UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI PSICOLOGIA GENERALE

INSTITUT DE RECHERCHE ET DE COORDINATION ACOUSTIQUE/MUSIQUE (STMS-IRCAM-CNRS)

DOTTORATO DI RICERCA IN PERCEZIONE E PSICOFISICA IN SUPERVISIONE CONGIUNTA

Ciclo XVII

SOUND SOURCE PERCEPTION IN IMPACT SOUNDS

Coordinatore: Prof. Clara Casco

Supervisore: Prof. Paola Bressan

Co-supervisore: Prof. Stephen McAdams

Dottorando: Bruno Lucio Giordano

DATA CONSEGNA TESI:
30 Giugno 2005

Sommario

La percezione permette agli animali di ottenere informazioni sull'ambiente a partire dalle informazioni in arrivo ai sistemi sensoriali. Nel caso dell'ascolto, la porzione dell'ambiente d'interesse è la sorgente sonora, ad esempio un solido in vibrazione, e l'informazione in arrivo è contenuta nel suono. Quali sono le capacità percettive degli esseri umani in condizioni quotidiane? Quali sono le proprietà della sorgente sonora rilevanti per l'ascoltatore? Quale è la natura dell'informazione acustica utile alla percezione della sorgente sonora? Rispondere a questi quesiti è stato l'obiettivo principale di questa tesi, che si è focalizzata su uno dei suoni più frequentemente incontrati durante la vita quotidiana: il suono d'impatto.

La tesi inizia con la presentazione di due degli approcci teorici rilevanti per la comprensione della percezione della sorgente sonora: l'approccio dell'elaborazione dell'informazione e l'approccio ecologico. Una rassegna esaustiva della letteratura sulla percezione della sorgente viene utilizzata come punto di partenza per la discussione delle assunzioni dell'approccio ecologico. Viene quindi presentata una rassegna della letteratura sulla percezione del timbro, cioè sulla percezione dei toni degli strumenti musicali, accompagnata dalla rianalisi di dati precedentemente pubblicati. I risultati di questa rianalisi evidenziano una forte rilevanza delle proprietà fisiche degli strumenti musicali nello spiegare la percezione del timbro, collegando così la percezione di materiali sonori musicali e non.

Sono state condotte tre indagini empiriche. In ciascuna di esse i suoni di impatto sono stati generati dall'interazione tra due oggetti reali o simulati: un oggetto altamente smorzato - il martello - e un oggetto vibrante - l'oggetto sonoro. In ciascuno di questi studi il dato comportamentale è stato spiegato sia in base alle proprietà della sorgente che in base alle proprietà acustiche. Ancora, la relazione tra proprietà della sorgente e struttura acustica è stata indagata, evidenziando quei parametri acustici che specificano univocamente le proprietà della sorgente. In questo modo è stata verificata la presenza di sufficiente informazione acustica per il perfetto allineamento tra percezione e proprietà della sorgente sonora.

Il primo studio ha investigato l'identificazione del materiale dell'oggetto sonoro. Sono state osservate buone capacità di identificazione quando erano implicate discriminazioni tra materiali di proprietà meccaniche altamente differenti, ad esempio metallo e plastica. D'altra parte, la discriminazione tra materiali caratterizzati da proprietà meccaniche simili, come nel caso del legno e della plastica, è stata trovata altamente deficitaria, essendo basata sulla grandezza dell'oggetto sonoro piuttosto che sul suo materiale. Diverse ipotesi

sono state avanzate per spiegare il risultato sperimentale sulla base delle proprietà delle sorgenti dei suoni d'impatto che popolano l'ambiente acustico quotidiano.

Il secondo studio ha indagato la rilevanza delle proprietà dell'interazione tra martello e oggetto sonoro nel determinare le proprietà percepite del martello, e ha testato la capacità di percepire indipendentemente martelli e oggetti sonori. Agli ascoltatori è stato chiesto di stimare la durezza del martello o dell'oggetto sonoro. È stato trovato limitato supporto empirico a favore dell'indipendenza percettiva dei due oggetti, sebbene i partecipanti dessero un peso maggiore alle proprietà dell'interazione nello stimare la durezza del martello, e un peso maggiore alle proprietà dell'oggetto sonoro nello stimare la durezza dell'oggetto sonoro.

Nel terzo studio sono state comparate la rilevanza percettiva del martello, dell'oggetto sonoro e delle proprietà della loro interazione. A questo scopo è stato utilizzato un compito sperimentale che non richiedesse l'uso di etichette linguistiche relative alla sorgente sonora: la stima della dissomiglianza. A questo modo sono stati eliminati degli eventuali effetti di disturbo che avrebbero limitato la validità ecologica dei risultati sperimentali. Si è trovato che il giudizio sperimentale si basa sulle proprietà dell'oggetto sonoro e, secondariamente, sulle proprietà dell'interazione. Non sono state trovate prove a favore della rilevanza percettiva del martello.

La comparazione dei risultati ottenuti nei diversi studi ha permesso di ottenere informazioni circa le proprietà della sorgente e del suono rilevanti per la percezione quotidiana dei suoni d'impatto e delle loro sorgenti. Si può concludere che la percezione quotidiana è basata sulla grandezza e sul materiale dell'oggetto sonoro e, secondariamente, sulle proprietà dell'interazione tra martello e oggetto sonoro. Parallelamente dal punto di vista acustico la percezione quotidiana si fonda sulla frequenza e durata del segnale e, in maniera secondaria, sul centro di gravità spettrale o centroide spettrale della porzione d'attacco.

Infine, in tutti gli studi sono stati trovati diversi parametri acustici che, limitatamente al contesto sperimentale, specificano univocamente differenti proprietà della sorgente sonora. È stata perciò trovata sufficiente informazione acustica per il perfetto allineamento tra proprietà fisiche e percepite della sorgente. Cionнамostante i soggetti sperimentali hanno mostrato un disallineamento tra proprietà fisiche e percepite della sorgente. Questo risultato è stato quindi contrastato con l'assunzione dell'approccio ecologico in base alla quale la percezione è basata sulla detezione di proprietà invarianti dell'informazione sensoriale, ovvero sulla detezione di quei tratti della struttura acustica che specificano univocamente le proprietà della sorgente sonora. Si può concludere che, ai fini della spiegazione della sorgente nei suoni di'impatto, il concetto di invariante è inutile.

Summary

Perception allows animals to gather information concerning the environment starting from information input to the sensory systems where, in the case of audition, the relevant portion of the environment is the sound source (e.g., a vibrating solid) and the incoming information is contained in the sound wave. What are the perceptual capabilities of humans under everyday conditions? Which properties of a sound source are relevant to the perceiver? What is the nature of the acoustical information useful for sound source source perception? Answering these questions has been the main goal of this thesis, which focused on one of the most frequently encountered types of sound during everyday life: impact sounds.

The thesis begins with a presentation of two of the theoretical approaches relevant to the understanding of sound source perception: the ecological and information-processing approaches. A comprehensive review of the literature on source perception is then taken as a starting point for the discussion of the assumptions of the ecological approach. A review of the literature on timbre perception, i.e., on the perception of musical instrument tones, along with a reanalysis of previously published data, is also presented. Results of the reanalysis point out the strong relevance of the physical properties of the musical instruments in explaining timbre perception, thus linking perception of nonmusical and musical acoustical materials.

Three empirical investigations were conducted. With each of them, impact sounds were generated by the interaction between two real or simulated objects: a highly damped object - the hammer - and a vibrating object - the sounding object. Across studies behavioral data were explained both in terms of source and acoustical properties. Also, the relationship between source properties and acoustical structure was studied, pointing out those acoustical parameters that uniquely specified the source properties. In this way the presence of sufficient acoustical information for unbiased perception of the source properties of interest was ascertained.

The first study investigated identification of the material of the sounding object. Good identification capabilities were observed when discrimination among materials of vastly different mechanical properties were involved (e.g., metal and plastic). On the other hand, discrimination between materials of similar mechanical properties (e.g., wood and plastic) was highly impaired, being based on the size of the sounding object rather than on its material. Several hypotheses were formulated, which explained the observed response biases with the properties of the impact sound sources that populate our everyday acoustical environment.

The second study investigated the relevance of the properties of the interaction between hammer and sounding object in determining perceived hammer properties, and ascertained the ability to perceive hammers independently from the sounding objects. Listeners were asked to rate the hardness of either the hammer or the sounding object. Limited support for perceptual independence between the two was pointed out, although participants weighted interaction properties more heavily when rating hammer hardness, and weighted sounding object properties more heavily when rating sounding object hardness.

The third study investigated the relative perceptual relevance of hammer, sounding object and interaction properties. A task that did not require the use of source-related linguistic labels was used: dissimilarity rating. Thus possible biases in the experimental procedure limiting the ecological validity of results were eliminated. Participants based their judgments on the properties of the sounding object, and, to a limited extent, on the properties of the interaction. No evidence to support the perceptual relevance of the hammer was found.

Comparison of results across studies allowed us to gather knowledge on the source and acoustical properties relevant to everyday perception of impact sounds and sources. Thus, everyday perception was found to be most likely based on the size and material of the sounding object and, to a secondary extent, on the properties of the hammer/sounding object interaction. From the acoustical point of view, everyday perception was found to be based on signal frequency and duration and, to a secondary extent, on the attack spectral center of gravity or spectral centroid.

Finally, in all studies, acoustical parameters were found, which uniquely specified different properties of the sound source within the experimental context. Thus acoustical information for unbiased perception of the source properties was found. Nonetheless participants revealed perception to be misaligned with respect to the actual physical properties. This result was contrasted with the assumption of the ecological approach according to which perception is based on the detection of invariant properties of the sensory information, i.e. on the detection of those traits of the acoustical structure that uniquely specify the properties of the sound source. As a result it was concluded that the concept of structural invariant is not useful to the understanding of source perception in impact sounds.

To William in Melk

Acknowledgments

This thesis and especially the writing of the acknowledgments represent the end of quite a long journey, spent roaming around different countries. Such a journey would hardly have been possible without the support and encouragement of my family: Sergio, Valentina and Thea.

In Italy, I would like to thank especially Giovanni Bruno Vicario for letting me absorb through countless conversations his evergreen astonishment for the facts of perception; Davide Rocchesso for involving me in the Sounding Object project, which represented the starting point of this thesis, and for many other suggestions; Giulio Vidotto and Massimiliano Pastore for their help with data analysis and encouragement. Many thanks also to Clara Casco for the initiative to make this Ph.D. cosupervised, and to the Laboratorio di Acustica Musicale e Architettonica of the Fondazione Scuola di San Giorgio (Davide Bonsi, Diego Gonzales, Domenico Stanzial) for letting me use their facilities to record the stimuli used in Chapter 6. I would also like to thank Massimo Grassi for many occasions for exchange of ideas on source perception, in the lab as well as in front of a spritz, and for his risotti; Matthias Rath for the help with the use of the sound synthesis model used in Chapter 7 and for answering hundreds of questions on signal analysis and vibration without showing any violent reaction. Federico Avanzini, Matthias Rath and Laura Fontana for staying for hours with me inside a sound recording booth without facing psychotic disorders, while I was striking more than 25 kg of plates. Again Matthias Rath and Cosmo Trestino for all the time spent with me in Stockholm, getting acquainted with the traditions of Sweden. Finally, I would like to thank Mauro Borgato, Stefano Migliore, Simone Nastasi and Mario Barrile for the night-life in Padova, for an absolutely entertaining though sometimes autistic Unreal winter, and for all the constructive discussions on the meaning of life; Andrea Turetta for the climbing training.

In Sweden, I would first like to thank Roberto Bresin for giving me the opportunity to continue my research at the TMH, for his generous help with the preparation of the setups used in Chapter 8, and for his Cicerone function in Stockholm. Many thanks also to Anders Askenfelt for the suggestions concerning the tests used in Chapter 8, and for reviewing part of Chapter 3; Markku Haapakorpi for his patience and help with the preparation of the sound generation device used in Chapter 8 and for tracing anything I needed inside the Department; Erik Jansson for useful suggestions on wood and Chladni patterns; Giampiero Salvi for the discussions on cluster analysis, which replaced any other topic of conversation for several months, and for the many pool games; Mikael Bohman and Svante Granqvist for the advice on sound recording; Kjetil Hansen for sharing my opinions on music and multimedia in general and for the CinemaTMHeket; Sofia Dahl for Aoria, and for luciafirande; Erwin Schoonderwaldt for his subtle sense of humor and for the discussions on musical acoustics. I would also like to thank all the Music Acoustics group of the TMH and Sten Ternström for generously hosting me in these last months while writing the thesis.

In France, I would like to thank above all Stephen McAdams for his support, suggestions and hints, and for welcoming me in his house putting a copy of a Ph.D. thesis in one of my hands, and a glass of Cognac in the other. I would also like to thank Alain de Cheveigné for his generous help with the signal analysis tools used throughout this thesis; Antoine Chaigne for the precious advice and comments on the tests applied in Chapter 8; Bennet Smith and Daniel Pressnittzer for all the chats with me, and for their absolutely entertaining way of chatting with each other; Suzanne Winsberg for suggestions on data analysis, and for letting me use the MDS models presented in Chapter 8; Anne Caclin and Olivier Houix for hints, tricks, tex and matlab files and for Fontainebleau.

Finally, I would like to thank Christine Englund for all the dinners, piroger and semlor, movies, walks, chats, discussions, laughter, help and support throughout the last two years, and for showing me the soul of her country.

Contents

Chapter 1

Introduction

Our everyday environment is populated with widely diverse species of sounds (cracks, drippings, etc.), originating from equally diversified sources: breaking objects, vibrating liquids, etc. When we encounter these sounds, we are likely to become aware of the nature of the physical events that generated them, thus demonstrating one of the most important tasks of our perceptual system: the use of the incoming information for the extraction of properties of the environment.

We rarely make the effort to realize consciously, just from hearing, that what is approaching is an automobile rather than a motorbike, or that a wood log is being sawed rather than struck. Nonetheless, solving this problem represents a challenge to researchers in audition. Many issues are still open, from the delineation of auditory capabilities in recognition of the properties of the sound source, to that of the properties of the sounds used by humans to these purposes, and arriving at a theoretical framework capable of giving a good account of the empirical data. These problems are part of what constitutes the field of sound source perception and have been the focus of this thesis.

Chapters 2-4 present the conceptual tools needed to carry research in this field. Chapter 2 presents and contrasts two theories of perception relevant to the understanding of sound source perception. Chapter 3 outlines the necessary methodological tools. Chapter 4 reviews the published literature on sound source perception, outlining possible future directions for research and testing part of the theoretical assumptions presented in Chapter 2 with available empirical data. Within this same chapter the literature on timbre perception is also reviewed, and previously published data are reanalyzed in order to point out the common traits between source and timbre perception.

Chapters 6-8 present three experimental studies. These are written in the form of stand-alone papers and can be read independently of the rest of the thesis. All studies investigated one of the simplest and yet most ubiquitous sound sources in our acoustical environment, the impacted sound source (e.g., fingers tapping on a keyboard). The main goal has been to outline both the source and acoustical properties relevant to the perception of impact sounds and sources.

Particular care has been taken with respect to the ecological validity of the experimental results. Thus, through methodological choices and the comparison of results gathered in the different experimental studies, knowledge useful to the understanding of the perception of impact sounds and sources in everyday conditions have been generated. Finally, on the basis of experimental results, a few conclusions are drawn concerning the future directions for the theoretical development of this field.

Chapter 2

The theoretical framework

Perception is the act of gathering information about the environment. The survival of a species is a good proof of the efficiency of such an act. How is this possible? Different theories have different assumptions, and explanations, concerning the nature of perception.

To better point out these differences we shall briefly outline the path that links the environment to perception. In the case of audition the portion of the environment with auditory relevance is the *sound source*, most of the times a vibrating object. In normal conditions objects are not located in empty space, but are surrounded by air. The vibrations of an object are then transmitted to the surrounding air particles and spread through the medium, where they travel under the form of a wave, eventually reaching our ears. We refer to the sound source as a *distal stimulus*, and to the acoustical wave that reaches our ears as the *proximal stimulus*. Changes in the source determine changes in the structure of the sound, i.e., a variation in the distal stimulus causes a variation in the proximal stimulus (see Section 3.2). Different theories of perception make different assumptions concerning the nature of the relationship between properties of the distal and proximal stimuli.

Once the sound reaches our ears and causes the tympanum to vibrate, a complex chain of reactions in the nervous system is triggered, which results in the emergence of the sound as we experience it perceptually. A second source of disagreement among theoretical approaches is found in the assumptions concerning the nature of the processes that link the properties of the proximal stimulus with the perceived properties of the environment.

This chapter outlines the main issues of two theoretical approaches to perception: the ecological approach and the information processing approach. Theoretical contributions focusing explicitly on sound source perception are highlighted. Several examples are drawn from studies on visual perception.

2.1 The ecological approach to perception

The ecological approach to perception is the fruit of the work of J. J. Gibson (Gibson, 1966, 1979), who dedicated his efforts to develop a theory of how the organism comes to know the world. The main assumptions of this approach are:

1. the proximal stimulus specifies uniquely the distal stimulus;

2. perception of the properties of the distal stimulus is direct and does not require additional computations carried out within the perceptual system;

3. perception is veridical.

An excellent overview of the ecological approach to perception is given by Michaels and Carello (1981). This section draws much from their work.

2.1.1 Invariants

When an object is observed from a variable distance, several properties of the stimulation arriving at the retina change, such as the size of the retinal projection. Despite these variations, the perceived size of the object remains constant. When the voice of a familiar person is heard over the phone or in person, the structure of the acoustical signal changes drastically. Despite these changes, identification of the talker remains unchanged. In these examples perception of properties of the environment (the size, the talker) remains constant despite variations in the structure of the proximal stimulus (the retinal projection, the acoustical signal) caused by variations in the distal stimulus (the distance of the object, the system that transmits the acoustical signal). These phenomena are referred to as *perceptual constancies* and are of central relevance for the development of theories of perception.

The ecological approach explains perceptual constancies on the basis of the properties of the stimulation. The fact that constant perceptions are possible despite variations in the proximal stimulus means that part of the stimulus structure remains unchanged. In other words, the proximal stimulus is characterized by *invariant* properties, and constant perceptions are based on the detection of such stable structures. Invariants are defined as higher-order variables of stimulation that specify the properties of the environment, often through a pattern of variability over time. The link between invariants and temporal variability of stimulation is well exemplified by the work of Johansson (1973) on biological motion. In this research several actors were asked to walk in a dark room while wearing light emitting/reflecting spots, attached to the joints of their body (e.g., knees) in a dark room. Walking humans were unanimously recognized in the resulting dynamic point light displays. With static displays, however, recognizing humans was impossible: information for perception was thus found in the pattern of temporal variability.

It is possible to distinguish between two types of invariants: *structural invariants* specify the properties of the objects in our environment; *transformational invariants* specify changes in the properties of the objects. Structural invariants allow us to recognize that an object is a chair, for example, despite variations in its size, shape, in its distance from the observer, or in the viewpoint, or to recognize the same melody even though it is played with different musical instruments, at different intensities, or in different rooms. In this latter case the invariant structure is found in the ratios among the fundamental frequencies of the tones that compose the melody (see Section 3.2.2 for a definition of fundamental frequency). Transformational invariants allow us, for example, to recognize that a sound source is approaching us or moving away from us, independently of whether it's the whistle of a train or the horn of a car. The invariant is found in the frequency and intensity variations of the signal reaching the listener where, independently of the particular source, both increase when the source is approaching and decrease as the source passes and moves away[1]. Another example of a transformational invariant in audition is found in the structure of the acoustical signal that allows us to distinguish between a bouncing and a breaking object (Warren & Verbrugge, 1984). A bouncing object generates a series of impact sounds progressively closer in time, while with breaking, after an initial burst of noise which is due to the rupture of the object, the signal is given by a superimposition of many independent bouncing sound sequences, originating from each of the pieces of the broken object (see section 4.3 for a more detailed presentation of this study).

Using invariants to explain perceptual constancies is equivalent to assuming that at some level of analysis the properties of the environment are uniquely specified by the structure of the proximal stimuli: I recognize that Luciano is talking over the phone, or in the next room, therefore some property of the acoustical signal specifies that the talker is Luciano across these circumstances, i.e. uniquely. Unique specification of the distal stimulus in the structure of the proximal stimulus is a key assumption of ecological theory. An important consequence of this is that the most sensible approach for the specification of the relevant information for perception, and of invariants, is based on the analysis of the relationship between the properties of the distal stimulus and those of the proximal stimulus. Such an analysis takes the name of *ecological optics*, if we are concerned with vision, or *ecological acoustics* if we are concerned with audition.

2.1.2 The animal and the environment

Traditionally psychology has treated the animal and the environment as two distinct units, and research has focused on the study of the former. The ecological approach rejects this dualism and claims that the study of perception should consider jointly the animal and the environment. This choice has several relevant consequences.

First, research within the ecological framework focuses on *everyday perception*, investigating stimuli and conditions that characterize the interaction of the animal within its usual environment. As such, auditory perception should be studied using, for example,

[1] The change of the frequency of the signal emitted from a moving source, reaching a stationary listener is known as the Doppler effect and is due to the fact that the number of wavefronts reaching the listener increases as the source approaches and decreases as it moves away.

slamming doors or jingling key sounds, rather than white noises or sinusoids (Gaver, 1988, 1993a). Also, experimental settings where stimulation is impoverished with respect to everyday conditions should be avoided (e.g., tachistoscopic presentation of visual stimuli, too brief to be informative of everyday vision), as providing no information about the nature of everyday perception.

Second, the study of the information for perception has to consider the properties of the sensory systems of the animal. Different animals are indeed sensitive, or *attuned*, to different properties of the stimulation. Bees, for example, are sensitive to the polarization of light and detect ultraviolet wavelengths; bats have developed a highly sophisticated system for the analysis of echoes, which serves perception of the properties of the surfaces that reflect the acoustical waves. In the case of research on human audition, this would imply that the characterization of the investigated stimuli should consider at least the most basic properties of the peripheral auditory system.

Third, research must consider that information is not passively imposed on the animal. The animal, in fact, *explores* the environment, actively determining the stimulation it is exposed to. It should be noted, however, that at least in the case of everyday audition, many of the acoustical signals that reach our ears are not generated by our actions or explorations but are generated by external events and are thus passively imposed on us.

