease in haptic stiffness). Only the behavioral effects of the change in auditory stiffness were modulated by congruence. In particular, when auditory stiffness changed in the opposite direction from haptic hardness, auditory information appeared to have no effect on striking velocity. Further paradigms in the study of multimodal congruence are illustrated by Marks [99]. From a design point of view, measurements of the effects of multimodal congruence can, for example, be used to predict more accurately the perceptions of users or can inform the process of linking together modality-specific displays into a single multisensory interactive object.

5.11 Measurement of Acoustical Information

One of the frequent goals of studies on the perceptual processing of complex sounds is to pinpoint what sound features influence the answers of the experimental participant. From the scientific perspective, this analysis stage aims at quantifying the acoustical information relevant to the perception and sensory processing of complex sounds. From the applied point of view, this analysis makes it possible to discover those sound properties that should receive the attention of the sound designer in order to achieve a target perceptual result (e.g., maximize preference, simulate variations in the perceived hardness of struck objects; see Gygi & Shafiro [58] for various applications of research on complex naturalistic sounds).

With most of the approaches to the characterization of the perceptually relevant acoustical information, a priori hypotheses about what properties of the sound might influence the judgments of experiment participants are necessary. To this purpose, each of the experimental stimuli is often characterized by a short sequence of numbers, each of the numbers measuring one attribute of the sound or acoustical feature (e.g., attack time, approximately the speed of level increase at sound onset [102]). It is important to note that our knowledge of the psychophysics of complex everyday sounds is still far from complete. Thus, the problem of defining sound features is largely unconstrained and can be practically endless: a large number of mathematical operations can indeed be adopted to describe a sound with a single number [41]. For these reasons, we omit from this presentation a definition of the various sound features previously used to study the perception of complex sounds. We nonetheless refer the interested reader to the work of Peeters, Giordano, Susini, Misdariis, and McAdams [113], for a recent and extensive system for the extraction of acoustical features.

In general, two different strategies can be adopted to discover perceptually relevant sound features. The first operates on unmodified recordings of the sound events. We can term this approach *correlational* because it often relies on tests of the statistical association between the sound features on the one hand and some behavioral variable on the other. The behavioral variables can be either the responses given by the experiment participant (e.g., probability of choosing the response "female" in gender categorization of walking sounds [89]) or the parameters of a mathematical model of the behavioral data (e.g., the dimensions of an MDS model of dissimilarity ratings [102], see section 5.4; the parameters of a spatial model of preferences, see section 5.8). One of the main issues associated with the correlational approach is that different sound properties can be statistically associated (e.g., natural impact sounds whose level decays more slowly also have a longer duration [39]). Because it is not possible to tell which of two strongly correlated features has the strongest influence on a target behavioral variable, a reasonable answer to this issue has been to recognize this ambiguity in the data-analysis process and to eliminate the need to choose arbitrarily among strongly correlated features by reducing them to one single variable [41].

The second approach adopted to assess perceptually relevant sound features is based either on the manipulation of sound recordings or on purely synthetic sounds. Notably, this approach somehow mitigates the features-correlation problem that characterizes studies of unmodified sound stimuli. The sound-manipulation approach was for example adopted by Li, Logan, and Pastore [89] to investigate the identification of the gender of a walker in sounds where the spectral mode (i.e., the most prominent spectral frequency) and the spectral slope were actively manipulated. Based on these stimuli, Li et al. [89] confirmed an influence of both acoustical factors, as previously revealed by the correlational analysis of the data collected for unmodified sounds. Grassi [46] investigated the extent to which various modifications of the sounds of a ball bouncing on a dish (e.g., removal of bounces) alter the ability to correctly estimate the size of the ball. Another interesting sound-manipulation paradigm has been adopted by Gygi et al. [55, 56] and Shafiro [128] to assess the relevance of temporal information for the identification of environmental sounds. In these studies the spectrum of the sound signals was progressively smeared while keeping the amplitude envelope untouched. After spectral smearing, sounds whose identification relies on the temporal variation of amplitude are comparatively better recognized than sounds whose identification relies on spectral factors. A final sound-manipulation strategy relies on the synthesis of novel sound stimuli. This approach was adopted by Caclin [11] to investigate the sound properties affecting timbre judgments. Importantly, the dimensions of acoustical variations of the synthetic stimuli corresponded to the acoustical attributes associated with the perception of unmodified musical stimuli in a previous correlational study [102]. The study by Caclin thus exemplifies the potential confirmatory function of sound-manipulation studies: whereas correlational studies can be used to generate hypotheses about what features influence the perception of complex sounds, sound-manipulation studies can be carried out to explicitly test these hypotheses.

A final interesting sound-manipulation approach is based on the method of *perturbation analysis* [149], frequently adopted by Lutfi to investigate the perception of the properties of sound sources [91–94]. In short, perturbation analysis relies on a trial-by-trial random perturbation of acoustical parameters of interest: if an acoustical parameter is relevant to the perceptual task, statistical analyses will reveal a significant association between the trial-specific responses and the trial-specific value of the perturbed acoustical parameters. In Lutfi and Oh [93] participants were presented with synthetic impact sounds and were asked to identify the material of the struck object. Sounds were generated by perturbing independently from trial to trial three parameters that characterize the spectral components of an impact sound: frequency, starting amplitude, and decay modulus, a measure of the temporal velocity of amplitude decay. A statistical analysis of the association between trial-specific responses and synthesis parameters made it possible to measure the extent to which each of the participants identified materials by focusing on frequency, amplitude, or decay modulus. It should be noted that whereas the majority of the studies mentioned in the previous paragraphs focus on global acoustical properties that characterize the entire sound signal (e.g., measures of the spectral distribution of energy), the molecular approach of Lutfi focuses on the attributes of the single spectral components that constitute a sound. The method of perturbation analysis has indeed never been applied to investigate the perceptual processing of global sound features and should represent an interesting methodological tool for future studies on this topic.

5.12 Motion Capture

Motion-capture methods are a natural choice for the analysis and evaluation of sonic interactions because they allow direct measurement of the motor interaction of the experiment participant with the sound-producing system. Being in general nonintrusive (no need of wires, and in some cases only very light sensors are placed on objects or on the body of the experiment participant), these methods allow an accurate measurement of the motor interaction of the user with the sound-producing system (e.g., “How do we use our body when interacting with a sonic artifact?” or “How do gestures and artifacts mutually influence a sonic interaction?”). Motion-capture techniques are often versatile and simple, and can also be used in real time to modify the structure of the sound signal reaching the user.

Motion-capture methods are frequently used in the study of music performance. Playing a musical instrument is indeed a complex form of sonic interaction in which the gestures of a user, the musician, ultimately trigger the generation of a musical sound and often influence the properties of ongoing sound events. Methods developed in this field can thus be easily generalized to the study of any type of sonic interaction. Within this research domain, sensors of different nature (e.g., accelerometers, cameras) have been used to collect high-temporal-resolution data on the movements of objects involved in the sound-generation process (e.g., hammer movements in a piano) and on the motor activity of the performer (e.g., motion of the arms and fingers of a pianist). In its most frequent application, motion capture data are collected for an off-line analysis. Goebl and Bresin [43] measured the effect of different touches and of striking speed on the acceleration of the keys and hammers in a piano. Dahl [22] used motion capture to uncover different percussion strategies in drumming performance. Schelleng [122] used these techniques to measure the parameters of a bowing gesture necessary to optimize the quality of violin sounds. Dahl and Friberg [23] used motion-capture methods to analyze the quality of movements interacting with sounding objects. Cameras of various kinds have been used to track user gestures in studies of music performance. For example, Schoonderwaldt