A final consequence of the rejection of the animal-environment dualism is found in the concept of *affordance*. Affordances are the "acts of behavior permitted by objects, places, and events" (Michaels & Carello, 1981, p. 42). Affordances, for example, are those properties of the stimulation that specify that we can sit on a surface, grasp an object, or eat a fruit. According to the ecological approach, it is affordances that are perceived by the animal. The notion of affordance reveals the centrality of action to the ecological approach, which assumes action itself to be the most appropriate test for perception. Accordingly, Carello, Anderson, and Kunkler-Peck (1998) studied auditory perception of the length of a rod dropped on the floor, asking participants to match the perceived length of the rod to their distance from a visible surface, as if they could reach this with a rod of the estimated length (see Section 4.3 for a more detailed presentation of this study). It should be pointed out, however, that it is not always possible to highlight proper tests of perception based on action: which action should be used to test whether one perceives that an object is made of plastic or wood, for example?

2.1.3 Direct perception

The ecological approach explains perceptual constancies with the properties of the proximal stimulus. Alternative explanations might however be formulated based, for instance, on previous experience. For example, one might hypothesize that visual size constancy (i.e., perception of constant size despite variation of the distance of the object from the observer) is achieved on the basis of previous experiences of the same object. This hypothesis was tested by Epstein (1965). In this study observers were asked to judge the size of three pictures of differently sized coins (a dime, a quarter, half a dollar), pictures being enlarged/reduced to the same size, and being located at a constant distance from the observer. Configurations were presented in a dark room illuminated with a spotlight, and were observed monocularly through an aperture. Results showed perceived size to be scaled to the actual size of the coins, demonstrating size perception to be influenced by previous experience with coins. These results, however, would not be accepted by those embracing the ecological approach, as the experimental methodology adopted provided the observer with impoverished information with respect to that available in everyday conditions (for instance monocular observation through a fixed aperture eliminates the differences among the retinal images at the two eyes, known as retinal disparity).

According to the ecological approach, the only information upon which perception is based is contained in the proximal stimulus, and no information coming from previous experiences or from internal computations is added. As a consequence the concept of memory is discarded as not useful to the understanding of perception. Also discarded is the concept of unconscious inference, according to which internal processes reconstruct the most probable distal stimulus from incomplete sensory evidence (see Section 2.2.2). These assumptions define perception as *direct*, a key concept in the ecological theory.

As a consequence of the rejection of internal computations, the ecological approach states that the animal simply detects, or picks up, the relevant information contained in the incoming stimulation. The concept of information pickup is exemplified by the metaphor of *resonance*. Acoustic resonance was, for example, used by H. L. F. von Helmholtz (1821-1894) to analyze musical sounds. Helmholtz resonators were hollow spheres with two opposite openings. The air inside the resonator had a natural frequency of vibration, determined by the geometry of the resonator (for example the larger the volume of the spherical cavity, the lower the natural frequency). The air in the resonator vibrates in response to the sound entering the opening, and its response is maximal when the incoming sound has significant energy at the natural frequency. Thus the Helmholtz resonator operates a sort of spectral analysis without needing complex intermediate processing, responding selectively to frequencies thanks only to its structure. According to the ecological approach perceptual systems pick up information in a manner similar to Helmholtz resonators, respond selectively to particular aspects of the incoming stimulation.

2.1.4 Veridicality of perception

Does a physical reality exist? What is its relationship with perception? The ecologist adopts the position of *realism*, which posits the existence of a physical reality whose properties are independent of perception. If the physical reality is independent of perception, then perception has to reflect its properties, i.e., perception should be *veridical*. In opposition, *idealism* gives primacy to perception, stating that what we might know of the physical reality is determined by perception, and that what we have access to is the product of perception and not a physical reality.

Realism is often criticized with reference to the assumption of veridicality of perceptions, sometimes found to be in error, i.e., they do not necessarily reflect the properties of physical reality. Different instances of error can be outlined. None of them, according to the ecological approach, disprove the assumption of veridical perception. Misperceptions might arise in the absence of sufficient information for the specification of the objects and events in the environment. This includes both conditions representative of the normal interaction of the animal with the environment (e.g., viewing in presence of fog), as well as artificial laboratory conditions (e.g., tachistoscopic presentation of visual stimuli). Talking about errors of perception in this case would, in principle, be equivalent to asserting that we are in error if we are not able to see through walls, an evident case of absence of sufficient information. In other cases, sufficient information might be available, but it might go undetected because of limits in the sensory apparatus, or because the perceptual system is still not tuned to detect it. Accordingly, it is not appropriate to talk about errors of perception if we are not able to perceive the ground vibrations that come with relatively weak seismic phenomena. Finally, misperceptions might be erroneously identified in those cases when, despite sufficient and detectable information, the perceived properties of the environment are not consistent with those we might measure with devices other than the perceptual system. This is the case, for example, of visual illusions (see Section 2.2.2 and Figure 2.1). In such cases, however, we might have failed in identifying the information that is relevant for the perceptual system, which is ultimately the information assumed to be detected veridically. Thus the error would not be at the level of the perceptual system, but would be with those who mistakenly defined the variables for perception.

After these considerations, it seems that the assumption of veridicality might still be tested empirically, provided that the variable under investigation is correctly defined, is transparent to the sensory system of the animal, and provided that sufficient information is given. But even if misalignments were found, the theory wouldn't be undermined. Indeed the ecological approach ultimately rejects the appropriateness of the application of the term error to perceptions. This label is indeed properly applied to propositions (e.g., the earth is flat), which can be either true or false. According to the ecological approach, then, perceptions, as well as actions, are not propositions, but "states of affair" (Michaels & Carello, 1981, p. 109), and should be conceived in the same way as one might conceive anatomic parts of the animal. Thus, because the five fingers of our hands can't be labeled as erroneous or true, but as useful or not, then perceptions should be evaluated with reference to their usefulness to the animal-environment interaction.

2.1.5 Contributions to sound source perception

Both the contributions presented here share with the ecological approach the idea that a better understanding of auditory perception should focus on the distal stimulus.

Balzano (1986) proposed a new approach for the study of timbre perception, and identification of musical instruments. Since the time of von Helmholtz (1877/1954), the peripheral auditory system has been conceived as a Fourier analyzer, and timbre perception has been thought to be based on the distribution of energy across the spectrum. According to Balzano (1986) this view can't account for several phenomena, such as the supposed constancy of musical instrument identification over the pitch range, and across variations in dynamics (see Section 4.7.1 for a review of the empirical results concerning pitch effects on musical instrument identification). Consequently "an alternative theory that [...] more nearly captures what we perceive" is proposed. Following Gibson's approach, Balzano (1986) suggests that a more proper framework for the study of timbre perception, and musical instrument identification, is based on the *wave equation* which specifies the vibratory behavior of the sound sources, and is invariant over variations in pitch and dynamics. Thus timbre perception is better understood with reference to the properties of the source, and source identification relies on the detection of those invariant properties of the acoustical information that specifies the source. A similar position is adopted by Gaver (1988, 1993b). Algorithms for the synthesis of environmental sounds (e.g., impact sounds, liquid sounds), based on an analysis of the physical behavior of the sound source, are conceived as "instantiated hypotheses about the acoustic information for events", i.e. for perception of source properties.

Another contribution by Gaver (1988, 1993a) concerns the definition of two different listening modalities. With *musical listening* the focus of perception is on the properties of the proximal stimulus per se, and the relevant dimensions of the percept are independent of the source. With *everyday listening*, instead, the focus is on the distal stimulus, and the perceptual dimensions of interest are related to the sound source. According to Gaver (1988, 1993a), both listening attitudes can be used for the same sound, although the second of them would be most frequently adopted during everyday life. A consequence of this hypothesis is that when musical listening is involved, the properties of the percept should be closely tied to the properties of the acoustical signal, and show little or no relationships with those of the source. This consequence of the distinction between musical and everyday listening drawn by Gaver (1988, 1993a) was tested in Section 4.7, analyzing previously published data on perception of musical instrument sounds.

Finally, a point of strong commonality between the ecological theory and the approach of Gaver (1988, 1993a, 1993b) is found in the definition of the object of study for the understanding of perception. As for the ecological theory perception should be studied using sounds representative of the environment of the perceiver, Gaver conceives research on source perception as necessarily focusing on the so-called "everyday sounds", i.e., all those sounds other than music or speech as breaking and rolling sounds, elsewhere defined as environmental sounds (cf. Gygi, 2001). Such a position is pushed further by Gaver in denying the relevance of musical sounds to the understanding of source perception. This choice is based on two assumptions. Firstly, musical sounds "reflect only a small part of the range of sounds encountered by people in their everyday lives" (Gaver, 1988, p. 10). Secondly, given their "quasi-harmonic" nature, musical sounds would "provide relatively little information about their sources" and would be "too simple to permit easy identification of their sources" (Gaver, 1988, p. 10). It should be noted that the first of these assumptions can not be tested on the basis of published data, and is plausible only if we conceive the "average listener" as having no interest in music. Concerning the second of these assumptions, it is unclear why the musical "quasi-harmonic" sounds would offer less source-specific information than the everyday inharmonic sounds. Also, the corpus of published data on perception of musical instrument sounds can be used to test the relevance of source properties to audition of these particular class of signals. Such analysis is presented in Section 4.7.

2.2 The information processing approach

The ecological approach has been opposed to the information processing approach to perception (Lindsay & Norman, 1977; Anderson, 1980). The main assumptions of the latter are:

1. the proximal stimulus is processed in a series of sequential stages, which alter the information systematically;
2. the proximal stimulus is ambiguous with respect to the distal stimulus;
3. internal processing compensates for ambiguity in the proximal stimulus by way of additional information provided by internal processes, based on past experience.

2.2.1 Stages in the processing of information

The information processing approach assumes perception to be the end-product of a multi-stage process which, starting with the proximal stimulus, analyzes and manipulates the incoming information in specific ways. Thus, the purpose of those adopting the information processing approach for the study of perception is to trace the sequence of stages resulting in perception, and the properties of the representations on which they operate.

A popular example of application of information processing to perception is found in the work of Marr (1982), who modelled recognition of visual objects as a series of consecutive stages, each operating on representations of a different nature. The input to the process is found in the retinal image, the *raw primal sketch*, which contains information about the distribution of intensities. The *primal sketch* operates on this representation, generating information on the distribution of edges, contours and blobs. The *2½-D sketch* generates information about the depth and orientation of visible surfaces, and is centered on the viewing position. Finally, the *3-D sketch* generates a three-dimensional model of the scene, independent of the viewing position. The recognition process would operate matching the final result of this elaboration to three-dimensional models stored in memory.

Matching of a representation resulting from the analysis of the incoming stimulation with information stored in memory is the basis of *pattern recognition*. Two types of pattern recognition models are found: *template models* and *feature-based models*. Template models assume recognition to be based on the comparison of the pattern of the incoming stimulation to a copy or template stored in memory. This mechanism would require, for example, 6 different templates to explain why we recognize the same letter in the following patterns: "A", "A" "A", "a", "a", and "a". This model is much too complex to be able to explain human recognition capabilities as it requires a separate template for each pattern that leads to the same recognition. This limit does not characterize feature-based models, which assume the incoming pattern to be analyzed in terms of elementary components or features (in the case of visual patterns the number of lines, of acute angles, of curves, etc..). Recognition is assumed to be based on the detection of those basic features, and higher level features that code for the relationships among the basic ones, that allow constant recognition despite variations associated with a change in font, size, etc..

2.2.2 Empiricism

According to the information processing approach the sensory information provides only an impoverished description of the world. Impoverished information is in contrast with our ability to interact efficiently with the environment, for which sufficiently veridical perception can be demonstrated. To solve this contradiction, the information processing approach has to assume that disambiguation of the input is achieved by virtue of additional information provided by the perceiver. Such an assumption is rooted in the stances of *empiricism*, the main paradigm for the study of perception in this century (Gordon, 1989).

Empiricism was spread in perception by Helmholtz, who thought perception to be the result of the application of a constructive process operating on sensory data, which adds information to an ambiguous input. The constructive process takes the form of an *unconscious inference*, on the basis of which sensory data are interpreted with reference to previous experiences. A modern version of the unconscious inference theory is found in the work of Gregory (e.g., Gregory, 1970). The principle can be explained examining a famous visual illusion: the Ponzo illusion (see Figure 2.1).

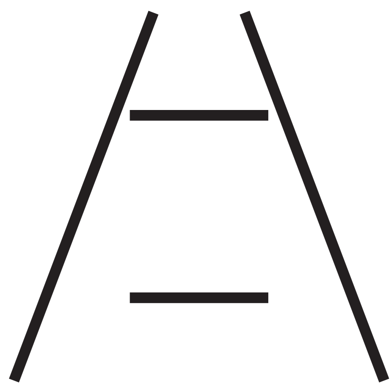

Figure 2.1: Ponzo illusion

In this configuration, despite the fact that the two horizontal segments have the same physical length, the upper segment is perceived as longer than the bottom one. According to Gregory, several visual illusions result from the wrong application of rules for the perception of three-dimensional objects to the perception of two-dimensional configurations. In the case of the configuration in Figure 2.1, the mistakenly applied rule takes the name of *size-constancy scaling.*

When the same object is viewed from two positions, one farther than the other, it's retinal projection changes, becoming smaller when the object is observed from a greater distance. Despite the difference in size of the retinal images, the size of the object is perceived as constant. This occurs because the visual system compensates for the variation in the size of the retinal image caused by a variation in the distance of the object from the observer (size-constancy scaling). Consequently the small retinal image of an object perceived at a farther position results in the same perceived size as the larger retinal image of the same object perceived at a closer position. This process is triggered by cues for depth. One of these cues is found in linear perspective, according to which parallel lines converge as their distance from the observer increases. In the case of Figure 2.1, the visual system treats the obliqueness of the two flanking lines as a cue to depth, where the closer the point on the line is to the convergence point, the farther its position from the observer appears to be. This depth information is extended to the two horizontal lines, so that the upper one, closer to the convergence point, is assigned a farther position in the three-dimensional space. According to the size-constancy scaling this line is also perceived as longer than the lower one, which is assigned a closer position in the three-dimensional space.

Gregory's explanation reveals the role that the so-called errors of perception are given for the understanding of the functioning of the perceptual systems. Indeed, the information processing approach conceives errors as "valuable to us because the mechanics of a system are frequently revealed primarily through its errors and distortions" (Lindsay & Norman, 1977).

2.2.3 Top-down vs bottom-up

It is possible to distinguish among processes for perception depending on whether they are based on the information input to the sensory system or not. In the first case we talk about *bottom-up* or *data-driven processing.* The approaches to recognition presented in Section 2.2.1 are based exclusively on bottom-up processing.

Several studies provide evidence that sensory information alone is not able to explain perception. Warren (1970) presented listeners a speech sequence, replacing the central phoneme in one of the words with either a cough or a tone. Listeners reported hearing the missing phoneme despite the absence of information in the incoming stimulation (*phonemic restoration*). When perception is determined by information not present in the stimulation we talk about *top-down*, or *conceptually-*, *schema-*, *knowledge-driven processing.* With the phonemic restoration effect the source of information for the perception of the missing phoneme is found in the phonemes flanking the cough/tone and in the listeners knowledge of the language. In other words, the additional information comes from the "*context* of the sensory event" (with Warren's configuration the sensory event evoked by the cough/tone), where context is the "information that is accumulated and routinely used to understand events", "as well as the overall environment in which experiences are embedded"(Lindsay & Norman, 1977).

When embedded in a meaningful context, recognition might require a lower amount of information. For example, when embedded in the drawing of a face, the drawing of an eye needs much less detail to be properly recognized than when presented out of this context (Palmer, 1975), as shown in figure 2.2.

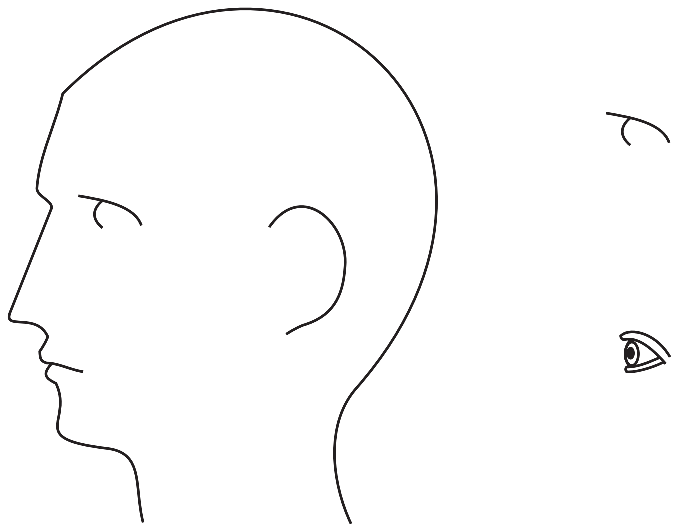

Figure 2.2: Effects of context on the amount of information needed for recognition. Adapted from Palmer (1975).

Thus with meaningful context, information for perception becomes *redundant*: the detailed drawing of the eye embedded in the context of the face drawing contains much more information than is actually needed for proper recognition. Redundancy of information is far from being useless. Indeed in the presence of redundancy, recognition is made *robust*, as distortions, or missing detections of portions of the signal in input do not cause a failure in recognition.

2.2.4 Contributions to sound source perception

In this section two sound source perception models are presented. Each of them, in accordance with the information processing approach, conceives perception as the result of a multi–stage process. However, in both cases some traits of the models reveal influences of the ecological approach too.

Figure 2.3 shows the stages of the sound source recognition model outlined by (McAdams, 1993). The *sensory transduction* stage involves the processing of the incoming signal by the cochlea, and the conversion of the movements of the basilar membrane in electrical signals transmitted through the fibers of the auditory nerve. As a result of this pro-

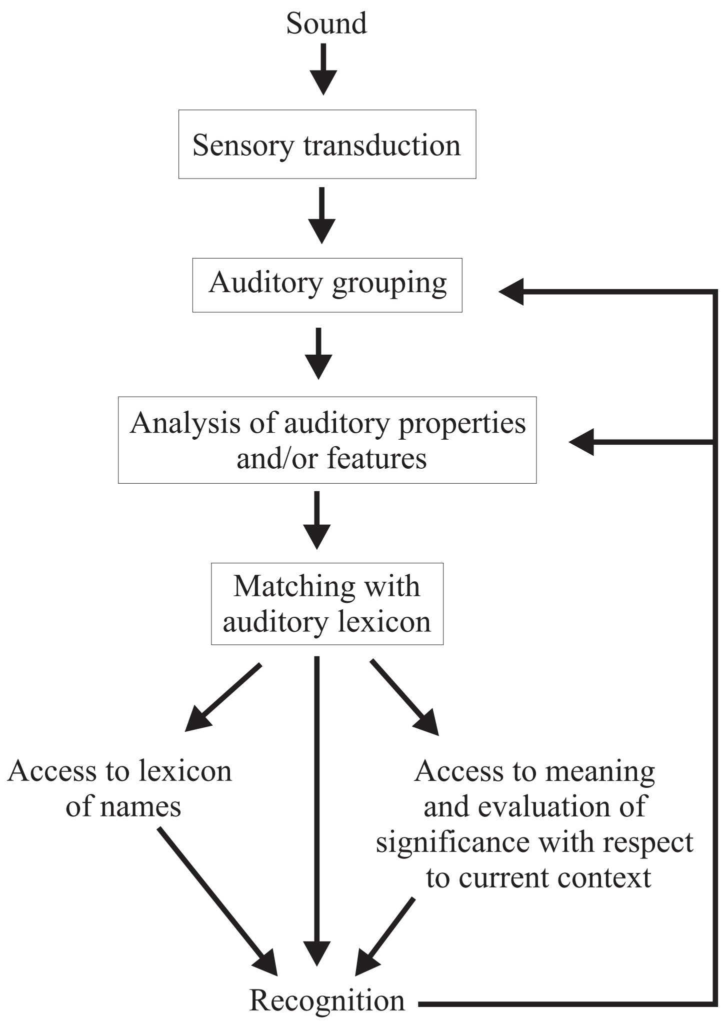

Figure 2.3: Stages of processing of information for sound source recognition, after McAdams (1993)

cessing, two different representations might be generated, one which codes the spectral evolution of the signal in time, the other its temporal fine-structure. *Auditory grouping* processes operate on the representations output by the sensory transduction process. The task of this process is to link those portions of the incoming information originating from the same sound source (Bregman, 1990), resulting in the formation of auditory objects or streams. Grouping processes are distinguished on the basis of the time scale on which they operate. Simultaneous integration links together simultaneous portions of the incoming representation and allows perceiving the note of a musical instrument as a single perceptual entity or auditory object, rather than as a mixture of unrelated spectral components. Sequential integration processes group together auditory events presented successively in time, and allows the piano tones in an orchestral composition to be perceived as a single stream. A second distinction among grouping processes is drawn according to whether they are based on the incoming information alone (primitive or bottom-up processes), or on previous knowledge of the listener (schema-based or top-down processes). Schema-based processing facilitates the segregation of the voice of someone pronouncing our name from the mixture within which it appears. The result of the grouping processes is then analyzed in terms of its relevant *features*. One can distinguish among features depending on whether their extraction operates over short or long time spans. Microtemporal properties are extracted over short time spans (milliseconds to centiseconds), and allow us to perceive whether a given signal has been generated by a large, rather than a small, struck object. Macrotemporal properties are extracted over longer time spans (up to a few seconds), and allow us to distinguish between breaking and bouncing objects (Warren & Verbrugge, 1984). According to McAdams (1993), extracted properties can also be conceived as invariants, properties of the stimulation that specify the source. Structural invariants, for example, would be part of the class of the microtemporal properties. The extracted features are then compared with the *auditory lexicon*, representations stored in long-term memory which define different classes of sound sources. Recognition is conditional on the activation of one representation in the auditory lexicon. Finally, items in the lexicon of names, concepts, and meanings, might be activated, allowing identification (naming), and programming of appropriate actions.

Richards (1988) proposed a sound source recognition model, closely connected to the model by Marr (1982) for visual object recognition. The model comprises three sequential stages, each operating on different representations. The vibration of the tympanum, in response to the incoming sound, is transduced into a spectrotemporal representation that codes the temporal variation of the vibration of the different portions of the basilar membrane in terms of neural activity. This representation is further elaborated in the *primal sketch*. Comparison of the information coming from the two inner ears allows portions of the representations originating from different sources to be separated. Potential features for the representation of the primal sketch are onset and offset time, tonal quality, frequency change, and harmonic structure. The central stage of the model is the *2½-D sketch*, which codes the information in a listener-position centered framework. Sound sources are conceived as comprising four components: a power source P, an oscillator O, a resonator R, and a coupler C. For example, in the case of voice the source of power is the lungs, the vocal folds are the oscillator, the resonator is the vocal tract, and the mouth is the coupler. Thus the signal A input to the auditory system, with dimensions time t and frequency f, is given by the properties of the P, O, R, and C components, and by the properties of the environment k, as mediated by phenomena such as reflections or echoes. This characterization is formalized in equation 2.1:

$$A\,(f,t) = k\sum_{i}^{n} g_i(P,O,R,C) \tag{2.1}$$

where n is the number of sources that generate the incoming signal, and g_i is the function that links the signal generating from each source to its constitutive components. Equation 2.1 is proposed as potentially solvable by the auditory system, and source perception is suggested to be better understood by means of the study of the influences of the

P, O, R, and C components, as well as of their interaction, on the structure of the acoustical signal/auditory representation. Finally, the *3-D sketch*, by analogy with the model outlined by Marr (1982), would represent the acoustically meaningful environment independently of the position of the listener. The usefulness of this latter representation is however questioned by Richards (1988), as all information for recognition would already be contained in the $2^1/_2$-D sketch representation. The central stage of the model by Richards (1988) expands the approach that Balzano (1986), inspired by the ecological approach, proposed as a new framework for the study of timbre perception. In both cases the study of source behavior is of primary relevance to the understanding of perception. However Richards (1988) considers a somewhat more detailed characterization of the source, as well as the effects of the environment on the signal/information and the presence of simultaneous sources.

2.3 Converging and diverging traits in the ecological and information processing approaches

Contrary to the information processing approach, the ecological approach seems to bypass the question of how perception is achieved (McAdams, 1993), as it assumes perception to be based on resonance to information, without further specifying the nature of this process. Thus it is not surprising that the ecological approach has been defined as analogue to the behaviorist stimulus-response approach, which discarded explanations based on internal processes and representations (Epstein, 1982). Indeed, for the ecological approach the question concerning how perception is achieved seems to reduce to the problem of highlighting the stimulus properties upon which it is based. In seeking an explanation for perception in the stimulus information alone, in the detected invariants, the ecological approach has a trait in common with the feature-based pattern-recognition models outlined in Section 2.2. Indeed both can be conceived as bottom-up approaches to perception (Eysenck, 2001, p.29). Perception being explained in both cases on the basis of properties of the incoming information, it is not surprising to find a commonality between the concept of invariant and that of feature. Indeed, in both cases the perceptually relevant variables are assumed to be independent of extraneous variations in the configuration, such as the size, the orientation, and the font for letters.

One point of strong opposition concerns the veridicality of perception, assumed by the ecological approach, but not by the information-processing approach. Although this assumption appears easily tested empirically, the considerations outlined in Section 2.1.4 show that this might not be the case. Another point of strong opposition is found in the conceptions of direct, and indirect perception. Epstein (1982) presents convincing arguments in favor of the indirect approach, based on experimental evidence. As stated several times in this chapter, the direct approach asserts that perception is a function of the proximal stimulus alone. As such, the theory assumes that no internal representations can influence the properties of the derived percept. Such an eventuality is however found in what Epstein (1982) defines as *percept-percept couplings*, i.e., those cases in which one property of the percept influences another perceptual property. A percept-percept coupling has been outlined, for example, by Gilchrist (1977), where, despite the constant luminance of a target patch in a visual display, it's perceived lightness depended on whether it was perceived as coplanar to one region or another of the display. Cases like this point toward the relevance of internal representations in determining the properties of the percept, and strongly contradict the direct perception theory. None of these cases are found, to my knowledge, in the literature on sound source perception, and the question of whether source perception is direct or not still remains open.

Chapter 3

Methodological issues

This chapter presents the methodological instruments needed to carry out research on sound source perception. The first section defines the research design in this field. The second section describes the acoustically relevant properties of the sound source, focusing on struck bars and plates, and highlights relevant acoustical parameters associated with the different source properties. The third, and final section summarizes briefly the main methodologies used for behavioral data collection in this field.

3.1 The research design in sound source perception

The research design in sound source perception has been formalized by Li, Logan, and Pastore (1991). The object of study can be described at three different levels: the physical or mechanical level, the properties of the sound source; the acoustical level, the structure of the signal; the perceptual level, the properties of the percept evoked by the incoming signal, as inferred from the behavioral response (see Figure 3.1).

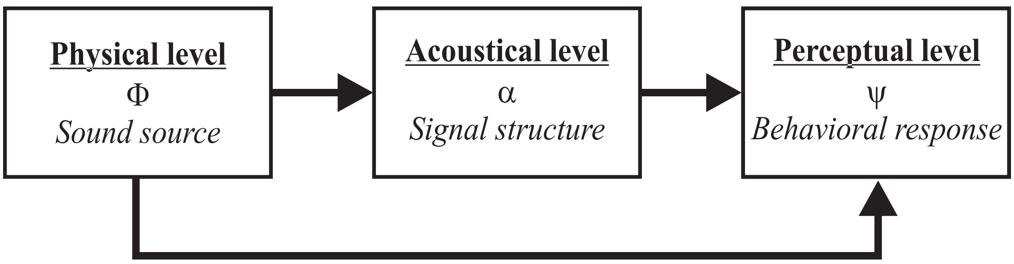

Figure 3.1: The research design in sound source perception, after Li et al. (1991).

A complete research design should analyze all the pairwise relationships among these three levels.

Concerning the physical level, a wise approach is to adopt the most complete, and finely grained characterization possible, in order to avoid assumptions concerning which properties of the source are perceptually relevant. This suggestion shouldn't be taken literally, so that in a hypothetical study on the auditory discrimination of metals one shouldn't consider the melting temperature as a descriptor variable. Instead, the selection of the set of physical descriptors should be ideally based on knowledge coming from the field of physical acoustics, i.e. from previous studies concerning the acoustically relevant physical variables.

Concerning the acoustical level, three indications should be followed. First, the computation of the selected descriptors should take into account knowledge concerning the processing of the signal in the auditory system (see Section 2.1.2). Second, the selection of acoustical descriptors should take into account, when available, previous knowledge of how source properties structure the acoustical signal (see Section 2.1.1). Third, even though not strictly linked with source properties, acoustical variables found in previous research to be associated with judgments of similar signals should be included.

Concerning the experimental methodology used to investigate source perception, one indication should be followed. When investigating perception of a given source property using direct judgments, say identification of the material of a struck object, the experimental set should include variations in material, as well as variations in other source properties, such as size. In other words one should also include variations in extraneous properties, referred to as *perturbation variables*. As pointed out by Gaver (1988), this "allows a more strict test of whether subjects can actually perceive the attribute in question despite irrelevant changes in the sounds". Another advantage of the use of perturbation variables is that of increasing the generalizability of experimental results to everyday listening conditions, in which sources are not constrained to vary along only one property. The generalizability issue suggests also that when investigating source perception with techniques that do not require direct judgments (e.g., dissimilarity ratings or classification, see Section 3.3), one should include variations of as many source properties as possible.

Analysis of all the pairwise relationships among the three levels allows increasing knowledge on how source properties structure the acoustical signals, and on the physical, and acoustical determinants of source perception. In my opinion, both the physical, and acoustical levels should be conceived as alternative and equally important ways of characterizing the stimulation. Each of them serves different purposes: considering the physical level allows for the gathering of knowledge directly related to the interaction of the animal with the environment while considering the acoustical level allows us to understand the medium upon which this interaction is based.

As pointed out by Li et al. (1991), the analysis of the acoustical determinants of perception should be complemented with a final experimental phase where those variables found to be associated with participants' judgments are explicitly manipulated by the experimenter. This additional stage in the research design has the purpose of highlighting a causal link between acoustical properties and source perception, a link that cannot be established with the correlational procedures usually used to investigate the relationship between the acoustical and perceptual levels. Manipulation of the stimuli might be necessary for another reason: usually with real signals, or with signals synthesized on the basis of a physical model of the source, some acoustical variables may be correlated with one another. Thus it might happen that more than one acoustical descriptor can explain equally well the same behavioral variable. In this case the experimenter might decorrelate them in a final experiment, manipulating the recorded or synthesized signals. Correlations might be found among source properties as well. For example, it is already known that with solid objects density and elasticity are correlated, where denser materials are also stiffer (Waterman & Ashby, 1997). Then a final decorrelation experiment might be necessary when strongly associated physical variables explain participants' responses equally well. In this case, then, it might be useful, or even necessary, to use simulated sound sources. Indeed, with real objects it might be difficult or even impossible to build experimental sets where these variables are not correlated.

3.2 Source properties and structure of the acoustical signal

This section focuses on the physics of the two sources that have been studied in this thesis, and that have been the main focus of research on impact sounds: bars (Gaver, 1988; Lakatos, McAdams, & Caussé, 1997; Lutfi & Oh, 1997; Carello et al., 1998; Houix, McAdams, & Caussé, 1999; Roussarie, 1999; Lutfi, 2001; Klatzky, Pai, & Krotkov, 2000; McAdams, Chaigne, & Roussarie, 2004; Houix, 2003), and plates (Roussarie, 1999; Kunkler-Peck & Turvey, 2000; Giordano, 2003; Tucker & Brown, 2003). The physical variables that influence the properties of the acoustical signals generated by these vibrating systems are presented. When appropriate, these are linked with acoustical descriptors used in previous research on source perception or in the investigations presented in this thesis. The following presentation is based on three monographs, Handel (1989), Rossing, Moore, and Wheeler (2002), Fletcher and Rossing (1991), and on additional material cited throughout the section.

3.2.1 Simple harmonic motion

The simplest vibrating mechanical system is composed of a mass m attached to a spring, and free to move longitudinally, i.e. along the direction of the spring (Figure 3.2).

When displaced a distance x from the equilibrium position and released, the mass moves back to the initial position thanks to *restoring forces* of the spring. However, its inertia causes the mass not to stop its motion at the initial position, but to pass after it, or overshoot. Again, restoring forces will move the mass back to the equilibrium position, and inertia will cause the mass to overshoot again. In other words, the mass will oscillate around the equilibrium position. With this system the relationship between the restoring force F, and the displacement x of the mass is governed by Hooke's law: $F = -Kx$, where K is called *spring constant* or *spring stiffness*. When F is proportional to x, the motion of the mass is a sinusoidal function of time, and is called *simple harmonic motion*. The *natural frequency* f of the vibrating system, with which the mass oscillates around the equilibrium position in absence of external driving forces, is given by Equation 3.1.

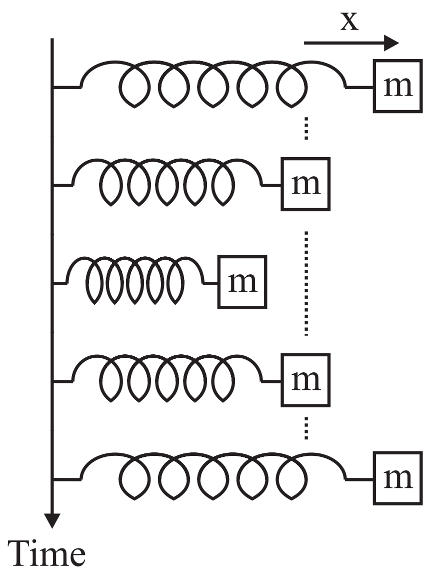

Figure 3.2: Simple harmonic motion: mass-spring system.

$$f = \frac{1}{2\pi}\sqrt{\frac{K}{m}} \tag{3.1}$$

In absence of energy dissipation, the oscillation of the mass would go on forever. However, for a number of different physical causes, generally referred to as *damping*, the amplitude of the vibration decreases with time, and the mass eventually stops at the equilibrium position again.

3.2.2 Modes of vibration

A slightly more complex system can be created by linking the mass to two springs clamped at their ends, and allowing the mass to move in one plane only. In this case the mass can vibrate both longitudinally, i.e. along the line connecting the springs at equilibrium, as well as transversally, i.e. perpendicular to this line (Figure 3.3). The system, thus, will have two *vibrational modes*, one longitudinal, and one transversal.

As we add more masses, and springs to the system, the number of vibrational modes increases. With two masses and three springs the system has two transversal and two longitudinal modes; with three masses and four springs both the transversal and longitudinal modes will be three, and so on. A system with an infinite number of masses, which approximates the case of a vibrating string, has infinite transversal and longitudinal modes.

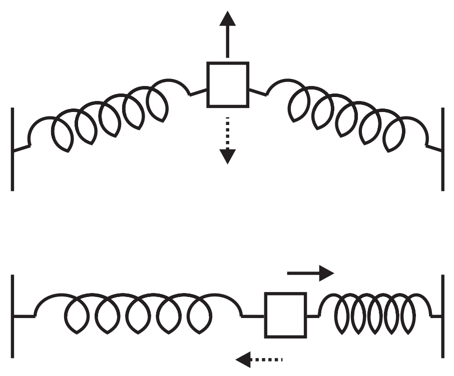

Figure 3.3: Vibrational modes of a two springs/one mass system.

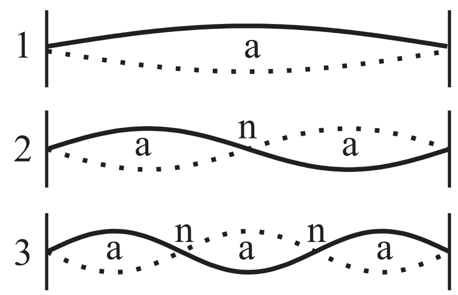

Figure 3.4: First three transversal modes of a string fixed at both ends (a = antinodes; n = nodes.

Figure 3.4 shows the first 3 transversal modes of vibration of a string fixed at both ends. The frequency of the n^{th} mode is given by Equation 3.2.

$$f_n = \sqrt{\frac{T}{\rho}}\,\frac{n}{2L} \text{ with } n = 1, 2, 3, ..., \tag{3.2}$$

where T is the tension of the string in newtons (N), which provides the restoring force to the vibrating system, ρ is the linear density of the string in kg/m^3, and n is a positive integer indexing the transversal vibrational mode. The frequencies of the vibrational modes are integer multiple of that of the first mode, called the *fundamental frequency*. In this case the frequencies of the vibrational modes constitute an *harmonic series*.

As can be seen in Figure 3.4, for each mode, there will be positions on the string where the displacement is always zero. For the second vibrational mode this is the case for the two extremes of the string and for its center. These positions are called *nodes*. Conversely, for each mode there will be positions were displacement reaches its maximum value, corresponding to the center of the string for the first mode. These positions are called *anti-nodes*.

Vibrational modes are independent of each other, and for this reason they are also called *normal modes of vibration*. Vibrational modes are independent because it is possible to excite each of them selectively. This, in general, is done by driving the vibrating system to oscillate at the frequency of the particular vibrational mode, using one of a variety of methods, such as mechanical or acoustical excitation. When the frequency of the external driving force equals the frequency of a specific vibrational mode, the amplitude of vibration of the system will increase significantly, and the system is said to *resonate*.

3.2.3 Vibration of bars and plates

Bars, as well as strings, are conceived as one-dimensional vibrating systems. In strings the restoring forces are supplied by the tension, while in bars restoring forces are supplied by the rigidity of the material. In bars three types of vibrations are possible: longitudinal, transversal, and torsional. As in strings transversal vibrations are perpendicular to the length of the bar, along which bending takes place. In longitudinal vibrations the bar expands and contracts in length. In torsional vibrations the bar twists around a longitudinal symmetry axis. In the following presentations only transversal modes of vibration are considered.

Many properties of the bar affect the frequency of its vibrational modes. First, how the bar is supported at each of the two extremes, which define the so-called *boundary conditions*, second the geometry of the cross section, third the length, and fourth its density and elastic properties, which jointly determine the speed of waves within the bar.

For each of the two extremes, three different boundary conditions are possible: the bar can thus be free to vibrate at both ends, or be rigidly clamped, or simply supported (hinged) (see Figure 3.5).

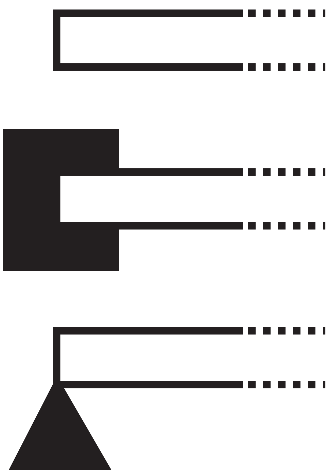

Figure 3.5: Boundary conditions for a vibrating bar. Top: freely vibrating; middle: clamped; bottom: simply supported.

Equations 3.3, 3.4, and 3.5 give the frequency of the n^{th} transversal mode when both extremes are, respectively, free, clamped, or simply supported (hinged).

$$\text{Free-free: } f_n = \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}m \text{ with } m = 3.011^2 \text{ for } n = 1 \text{ and } m = (2n+1)^2 \text{ for } n > 1 \tag{3.3}$$

$$\text{Clamped: } f_n = \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}m \text{ with } m = \left\{1.194^2, 2.988^2\right\} \text{ for } n = \{1,2\} \text{ and } m = (2n-1)^2 \text{ for } n > 2 \tag{3.4}$$

$$\text{Simply supported (hinged): } f_n = \frac{\pi K}{2L^2}\sqrt{\frac{E}{\rho}}n^2 \text{ with } n = 1, 2, 3, ..., \tag{3.5}$$

where L is the length of the bar; K, in meters, is a constant called *radius of gyration* which depends on the geometry of the section of the bar ($K = h/\sqrt{12}$ for a bar of rectangular cross-section, and thickness h; $K = r/2$ for a bar of circular cross-section of radius r); ρ is the density of the bar, in kg/m^3; E is the Young modulus, which characterizes the elasticity of the material and is measured in N/m^2. In particular, in a bar subjected to a longitudinal stress, the higher the value of E, the lower the longitudinal strain. The quantity $\sqrt{E/\rho}$ gives the speed of longitudinal waves.

A variation of all the terms in Equations 3.3, 3.4, and 3.5 determine a change in the absolute frequency of the vibrational modes, but not of their ratios. One of the acoustical measures used by Klatzky et al. (2000) and McAdams et al. (2004) to characterize bar sounds takes into account these dependencies: the frequency of the lowest spectral component. If all the other source properties are kept constant, indeed, variations in the frequency of the lowest spectral component unambiguously specify the length of the bar, as in Klatzky et al. (2000) and in two of the stimulus sets investigated by McAdams et al. (2004), or its density, as in one of the sets investigated by McAdams et al. (2004). Variations of the boundary conditions determine a change in both the frequency of the vibrational modes and in their ratios. In particular, when both extremes are simply supported, frequencies are less widely spaced. For this reason, it is likely that a variation in the boundary condition would result in a change in the amount of inharmonicity of the generated signals (i.e. deviation from the harmonic relation).

Plates can be considered as two-dimensional bars or as stiff membranes. As with bars, the frequencies of the vibrational modes are influenced by the boundary conditions, by the elastic properties, by the density, and by the geometrical properties. If the plate is simply supported at all four edges, the frequencies of the vibrational modes are given by Equation 3.6.

$$f_{mn} = 0.453\sqrt{\frac{E}{\rho\left(1-\nu^2\right)}}h\left[\left(\frac{m+1}{L_x}\right)^2 + \left(\frac{n+1}{L_y}\right)^2\right] \tag{3.6}$$

where ν is a dimensionless constant called Poisson's ratio, which gives a measure of the lateral contraction of a solid subjected to a longitudinal strain; L_x and L_y are the width and the length of the plate, respectively, and m, and n are positive integers which give the number of nodal lines along the length, and the width of the plate, respectively, without counting the nodes at the edges.

Equation 3.6 highlights how variations in the elastic properties, in the thickness, and in the density of the plate determine a change in the modal frequencies, but not in their ratios. The same effect is achieved when the size of a plate (height $\times$ width) is varied, keeping shape constant (height/width ratio). A variation in each of these parameters alone is uniquely specified by the frequency of the lowest spectral component of the generated signal. This acoustical descriptor was used in the experiments presented in Chapters 6 and 8, where plates of constant shape, variable size and variable material were investigated. It should be noted, however, that in these cases the frequency of the lowest spectral component did not uniquely specify the size or the material properties (density, elasticity) of the plates, as both were varied within the stimulus sets investigated.

More complex effects are associated with a variation in the shape of the plates, which alter the absolute value, as well as the ratios among modal frequencies. Following the same considerations as for the variation in the boundary conditions in bars, it is likely that variations in the shapes of the plates determine variations in the inharmonicity of the modal frequencies. Inharmonicity measures, however, were not considered in previous studies on shape perception (e.g., Lakatos et al., 1997). Performance was instead explained measures related to the modal frequencies. Lakatos et al. (1997) investigated recognition of the height/width ratio of struck bars, large enough to be considered as two-dimensional systems, and thus capable of transversal bending modes along the length and width. Accordingly, one of the descriptors used to explain participants' performance was the ratio of the frequency of the lowest transversal mode related to the height, to the frequency of the lowest transversal mode related to the width.

Up to this point it has been assumed that the material of the plates and bars have constant elastic properties in all directions. This condition, which characterizes for example metals and glass, is called *isotropy*. The elastic behavior of isotropic materials can be fully characterized using two constants: the Young modulus, and the Poisson's ratio. For other materials, however, the elastic properties are not constant in different directions, and more than two elastic coefficients are needed. With *orthotropic* materials, such as wood, different elastic properties are found along three orthogonal axes: a longitudinal axis, parallel to the direction of the fibers, or grain; a radial axis, in the direction of growth of the rings, and a tangential axis. A variation in the elastic properties along the different directions determines a variation in the waves velocity, and thus in the frequency of the vibrational modes. Although several previous studies have investigated sets generated by striking both isotropic and orthotropic bars or plates, none of them has explicitly related this distinction to the structure of the acoustical signals.

3.2.4 Damping

Damping refers to the dissipation of energy in the vibrating system. Damping is responsible in first place for the decay of the amplitude of vibration, and consequently for the decrease in the overall amplitude of the radiated sound. Further, damping causes a decrease in the modal frequencies, although this effect often may be considered as secondary. It is possible to distinguish among different types of damping, and thus, among different mechanisms responsible for energy dissipation. *Internal damping* refers to energy losses resulting from various processes within the material of the vibrating object. *Structural damping* refers to energy losses resulting from the relative motion of the vibrating element with respect to the structures that eventually support it. Additional damping is provided by the interaction of the vibrating element with the surrounding medium. When the surrounding medium is the air we talk about *air damping* (Fletcher & Rossing, 1991); when the medium is a fluid we talk about *fluid damping* (de Silva, 2000).

Wildes and Richards (1988) developed an internal damping model with the intent of extracting an acoustical measure that uniquely specifies the material of an object. Following the work by Zener (1948), Wildes and Richards (1988) proposed material to be specified by a measure of internal damping, the coefficient of internal friction $tan\phi$, given by Equation 3.7.

$$\tan\phi = \frac{\alpha}{\pi f} \tag{3.7}$$

where f is the frequency of vibration, and α is a quantity termed the *damping factor*, defined as the reciprocal of the time t_e required for the amplitude to decay to $1/e$ of its starting value. Damping increases with internal friction, so that highly damped materials, such as rubbers, are characterized by higher $tan\phi$ values. The model by Wildes and Richards (1988) assumes a linear relationship between damping factors and frequencies, where α increases with increasing frequency, i.e. the amplitude of higher frequency vibrations drops faster.