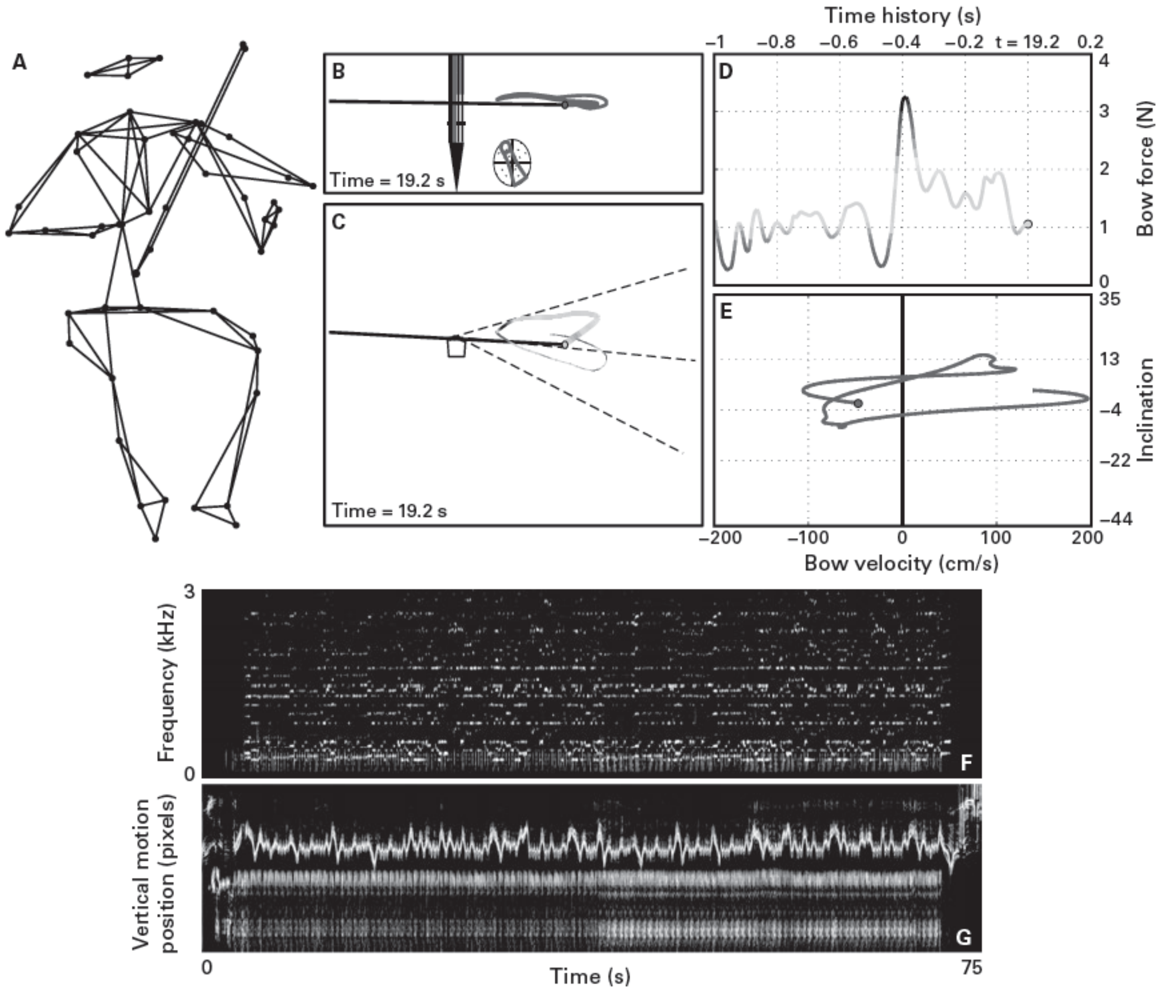

Figure 5.10

Motion capture analysis of violin performance (panels A–E, courtesy of Erwin Schoonderwaldt), and analysis of audio and video recordings (panels F–G, courtesy of Alexander Refsum Jensenius). An extended video of this violin performance can be seen at http://youtu.be/Y2zkrT76CvI. (A) Reflective markers (black dots) attached to the violinist's body and to the violin. (B) Top view of the motion of the bow (horizontal line) on the violin strings (vertical lines). The blue line displays the past bow trajectory. The bottom polar plot displays the rotation of the bow frog (keyhole shape), that is, the tilt of the bow. (C) Same as B, seen from the player perspective (bow trajectory = colored line). Each of the colored fanshaped areas correspond to one of the four violin strings, selected by changing the bow inclination. (D) Time-varying bow force. (E) Plot of time-varying bow inclination as a function of bow velocity, useful for assessing the coordination of string crossings and bow reversals. (F) Spectrogram of the sound signal resulting from the motions measured in panels B–E. (G) Motiongram of the violin performance analyzed in panels B–E. The vertical axis shows the vertical motion of the player. The grayscale measures the activity level (black = no activity; white = peak activity).

and Wanderly [124] used cameras to measure violin bowing gestures and to develop visual feedback methods for students in violin performance. Schoonderwaldt and Demoucron [123] designed a nonintrusive system combining optical motion capture with sensors. This system allows accurately measurement of all bowing parameters in bowed-string instrument performance. Video data have been captured by Dahl [23] to investigate the influence of visual information on the perception of emotional expression and gesture quality in music performance. Further methods for the study of body movement in music performance based on the real-time analysis of video data have been developed by Jensenius [69]. These methods are based on a collapsed visualization of vertical and horizontal motion of musicians (motiongram) as registered in video recordings and on the synchronization of the motiongram with the corresponding audio spectrogram (see figure 5.10 for further explanation).

Promising and largely unexploited applications of motion-capture data, particularly for the study of sonic interactions, are based on their on-line use. Indeed, motion-capture data can, for example, be used to control in real time the parameters of a sound-generating algorithm based on measures of motor activity (i.e., the spatial position, speed, and acceleration of a hand hitting and grabbing a sounding object). For example, in a study on percussion performance Giordano et al. [37] used measures of the velocity with which participants struck a virtual object to control the parameters of a real-time model for the generation of synthetic impact sounds. In another study DeWitt and Bresin [26] used the real-time gestures (speed and pressure) of a pen on a tablet to investigate the control of sound models mimicking the sound of a pen scribbling and its emotional content.