The $tan\phi$ model has been used both for synthesis (van den Doel & Pai, 1998), and analysis (Tucker & Brown, 2003) of impact sounds. In the latter case the signal was processed with a bank of bandpass filters, and, roughly speaking, an average $tan\phi$ value was computed, applying Equation 3.7 to the amplitude of the output from each filter, f being their center frequency. It should be noted, however, that for recorded signals, the extracted $tan\phi$ values are not determined by internal damping alone. Indeed, in this case damping of vibration is determined also by the way the plate is supported (structural damping) and by the interaction with the air medium, or, in the case of underwater recordings (Tucker & Brown, 2003), with the surrounding fluid. Thus the $tan\phi$ measure extracted from recorded impact sounds is to be conceived as a rough damping measure.

More complex damping models have been used for the synthesis of impact sounds (Doutaut, Matignon, & Chaigne, 1998; Lambourg, Chaigne, & Matignon, 2001) investigated in sound source perception research (McAdams et al., 2004; Roussarie, 1999). In Doutaut et al. (1998), the relationship between damping factors and frequency is quadratic and not linear as in the $tan\phi$ model. The perceptual relevance of the quadratic damping model was tested by (McAdams et al., 2004; Roussarie, 1999). Lambourg et al. (2001) modeled three independent sources of damping: two internal damping mechanisms (thermoelastic and viscoelastic damping), as well as one mechanism of air damping, the energy losses due to acoustic radiation (radiation damping). The perceptual relevance of this damping model was tested in Roussarie (1999). Interestingly, an acoustical measure closely related to $tan\phi$ was found to be associated with the behavioral outcome: the slope of a linear function relating the frequency of the spectral components to their damping factors α.

In chapters 6, 7, and 8 an approach similar to that adopted by Tucker and Brown (2003) was used to extract a rough damping descriptor from the investigated sounds according to the $tan\phi$ model. The choice of the simplest of the damping models, among those tested in previous perception studies, was based on the assumption of at-best-limited capabilities of the auditory system in the discrimination between a linear and quadratic damping model (cf. McAdams et al., 2004; Roussarie, 1999) and between these and more complex models (cf. Lambourg et al., 2001). Such an assumption was in part supported by the above-highlighted relevance of a $tan\phi$-related acoustical measure in explaining performance with stimuli generated according to the three-components damping model by Lambourg et al. (2001), whose perceptual relevance was tested by Roussarie (1999).

3.2.5 Interaction properties

All studies that investigated vibrating plates and bars have adopted the *hammer - sounding object paradigm* according to which signals were generated by a highly damped object - the hammer - striking a vibrating object - the sounding object (i.e., the bar or plate). As a result of the high damping of the hammer, it's vibrations are so negligible that the acoustical energy they generate can be assumed as perceptually irrelevant. This section focuses on the acoustically relevant parameters of the hammer-sounding object interaction.

Vibrating systems theoretically have an infinite number of vibrational modes. Potentially, for each of them a spectral component should be found in the generated signal. However, depending on the position where the hammer stroke is applied, only some of the vibrational modes are excited. In particular, those modes that have a node in the position of the stroke are excited minimally, while those that have an antinode in the position of the stroke are excited maximally. Figure 3.6 shows the nodal lines for the lowest three vibrational modes of a square isotropic plate. The dots mark the striking positions for the selective maximal excitation of each of them. Different striking positions are also more or less efficient in exciting torsional, transversal, or longitudinal modes of vibration. For example, if a bar is struck on a surface perpendicular to its length, longitudinal modes are maximally excited (c.f. Houix, 2003).

A hammer stroke introduces kinetic energy into the vibrating system, the sounding object. The higher the amount of kinetic energy introduced into the system, the larger the amplitude of oscillation and the louder the resulting sound. In general, the higher the striking force, or, conversely, the higher the mass and acceleration of the hammer, the higher the kinetic energy introduced in the vibrating system and the louder the resulting

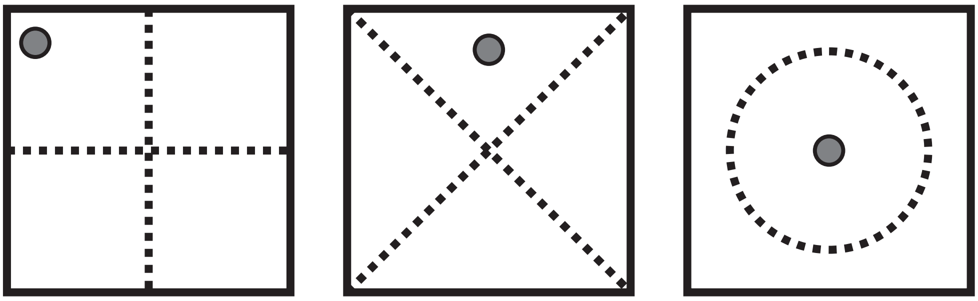

Figure 3.6: Nodal lines for the lowest three vibrational modes of a square isotropic plate. Gray circles show the striking position for the selective, maximal excitation of each of them.

sound. Similar considerations were made by Grassi and Burro (2003) and Grassi (2005) to motivate the use of signal power as acoustical the acoustical correlate of hammer mass.

During the stroke, the hammer and sounding object remain in contact for a finite time τ. The contact duration of τ has a direct influence on the vibration of the sounding object as those vibrational modes whose period is longer than τ will not be excited efficiently (Benade, 1979). Thus for increasing τ the resulting sound will be characterized by decreasing high frequency energy. Also, a decrease in τ will determine a decrease in *spectral centroid*, or spectral center of gravity, defined as the amplitude weighted average of the frequencies of the spectral components (see Marozeau, de Cheveigné, McAdams, & Winsberg, 2003, for a brief discussion of the different definitions of spectral centroid).

Chaigne and Doutaut (1997), following the work by Landau and Lifshitz (1981), derived Equation 3.8, relating contact time τ and the properties of the hammer, the sounding object, and to the initial velocity of the hammer (striking velocity). A basic assumption is that the hammer strikes a flat sounding object, i.e. with infinite radius of curvature. This assumption is valid for almost all the impact sound sources investigated in perceptual research.

$$\tau = 3.2181 \left(\frac{\mu^2}{K^2 V_0} \right)^{1/5} \tag{3.8}$$

$$\mu = \frac{m_h m_S}{m_h + m_S} \tag{3.9}$$

$$K = \sqrt{R_h} \frac{4}{3} \left(\frac{1-\nu_h^2}{E_h} + \frac{1-\nu_S^2}{E_S} \right)^{-1} \tag{3.10}$$

μ is the *reduced mass* of the hammer and the sounding object, a quantity introduced to simplify notation, given by Equation 3.9; m_h is the *dynamic mass* of the mallet, given by the ratio of the impact force to the resulting acceleration; m_S is the mass of the sounding object. In practice, for all impact sounds studied in previous works, m_h should be slightly higher than the static mass of the hammer. Also, $m_h \ll m_S$, so that to a first approximation μ, and consequently τ, are independent of m_S. K is the *force stiffness coefficient*, which, according to Hertz's law of contact, relates the striking force F to the compression δ of the mallet during the contact phase ($F = K\delta_h^{3/2}$ in the analysis developed by Chaigne and Doutaut (1997)). K is related to the Young modulus E, and Poisson's ratio ν of the hammer and sounding object, and the radius of curvature of the hammer R_h (Equation 3.10). V_0 is the initial velocity of the hammer at the beginning of the contact (striking velocity).

This analysis highlights the dependence of contact time on multiple properties of the hammer/sounding object source. From equations 3.8, 3.9, and 3.10, and the above-mentioned considerations, it follows that variations in τ characterize the sources investigated by Freed (1990), McAdams, Kudo, and Kirchner (1998), Grassi (2005), Grassi and Burro (2003), Giordano (2003), and Tucker and Brown (2003). Also, variations in the simulated force stiffness coefficient, associated with a change in contact time, have been investigated by Roussarie (1999). Among these studies, however, only Grassi and Burro (2003) and Grassi (2005) explicitly linked τ variations to acoustical properties of the generated signals, namely spectral centroid. It should be noted, however, that both Freed (1990) and Roussarie (1999) used this descriptor to explain their experimental results. Following these considerations, spectral centroid-related descriptors will be used to describe the signals investigated in Chapters 6, 7, and 8.

3.3 Experimental techniques for behavioral data collection

With a few exceptions, all experimental techniques used to investigate timbre perception (McAdams, 1993; Hajda, Kendall, Carterette, & Harshberger, 1997) have also been adopted to study source perception. In this section, the most common among those used in the studies summarized in Chapter 4 are described.

With *identification*, listeners are required to assign a verbal label to a given sound. When the listener is not given a predefined list of labels, *free-identification* is used (e.g. Vanderveer, 1979). When the listener is given the list of possible labels to be used, *forced-choice identification* is used (e.g. Klatzky et al., 2000). Identification data can be represented in a confusion matrix, where the frequency, or proportion of times that a given sound has been assigned a given label is shown. These data can be modelled using linear regression techniques (e.g. Klatzky et al., 2000), or, preferably, log-linear and logistic regression models (e.g. Kunkler-Peck & Turvey, 2000; Giordano, 2003), in order to highlight the physical and/or acoustical determinants for the choice of a given label.

With *classification*, the listener is with presented the set of sounds and is asked to sort them into classes according to a given criterion (e.g., make classes of similar sounds). The listener might be asked to group stimuli in a predetermined number of classes or might not be given constraints concerning the number of classes (*free-classification*). Classification tasks have been used by Houix et al. (1999)/Houix (2003). With this research, the frequencies with which sounds were placed in the same category were transformed into measures of the distances among them (the higher the frequency, the lower the distance). This distance measure was then analyzed using clustering algorithms (Gordon, 1999; Arabie, Hubert, & De Soete, 1996).

With *unidimensional scaling*, or *rating*, participants are asked to rate the signals with respect to a given attribute (e.g., rod length, Carello et al., 1998). Participants' responses can be numerical format, or in analog format, as when they are asked to rate the stimuli by moving a slider along a continuous scale. These data are analyzed using ANOVA models or linear regression techniques.

With the *semantic differential* (Osgood, Suci, & Tannenbaum, 1957), the listener is asked to estimate the position of a given sound along multiple judgment scales defined by opposing bipolar adjectives (e.g. pleasant-unpleasant, Kidd & Watson, 2003). Data reduction techniques, namely factor analysis, are then applied to responses in order to reduce the set of scales to a limited number of salient factors used to judge the sound set. The factors extracted with this analysis are assumed to correspond to the basic dimensions used to judge the stimuli.

With *dissimilarity ratings*, listeners are presented with sounds in pairs, and are asked to rate the similarity or dissimilarity between them. This method presents the advantage of avoiding a priori assumptions concerning the nature of the perceptually relevant source properties and frees the methodology from linguistic issues. For example, free-identification can highlight relevant source properties only to the extent to which listeners possess an adequate vocabulary to describe them. Also, with unidimensional scaling, the perceptual relevance of physical parameters unknown to the listener (e.g., force stiffness coefficient) cannot be directly investigated. These data can be analyzed using ANOVA models (e.g., Marozeau et al., 2003) to test specific hypotheses concerning the determinants of dissimilarity ratings or can be related directly to acoustical or source properties using correlational or regressive techniques (cf. Iverson & Krumhansl, 1993). However, the statistical models of choice for the analysis of dissimilarity rating data are multidimensional scaling (MDS) models (Borg & Groenen, 1997; Cox & Cox, 1997). In general, MDS techniques map ratings of the dissimilarities between experimental stimuli to their distance within a geometrical structure, a space defined by a given number of dimensions. The dimensions of the space are then assumed to represent the perceptual dimensions attended to by listeners to perform the rating (cf. McAdams, 1993) and are interpreted with reference to known stimulus properties. The main MDS models used in timbre and source perception research are summarized in the following section.

3.3.1 Multidimensional scaling models

The classical MDS model (Torgerson, 1958; Gower, 1966), maps dissimilarities d_{ij} among stimuli i and j to Euclidean distances in an R-dimensional space:

$$d_{ij} = \left[\sum_{r=1}^{R} (x_{ir} - x_{jr})^2 \right]^{1/2} \tag{3.11}$$

where x_{ir} is the location of stimulus i along dimension r. Distances in this model are rotationally invariant. For this reason, in interpreting the configuration, the analyst is free to rotate the MDS solution until the location of the objects along the dimensions reflects known object properties. This ambiguity in the choice of the dimensions of the MDS model is eliminated in the INDSCAL, or weighted Euclidean, model (Carroll & Chang, 1970). With INDSCAL, psychologically meaningful dimensions are postulated to be weighted differently by the different subjects.

$$d_{ijn} = \left[\sum_{r=1}^{R} w_{nr} \left(x_{ir} - x_{jr} \right)^2 \right]^{1/2} \tag{3.12}$$

where w_{nr} is the weight of the r^{th} dimension for the n_{th} subject.

The classical model has been extended by Winsberg and Carroll (1989a), with the extended Euclidean model, EXSCAL, which characterizes stimuli with their location in a common Euclidean space, as well as with dimensions specific to each of them (specificities s):

$$d_{ij} = \left[\sum_{r=1}^{R} \left(x_{ir} - x_{jr} \right)^2 + s_i + s_j \right]^{1/2} \tag{3.13}$$

Specificities may be conceived as the "square of the perceptual strength of a feature possessed by the stimulus" (McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). The EXSCAL model have then been extended with the extended INDSCAL model by Winsberg and Carroll (1989b), where each of the n subjects weight differently both the dimensions common to all stimuli (w_{nr}), and the set of specificities (v_n):

$$d_{ijn} = \left[\sum_{r=1}^{R} w_{nr} \left(x_{ir} - x_{jr} \right)^2 + v_n \left(s_i + s_j \right) \right]^{1/2} \tag{3.14}$$

With the INDSCAL model the cost of the removal of the rotational invariance is the introduction of many additional parameters, one for each subject and for each dimension, which are rarely interpreted, and which often contribute marginally to the improvement of the ability of the model distances to explain the dissimilarity ratings. This problem was faced by Winsberg and De Soete (1993) with the CLASCAL model, where application of the latent-class approach allows a considerable reduction in the number of parameters needed to model interindividual differences. According to this model, each of the n subjects is assumed to belong to one of $T \ll n$ latent classes, where different latent classes weight the dimensions of the MDS space differently. Thus the CLASCAL model retains the advantage of the rotational invariance of the configuration, but avoids introducing a high number of additional parameters. The model distances are given by the following equation:

$$d_{ijt} = \left[\sum_{r=1}^{R} w_{tr} \left(x_{ir} - x_{jr} \right)^2 \right]^{1/2} \tag{3.15}$$

where each subject is characterized by the probability λ_n of belonging to each of the T latent classes, using post-hoc Bayesian techniques.

A final extension of the CLASCAL model (McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995), used in Chapter 8, allows the computation of both common and specific dimensions:

$$d_{ijt} = \left[\sum_{r=1}^{R} w_{tr} \left(x_{ir} - x_{jr}\right)^2 + v_t \left(s_i + s_j\right)\right]^{1/2} \quad (3.16)$$

It should be noted that this last model includes all those presented above as special cases.

Chapter 4

The empirical framework

In this chapter, several experimental results are summarized that are relevant to the understanding of sound source perception. Ideally these should include research that has focused on spatial location or on the movement of the source, as well as those studies that have focused on properties of static sources other than their location. Also included should be research on the perception of animate sources, as well as perception of inanimate sources, i.e., is, studies that have focused on speech or non-speech vocalizations and, if existent, on perception of animal vocalizations, as well as studies which investigated signals generated by the vibration of inanimate objects. In the following, research on animate sources, and on the spatial attributes of sources are not considered. The focus is, instead, on the perception of static inanimate sources.

Section 4.1 presents studies carried out with heterogeneous sound sets, focusing mainly on *environmental sounds*: all those sounds that are classified neither as speech nor as music (see Gygi, 2001, for a thorough discussion on the definition of the class of environmental sounds).

Inspired by the ecological approach, Gaver (1988, 1993b) developed a taxonomy of environmental sounds, which provides a useful framework for organizing already collected results, as well as for directing future research in this field. A hierarchical organization is drawn, which distinguishes first among the materials involved in sound generation, and then on the type of interaction. The higher partition distinguishes between sounds generated by interacting solid objects (*solid sounds* such as keyboard typing), liquids (*liquid sounds* such as pouring), and by direct introduction of a pressure variation in the atmosphere (*aerodynamic sounds* such as a bursting balloon). Within these categories several basic events are distinguished, which involve different types of interactions among materials. For example, basic solid events are impacts, rolling, deformation, and scraping, while basic liquid events are dripping, pouring, splashing, and rippling. A distinction is then made among events resulting from a combination of these basic categories. Bouncing, for example, results from a temporal patterning of impacts. Compound events are given by a temporal patterning of different basic events (bowling is given by a rolling followed by an impact). Hybrid events result from the interaction of materials belonging to different basic categories (rain is given by the interaction between liquids and solid objects).

The taxonomy delineated by Gaver (1988, 1993b) is used as a basis to organize the presentation of results collected with smaller sets of sources, given by the experimental manipulation of a limited number of properties, in Sections 4.2-4.6.

Another class of inanimate sound sources is that of musical instruments. As previously outlined in Section 2.1.5, Gaver (1988) considered the focus on musical instrument tones to determine a limited usefulness of research on timbre for the understanding of source perception. He motivated his position assuming that the "quasi-harmonic" nature of musical sounds, and the simplicity of their structure, makes them a bad vehicle for transmitting information on source properties. Also, Gaver (1988, 1993a) distinguished between two listening modalities: everyday listening, focused on the properties of the source, and musical listening, focused on the properties of the acoustical signal. A consequence of this definition is that when musical listening is involved the structure of the percept should present little or no relationship with the properties of the source. This should be the case for perception of musical instrument tones too. With Section 4.7, studies on the perception of musical instrument sounds, defining the field of timbre perception research, are reviewed. In particular, a substantial part of previously published data is analyzed in order to test for the correctness of the position adopted by Gaver (1988, 1993a), concerning the limited usefulness of musical sounds for the understanding of source perception, and the absence of relevance of source properties to musical listening.

4.1 Sets of heterogeneous environmental sounds

4.1.1 Identification studies

Vanderveer (1979) carried out several investigations on heterogeneous sets of everyday sounds. A first free-identification experiment was carried out with adults, who were presented mainly solid sounds, but also some non-speech human sounds (e.g., coughing), and were asked simply to describe what they heard. Descriptions addressed mainly the sound source, in terms of actions assumed to be specified by temporal properties of the sounds, and in terms of objects assumed to be specified by spectral properties of the sounds. Descriptions based on auditory properties were much less frequent, and seemed to be used only in case of uncertainty with respect to the sound source. The most mentioned source properties were, from the most to the least frequent, the kind of action, the instrument used to execute the action, and the recipient of the action. Identification performance was particularly high with non-speech human sounds. Errors seemed to be based mainly on the temporal properties of the signals (e.g., fingernail scratching on a book was confused with writing, knocking with hammering). With impact sounds, more or less detailed information on material could be extracted, where wood and metallic objects were consistently identified, although with signals characterized by a particularly fast decay (e.g., dropped book), material identification was seldom possible. Impaired performance with impact sounds was explained in terms of the absence of source-specific acoustical information. Free-identification was investigated also with children (age 4-5 years). As with adults, descriptions focused on the sound source. Worse recognition performance than for the adults was observed. It was unclear whether this was due to an inability to attend to source-specific acoustical information or to limitations in the available vocabulary. Three additional experiments, conducted with adults, investigated the similarity of environmental sounds. Two experiments, conducted on different stimulus sets, used the free-classification technique, and one experiment was based on dissimilarity ratings. The main determinant of similarity seemed to be the temporal patterning of the sounds, a result that explains why confusions in the free-identification experiments were based mainly on the temporal properties of the sounds. In other cases classification and similarity appeared to be based on the materials involved (e.g., paper sounds were grouped together by most listeners).

Lass, Eastham, Parrish, Schebrick, and Ralph (1982) investigated free-identification of animal, inanimate, and musical sounds, as well as speech. High recognition rates were found. In general, animal sounds were the least correctly identified, followed by inanimate, musical, and speech sounds. Careful examination of the data reveals that differences among classes are caused by low recognition for isolated signals. In particular, the lower performance with animal sounds was due to the inability of listeners to correctly identify sheep and pig sounds, result probably related to the reduced experience of participants with countryside acoustical environments.

Gaver (1988) studied free-identification of liquid and solid sounds, such as paper sounds, impact sounds, walking sounds and machine sounds (electric razor). Participants were asked to describe what they heard. Consistently with the results ofVanderveer (1979), descriptions focused on the sound source. In general, the material of the interacting objects seemed one of the most easily recovered source properties. Solid sound sources were never recognized in liquid sounds, and vice versa. With impact sounds participants were also able to recover the size and shape of the interacting objects. Mutual constraints between materials and actions were observed: crumpling sounds were sometimes confused with multiple impacts, especially when the identified material was hard (metal can) rather than soft (paper), while when the material was identified as paper, it was more likely that crumpling, rather than multiple impacts were described.

Ballas and Mullins (1991) investigated the effects of context on the identification of everyday sounds using a variety of different conditions. Pairs of “nearly homonymous” sounds were selected, generated by different sources (e.g., the sound of a fuse burning and that of food frying) and, despite being aurally discriminable, they were highly confused in the identification tasks. Context was operationalized either by means of linguistic labels presented along with the sounds, or by means of sequences of sounds belonging to a given acoustical environment (e.g., kitchen sounds). Context could be either absent, consistent, or, when related to the other sound of the nearly homonymous pair, inconsistent. Free and forced-choice identification were used, where in this latter case participants had to choose between the linguistic labels for the two nearly homonymous sounds. Inconsistent context impaired identification, biasing listeners to identify a source consistent with the given context. Consistent context, however, did not enhance identification, leading to the same performance as when the sound was presented without contextual information.