Acknowledgments

The authors wish to thank Richard E. Pastore, Carl Gaspar, and Philip McAleer for comments on earlier versions of this chapter and Erwin Schoonderwaldt and Alexander Refsum Jensenius for providing the materials displayed in figure 5.10.

Notes

1. The absolute threshold can be considered as a special case of the differential threshold. For example, the absolute threshold for sound level can be conceived as the lowest sound level that can be discriminated from the lowest possible level: silence.

2. In this section we focus on so-called direct scaling methods which measure scales of sensation directly, based on judgments of sensory magnitudes.. In contrast, indirect scaling methods derive scales of sensory magnitude from measures of the discrimination of the stimuli [35, p. 185].

Direct and indirect scaling methods are largely based on the work of S. S. Stevens and L. L. Thurstone, respectively [135,142].

3. In common practice the paired stimuli are always different (i.e., participants never rate the dissimilarity between one stimulus and itself) because the mathematical models most frequently used to analyze the dissimilarity ratings (e.g., multidimensional scaling models, see below) do not take into account same-stimulus dissimilarities. It should be nonetheless noted that self-dissimilarity data are useful for assessing whether participants correctly understood the use of the response scale, in which case same-stimulus dissimilarities should be on average lower than different-stimulus dissimilarities (see McAdams, Roussarie, Chaigne, & Giordano [100] for an example).

4. Although the most popular, MDS is one of several different distance models available for analysing dissimilarity data; see the appendix of Giordano et al. [38].

5. In a Euclidean space, the distance between two points A and B equals

$$\sqrt{\sum_{i=1}^{N}(A_i - B_i)^2} \,,$$

where N equals the number of dimensions of the space. The Euclidean distance is a special case of the Minkowsky distance:

$$\left(\sum_{i=1}^{N}|A_i - B_i|^p\right)^{1/p} ,$$

where p is the power of the Minkowski metric, and the Euclidean distance equals the Minkowski distance for $p = 2$. Not all MDS algorithms assume a Euclidean distance: measuring the power of the Minkowsky metric that best accounts for the dissimilarity ratings has indeed been a research topic in several previous studies [96, p. 149].

6. We can distinguish between metric and nonmetric MDS depending on whether model distances approximate a linear or monotonic transformation of the input dissimilarities, respectively.

7. See Calvert, Spence, and Stein [12] for an excellent overview of research on multisensory processes.

References

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

2. Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*(1), 124–150.

3. Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology. Human Perception and Performance*, *19*(2), 250–267.

4. Bijmolt, T. H. A., & Wedel, M. (1995). The effects of alternative methods of collecting similarity data for multidimensional scaling. *International Journal of Research in Marketing*, *12*(4), 363–371.

5. R. Bisping. (1997). Car interior sound quality: Experimental analysis by synthesis. *Acta Acustica united with Acustica*, *83*(5), 813–818.

6. Björk, E. A. (1985). The perceived quality of natural sounds. *Acustica*, *57*(3), 185–188.

7. Böckenholt, U. (2005). Scaling of preferential choice. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1790–1794). Chichester, UK: John Wiley & Sons.

8. Bonebright, T. L. (1996). An investigation of data collection methods for auditory stimuli: Paired comparisons versus a computer sorting task. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 275–278.

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

10. Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs. *Biometrika, 39,* 324–345.

11. Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, *118*(1), 471–482.

12. Calvert, G., Spence, C., & Stein, B. (2004). *Handbook of multisensory processes*. Cambridge, MA: MIT Press.

13. Carroll, J. D., & Green, P. E. (1995). Psychometric methods in marketing research: Part I, conjoint analysis. *JMR, Journal of Marketing Research*, *32*(4), 385–391.

14. Chang, J. J., & Carroll, J. D. (1968). *How to use MDPREF, a computer program for multidimensional analysis of preference data. Technical report*. Murray Hill, NJ: Bell Telephone Laboratories.

15. Chouard, N., & Hempel, T. (1999). A semantic differential design especially developed for the evaluation of interior car sounds. *Journal of the Acoustical Society of America*, *105*(2), 1280.

16. CLA2T research group, UPRES SYLED. (n.d.). Lexico3. URL: http://www.tal.univ-paris3.fr/lexico/index-gb.htm [retrieved June 21 2012]. Université de la Sorbonne Nouvelle, Paris 3.

17. Coombs, C. H. (1964). *A theory of data*. New York: John Wiley & Sons.

18. Cornsweet, T. N. (1962). The staircase-method in psychophysics. *American Journal of Psychology*, *75*(3), 485–491.

19. Cowie, R., Douglas-Cowie, E., Savvidou, S., Mcmahon, E., Sawey, M., & Schröder, M. (2000). “feeltrace”: An instrument for recording perceived emotion in real time. In *ISCA tutorial and research workshop (ITRW) on speech and emotion* (pp. 19–24). Newcastle, Northern Ireland.

20. Cox, T. F., & Cox, A. A. (1997). *Multidimensional scaling* (2nd ed.). New York: Springer-Verlag.

21. Coxon, A. P. M. (1999). *Sorting data: Collection and analysis. Sage university papers on quantitative applications in the social sciences*. Thousand Oaks, CA: Sage Publications.

22. Dahl, S. (2004). Playing the accent-comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica united with Acustica*, *90,* 762–776.

23. Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, *24*(5), 433–454.

24. de Leeuw, J. (2005). Multidimensional unfolding. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science*. Chichester, UK: John Wiley & Sons.

25. De Soete, G., & Carroll, J. D. (1992). Probabilistic multidimensional models of pairwuse choice data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 61–88). Hillsdale, NJ: Lawrence Erlbaum Associates.

26. DeWitt, A., & Bresin, R. (2007). *Sound design for affective interaction* (vol. 4738, pp. 523–533). Berlin/Heidelberg: Springer.

27. Ellermeier, W., & Daniel, P. (2002). Tonal components in tire sounds: Refined subjective and computational procedures. In G. Ebbitt & P. Davies (Eds.), *Proceedings of the sound quality symposium (SQS2002) at Inter-noise 2002*, Ames, IA: Institute of Noise Control Engineering (INCE-USA), Iowa State University.

28. 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*(1), 101–107.

29. Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(24), 429–433.

30. Fastl, H. (1989). Average loudness of road traffic noise. In *Proceedings of the international congress of noise control engineering* (Vol. II, pp. 815–820). Newport Beach, CA.

31. Fastl, H. (1991). Loudness versus level of aircraft noise. In A. Lawrence (Ed.), *Proceedings of Inter-noise '91* (Vol. 1, pp. 33–36). Sydney, Australia.

32. Faure, A. (2000). *Des sons aux mots, comment parle-t-on du timbre musical?* PhD thesis, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

33. Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *Journal of the Acoustical Society of America*, *87*(1), 311–322.

34. García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research*, *38*(12), 1861–1881.

35. Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

36. Gingras, B. T., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. (2011). Perceiving musical identity: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, *40*(10), 1206–1220.

37. Giordano, B. L., Avanzini, F., Wanderley, M., & McAdams, S. (2010). Multisensory integration in percussion performance. In *Actes du 10eme congrès francais d'acoustique, Lyon* [CD–ROM], Paris, France: Société Française d'Acoustique.

38 Giordano, B. L., Guastavino, C., Murphy, E., Ogg, M., Smith, B. K., & McAdams, S. (2010). Comparison of methods for collecting and modeling dissimilarity data: applications to complex sound stimuli. *Multivariate Behavioral Research*, *46*(5), 779–811.

39. Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America*, *119*(2), 1171–1181.

40. Giordano, B. L., McDonnell, J., & McAdams, S. (2010). Hearing living symbols and nonliving icons: Category-specificities in the cognitive processing of environmental sounds. *Brain and Cognition*, *73*, 7–19.

41. Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(2), 462–479.

42. Giordano, B. L., Visell, Y., Yao, H. Y., Hayward, V., Cooperstock, J., & McAdams, S. (2012). Identification of walked-upon materials in auditory, kinesthetic, haptic and audio-haptic conditions. *Journal of the Acoustical Society of America*, *131*(5), 4002–4012.

43. Goebl, W., & Bresin, R. (2003). Measurement and reproduction accuracy of computer-controlled grand pianos. *Journal of the Acoustical Society of America*, *114*, 2273.

44. Goodhill, G. J., Simmen, M. W., & Willshaw, D. J. (1995). An evaluation of the use of multidimensional scaling for understanding brain connectivity. *Philosophical Transactions of the Royal Society B. Biological Sciences*, *348*(1325), 265–280.

45. Gottschling, G. (1999). On the relations of instantaneous and overall loudness. *Acta Acustica united with Acustica*, *85*(3), 427–429.

46. Grassi, M. (2002). Recognising the size of objects from sounds with manipulated acoustical parameters. In J. A. Da Silva, E. H. Matsushima, & N. P. Ribeiro-Filho (Eds.), *Proceedings of the 18th annual meeting of the international society for psychophysics (Fechner day 2002)* (pp. 392–397). Rio de Janeiro, Brasil.

47. Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, *67*(2), 274–284.

48. Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons.

49. Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *JMR, Journal of Marketing Research*, *8*(3), 355–363.

50. Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, *5*(2), 103.

51. Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, *61*(5), 1270–1277.

52. Gros, L., & Chateau, N. (2001). Instantaneous and overall judgements for time-varying speech quality: Assessments and relationships. *Acta Acustica united with Acustica*, *87*(3), 367–377.

53. Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology*, *61*(1), 54–63.

54. Gustaffson, A., Herrmann, A., & Huber, F. (2007). Conjoint analysis as an instrument of marketing research practice. In A. Gustaffson, A. Herrmann, & F. Huber (Eds.), *Conjoint measurement: Methods and applications* (4th ed., pp. 3–30). Berlin: Springer-Verlag.

55. Gygi, B. (2001). *Factors in the identification of environmental sounds*. PhD thesis, Indiana University, Department of Psychology.

56. Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America*, *115*(3), 1252–1265.

57. Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics*, *69*(6), 839–855.

58. Gygi, B., & Shafiro, V. (2007). General functions and specific applications of environmental sound research. *Frontiers in Bioscience*, *12*, 3152–3166.

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

60. Hansen, M., & Kollmeier, B. (1999). Continuous assessment of time-varying speech quality. *Journal of the Acoustical Society of America*, *106*(5), 2888–2899.

61. Harbke, C. R. (2003). *Evaluation of data collection techniques for multidimensional scaling with large stimulus sets*. Master's thesis, Washington State University, Department of Psychology.

62. Hedberg, D., & Jansson, C. (1998). *Continuous rating of sound quality. Technical report*. Technical Audiology, Karolinska Institutet.

63. Hellbrück, J. (2000). Memory effects in loudness scaling of traffic noise: How overall loudness of short-term and long-term sounds depends on memory. *Acoustical Science and Technology*, *21*(6), 329–332.

64. Hellbrück, J., & Zeitler, A. (1999). Evaluating sequences of environmental noise using the method of absolute judgment in laboratory and outdoor situations. Some methodological considerations. *Journal of the Acoustical Society of America*, *105*, 1083.

65. Hempel, T., & Chouard, N. (1999). Evaluation of interior car sound with a new specific semantic differential design. *Journal of the Acoustical Society of America*, *105*(2), 1280.

66. Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. New York: John Wiley & Sons.

67. Ih, J. G., Lim, D. H., Shin, S. H., & Park, Y. (2003). Experimental design and assessment of product sound quality: Application to a vacuum cleaner. *Noise Control Engineering Journal*, *51*(4), 244–252.

68. Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*(8), 345–350.

69. Jensenius, A. R. (2007). *Action-sound: Developing methods and tools to study music-related body movement*. PhD thesis, Department of Musicology, University of Oslo.

70. Jeon, J. Y., You, J., & Chang, H. Y. (2007). Sound radiation and sound quality characteristics of refrigerator noise in real living environments. *Applied Acoustics*, *68*(10), 1118–1134.

71. Kendall, R. A., & Carterette, E. C. E. C. (1993). Verbal attributes of simultaneous wind instrument timbres: I. Von Bismarck's adjectives. *Music Perception*, *10*(4), 445–468.

72. Kidd, G. R., & Watson, C. S. (2003). The perceptual dimensionality of environmental sounds. *Noise Control Engineering Journal*, *51*(4), 216–231.

73. Kitagawa, M., & Windsor, B. (2008). *MoCap for artists: Workflow and techniques for motion capture*. Burlington, MA: Focal Press.

74. Kolesnik, P., & Wanderley, M. (2004). Recognition, analysis and performance with expressive conducting gestures. In *Proceedings of the international computer music conference (ICMC '04)*. Miami, FL: ICMA.

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

76. Krumhansl, C. L. (1998). Topic in music: An empirical study of memorability, openness, and emotion in Mozart's String Quintet in C Major and Beethoven's String Quartet in A Minor. *Music Perception*, *16*, 119–134.

77. Kuwano, S., Hayakawa, Y., & Namba, S. (1993). Psychological evaluation of noise in passenger cars using the method of continuous judgment by category: Analysis in different noise sources. In Chapelle, P., Vermier, G. (Eds.) *Proceedings of Inter-Noise 93* (Vol. II, pp. 919–922). Leuven, Belgium.

78. Kuwano, S., & Namba, S. (1978). On the loudness of road traffic noise of longer duration (20 min.) in relation to instantaneous judgment. *Journal of the Acoustical Society of America*, *64*, S127–S128.

79. Kuwano, S., & Namba, S. (1985). Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness. *Psychological Research*, *47*(1), 27–37.