Ballas (1993) studied the cognitive, perceptual, and acoustical determinants for the identification of environmental sounds. A first free-identification experiment was carried out. Identification responses were used to compute a measure of the uncertainty of participants concerning the source of the sound (Hcu), the higher the value of Hcu the worse the ability of participants to identify the source. Hcu explained identification times better than a multiple regression predictor based on acoustical descriptors. In particular, the higher the Hcu value, the larger the identification time. A second experiment investigated the relationship between identification and frequency of occurrence of sounds in everyday conditions (ecological frequency, Ef). Participants were asked to describe the cause of the sounds they heard at various times of the day, and on various week days. Sounds with higher Hcu were encountered less frequently during everyday life (had lower Ef), and lower identification characterized sounds with higher Ef. A linear combination of spectral, temporal, and signal envelope descriptors, along with Ef explained a high portion of the variance in identification. In a third experiment, participants were asked to rate sounds along several scales, describing their auditory properties, several cognitive factors antecedent to identification (e.g. familiarity with the sound), and several cognitive factors related to identification (e.g., "how easy is it to describe in words the action that generated this sound?"). Three factors were extracted from the ratings. The first factor, strongly correlated with identification and Hcu, measured sound's ease of identification. The second factor measured a sound's timbre (e.g., dull vs. sharp), and the third a sound's oddity (e.g., number of events that would produce the sound). The sounds' loadings on these factors were analyzed with a clustering procedure. Four groups were extracted. The first contained mainly water sounds, the second signals, the third door and engine sounds, the fourth sounds made up of multiple transients. Two more experiments investigated listeners' performance in accepting or rejecting a given cause for a sound. Results indicated high probability causes to be accepted faster. Also the typicality of sounds with respect to their causation was investigated. Results indicated typical sounds to be associated with faster decision times in cause rejection than non-typical sounds.

Gygi, Kidd, and Watson (2004) (see also Gygi, 2001) investigated the acoustical determinants for the forced-choice identification of vast sets of environmental sounds, including nonverbal human sounds (e.g., infant cries), or animal vocalizations. A first experiment investigated the effects of high- and low-pass filtering, varying the cutoff frequency of the filters. Listeners received training with unfiltered sounds. Identification performance was found to be particularly high (60% correct at worst) even in the most extreme filtering conditions. The crossover point, defined as the filter cutoff frequency for equal recognition performance with high and low-pass was measured. This measure is assumed to give an estimate of the redundancy of information for identification, where redundancy increases with decreasing crossover frequency. An estimate of around 1300 Hz was derived from the data, a frequency close to the lower estimates for speech, which highlighted the high redundancy of acoustical information in environmental sounds. A second experiment investigated the effects of band-pass filtering. Trained listeners were used. Performance decreased drastically for filter center frequencies above 6800 Hz, and below 850 Hz, while high performance was observed for intermediate center frequencies. A second group of experiments tested identification in signals that kept the temporal information of the original signals, and little or no spectral information. Both trained and naive listeners were tested. Temporal information was found to be sufficient for above-chance-level correct performance: the higher the amount of preserved spectral information, the higher the identification performance. Not surprisingly, training led to significant improvements in performance. Finally, a large set of acoustical features was used to explain percent correct identification data for this last set of experiments. Several descriptors (from 5 to 9 across datasets) were needed to account for limited amounts of variance in the original data. Among the datasets, one of the best predictors was the number of peaks in the autocorrelation function, a measure of signal periodicity. In particular, the higher the number of the autocorrelation peaks, the better the identification.

Across studies good source identification abilities were observed. In free-identification studies by Vanderveer (1979), and Gaver (1988), participants described sounds in terms of sources and did not use terms referring to perceptual qualities such as pitch or loudness. Care should be taken in interpreting this result, as it might reflect simply a higher availability of source-related terms, rather than the fact that we hear sources and not sounds, in line with the assumptions of the ecological approach. On the basis of studies by Vanderveer (1979) and Gaver (1988), it can be concluded that the materials involved in the sound generation process, as well as the properties of the interaction, were the most easily identified source properties. Concerning the acoustical determinants for identification, temporal factors were found to be highly relevant both in studies by Vanderveer (1979) and by Gygi et al. (2004), where in this latter case the amplitude envelope was sufficient for above-chance identification. Also, Gygi et al. (2004) demonstrated that acoustical information for sound source identification is highly redundant. Identification was found to be influenced by the level of experience with the sounds. Sounds experienced particularly frequently, as with non-speech human sounds in Vanderveer (1979), were particularly well identified, while the sheep and pig sounds investigated by Lass et al. (1982) were frequently misidentified, most likely because of the low degree of experience participants had with them. The relationship between the likelihood of encountering sound sources in everyday conditions and identification performance was indeed proved empirically by Ballas (1993), highly frequent sources being better identified than less likely sources, and vice versa. The relevance of previous experience was further highlighted by Gygi et al. (2004), who showed training to yield higher identification performance with sounds where the amount of acoustical information was diminished. Finally, results by Ballas and Mullins (1991) and Ballas (1993) highlight the relevance of cognitive factors in identification. Particularly interesting are the context effects reported in Ballas and Mullins (1991), which, inconsistently with the assumptions of the ecological approach, reveal that source identification can be influenced by information not contained in the proximal stimulus.

4.1.2 Semantic differential studies

Björk (1985) investigated a small set of environmental sounds, also comprising animals and human vocalizations. Although sounds were selected in order to investigate a set of signals more representative of our natural acoustical environment, these were presented in the highly artificial condition of reversed playback, i.e., recorded signals were played backwards. Five main factors were extracted from the rating data. The first, correlated with signal roughness, represented an evaluation dimension (pleasant vs. unpleasant). The second factor, correlated with sharpness and pitch, was interpreted in terms of activity (sharp vs. dull). The third factor, correlated with signal loudness, was interpreted in terms of potency (powerful vs. weak). The fourth factor was interpreted as reflecting auditory complexity. The fifth factor was related to the fast-slow rating scale.

Ohta, Kuwano, and Namba (1999) studied a set comprising only impulsive sounds, including sport sounds, explosion and construction sounds, and musical sounds. Four factors were extracted from the rating data. The first was interpreted as related to power (clamorous vs. quiet), and separated musical sounds (quiet) from construction and explosion sounds (clamorous). Clamorous sounds were characterized by a higher overall loudness and by broader spectral peaks. The second factor was interpreted as related to reverberation (diffused vs. not diffused) and was explained in terms of level decay time. Sounds judged as reverberant, mostly musical sounds, were characterized by a slower decay. The third factor was interpreted as related to metallicness (metallic vs. deep), and distinguished between sounds generated by striking metallic objects and sounds generated striking wooden objects, along with explosion sounds. Metallic sounds were characterized by higher maximum sharpness values and by lower loudness. Interestingly the hard vs. soft judgment scale had a high loading on this factor, where sounds judged as hard had the same acoustical properties as those judged as metallic. The fourth factor was interpreted as related to pleasantness (pleasant vs. unpleasant). Judgment scales with a high loading on this factor (pleasant vs. unpleasant; refreshing vs. not refreshing; beautiful vs. ugly) correlated strongly with the sharpness of the spectral peak. This acoustical measure was interpreted in terms of pitch clarity, more definite pitches being associated with sharper spectral peaks, where pleasant sounds had a sharper peak.

Kidd and Watson (2003) investigated a vast set of environmental sounds, intended to cover the range of sounds encountered every day (ambiances, appliances, bouncing, cars, doors, scraping, water, and wind sounds). Four factors were extracted from the rating data. The first factor, harshness, was related to pleasantness (breaking sounds were the least preferred, water sounds the most preferred). The second factor coded for sound complexity, single impacts (e.g., doors closing) being the least complex sounds, ambience sounds the most complex (e.g., casino sounds). The third factor coded for appeal (attractiveness), and quality, water sounds having the highest appeal, scraping sounds the lowest. The fourth factor coded for object size or signals loudness, explosion sounds having the highest scores for this factor and sounds generated with small objects (e.g., bouncing ball) the lowest. Several acoustical descriptors were used to explain the factor scores. Harsh sounds were associated with a concentration of energy in the higher frequencies, complex sounds had a lower percentage of within-event silences, appealing sounds were characterized by high pitches, and a higher spectral variation, and “big” or loud sounds by a higher total energy, and by a concentration of energy in the lower frequencies. Interestingly, if separated groups of sounds were considered, the acoustical explanations for the factors differed. For example, while for the impact class "big" sounds were characterized by a higher total energy, the same judgment in the wind class was given for signals characterized by a higher temporal variability of energy. Also, while the most preferred water sounds were characterized by a lower energy, and by a higher percentage of audible silences, the most preferred wind sounds were less repetitive, and had a higher concentration of energy in the lower portions of the spectrum.

In summary, all of these studies found a loudness-related dimension of judgment, interpreted as potency/power in Björk (1985) and Ohta et al. (1999), and as object size in Kidd and Watson (2003). Also, all of the researches highlighted a pleasantness-related dimension, explained by roughness in Björk (1985), pitch salience in Ohta et al. (1999), and presence of high-frequency energy in Kidd and Watson (2003). Finally, both Björk (1985) and Kidd and Watson (2003) found a dimension related to sound complexity. The variability in the acoustical explanation of similar dimensions of judgment across studies might be related to variations in the considered acoustical descriptors, but also, and more interestingly, to the variability within the investigated sound sets. Indeed, Kidd and Watson (2003) suggested that the same judgment is based on different acoustical properties depending on the nature of the stimulus set. This latter result strongly motivates source perception research carried with sets of relatively homogeneous sounds.

4.2 Isolated impact sounds

All studies conducted on isolated impact sounds have made use of the hammer-sounding object paradigm for signal generation (see Section 3.2.5). These can be grouped according to the gross property under investigation. Thus, a main distinction can be drawn between the studies on the perception of geometric properties of the sounding object, and those on the perception of material properties of either the sounding object or the hammer. In this section, studies based on direct judgments on source properties are presented separately from studies based on dissimilarity ratings.

4.2.1 Perception of geometrical properties

Lakatos et al. (1997) studied discrimination of shape (height/width ratio) in steel and wood bars. Training on the acoustical effects of shape variation was given prior to the experimental phase, allowing participants to strike different bars not used in the testing phase. Sounds were presented pairwise along with two pairs of figures, representing the two bar shapes. Participants had to indicate which of the two pairs of figures corresponded to the presented pair of sounds (see Figure 4.1). No feedback on response correctness was given. Stimuli were equalized in loudness. Steel and wood bars were tested in separate sessions. A cutoff 75% correct criterion across all stimuli was not reached only by 8.3% of the participants with steel bars and by 16.6% of the participants with wood bars. Their data were not further analyzed. Percent correct scores were converted to a dissimilarity measure and were analyzed with an MDS algorithm. Two and one-dimensional solutions

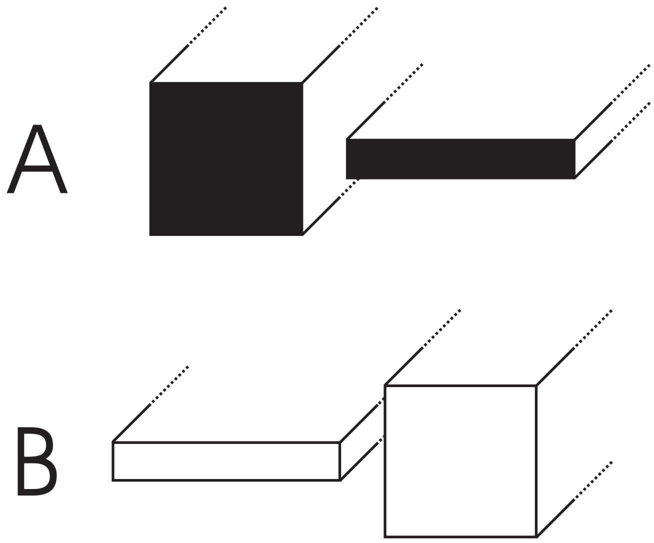

Figure 4.1: Experimental method used by Lakatos et al. (1997) to investigate perception of the shape of struck bars. Participants were presented consecutively two sounds, generated striking bars with a different heigth/width ratio. Participants had to indicate which of two pairs of figures (A, B) represented correctly the order of the bar shapes for the pair of sounds.

were extracted, respectively, for the steel, and wood bars. In both cases one dimension correlated strongly with the height/width ratio of the bars, with the ratio of the frequency of lowest transverse modes related to height and width, and with the frequency of the lowest torsional mode. The second dimension of the steel-bar space correlated with spectral centroid. Interestingly, a cluster analysis conducted on steel-bar data revealed a distinction between plate-like and block-like bars that depended on both dimensions.

Houix et al. (1999), Houix (2003) expanded the shape perception study by Lakatos et al. (1997) using steel bars with variable height/width ratio, struck at different locations. A first experiment used a classification task. Participants were asked to group sounds either 1) on the basis of their similarity, 2) on the basis of the shape of the struck bar, or 3) to create one group per bar, each containing the 6 sounds generated by striking them at 6 different locations. No constraints on the number of groups were imposed in conditions (1) and (2). In spite of the fact that acoustical information for shape was available, partitions were found to reflect only in part the bars' shapes, and groupings were explained by the most salient pitch evoked by the sounds, related to the transverse flexural modes of vibration along the height or width. A second experiment investigated the relative contribution of flexural and torsional vibrational modes to shape discrimination. The same procedure and the same steel bar signals used by Lakatos et al. (1997) were adopted. In three additional conditions, synthetic signals were investigated, extracting from the original sounds either the flexural and torsional mode components, or the flexural components only, or the torsional components only. Proportion correct scores were analyzed with MDS and cluster algorithms. Consistently with Lakatos et al. (1997), in all conditions plate-like bars were partitioned from block-like bars. Two-dimensional MDS spaces were extracted for all conditions. Dimension 1 in all spaces, excepting the torsional mode condition, was related to the height of the bars. For the torsional mode condition, the first dimension was explained by the height/width ratio of the bars. Results were explained in terms of pitch and amplitude decay.

Lutfi (2001) studied discrimination of hollowness in synthetic struck bar signals. Participants were asked to indicate which of two presented sounds was generated striking a hollow (vs. solid) bar. Length and material of the bars (wood, iron, aluminum) were varied across pairs. Feedback was given concerning response correctness. Listeners were found to adopt two different strategies, one based on frequency and decay time, which allowed optimal discrimination between hollow and solid bars, the other based on frequency alone. The optimal decision strategy was shown to yield at best a small advantage over the decision rule based on frequency alone.

Kunkler-Peck and Turvey (2000) studied perception of the shape of struck plates. In a first experiment steel plates with constant area and variable height/width ratio were investigated. Participants estimated their height and width by mechanically adjusting the position of two response bars. Estimates were found to be ordered according to the actual height and width of the plate, and in the approximate range of the actual values. An underestimation bias was observed however. Performance was explained in terms of the corresponding modal frequencies of the plates. A second experiment tested the effect of plate material (steel, wood, and plexiglas) on height/width estimation, using the same geometrical properties and procedure as for the first experiment. Height and width estimates decreased as material changed from steel to wood to plexiglas and were well accounted for by the modal frequencies of the struck plates, which decreased accordingly. In a third experiment, participants identified the shape of circular, triangular, and rectangular steel plates. Identification performance was found to be significantly above chance. In the last experiment, participants identified both shape (circle, triangle, rectangle) and material (steel, wood, plexiglas). Again shape identification performance was significantly above chance. Material identification was almost perfect. A secondary interaction between shape and material was found, steel being associated with triangular shapes, wood with circular shapes, plexiglas with rectangular shapes.

Gaver (1988) studied length estimation with iron and wood bars. Participants rated the length of the bars. They were informed of the material variation and, before the beginning of the experimental phase, were shown the correct response for the longest and shortest bars. Estimates pooled across participants were scaled with actual length and were nearly independent of material. However, much interindividual variability was observed. Only a few participants showed estimates independent of material, while part of them judged wood bars to be shorter than iron bars, and part of them made the opposite judgment. Analogous results were obtained with synthetic struck bar sounds. It should be noted that lower length estimates for wood, rather than iron bars, are highly inconsistent with Kunkler-Peck and Turvey's (2000) results, where plates were judged as larger and wider when their material was wood, rather than steel. A last experiment studied the effect of training. Participants were trained to recognize length in real bar sounds, being afterwards tested with synthetic sounds. A substantial improvement in size estimation abilities was observed. Interestingly enough, despite the extensive training received, participants showed a material effect, this time in agreement with results by Kunkler-Peck and Turvey (2000). Indeed, synthetic metal bars were judged as significantly shorter than wood bars. It was concluded that the impairment in the absence of training was not caused by the absence of acoustical information for length, but by the fact that length judgments are an unfamiliar task to everyday listeners. This conclusion, however, contrasts with the abilities of untrained listeners reported by Kunkler-Peck and Turvey (2000), and Carello et al. (1998) (see below).

Tucker and Brown (2003) investigated shape and size perception in wood, plastic, and aluminum plates struck in open air or underwater. In a first experiment, participants were asked to identify the shape of constant-area square, circular and triangular plates. In contrast with results by Kunkler-Peck and Turvey (2000), highly impaired performance was observed for both the open-air and underwater conditions. Consistently with results by Kunkler-Peck and Turvey (2000), significant shape-material associations were observed in the open-air condition, square shapes were more frequently identified in plastic sounds, circles in wood sounds. In a second experiment, size estimation in struck square-plate sounds of variable material and area (three levels) was investigated. Participants were presented pairs of signals generated by striking plates of the same material and were asked to estimate the size of the plates by adjusting the size of squares presented on a screen. Better performance was observed with open-air than with underwater recordings and with aluminum than wood and plastic plates. However, strong estimation biases were found: size estimation was better when the first sound of the pair was generated with a medium-size plate, while assimilation effects were found when the first sound of the pair was either a small or large plate, size estimates for the second sound being biased toward the size of the first sound. Impaired performance in this last experiment cannot be explained in terms of an absence of acoustical information for the task, as the difference in signal frequency between small- and medium-sized plate sounds of the same material, and between medium and large plates, should have been of the order of two octaves.

Lakatos et al. (1997), Kunkler-Peck and Turvey (2000), Lutfi (2001), one experiment in Houix (2003), and, to a limited extent, Gaver (1988) demonstrated perception of geometrical properties to map veridically the actual geometry of the sounding objects. Variability on veridicality of perception is, however, found across studies. Explaining these differences might involve differences in the task, in the investigated stimuli, and on the level of expertise of participants in perception of geometrical properties, as modulated by pre-experimental training or by feedback on the correctness of responses.

An explanation based on the task seems likely when comparing inconsistencies between results by Lakatos et al. (1997) and by Houix et al. (1999) and Houix (2003). Indeed, the accurate shape perception results obtained by Lakatos et al. (1997) were replicated by Houix et al. (1999) and Houix (2003) using the same discrimination task and the same stimuli, but were not replicated with a classification experiment carried out with similar stimuli. Differences between the additional results by Houix (2003) collected with synthetic stimuli and those by Lakatos et al. (1997) might, instead, be explained by differences among stimulus sets, where the synthetic stimuli used by Houix (2003) might have provided listeners with impoverished acoustical information for shape. An explanation based on differences among stimuli might also explain the differences between the accurate shape identification results by Kunkler-Peck and Turvey (2000) and the impaired performance reported by Tucker and Brown (2003). Kunkler-Peck and Turvey (2000) generated stimuli live and struck plates repeatedly before the response was given. Given the plausible acoustical differences among repetitions, extraction of shape-related invariants might have been facilitated in the latter case. On the contrary, Tucker and Brown (2003) investigated recorded stimuli, where for each shape only one signal was given, so that impoverished shape information might have been available to the listeners. Concerning size perception, the good performances observed by Kunkler-Peck and Turvey (2000) in auditory estimation of the length and width of struck plates is consistent with results of Gaver (1988) on estimation of bar length, but not with those by Tucker and Brown (2003). It should be noted, however, that Gaver (1988) trained listeners, showing them the correct responses for the longest and shortest bar sounds, so that good performance is not surprising. An explanation of differences among results of Kunkler-Peck and Turvey (2000) and those of Tucker and Brown (2003), instead, might be similar to that formulated above concerning shape identification, but might also involve differences in the task. Choosing among these alternatives is not possible with the available information.

More consistent across studies, instead, is the effect of material on geometry perception. With size perception experiments, both Kunkler-Peck and Turvey (2000), and, with the majority of participants, Gaver (1988) found wood objects to yield higher size estimates than metallic objects. With shape perception experiments, both Kunkler-Peck and Turvey (2000) and Tucker and Brown (2003) found circular shapes to be recognized more often in wood sounds and rectangular shapes in plexiglas sounds. These results show that geometry perception is influenced also by source properties other than geometry itself. Geometry perception performance has been associated with measures of the frequencies of the vibrational modes and of signals frequency (Lakatos et al., 1997; Lutfi & Oh, 1997; Kunkler-Peck & Turvey, 2000; Houix, 2003). Given the influence of density and elasticity on the frequency of vibrational modes (see Section 3.2.3), effects of material on geometry perception do not appear surprising. It remains however unclear which, among the material-related properties that influence modal frequencies (density, elastic properties), is responsible for the material-geometry perceptual coupling, or, alternatively, whether both should be taken into account.

Lakatos et al. (1997) and Tucker and Brown (2003) found geometry perception performance to be impaired with wood, rather than metal sounding objects. These effects might be explained with another material-related property: damping. Wood vibrations are generally more damped than with metals, so that the information attended to by listeners to judge geometry is impoverished, i.e. the vibrations last for shorter times. This would explain also why Tucker and Brown (2003) found impaired performance with underwater sounds as compared with performance with open-air recordings where plates vibrations were more damped when the surrounding medium was water rather than air. Thus a distinction can be drawn among the perceptual effects of the different material related properties. While density and/or elasticity modulate the acoustical information used to perceive the geometry, damping modulates its salience, affecting the ability of the auditory system to detect such information. A psychophysical metaphor might be used to explain further these effects. Given a certain acoustical signal, perception of geometrical properties, say size, is not fixed over presentations and might be modeled with a normal distribution with a given average value and with a given spread or standard deviation. Therefore it might be hypothesized that while a variation in density and/or elasticity, along with size variations, affect the average value of the perceived size distribution, damping variations affect the spread of the distribution. In particular increases in damping are associated with increases in the spread of the distribution of perceptual effects and with a higher probability of observing impaired performance.

Finally it is interesting to note that both Lakatos et al. (1997), Houix et al. (1999) and Houix (2003) found a perceptual partitioning of bars in plate-like (i.e. thin bars) and block-like (i.e., thick) objects. This result points toward the perceptual relevance of a thickness-related dimension, which remains to be investigated.