80. Kuwano, S., & Namba, S. (1990). Temporal change of timbre of helicopter noise. *Noise Control Engineering Journal*, *2*, 1177–1180.

81. Kuwano, S., & Namba, S. (2001). Dimension of sound quality and their measurement. In *Proceedings of the 17th international congress on acoustics (ICA)*, Rome, Italy.

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

83. Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

84. Lawless, H. T., Sheng, N., & Knoops, S. S. C. P. (1995). Multidimensional scaling of sorting data applied to cheese perception. *Food Quality and Preference*, *6*(2), 91–98.

85. Lederman, S. J. (1979). Auditory texture perception. *Perception*, *8*, 93–103.

86. Lederman, S. J., & Klatzky, R. L. (2004). Multisensory texture perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (pp. 107–122). Cambridge, MA: MIT Press.

87. Lemaitre, G., Susini, P., Winsberg, S., McAdams, S., & Letinturier, B. (2007). The sound quality of car horns: A psychoacoustical study of timbre. *Acta Acustica united with Acustica*, *93*(3), 457–468.

88. Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*(2), 467–477.

89. Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America*, *90*(6), 3036–3049.

90. Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: John Wiley & Sons.

91. Lutfi, R. A. (2001). Auditory detection of hollowness. *Journal of the Acoustical Society of America*, *110*(2), 1010–1019.

92. Lutfi, R. A., & Liu, C. J. (2007). Individual differences in source identification from synthesized impact sounds. *Journal of the Acoustical Society of America*, *122*, 1017–1028.

93. Lutfi, R. A., & Oh, E. L. (1997). Auditory discrimination of material changes in a struck-clamped bar. *Journal of the Acoustical Society of America*, *102*(6), 3647–3656.

94. Lutfi, R. A., & Wang, W. (1999). Correlational analysis of acoustic cues for the discrimination of auditory motion. *Journal of the Acoustical Society of America*, *106*, 919.

95. Macmillan, N. S., & Creelman, C. D. (1991). *Detection theory: A user's guide* (2nd ed.). Cambridge, MA: Cambridge University Press.

96. Maddox, W. T. (1992). Perceptual and decisional separability. In G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Lawrence Erlbaum Associates.

97. Madsen, C. K. (1996). Empirical investigation of the "aesthetic response" to music: Musicians and nonmusicians. In *Proceedings of the 4th international conference for music perception and cognition* (pp. 103–110). Montréal, Canada: Society for Music Perception and Cognition.

98. Marcell, M. E., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology*, *22*(6), 830–864.

99. Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (pp. 85–105). Cambridge, MA: MIT Press.

100. McAdams, S., Roussarie, V., Chaigne, A., & Giordano, B. L. (2010). The psychomechanics of simulated sound sources: Material properties of impacted thin plates. *Journal of the Acoustical Society of America*, *128*(3), 1401–1413.

101. McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception*, *22*(2), 297–350.

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

103. McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

104. McNicol, D. (2004). *A primer of signal detection theory*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

106. Namba, S., Kato, T., & Kuwano, S. (1995). Long-term evaluation of the loudness of train noise in laboratory situation. In M. Newman (Ed.), *Proceedings of the international congress on acoustics* (pp. 215–218). Trondheim, Norway.

107. Namba, S., Kuwano, S., Hatoh, T., & Kato, M. (1991). Assessment of musical performance by using the method of continuous judgement by selected description. *Music Perception*, *8*(3), 251–276.

108. Nosulenko, V. N., & Samoylenko, E. S. (1997). Approche systémique de l'analyse des verbalisations dans le cadre de l'étude des processus perceptifs et cognitifs. *Information sur les Sciences Sociales*, 36, 223–261.

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

110. Osgood, C. E., Suci, G. H., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

111. Parizet, E., Hamzaoui, N., Segaud, L., & Koch, J. R. (2003). Continuous evaluation of noise uncomfort in a bus. *Acta Acustica united with Acustica*, *89*(5), 900–907.

112. Pastore, R. E., Flint, J. D., Gaston, J. R., & Solomon, M. J. (2008). Auditory event perception: The source-perception loop for posture in human gait. *Perception & Psychophysics*, *70*(1), 13–29.

113. Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting acoustic descriptors from musical signals. *Journal of the Acoustical Society of America*, *130*(5), 2902–2916.

114. Peters, R. W. (1960). *Research on psychological parameters of sound*. Wadd report 60–249, ordered number ad-240814, Wright Air Development Division, U.S. Department of Commerce.

115. Poitrenaud, S. (1995). The Procope semantic network: An alternative to action grammars. *International Journal of Human-Computer Studies*, *42*(1), 31–69.

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

117. Rao, V. R., & Katz, R. (1971). Alternative multidimensional scaling methods for large stimulus sets. *JMR, Journal of Marketing Research*, *8*(4), 488–494.

118. Repp, B. H. (1987). The sound of two hands clapping: In exploratory study. *Journal of the Acoustical Society of America*, *81*(4), 1100–1109.

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

120. Rosch, E. (1977). *Human categorization* (Vol. 1, pp. 1–49). London: Academic Press.

121. Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, *10*(4), 489–502.

122. Schelleng, J. C. (1973). The bowed string and the player. *Journal of the Acoustical Society of America*, *53*, 26.

123. Schoonderwaldt, E., & Demoucron, M. (2009). Extraction of bowing parameters from violin performance combining motion capture and sensors. *Journal of the Acoustical Society of America*, *126*(5), 2695–2708.

124. Schoonderwaldt, E., & Wanderley, M. M. (2007). Visualization of bowing gestures for feedback: The Hodgson plot. In *Proceedings of the 3rd international conference on automated production of cross media content for multi-channel distribution (AXMEDIS07)* (Vol. 2, pp. 65–70). Barcelona, Spain.

125. Schubert, E. (1996). Measuring temporal emotional response to music using the two dimensional emotion space. In *Proceedings of the 4th international conference for music perception and cognition* (pp. 263–268). Montréal, Canada.

126. Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 393–414). New York: Oxford University Press.

127. Schubert, E. (2010). Continuous self-report methods. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 223–253). New York: Oxford University Press.

128. Shafiro, V. (2004). *Perceiving the sources of environmental sounds with a varying number of spectral channels*. PhD thesis, City University of New York.

129. Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788.

130. Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulations of visual perception by sound. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (pp. 27–33). Cambridge, MA: MIT Press.

131. Siekierski, E., Derquenne, C., & Martin, N. (2001). Sensory evaluation of air-conditioning noise: Sensory profiles and hedonic tests. In *Proceedings of the 17th international congress on acoustics (ICA)* (p. 326). Rome, Italy.

132. 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*(1), 87–120.

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

134. Stepánek, J. (2006). Musical sound timbre: Verbal description and dimensions. In *Proceedings of the international conference on digital audio effects (DAFx-06)* (pp. 121–126). Montréal, Canada.

135. Stevens, S. S. (1956). The direct estimation of sensory magnitude—loudness. *American Journal of Psychology*, *69*, 1–25.

136. Stevens, S. S., & Guirao, M. (1962). Loudness, reciprocality, and partition scales. *Journal of the Acoustical Society of America*, *34*, 1466.

137. Susini, P., McAdams, S., & Smith, B. K. (2002). Global and continuous loudness estimation of time-varying levels. *Acta Acustica united with Acustica*, *88*(4), 536–548.

138. P. Susini, S. McAdams, and B. K. Smith. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica united with Acustica*, *93*(4), 623–631.

139. Susini, P., McAdams, S., & Winsberg, S. (1999). A multidimensional technique for sound quality assessment. *Acta Acustica united with Acustica*, *85*(5), 650–656.

140. 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.

141. Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*(1), 7–67.

142. Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286.

143. Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*(17), 2503–2522.

144. Upham, F., & McAdams, S. (2010). Approaching an audience model of listening experience. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.), *Proceedings of the 11th international conference on music perception and cognition (ICMPC11)* (pp. 489–490). Seattle, WA.

145. Vanderveer, N. J. (1979). *Ecological acoustics: Human perception of environmental sounds*. PhD thesis, Cornell University. *Dissertation Abstracts International, 40,* 4543B. (University Microfilms No. 80–04–002).

146. von Békésy, G. (1947). A new audiometer. *Acta Oto-Laryngologica*, *35*(5–6), 411–422.

147. Von Bismarck, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, *30*(3), 146–159.

148. Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(5), 704–712.

149. Watson, C. S. (1992). *Signal detection and certain physical characteristics of the stimulus during the observation interval*. PhD thesis, Indiana University, Bloomington, IN.

150. Weber, R. (1991). The continuous loudness judgement of temporally variable sounds with an "analog" category procedure. In A. Schick, J. Hellbrück, & R. Weber, (Eds.). *Fifth Oldenburg symposium on psychological acoustics* (pp. 267–294). Oldenburg, Germany: BIS.

151. Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology*, *13*(3), 228–240.

152. Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313.

153. Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–1329.

154. Zampini, M., & Spence, C. (2004). The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of Sensory Studies*, *19*(5), 347–363.

155. Zeitler, A., & Hellbrück, J. (2001). Semantic attributes of environmental sounds and their correlations with psychoacoustic magnitudes. In *Proceedings of the 17th international congress on acoustics* (Vol. 28), Rome, Italy.