4.2.2 Material perception

Freed (1990) studied hammer hardness scaling, with sounds generated by striking variable size cooking pans with mallets of variable hardness (felt covered rubber, felt, cloth covered wood, rubber, wood, and metal). Sounds were equalized in loudness. Some training was given, demonstrating to participants the sounds of the metal, and felt mallets striking two pans of different sizes. Ratings were found to be scaled with actual hardness and independent of pan size. A preliminary experiment showed the first 300 ms of the signals to be critical to hardness estimation. Consequently acoustical predictors were extracted from this portion of the signal: average spectral level, spectral level slope (i.e., rate of change in spectral level), average spectral centroid, and spectral centroid time-weighted average. Average spectral centroid was the best predictor. Altogether predictors accounted for a high portion of variance in the ratings.

McAdams et al. (1998) studied the perceptual independence of hammer and sounding object properties with stimuli generated by striking hard and soft wood xylophone bars and large and small tympani membranes with medium and soft rubber mallets. Pitch was kept constant. The Garner paradigm was used (Garner, 1974), participants being asked to identify the hammer or instrument used to generate the sound. Feedback on response correctness was given. Both speeded, and unspeeded classifications, were investigated. Perceptual interaction was found in the speeded condition, perceptual independence in the unspeeded condition, indicating an interaction among the properties of hammer and sounding object only at an early stage of processing. This results might point toward the ability of listeners to tell apart hammer and sounding object hardness in normal listening condition, to the limit where these are characterized by temporally unconstrained judgments.

Lutfi and Oh (1997) studied the discrimination of material with synthetic struck bar sounds. Signals were synthesized so as to perturb the elasticity and density of the simulated bars around values typical for iron, silver, steel, copper, glass, crystal, quartz, and aluminum. Perturbations were applied either to all the frequency components or to each component independently. Participants were asked to tell which of two stimuli was generated by striking an iron bar, silver, steel, and copper being the alternatives, or with a glass bar, the alternatives being crystal, quartz, and aluminum. Feedback was given. Far from ideal performance was observed, because of the strong weight given to frequency; amplitude and decay time played only a secondary role.

Gaver (1988) studied material identification with variable length iron and wood bars. Participants were informed of the length variation and, before the experiment, were shown the correct responses for the shortest and longest iron and wood bars. Not surprisingly performance was almost perfect. Similar results were obtained with synthetic struck bar signals.

Klatzky et al. (2000) investigated material identification with synthetic struck bar sounds, generated by varying a damping parameter related to $tan\phi$ (see Section 3.2.4) which modeled the bars material, and the length of the bars. The length manipulation was associated, within the experimental set, with a frequency variation of 3.3 octaves. Four response categories were used: rubber, wood, steel, and glass. Both experimental variables influenced participants' responses, glass and wood being associated with higher frequencies than rubber and steel, glass and steel being associated with lower damping values than wood and rubber.

Roussarie (1999) investigated material identification with synthetic struck plate sounds. Damping coefficients, elastic properties, and the density of the simulated plates were varied around values typical of glass and aluminum. Hammer properties were manipulated as well, via the force stiffness coefficient (see Section 3.2.5), using parameters typical of either wood or rubber. Stimuli were equalized in loudness. Two response categories were adopted: glass, and aluminum. Responses were found to be influenced only by the damping properties of the plates, strongly correlated with an acoustical parameter analogous to $tan\phi$, and with the average spectral center of gravity. Variations in density and elasticity, associated with a frequency variation of a musical interval of one perfect fifth, had no effect.

Avanzini and Rocchesso (2001a) (see also Rocchesso, Ottaviani, Fontana, & Avanzini, 2003) investigated material identification with stimuli synthesized with a physical model employing a single vibrational mode. Both the frequency of the vibrational mode and damping were varied. Frequency, in particular, was varied over one octave. Four response categories were used: rubber, wood, glass, and steel. Steel and glass were associated with slower level decays than rubber and wood. Secondary frequency effects were found, rubber and glass being associated with higher pitches than wood and steel.

Giordano (2003) carried out three experiments on material identification using struck plate sounds. In all experiments both plate material (steel, glass, plexiglas, and wood) and size/area were varied. In the first and second experiments the shape (height/width ratio) of the plates was also varied. In the third experiment, the material of the hammer was varied (steel, glass, wood, and plastic). In the second experiment, plate vibration was externally damped with a low density plastic plate attached opposite to the struck surface. Plate shape and hammer material did not influence material identification. With the first and third experiments, perfect material identification was found with respect to gross categories (wood and plastic vs. glass and steel). Within the same gross category identification was based on plate size, steel and plastic being identified more often in larger plates than wood and glass. Similar effects were found with the second experiment, where external damping caused wood and plastic to be identified in glass sounds. Preliminary acoustical analyses explained identification with signal frequency and amplitude decay.

Tucker and Brown (2003) investigated material identification with plates of variable materials (aluminum, plastic, and wood plates), constant size, and different shapes (square, circle, and triangle). Plates were struck either in the open air or underwater. In this latter case vibrations were more damped, and signals duration was shorter. Aluminum was perfectly discriminated from wood and plastic, while wood and plastic were strongly confused each other, although to a lesser extent in the freely vibrating condition. Shape did not influence material identification. A damping measure related to $tan\phi$ was found to account for large portions of variance in identification data.

Results by Freed (1990) and McAdams et al. (1998) and of the hammer variation condition in Giordano (2003) point toward the perceptual independence of hammer and sounding object properties, i.e. the absence of effects of hammer properties on the perception of the properties of the sounding objects and vice versa. Given the relative absence of relevant acoustical energy emitted by the hammer which should characterize the hammer-sounding object paradigm (see Section 3.2.5), it might be concluded that the dependence of the resulting acoustical signal on the hammer properties is mediated by the properties of its interaction with the sounding object (e.g., contact time, see Section 3.2.5). As a consequence hammer perception should be based on the properties of its interaction with the sounding object, and not on hammer properties itself. Then, the perceptual independence of hammer and sounding object should be better conceived as due to the relevance of interaction properties for perception of the properties of the former, sounding object properties for the perception of the properties of the latter. Consistently Roussarie (1999) found identification of the material of the sounding object to be independent of variations in the force stiffness coefficient. The relevance of interaction properties to hammer perception and the perceptual independence of hammer and sounding object were studied in Chapter 7.

Gaver (1988), Kunkler-Peck and Turvey (2000), Giordano (2003), and Tucker and Brown (2003) found material identification to be almost perfect when discriminations among gross categories were involved (i.e. wood or plastic vs. glass or metal), and, limited to this discrimination, to be fairly independent of the geometry of the objects. Disagreement is however found concerning performance within these gross categories. While Giordano (2003) and Tucker and Brown (2003) found plastic materials to be confused with woods, Kunkler-Peck and Turvey (2000) found this discrimination to be perfect. As in the case of perception of the geometrical properties, this inconsistency might be explained with the fact that Kunkler-Peck and Turvey (2000) generated stimuli live, and repeatedly struck the plates before participants emitted this response. These conditions might have provided listeners with additional source-specific acoustical information. Disagreement is also found concerning the effects of geometry on the discriminations within the gross categories. While Giordano (2003) found size influence material identification, Kunkler-Peck and Turvey (2000) and Tucker and Brown (2003) found shape to be not influent. It should be noted, however, that the size manipulation investigated by Giordano (2003) might have caused more drastic changes in the acoustical structure than the shape variation by Kunkler-Peck and Turvey (2000), and Tucker and Brown (2003). Indeed, the shape variation investigated by Giordano (2003) did not significantly influence material identification either. An alternative explanation for the emergence of size effects in the study by Giordano (2003) is related to the available amount of acoustical information for the discrimination between grossly similar materials (e.g., metal and glass). Thus, it might be hypothesized that participants focused on source properties irrelevant to the task, plate size, because sufficient acoustical information for the discrimination between grossly similar materials was not available. This hypothesis was addressed in Chapter 6.

The relevance of geometry to material perception is symmetrical to the observed relevance of material to geometry perception highlighted in Section 4.2.1. In analogy with what was hypothesized for the effects of material on geometry perception, this interaction should arise from the fact that acoustical variables modulated by perceptually relevant material properties, attended to by listeners to judge the material, are influenced also by the geometry of the sounding object. The most probable acoustical candidate is the frequency of the spectral components in the generated signal, determined by the geometry of the sounding object, as well as by two material-related properties: density and elasticity. It should however be pointed out that not all studies found evidence in favor of the perceptual relevance of elasticity and density: while Roussarie (1999) found material identification not to be influenced by these properties, these variables were almost surely relevant in determining listeners responses in the study by Lutfi and Oh (1997).

Much more consistent across studies, instead, is the significance of damping to material identification: measures of this property are significantly correlated with results by Klatzky et al. (2000), Avanzini and Rocchesso (2001a), Roussarie (1999), and, when extracted from the acoustical signals, by Tucker and Brown (2003).

Concerning the acoustical basis for material perception, more complete evidence has been found for the perception of hammer properties than for sounding object properties. Perceived mallet hardness has been found to be related to the loudness and spectral centroid of the initial portion of the signals, as well as to loudness decay and spectral centroid temporal evolution (Freed, 1990). Although not explicitly tested, the absence of the size of the sounding object on hardness estimates suggests that this judgment is independent of signal frequency.

Concerning frequency, effects of this variable were observed by Klatzky et al. (2000), Avanzini and Rocchesso (2001a), but not by Roussarie (1999), most probably given the smaller range of variation of this variable in his stimulus set. Also, material perception was associated with damping measures, which, given their nature, can also be considered as acoustical descriptors. However, before concluding that acoustical measures of damping are perceptually relevant it is necessary to compare its explanatory power with that of other related descriptors, such as loudness decay or duration. Given the relevance of spectral centroid measures to material identification highlighted by Roussarie (1999), further studies should consider this acoustical variable as well. With the study presented in Chapter 6 a large set of acoustical descriptors were considered to investigate the acoustical correlates of material identification. In particular the explanatory power of acoustical measures of damping was contrasted with measures of the loudness decay, and of signal duration. Also, several spectral centroid-related descriptors were considered.

4.2.3 Dissimilarity rating studies

From the methodological point of view one of the studies reported in this section is a hybrid between the dissimilarity rating and direct source judgment techniques (Klatzky et al., 2000). This study is reported in this section as data analysis is based on MDS models.

Gaver (1988) investigated separately real and synthetic wood and iron bars of variable length. Two-dimensional solutions were derived in both cases. The first dimension partitioned categorically wood sounds from iron sounds. The second dimension mapped almost perfectly to bar length.

Klatzky et al. (2000) investigated synthetic struck bar sounds of variable damping and length, associated with variations in frequency. Participants were asked to rate either dissimilarity in the bar materials or length. Two-dimensional solutions were derived in both cases. High interindividual agreement was found for material dissimilarity, based on both damping and bar length. Much interindividual variability was found in the length dissimilarity rating task, where length was relevant for the near totality of listeners, while damping was relevant for only half of the participants. Instructions thus modulated the perceptual relevance of damping, much more relevant to material dissimilarity ratings than to length.

Roussarie (1999) investigated synthetic struck plates sounds. Both plate properties (damping, density, and elasticity with values around those for aluminum and glass), and hammer properties (wood/rubber, manipulated with an interaction parameter, the force stiffness coefficient, described in Section 3.2.5) were varied. Wood and rubber hammer sounds were rated in separate sessions. Two and three-dimensional solutions were found for the wood and rubber hammer datasets respectively. Results for the two datasets were highly similar. In both cases, the first dimension strongly correlated with the damping parameter, with an acoustical parameter analogous to $tan\phi$, and with signal spectral centroid. The second dimension was found to be related to the flexural and torsional wave velocities, related to plate elasticity and density, and with two pitch-related measures: the frequency of the spectral peak and that of the lowest resonant mode.

McAdams et al. (2004) investigated synthetic struck bars. Two experiments were carried out varying bar damping, and either bar density or length. A two-dimensional solution was derived from data for the first experiment. The first dimension was related to the damping parameter, and to the temporal decay of the spectral centroid. The second dimension was related to bar density and to frequency, varying over 9.5 semitones. The damping-related dimension had a greater perceptual weight than the frequency-related dimension. The second experiment tested whether an increase in the range of variation of frequency and a decrease in the range of the damping parameter could change their relative perceptual relevance. Two stimulus sets were investigated in the second experiment, frequency spanning over 15 semitones in both cases, one set being two octaves below the other. Two-dimensional solutions were derived for both datasets. In both cases, one of the dimensions was related to bar length and signal frequency, while the other dimension was related to bar damping and to level decay descriptors for the higher-pitched set or to spectral centroid related descriptors in the lower-pitched sound set. Significant interindividual differences were found in the weighting of the two dimensions, where part of the listeners weighted more heavily the pitch-related dimension than the damping related one, while another part showed the opposite weighting. Data for all experiments were explained in terms of signal frequency and of a linear combination of the average spectral center of gravity and of level decay in the last portion of the signal.

Across studies two orthogonal dimensions of judgment were highlighted. A first dimension was related either with the material of the sounding object (Gaver, 1988), or with damping measures (Roussarie, 1999; Klatzky et al., 2000; McAdams et al., 2004). In Section 4.2.1 it was noted that the relevance of damping to geometry perception was indirect and secondary, simply affecting the detectability of the perceptually relevant information for geometry. Consistently, when listeners where asked to rate length dissimilarity, damping measures were much less relevant than when they were asked to rate material dissimilarity (Klatzky et al., 2000). The second dimension was related to size (Gaver, 1988; Klatzky et al., 2000; McAdams et al., 2004) or density/elasticity (Roussarie, 1999; McAdams et al., 2004). Only Roussarie (1999) tested the relevance of interaction parameters to dissimilarity rating, finding no effect of this variable.

It should be noted that none of these studies contrasted the perceptual relevance of hammer and sounding object properties in determining dissimilarity ratings, varying both in the same experimental set. Also, scarce evidence is found concerning the relevance of the properties of the interaction between hammer and sounding object to dissimilarity ratings, investigated only by Roussarie (1999). Indeed, the particular experimental design used by Roussarie (1999) might have led listeners to focus judgment on the only source properties that were varying within the experimental set, i.e., not on the interaction parameter force stiffness coefficient, which was varied across stimulus sets, but not within the same set. In Chapter 8 the relative relevance of hammer, sounding object and interaction properties in determining dissimilarity ratings was investigated.

From the acoustical point of view, the first dimension of the spaces was also correlated with other variables than the damping descriptors, such as the overall spectral centroid (Roussarie, 1999; McAdams et al., 2004), spectral centroid decay rate (McAdams et al., 2004) or loudness decay (McAdams et al., 2004). The second dimension, instead, was invariably related to signal frequency (Klatzky et al., 2000; Roussarie, 1999; McAdams et al., 2004). In line with the reasonings on the acoustical determinants for material identification outlined in Section 4.2.2, the perceptual relevance of damping to dissimilarity ratings points toward the relevance of signal duration. However, no previous study attempted to explain experimental data on the basis of this simple acoustical property. This was done in the study presented in Chapter 8.

4.3 Sequences of impact sounds

Repp (1987) studied clapping sounds, investigating listeners' abilities to identify the gender of the clapper, as determined by the size of the hands, and hand configuration. In a first experiment, listeners were presented with sequences of claps produced by different clappers without constraints imposed on hand configuration. Listeners knew each of the clappers and were asked to identify them. Although poor performance was observed (13% correct), recognition was above chance level (5%). Interestingly, self-recognition was much better (46%), confirming the relevance of the level of experience with a source in determining performance. In this case, facilitation might have originated both from memory of the recording session, as well as from the more frequent exposure to one's own claps in everyday life. Performance in gender recognition was the worst, being at chance level. Identified gender was determined on the basis of clapping rate (slower for perceived males), amplitude (louder claps for perceived males), and spectral shape, roughly the concentration of energy on low rather than high portions of the spectrum (signals with more energy in the lower portion of the spectrum were identified as coming from males). None of these variables differentiated among actual gender of the clappers. It was concluded that listeners used acoustic stereotypes to judge gender, although these criteria might have reflected a regularity in the population, of which the stimulus set was not representative. In a second experiment, participants were asked to identify clapping hand configurations in two stimulus sets. The first was generated by the author, who clapped in a number of different configurations. The second set was that used in the first experiment. While good recognition capabilities were found for the first set, performance for the second set was much lower. Further, the same acoustical properties used to judge gender in the first experiment were relevant in the second experiment.

Tousman, Pastore, and Schwartz (1989) tested recognition of clapping hand configuration under more favorable experimental conditions. Hand configuration discrimination was tested both within the same clapper and between clappers. Also, a training condition with feedback on the correctness of the response was included. In agreement with the limited capabilities documented by Repp (1987), impaired performance was found, even in conditions of high training.

Li et al. (1991) studied auditory identification of the walkers' gender, explaining listeners' performance both in terms of source properties, namely anthropomorphic measures of the walkers, and in terms of statistical properties of the signal spectra. A first experiment found recognition performance well above chance, although identification performance varied across walkers. Two source properties explained the judgments: walker height and shoe size, "male" responses being more likely for taller walkers and larger shoe sizes. A second experiment examined explicitly the effect of shoe size. Walking sounds of three females were investigated, walkers being asked to wear either their own shoes or two different types of male shoes. Wearing male shoes, wider and longer than females shoes, caused an increased probability of female walkers to be identified as males, although the strength of the effect varied across walkers. Thus gender recognition is better conceived as shoe size recognition. Several temporal and spectral descriptors were investigated to explain identification data. Among the temporal descriptors, pace (the number of steps per second) was strongly associated with identified gender, females being consistently recognized in faster walkers. Interestingly this property did not differentiate between walkers actual genders. Spectral descriptors were reduced to two uncorrelated factors, which discriminated among actual genders. Roughly, female walkers were found to be more probably identified in spectra with a higher concentration of energy in the upper frequencies. Acoustical explanations for walker gender identification were validated in a subsequent experiment based on the manipulation of recorded signals.

Warren and Verbrugge (1984) investigated bouncing and breaking sounds. A first experiment tested identification with natural sounds, generated dropping glass objects on the floor, which either bounced or broke. Almost perfect performance was observed. Additional experiments tested the role of the temporal patterning of the signals, controlling for variations in spectral properties. Starting from the recordings of isolated impact sounds of pieces of glass falling on the floor, several sequences were created, repeating isolated impact sounds progressively closer in time. Such sequences were termed damped, quasi-periodic pulse trains and reproduced the temporal patterning of bouncing sounds, keeping constant spectral properties across bounces. Bouncing sequences were created by overlapping four synchronized damped, quasi-periodic pulse trains. Breaking sequences were synthesized by overlapping pulse trains asynchronously, with or without a 50-ms noise burst added at sequences onset to simulate the initial rupture of the object. Independently of the presence of the initial noise burst, extremely high auditory identification rates were found, thus confirming the role of temporal patterning in the discrimination between the two events.

Carello et al. (1998) investigated length estimation in wood rods dropped on a surface. Two experiments were carried out, the former using longer rods than the latter. Participants judged rod length by adjusting the position of a visible surface, as that they could reach it with a rod of the estimated length. Overall participants were able to estimate rod length well, metrical precision being higher in the first than in the second experiment. Signal duration, amplitude, and spectral centroid were used to explain performance. With the exception of amplitude for the second experiment, none of the acoustical descriptors explained performance better than the actual rod lengths. Thus perception was more tightly linked with properties of the distal stimulus than with those of the proximal stimulus. Care should be taken in interpreting this result as supporting the assumptions of direct perception, as the analysis of the acoustical event cannot be considered complete. For example, no attempt was made to quantify either the temporal structure of the events or the signal frequency. A rod dropped onto a surface generates a sequence of impacts as it bounces over the floor and eventually rolls on it at the end of the bouncing phase. It is highly probable that the temporal distance among subsequent bounces of the rod is influenced by its length. Also, the longer the rod, the lower the signal frequency should be. An almost perfect explanatory variable for performance was found instead in the inertia tensor, which quantifies the resistance of the object to being rotated in different directions. Since this variable had previously been found to explain length estimation with dynamic touch, an interesting link between audition and action was suggested.

Warren, Kim, and Husney (1987) studied elasticity perception in bouncing. Both auditory and visual information where investigated. A first experiment tested the ability to control the bounce of a variable elasticity ball. Participants were asked to throw the ball to the floor reaching a target height with the bounce. Before the experimental trials, participants were exposed to different types of information concerning the balls: auditory (participants heard the balls bouncing), visual (participants saw the balls bouncing), or haptic alone (participants touched the balls), all modalities together when given the opportunity to practice the task, or auditory plus visual. Good bounce control abilities were observed. In particular, performance in the audiovisual and auditory information conditions did not differ. A final experiment studied the auditory and visual information for elasticity estimation. The auditory displays consisted of a brief constant-amplitude 190-Hz tone reiterated for each impact of the bouncing ball on the floor, with decreasing inter onset intervals. Results showed auditory elasticity estimates to be based on the duration of the periods separating consecutive bounces. Although obtained using highly simplified auditory displays, these results show the use of a temporal property of the auditory event for the estimation of an object property: the elasticity of the ball.

Guski (2000) conducted two experiments on signals generated by dropping balls of different material and equal size, and thus different weight, from a variable height onto a drum. Participants were asked to estimate the mass of the balls, the drop height, and the physical work, defined as force to the participants. Impaired performance was found for drop height estimation, but not for mass and force estimation. Force estimates were equally well correlated with the temporal distance between the first and second impact, as well as with peak signal level. A second experiment tested the perceptual relevance of peak level, equalizing signals with respect to this acoustical parameter. Participants were again asked to estimate the mass of the balls, the drop height, and the force. Peak level equalization only minimally affected height and mass estimates, but strongly influenced force estimates. It was concluded that height and mass estimation was based on the temporal distance of the first to the second bounce.

Grassi (2005) (see also Grassi & Burro, 2003) investigated estimation of the size of balls from bouncing sounds. In three experiments, listeners were asked to estimate the size of a wooden ball dropped on a circular ceramic dish, adjusting the diameter of a circle presented on a screen. In the first experiment only one diameter only was used for the dish, whereas two were used in the second, and three in the third. Results showed that participants were able to scale properly the size of the balls. Ball size estimates were not independent of the size of the dishes, a decrease in the size of the dish being associated with a decrease in the perceived size of the ball. Several acoustical predictors were used to explain performance: average RMS power, duration, peak amplitude, spectral centroid, and temporal distance between the first and second bounces. Across experiments, average RMS power was found to be the best predictor. The relevance of the average RMS power confirmed results from a previous study that investigated the effects of various signal manipulations on the ability of listeners to categorize ball size (Grassi, 2002); (see also Grassi & Burro, 2003). The manipulation that affected performance the least was the removal of the bounces successive to the first impact. The manipulation that affected performance the most was RMS power equalization, applied to signals deprived of all bounces but the first. It is unclear, however, whether RMS power equalization without removal of the successive bounces would have led to the same drop in performance. Finally, it is highly likely that listeners' size estimates were associated with signal frequency, as modulated by the size of the struck dish. However, no tests were carried out that allow verification of this hypothesis.

The investigations summarized in this section demonstrate, in general, good source perception capabilities. Impaired recognition was however observed with perception of clapper gender and of hand configurations, when the stimulus set included different clappers (Repp, 1987), in the subsequent experiments by Tousman et al. (1989) on clapper recognition, and in the experiment by (Guski, 2000) on perception of ball drop height. Concerning the acoustical determinants of perception, almost all the studies summarized here suggested that judgments were based on the temporal patterning of the impacts (Repp, 1987; Li et al., 1991; Warren and Verbrugge, 1984; Warren et al., 1987; and also height and mass perception in Guski, 2000). As pointed out, temporal patterning variables may explain performance also in the study by Carello et al. (1998). The only exception to this is found in the experiments carried by that showed perception of bouncing ball size (Grassi & Burro, 2003) and force estimation (Guski, 2000) to be independent of the distance between successive bounces. Experiments by Grassi and Burro (2003) also showed, differently from results from isolated impact sounds, that variations in sounding object properties affect perception of hammer properties.

4.4 Rolling sounds

Two experiments conducted by Fowler (1990) investigated the ability of listeners to judge the steepness of ramps on the basis of rolling sounds. The ball rolled down a sandpaper-covered tilted ramp, which could have one of five slopes, and then onto a steel track, which could either be flat or tilted up. The duration of the ramp rolling sound increased with decreasing ramp slope. The duration of the track rolling sound was influenced by the track slope as well as by the ramp slope. For the flat track, it increased with decreasing ramp slope, while for the tilted track it increased with increasing ramp slope. In a first experiment, participants were presented with several configurations created by concatenating ramp rolling sounds and flat or tilted tracks rolling sounds, generated after the ball rolled down the steepest and shallowest ramps. Participants were asked to tell whether the ramps were steep or shallow. They were presented the correct response for the isolated ramp sounds and were informed of the slope of the track sound. Participants were reminded that the duration of the track rolling was influenced by the steepness of the ramp, but were presented no examples. Across conditions the probability of identifying ramps as steep increased with their slope. However, steepness judgments were also influenced by the track rolling sound. In particular, in agreement with the physics of the investigated event, when the track was flat, the longer track sounds caused ramps to be identified as shallower and vice versa. Also, consistently with the physics of the investigated event, when the track was tilted up, the longer track sounds caused ramps to be identified as steeper and vice versa. A second experiment examined the relevance of the duration of the track sounds in determining ramp slope judgments. Only the flat track sounds were used, as was the procedure for the first experiment. Different track sounds were used. The first was a short track sound generated after rolling from a steep ramp. The second was the "long" track sound cut to the short duration of a track sound generated after rolling onto a steep ramp. Even though equalized in duration, similar effects to those obtained in the first experiment were observed. Indeed, ramps were judged as steeper when followed by a short, than by a "long" track sound. Thus, acoustical information for ramp steepness, which accounts for the effects of tracks rolling sounds, is not limited to duration.

Houben, Kohlrausch, and Hermes (2004) (see also Houben, 2002) studied estimation of the size and speed of wood balls rolling on wood tracks. Experiments were carried out with excerpts of the recorded rolling sounds with constant duration and loudness. A first experiment tested participants' abilities to discriminate and sort the size of rolling balls. Good discrimination performance was observed. A second experiment tested speed discrimination. Good performances was observed here as well, although strong differences among participants were found. A replication of the speed discrimination experiment with stimuli not equalized for loudness showed this variable to be relevant for speed estimation, as in this condition performance was close to perfect. A last experiment tested listeners' abilities to discriminate and sort the size of rolling balls with variation in their speed and vice versa. An interaction effect was found in which variation in speed impaired size discrimination performance and vice versa. Analysis of the acoustical properties of stimuli investigated in the first two experiments showed both size and speed to influence the spectral centroid of the signals. Thus the perceptual interaction between size and speed was explained in terms of the influence of these latter parameters on the same perceptually relevant acoustical feature.

Although the tasks and source properties investigated in studies by Fowler (1990) and Houben et al. (2004) were different, a similar conclusion can be reached concerning the acoustical determinants for the judgments. Fowler (1990) found signal duration to be of secondary relevance for event perception, and Houben et al. (2004) found reasonably good source perception performance even in stimuli equalized for duration.

4.5 Scraping sounds

Lederman (1979) studied estimation of the roughness, i.e. texture, of metallic plates, when haptic, auditory or both auditory and haptic information were available. The roughness of aluminum plates was manipulated by varying the distance between adjacent grooves of fixed width (Experiments 1-2) or by varying the width of the grooves (Experiment 3). Across experiments, tactile information was found to dominate as no differences were found between the multimodal and active-touch-only conditions. In the auditory condition, signals were generated by the experimenter moving fingers on the plates with variable force (Experiment 1) and also with variable speed (Experiment 3). Auditory information was sufficient for roughness estimation as in both experiments estimates increased fairly monotonically with plate roughness. Perceived roughness was also influenced also by force and speed. While increases in force were associated with increases in roughness independently of speed, increases in speed were associated with increases in roughness for low force levels and with decreases for high force levels. The eventual role of pitch and loudness in determining estimates was discussed, although no acoustical analyses were provided.

4.6 Liquid sounds

Cabe and Pittenger (2000) investigated listeners' abilities to estimate and control the filling of a vessel with liquid. When liquid is poured into a container, the resonating air column decreases in length, causing an increase in the fundamental resonance frequency. A first experiment tested the perceptual relevance of this variable. Participants were presented pouring sounds in three conditions: filling, emptying, and constant water level, where in the latter two cases water flowed out of the vessel at the a same or higher rate than it was poured in. Participants were able to identify correctly the three different events. A second experiment tested the control of vessel filling on the basis of acoustical information. Participants had to stop pouring when water reached one of two different levels: comfortable drinking level and to the brim. Participants were able to execute the task consistently. A third experiment tested filling control with variation in the size of the vessel and of the maximum pouring rate. Participants were asked to fill vessels to the brim. Good performance was found, although it decreased with increasing vessel size and with decreasing maximum flow rate. Size and flow rate effects were consistent with the perceptual relevance of the fundamental resonance frequency of the air column. Indeed, with lower flow rates, and larger vessels the temporal change of the fundamental resonant frequency was slower, thus providing less information for filling control, and leading eventually to impaired performance. A final experiment tested the ability of listeners to predict the time required for a vessel to be completely filled on the basis of acoustical signals generated by filling the vessel up to a variable level. Both the filling level and the flow rate were varied. Participants were able to predict filling time with reasonable accuracy, better performance being observed with faster flow rates.

Jansson (1993) studied estimation of the filling level of shaken vessels. Performance was compared across modalities. In the haptic condition, participants were asked to shake the vessel. In the visual condition, they saw an actor shaking the vessel. In the auditory condition they heard the sound generated by the actor shaking the vessel. Results suggest a limited usefulness of acoustical information in specifying vessel filling, although a general tendency for estimates to increase with actual filling level was found. In a second experiment, performance with all sensory modalities together was also studied. The full-information condition preceded the conditions based on only one modality. In this case, auditory performance increased. Thus it can be concluded that impaired performance in the first experiment was not caused by an absence of appropriate acoustical information, but by the inability of participants to attend to it or to employ it in this task. This ability increased after the training received when performing the full-information condition.

Source perception abilities were demonstrated also with liquid sounds. Comparison of results across studies is however not possible, given the diversity in acoustical signals and tasks.

4.7 Research on timbre perception

Timbre is defined as “that attribute of auditory sensation in terms of which a subject can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar” (American National Standards Institute, 1973). Timbre, then, allows listeners to tell that two tones with the same pitch, loudness, and duration, one played with an oboe, the other played with a violin, differ. Also, timbre contributes to the identification of musical instruments.

The standard definition of timbre has been often criticized, as it specifies only what this attribute of auditory sensation is not (e.g. Hajda et al., 1997). For this reason the concept of timbre has been defined as a “wastebasket category”, where “if two sounds are different though having the same pitch and loudness, then they must differ in timbre” (Ward, 1970). Following this definition, timbre perception studies have focused mainly on sets of isolated sounds, equalized in pitch, loudness, and duration, in particular on musical instrument tones.

Experimental evidence accumulated throughout the decades contributed much to the understanding of the nature of timbre (see Hajda et al., 1997; McAdams, 1993; Handel, 1995, for a review of research in this field). Several regularities emerged across studies, highlighting the multidimensional nature of timbre,i.e., its dependence on multiple perceptual attributes and underlying acoustical properties. Results concerning the main acoustical determinants of musical instrument identification and of dissimilarity ratings on musical tones are summarized in this section.

For both identification and dissimilarity ratings studies, evidence concerning the relevance of source-based distinctions for the explanation of perception of musical instrument sounds is also outlined. Part of previously published data has been analyzed to this purpose. As pointed out at the beginning of this chapter, the main goal of this analysis is to test for the low of relevance of source properties to the perception of musical content hypothesized by (Gaver, 1988, 1993b).

4.7.1 Musical instrument identification

4.7.1.1 The relevance of source properties

Families of musical instruments can be distinguished on the basis of the nature of the vibrating body that generates the sound (von Hornbostel & Sachs, 1914, 1961). With *idiophones*, it is the body of the instrument that vibrates, without need of additional tension, thanks to the rigidity of the material of which it is made. Many percussion instruments belong to this family (e.g. cymbals, xylophones, tubular bells, etc.). In *membranophones*, the sound is generated by a vibrating stretched membrane, usually placed over the aperture of an empty body which amplifies the generated sound. All drums (e.g., snare drum, tympani), and also kazoos, belong to this family. In *chordophones*, the sound is generated by vibrating stretched strings. Bowed string instruments (e.g., viola, violin), guitars, harps, and the piano belong to this family. Finally, with *aereophones*, also called wind instruments, the primary vibrating element is the air column itself. Aereophones can be further distinguished depending on whether they result from the blowing an *air jet* through an opening (e.g., flutes, organ), through the buzzing of *single reeds* (e.g., clarinets and saxophones) or *double reeds* (e.g., oboe and bassoon) or through the buzzing of *lip reeds* (vibrating lips such in brass instruments such as trumpet or trombone) (Fletcher & Rossing, 1991).

Previous studies on musical instrument identification reveals the perceptual validity of the distinction among families of musical instruments, highlighting high recognition performance, particularly with respect to these gross source-based classes (e.g. Martin, 1999; Srinivasan, Sullivan, & Fujinaga, 2002). Part of those results are summarized and further analyzed here. In particular, only research conducted on non-synthetic, unedited tones, and which reported the matrices of identification confusions, was considered:

- Clark, Luce, Abrams, Schlossberg, and Rome (1963), **CLA63**;
- Berger (1964), **BEG64**;
- Strong and Clark (1967): **STC67**;
- Martin (1999): isolated tone condition, **MAI99**; solo excerpt condition, **MAS99**;
- Srinivasan, Sullivan, and Fujinaga (2002): 27 instrument condition, **SS202**; 9 instrument condition, **SS902**.

For each of the selected studies, the following statistics were calculated: the probability of correctly identifying the musical instrument or the instrument family; the probability of confusing a musical instrument with another instrument of the same family (within-family errors) or with an instrument of a different family (between family errors). The definition of musical instruments used for these analyses is less broad than that used by

Table 4.1: Probability of correct identification of musical instruments and musical instrument families, and probabilities of within- and between-family errors for the considered datasets. Chance probabilities are reported in parentheses. Chance probabilities for the within- and between-family errors, as well as chance probabilities for correct identification of musical instruments families, are averaged across families, because in the datasets the different families are represented by a variable number of instruments.

Dataset	*Musical instrument*	*Family*	*Within-f. errors*	*Between-f. errors*
CLA63	0.705 (0.077)	0.982 (0.212)	0.281 (0.154)	0.015 (0.769)
BEG64	0.593 (0.100)	0.789 (0.250)	0.250 (0.150)	0.157 (0.750)
STC67	0.852 (0.111)	0.961 (0.250)	0.100 (0.139)	0.046 (0.750)
MAI99	0.437 (0.037)	0.877 (0.200)	0.443 (0.163)	0.120 (0.800)
MAS99	0.594 (0.037)	0.960 (0.200)	0.369 (0.163)	0.038 (0.800)
SS202	0.570 (0.037)	0.860 (0.250)	0.303 (0.222)	0.095 (0.750)
SS902	0.901 (0.111)	0.940 (0.200)	0.033 (0.089)	0.054 (0.711)

von Hornbostel and Sachs (1914, 1961), because the subfamilies of aereophones have also been considered for this purpose. The results of this analysis are shown in Table 4.1.

As can be noted, the probability across studies of correctly identifying musical instrument families is higher than that of correctly identifying musical instruments. This result per se is not surprising, given the increased chance probability of correctly identifying musical instrument families. Extremely interesting, instead, is the lower discriminability of instruments belonging to the same family, shown by the higher probability of observing within-family errors, as compared to the probability of between-family errors. Musical instruments based on similar physical principles are therefore also strongly perceptually confused. Finally it should be noted that the experimental sets in the examined research included neither membranophones nor idiophones. It is highly likely that the data gathered on these classes of instruments would show the same tendencies concerning within- and between-family identification errors. Partial evidence on this issue is found in dissimilarity rating data (see Section 4.7.2). Another reason for expecting at least limited confusions between membranophones and chordophones on the one hand, and aereophones on the other, stands on another source-related distinction, based on the nature of the excitation process. Indeed the nature of the excitation that makes membranophones and idiophones sound is highly different from that which leads to the generation of tones in aereophones (see below). Such a distinction has been proven to be of high perceptual relevance in dissimilarity rating studies (see Section 4.7.2).

4.7.1.2 The relevance of acoustical properties

The study of the acoustical basis for musical instrument identification has used a signal-partitioning paradigm, according to which variations in identification performance was investigated with portions of tones presented in isolation or with tones without different portions. Musical instrument tones can be distinguished depending on the temporal extent of the excitation process that causes the primary source of acoustical energy to vibrate (e.g., the string in the violin). Thus we distinguish between *continuant* and *impulsive* excitations, or, equivalently, between continuant sounds (e.g., flute tones) and impulsive sounds (e.g., piano tones). For continuant excitations, the supply of energy to the source is extended over time (with flute tones the player continuously blows, causing the air column inside the instrument to vibrate until the blowing stops), whereas for impulsive tones, the excitation is limited to a restricted temporal span (the strike of the hammer on the string for piano tones). Ideally this should be a timeless impulse, but in practice it lasts from few microseconds to a few milliseconds. The difference in the excitation process is reflected in the temporal structure of the generated tones. Continuant sounds are usually partitioned into three sections. The usual subdivision is based on their amplitude envelope, the temporal variation of the RMS (root mean square) amplitude (see Hartmann, 1997, for a definition of RMS, and for more refined envelope extraction procedures), although more strict criteria have been proposed based on both the temporal variation of the RMS amplitude and of the spectral centroid (cf. Hajda et al., 1997). The envelope of continuant sound can be divided into three portions. The attack is considered as that portion emitted when the vibrating system achieves stability and is characterized by a rise in amplitude. The steady state is the portion during which energy is continuously supplied to the vibrating system. The decay is the portion during which the supply of energy to the vibrating system has ceased and the amplitude decreases abruptly. With impulsive sounds, instead, only two portions are found, the attack, and the decay. Figure 4.2 shows the partitioning of the envelope of a continuant sound (a French horn tone) and of an impulsive sound (a marimba tone).

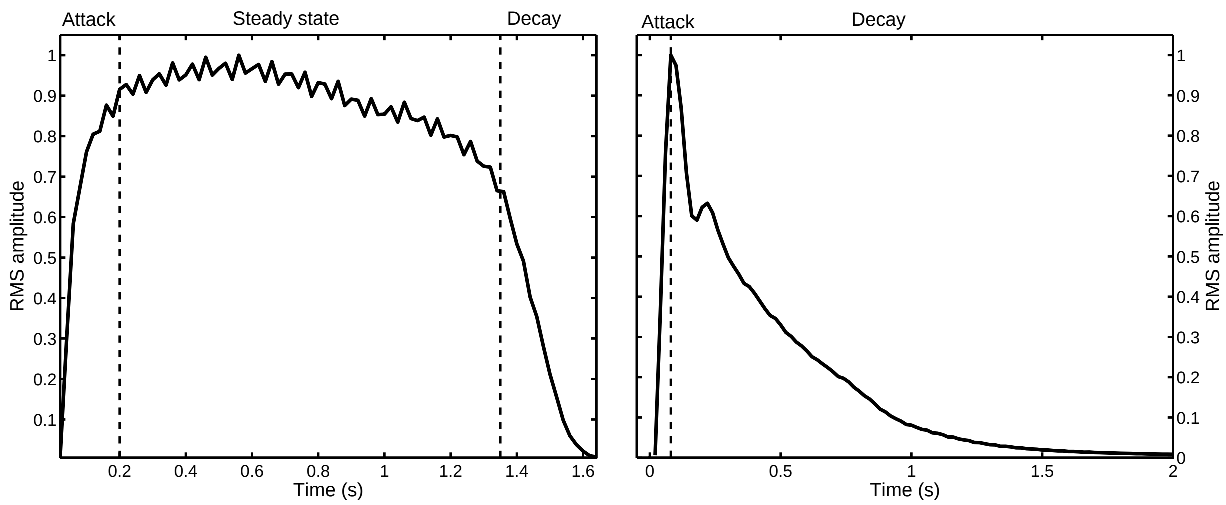

Figure 4.2: Partitioning of a continuant sound, a French horn tone (left) and of an impulsive sound, a marimba tone (right).

Several studies have investigated the relevance of tone partitions to the identification of musical instruments using isolated tones. Across the studies that investigated continuant tones (Clark et al., 1963; Berger, 1964; Elliott, 1975; Saldanha & Corso, 1964; Wedin & Goude, 1972), a drop in identification performance with the removal of the attack, but not with the removal of the decay was revealed. in these studies, the attack was defined simply by its temporal extent (i.e., fixed number of milliseconds from onset). Hajda, Kendall, and Carterette (1994) and Hajda (1999) applied a more rigorous model for the partitioning of the envelope constituents, based on the relationship between the temporal evolution of the spectral centroid and of the RMS amplitude. In contrast with previous research conducted on continuant tones, the attack was found to be neither necessary nor sufficient to the identification of continuant signals, yielding identification rates much lower than for the steady state alone. Indeed the attack portions of previous studies also included post-attack segments or portions of the steady state, which, in isolation, were found by Hajda (1999) to yield identification rates comparable to those for entire tones. This result, however, was valid only for long continuant tones, while with short continuant tones removal of the attack caused them to be confused with impulsive tones. Consistently with previous research, Hajda (1996, 1999) found decay portions not to be relevant for the identification of continuant tones. A higher relevance of the steady state to the identification of continuant signals was also revealed in musical contexts by Kendall (1986) with a matching procedure. Participants were presented a model melody played by a given musical instrument and were asked to choose which of three test melodies, different from the model, were played by the same musical instrument (to be exact the melody that was most similar to the model). Performance for steady states alone was not different from performance for unaltered stimuli and was superior to that for transients alone (attack and legato transients connecting the notes of the melody). Only one study investigated the relative relevance of tone partitions for the identification of musical instruments with impulsive sounds (Hajda et al., 1994; Hajda, 1999). Contrary to the major trend for the identification of continuant instruments, removal of the attack was not found to hinder identification and yielded performance levels close to the unaltered and decay-segment conditions. Thus, both attack transients and decay segments contain sufficient information for identification. Finally it should be noted that signal-partitioning studies provide evidence concerning where information for musical instrument identification is found throughout the tone. However, they are not useful for revealing precisely the nature of such information.

Another source of distinction between identification of impulsive and continuant instruments concerns the effects of reverse playback. While for long continuant tones reverse playback does not hinder identification (Berger, 1964; Hajda, 1996, 1999), it impairs performance for impulsive sounds. However, the distinction is not clear-cut, because with short continuant tones, reverse playback results in a decrease in identification performance, as with impulsive tones.

Two further effects are worth mentioning. First, identification performance of continuant tones is dependent on the pitch of the tested stimuli (Saldanha & Corso, 1964), where across the tones generated with the same musical instrument correct identification was found varying as a function of pitch. Although extensive data were not provided by the authors, it appears likely that pitches far from the normal playing range of a given musical instrument are more likely to be misidentified (e.g., a very high tone of a double bass may be misidentified as a violin tone). Pitch has also been found to be relevant to musical instrument discrimination (Handel & Erickson, 2001). In one experiment, listeners were presented two tones of different pitch generated by the same or by different instruments, and were asked to rate their impression that they were generated by different musical instruments. With pitch differences greater than one octave, tones generated by the same instrument were judged as being generated by different instruments, whereas when the separation was lower than one octave, the opposite, correct judgment was given. Conversely, with pitch separations lower than one octave, listeners made the error of judging different instrument pairs as generated by the same instrument, although correct performance was much higher than in the same instrument condition for pitch separations greater than one octave. Although directly relating this result to identification performance with isolated tones is not straightforward, this research highlights again the salience of pitch to the judgment of musical instrument tones, particularly when interstimulus pitch differences exceed one octave.

Another interesting effect with musical instrument identification concerns the change in performance with musical context. Both Kendall (1986) and Martin (1999) demonstrated that identification performance increases when listeners are presented with musical phrases rather than isolated tones. Two equally plausible explanations might account for these effects. First, with musical contexts players have greater freedom in using the expressive potentialities of musical instruments, thus revealing to a greater extent the acoustical signatures of a particular instrument compared with when they are asked to generate one isolated tone. Second, being presented musical phrases, composed by variable pitched tones, listeners are forced to ignore pitch when emitting their judgment, so that its biasing effect demonstrated with isolated tones fades away (Saldanha & Corso, 1964; Handel & Erickson, 2001). These explanations can also be reformulated using the view of the ecological approach, stating that when a listener is presented with different pitches generated by the same musical instrument, extraction of the invariant acoustical structure that specifies a musical instrument over its pitch range is made easier.

4.7.2 Dissimilarity rating

4.7.2.1 The relevance of acoustical properties

In this section the main results concerning the acoustical determinants of dissimilarity ratings of musical instrument sounds are presented. Both studies conducted on recorded and simulated musical instrument tones are presented. Also, results from two studies conducted on synthetic stimuli are considered (Caclin, 2004; Miller & Carterette, 1975).

Most of these studies made use of stimulus sets of constant, or perceptually equalized, duration, loudness, and pitch. As pointed out in Section 3.3, dissimilarity rating data are usually analyzed with MDS techniques, which map observed dissimilarity to the distance of stimuli within Euclidean spaces. MDS spaces are thus usually interpreted with reference to the dimensions that define them, assumed to represent the criteria used to estimate the dissimilarity. In timbre perception research, this has usually been done with reference to known acoustical properties of the stimuli. The first part of this section summarizes results on the acoustical interpretations of the dimensions of previously published timbre spaces. A different approach for the analysis of timbre spaces is used in the last part of this section, the so-called regional interpretations (Borg & Groenen, 1997), applied to test for the presence of structure in the spaces connected to the properties of the musical instruments as sound sources.

One of the most cited dissimilarity rating studies was conducted by Grey (1977) on synthetic emulations of continuant tones, created by decreasing the temporal variability of the frequency and amplitude of the spectral components of recorded signals. A three-dimensional space was extracted. A subsequent study by Grey and Gordon () found the first dimension to be strongly correlated with various measures of the spectral centroid. A qualitative interpretation was given to the other two dimensions by Grey (1977). The second dimension was related to the synchrony in the attack of the different spectral components and with the amount of spectral time-variance. The third was related to the presence of high-frequency inharmonic noise in the signals' attack.

Krumhansl (1989) investigated synthetic tones which either emulated different musical instruments or represented hybrids among pairs of musical instruments. Both impulsive and continuant sounds were used. A three-dimensional space was extracted. The dimensions were explained acoustically in a subsequent study by Krimphoff, McAdams, and Winsberg (1994). The first dimension, which distinguished continuant from impulsive tones, was strongly correlated with the *Log rise time*, defined as the logarithm of the time from perceptual threshold (2% of maximum amplitude) to maximum amplitude. The second dimension correlated with the spectral centroid. The third dimension was defined *spectral flux* by Krumhansl (1989) and was thought to reflect spectral time-variance, as the second dimension in Grey (1977). However, no measures of the spectral time-variance were found by Krimphoff et al. (1994) to explain this dimension. Instead this dimension was explained by a static spectral measure, *spectral irregularity*, which measured the deviation in amplitude across triads of adjacent spectral components.

McAdams et al. (1995) studied a stimulus set highly similar to that investigated by Krumhansl (1989). A three-dimensional space was found. Consistently with what was found for the space by Krumhansl (1989), the first dimension was strongly correlated with log rise time and distinguished impulsive from continuant tones, whereas the second strongly correlated with spectral centroid. Notably, the first dimension correlated also with a measure of signal duration, itself correlated with log rise time. The third dimension, differently from what was found for the space by Krumhansl (1989), was weakly correlated with a measure of spectral time-variance, or spectral flux, given by the average correlations among Fourier spectra computed in adjacent temporal windows along the entire signal.

Impulsive and continuant sounds were also investigated by Iverson and Krumhansl (1993) in a study focusing on the effects of signal partitioning on dissimilarity ratings. Three conditions were investigated (unedited tones, tones minus attack, and attack only).

Two-dimensional spaces were extracted for all the conditions. The position of stimuli within the spaces for the different editing conditions was highly correlated. This result led the authors to conclude that the perceptually salient attributes are found throughout the entire tone. For all spaces, one of the dimensions strongly correlated with spectral centroid. The other dimension, which tended to distinguish impulsive from continuant signals, was related to the structure of the amplitude envelope. An analytic parameter that correlated significantly with this dimension was found, however, only for the attack condition: the time from onset to maximum amplitude.

Sets including both impulsive and continuant stimuli where also studied in Lakatos (2000) and in Marozeau et al. (2003). In this latter study, sets also included variations in pitch. Across the studies two- to four-dimensional spaces were found. For all the spaces, one dimension strongly correlated with spectral centroid, and one dimension distinguished impulsive from continuant tones. Lakatos (2000) explained the impulsive vs. continuant dimension with log rise time, whereas Marozeau et al. (2003) used a measure of impulsiveness, probably correlated with log rise time. The third dimension in the Marozeau et al. (2003) spaces correlated with a measure of *spectral spread*, which distinguished between chordophones and aereophones. Spectral spread was found to account better for the spectral flux dimension in McAdams et al. (1995) than the previous measure of spectral time-variance. The fourth dimension in Marozeau et al. (2003) spaces appeared to be related to pitch, but a strong correlation was found only for the experimental set that included the highest pitch variation (11 semitones), and not for the other set (2 semitones of pitch range).

In summary, for all the studies that investigated sets of both continuant and impulsive tones, one of the extracted dimensions distinguished these two classes of sounds and was almost always related to log rise time, whereas another of the dimensions was strongly related to spectral centroid. Other dimensions, related to spectral time-variance, or to spectral spread, emerged with less regularity, thus pointing toward the primary perceptual relevance of spectral centroid and log rise time. A similar conclusion concerning the perceptual relevance of measures of spectral time-variance was also reached by Caclin (2004), with a study conducted on synthetic stimuli. Across the investigated conditions spectral centroid and attack time were found to mask variations in spectral flux, modeled as a rise in the spectral centroid in the first 100 ms of the signals.

Dimensions related to spectral time-variance emerged, instead, with studies conducted with sets comprising only continuant sounds. Kendall, Carterette, and Hajda (1999) conducted different experiments investigating sets of recorded natural-instrument tones or mixed sets of recorded and synthetic tones. Across studies and conditions, dimensions of the rotated classical MDS solutions were correlated with spectral centroid, with the standard deviation of the time-variant spectral centroid, and, to a lesser extent, with the *mean coefficient of variation*, a measure of spectral flux. Hajda (1999) investigated continuant stimuli resynthesized from recorded signals. One of the sets included the same time-variant properties of the original tones, while these were eliminated in the second set (steady-state set). For both sets a two-dimensional space was extracted. The first of the dimensions correlated strongly with the spectral centroid. The second dimension of the time-variant set correlated with a modified version of the mean coefficient of variation and with ratings of the amount of *vibrato*. The second dimension of the steady-state set remained unexplained. Handel and Erickson (2004) investigated a set of continuant stimuli generated with three different wind instruments, varying pitch over two octaves. A three-dimensional space was derived. The first dimension was strongly correlated with pitch, the second with spectral centroid, the third with the rate of frequency vibrato.

A higher variability is found in results of studies that investigated sets of impulsive signals alone. Serafini (1993) investigated Javanese gamelan instrument tones with isolated tones or melodies. In both cases, two-dimensional spaces were derived. For both conditions, the attack spectral centroid, extracted from the first 50 ms of the signals, explained the first dimension of the spaces. The second dimension was explained only in the melodic context and was correlated with the average amplitude in the middle part of the tones. Harsberger, Kendall, and Carterette (1994) investigated a set of Indonesian gong tones with variable pitch ranging over 4 octaves. A two-dimensional solution was derived. The first dimension was related to pitch. A second dimension was also identified, which reproduced variations in *amplitude modulation*. It should be noted that the angle between the first and the second dimension was approximately 45° and that, equivalently, the two dimensions were non-orthogonal. Hajda (1995) investigated sets of recorded and synthetic impulsive tones. Across the many investigated conditions, two-dimensional spaces were derived. In general, spaces were explained in terms of tone *duration* and in terms of the percent change in the spectral centroid over the first 100 ms of the signal. Across studies, then, different acoustical variables for dissimilarity ratings of impulsive signals emerged as perceptually relevant. It should be noted that part of this variability in results might be caused by a variation in the measures used to characterize the acoustical structure of the signals.

A summary should be made on pitch effects. Independently of whether stimulus sets comprised only continuant tones, only impulsive tones, or both, pitch-related dimensions emerged when the range of pitch variation within the experimental set was equal or greater than 11 semitones (Harsberger et al., 1994; Marozeau et al., 2003; Handel & Erickson, 2004), this independently of the fact that listeners were asked to ignore pitch in making their responses. In line with this trend are results from one of the experiments reported in Miller and Carterette (1975). The experimental set comprised pitch variations of two octaves, variations in the amplitude envelope (impulsive, trapezoidal, and impulsive with a sustained pre-offset portion), and variations in the rank of the most intense harmonic component. A three-dimensional MDS solution revealed one dimension related to pitch and two dimensions related to amplitude-envelope structure and perceived duration (despite constant physical duration, impulsive envelopes were perceived as shorter than the other envelope types).

Another summary should be made on duration effects. Of the above mentioned studies, Iverson and Krumhansl (1993), McAdams et al. (1995), Hajda (1995), Kendall et al. (1999) used stimuli of variable duration, and tested the correlation of this variable with experimental results. Duration was found to be relevant by Hajda (1995), ranging in different sets from 3.1 to 5.2 ms and from 3.1 to 7.6 ms. Further, McAdams et al. (1995)

reported an effect of duration, ranging from 495 to 1096 ms. It should be noted however that this variable was strongly correlated with log rise time. The relevance of this latter variable across many different studies makes it a more likely candidate to explain the duration-related dimension. Finally, both Iverson and Krumhansl (1993), and Kendall et al. (1999) found no effect of this variable, which ranged approximately from 2 to 3.29 sec in the study by Iverson and Krumhansl (1993). Also, perceived duration was hypothesized as a factor to explain one of the MDS dimensions by Miller and Carterette (1975). To summarize, it is plausible to hypothesize that, as with pitch, duration effects emerge with increases in its range of variation within experimental sets. However, available results do not allow strong conclusions to be drawn.

4.7.2.2 The relevance of source properties

In order to test for the relevance of source-based distinctions in explaining listeners' judgments, previously published analyses of dissimilarity rating data were considered. A broad criterion was used for this purpose, which falls within the category of regional interpretations of MDS (Borg & Groenen, 1997), to test whether stimuli belonging to different source-based classes occupy disjoint regions in the MDS spaces. A positive outcome for this test was considered as evidence for the relevance of source-based distinctions in explaining the structure of listeners judgments and perceptions. The distinction among main classes of musical instruments (aereophones, chordophones, membranophones, and idiophones), as well as the distinction between impulsive and continuant signals were considered. The following studies conducted on real and synthetic isolated musical sounds, which included at least two classes of musical instruments, were included in the analyses:

- Wedin and Goude (1972): entire tones condition, **WE72E**; tones minus attack, **WE72A**;
- Wessel (1973), **WES73**;
- Grey (1977), **GRE77**;
- Krumhansl (1989), **KRU89**;
- Serafini (1993), **SER93**;
- Iverson and Krumhansl (1993): entire tones condition, **IV93E**; attack only, **IV93A**; tones minus attack condition, **IV93R**;
- Hajda (1995), **HAJ95** (recorded tones only);
- McAdams, Winsberg, Donnadieu, De Soete, and Krimphoff (1995), **MAC95**;
- Kendall, Carterette, and Hajda (1999), **KEN99**;
- Lakatos (2000): harmonic set, **LA00H**; percussive set **LA00P**; combined set, **LA00C**;

- Marozeau, de Cheveigné, McAdams, and Winsberg (2003): pitch variation of 1 tone, **MA03A**; pitch variation of 11 semitones, **MA03B**.

Table 4.2 describes these studies, the nature of the sets of investigated stimuli, and the represented categories of musical instruments and types of excitation. For dissimilarity rating data by Wedin and Goude (1972), the MDS spaces computed by Hajda et al. (1997) were considered. Krumhansl (1989) and McAdams et al. (1995) investigated sets of synthetic musical instruments sounds, part which were hybrids between two different instruments. Data for the hybrids were not considered given the impossibility of assigning them to unique classes of musical instruments (e.g., hybrids between chordophones and aereophones). Kendall et al. (1999) studied dissimilarity ratings on natural and synthetic tones. Only data for the natural tones (Figure 4 in Kendall et al., 1999) were considered, which were associated with higher proportions of correct musical instrument identification than synthetic signals. Among the data collected by Hajda (1995), only those spaces computed for judgments of recorded instruments were considered. For data by Hajda (1995) and Lakatos (2000) one additional class for the type of excitation, referred to as *multiple impulses* was considered, according to which the vibrating body is struck repeatedly in a short time span, as with xylophones when a tremolo is simulated (Hajda, 1995), or where multiple vibrating bodies are struck close in time, as with the bamboo chimes investigated by Lakatos (2000). The martelé style of playing for the violin, investigated in Lakatos (2000), was considered as generating an impulse signal, given the extremely short duration of the bowing that excites the violin string.

Table 4.2: Datasets, number of dimensions of MDS spaces, and stimulus properties in the considered dissimilarity rating studies on musical timbres. Aereo. = aereophones; chordo. = chordophones; membr. = membranophones; idio. = idiophones; cont = continuant; imp = impulsive; m. i. = multiple impulses. The three rightmost columns show whether significant partitionings were found (Y) or not (N), and, in parentheses, the number of dimensions of the MDS spaces needed to highlight the partitioning. For those cases where partitionings were not found (N), relevant partitionings were highlighted discarding a variable number of data points. For these cases the musical instrument and the number of discarded data points are shown in parentheses.

		Musical instruments class				*Excitation type*			*Partitionings*		
Dataset	*N Dim.*	Aereo.	Cordo.	Membr.	Idio.	Cont.	Imp.	M. I.	Mus. ins.	Exc.	Mus. ins. + Exc.
WE72E	2	6	2			8			Y (2)		
WE72R	2	6	2			8			Y (2)		
WES73	2	6	3			9			Y (1)		
SER93	2		1		5		6		Y(1)		
GRE77	3	13	3			13			N (2, flute)		
KRU89	3	7	5		1	8	6		Y (2)	Y (1)	
IV93E	2	11	3		2	13	3		Y (1)	Y (1)	
IV93A	2	11	3		2	13	3		N (2, violin)	Y (2)	
IV93R	2	11	3		2	13	3		Y (2)	Y (2)	
HAJ95	2		3		5		6	2	Y (2)	Y (2)	
MAC95	3	6	5		1	7	5		Y (2)	Y (1)	
KEN99	2	10	1			11			Y (2)		
LA00H	2	12	5			13	4		N (1, organ)	Y (1)	
LA00P	3			6	12	3	13	2	N	Y (2)	Y (3)
LA00C	2	6	4	4	6	9	11		N	Y (2)	N (2, cuica; steel drum)
MA03A	4	10	8			12	6		Y (1)	Y (1)	
MA03B	4	10	8			12	6		Y (1)	Y (1)	

We tested, then, whether these distinctions were reflected in the distribution of stimuli within the associated MDS spaces. We assumed that evidence for the perceptual relevance of these distinctions existed if they defined disjoint regions of the MDS spaces. For example, in the case of a unidimensional space, the distinction among impulsive, and continuant sounds would be considered perceptually relevant if above a given location along the dimension all stimuli were impulsive, and below all stimuli were continuant. The boundaries between disjoint regions were assumed to be linear, i.e., when two dimensions were considered, the boundary was a line, and a plane/hyper-plane when three/four dimensions were considered. Such a choice was made for the ease of computations, and was not based on considerations of psychophysical nature. A better possibility was to consider also non-linear boundaries (e.g., quadratic functions in bi-dimensional spaces), where the probability of observing disjoint regions in MDS spaces would have been higher or equal to that when only linear boundaries were considered. As such, it should be then noted that the linear-boundary assumption makes the test for the hypothesis of the relevance of source-based distinctions to dissimilarity ratings even more conservative.

Logistic regression models were used (Hosmer & Lemeshow, 1989; Agresti, 1996). Disjoint regions in the MDS spaces, defined by the source-based criteria presented above, were sought, predicting the source categories using the MDS coordinates as independent variables. In particular, we tested whether the MDS coordinates allowed a perfect prediction of the source categories, a condition known as perfect separation (Albert & Anderson, 1984). When such a condition occurred, disjoint regions were found in the MDS spaces, containing only members of one of the categories, whose boundaries were defined geometrically as mentioned above. When complete separation was found using all the dimensions of the MDS spaces, we checked whether a lower number of dimensions was sufficient to yield the same outcome. The musical instrument classes, and the excitation type classes were tested both in isolation and, with reference to Lakatos' (2000) data, in conjunction, so that, for example, bowed and struck cymbal sounds belonged to different classes (idiophone-continuant and idiophone-impulsive). An example of this analysis is shown in Figure 4.3. Analysis of the MDS solution of McAdams et al. (1995) revealed partitions corresponding to musical instrument class (aereophones, chordophones, and idiophones), as well as partitions corresponding to type of excitation (impulsive vs. continuant). Both dimensions 1, and 3, were necessary to discriminate among classes of musical instruments, whereas dimension 1 was sufficient to perfectly separate impulsive sounds from continuant sounds. The boundaries among these classes, computed with logistic regression models, are also shown.

Table 4.2 summarizes the results of this analysis. For each of the considered datasets it is shown whether musical instruments belonging to different classes, and generated with different types of excitation occupied disjoint regions of the MDS spaces. The same outcome is shown, when appropriate, with reference to the classes defined considering jointly the type of excitation and the musical instrument class. Also shown is minimum number of MDS dimensions required to separate the investigated classes.

Out of the 17 considered MDS spaces, 12 revealed partitionings based on musical instrument classes. With 3 of the remaining studies, the failure to reveal disjoint regions occupied by musical instruments was caused by only one stimulus in the dataset. With the remaining two datasets (harmonic and combined sets; Lakatos, 2000), significant partitionings were found when musical instrument class and excitation type where con-

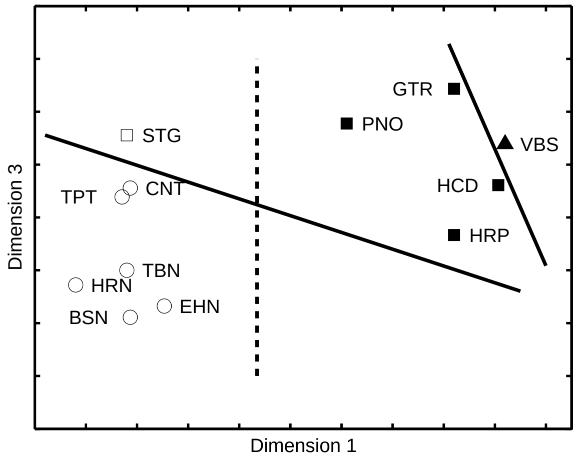

Figure 4.3: Partitions based on the musical instrument class and on the type of excitation revealed in the MDS solution of McAdams et al. (1995). BSN = bassoon; CNT = clarinet; EHN = English horn; GTR = guitar; HCD = harpsichord; HRN = French horn; HRP = harp; PNO = piano; STG = string; TBN = trombone; TPT = trumpet; VBS = vibraphone. Circle = aereophone; square = chordophone; triangle = idiophone. Black = impulsive excitation; white = continuant excitation. Continuous lines show the boundaries between musical instruments classes; dashed lines show the boundary between impulsive and continuant sounds.

sidered jointly. However, with the combined set in Lakatos (2000), a perfect partitioning was found only when data for the steel drum and the cuica were removed (see Figure 4.4). Concerning the type of excitation that initiates vibration in the musical instrument (impulsive, continuant, and multiple impulses), with all considered datasets different types of excitation occupied disjoint regions in the MDS spaces.

In summary, across studies the evidence supports the perceptual relevance of the distinction between aereophones, chordophones, membranophones, and idiophones, even though not all studies revealed clear cut partitionings. Instead, strong evidence for the perceptual relevance of the distinction between types of excitation was found, where all studies revealed significant partitionings based on this source property.

4.7.3 Discussion

Previous data on identification of musical instruments and dissimilarity ratings was analyzed in order to test for the relevance of source properties in explaining experimental results.

Independently of the methodology used to test perception distinctions based on the nature of the generators of the acoustical signals, which defined the musical instrument family, and/or based on the nature of the excitation of the generators were found, to a gross extent, mapped in perceptual data. In particular previous studies on identification

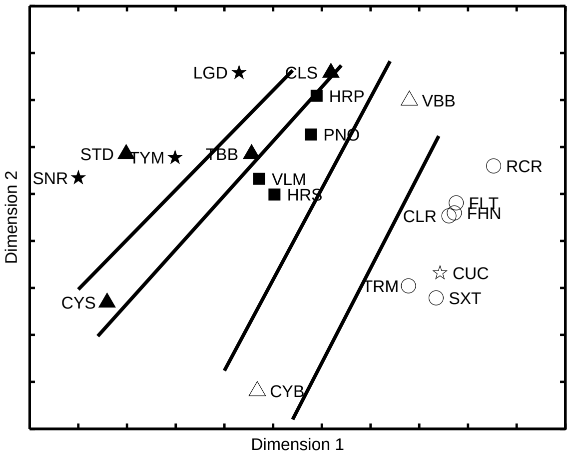

Figure 4.4: Partitions based on the musical instrument class, and on the type of excitation revealed in the MDS solution of Lakatos (2000) for the combined set. CLR = B flat clarinet; CLS = celesta; CUC = cuica; CYB = cymbal (bowed); CYS = cymbal (struck); FHN = French horn; FLT = flute; HRP = harp; HRS = harpsichord; LGD = log drum; PNO = piano; RCR = baroque recorder; SNR = snare drum; STD = steel drum; SXT = tenor sax; TBB = tubular bells; TRM = C trumpet; TYM = tympani; VBB = vibraphone (bowed); VLM = violin (martelé). Circle = aereophone; square = chordophone; triangle = idiophone; star = membranophone. Black = impulsive excitation; white = continuant excitation. Continuous lines show the boundaries between musical instrument categories.

found tones generated with musical instruments belonging to the same family to be highly confused, and tones generated with different-family instruments to be seldom confused. Source-based distinctions were found mapped to dissimilarity ratings data too. In particular categories of musical instruments and/or types of excitation were found, to a variable extent, occupying disjoint regions within the MDS spaces.

Overall these analyses revealed that source properties explain perception of musical instrument tones and, as a consequence, that despite their "quasi-harmonic nature" (cf. Gaver, 1988) this class of signals carry perceptually meaningful source-related information. This result points toward the usefulness of research conducted on musical instrument sounds to the understanding of source perception. For example, it is quite interesting to note that the totality of dissimilarity ratings studies revealed clear-cut distinctions based on the nature of the excitation, while less clear distinctions were found when only the nature of the generator (i.e., musical instrument family) was considered. It must then be concluded that the nature of the excitation has a higher perceptual salience than the nature of the generator. It should be finally noted that studies on musical instrument tones can also clarify issues concerning the acoustical determinants of source perception with environmental sounds. For example, studies on musical instrument tones found pitch, or, roughly, signal fundamental frequency, to be a strong determinant for listeners' judgments when its variation in the stimulus set was greater than one octave (see Section 4.7.2). Consistently, results concerning material identification highlighted a relevance of signal frequency to judgments only when the stimulus set included large frequency variations (see Section 4.2.2). Also, one of the major acoustical determinants for timbre perception, spectral centroid, has found to be associated with judgments of source properties by several of the studies conducted with environmental sounds.

The relevance of source properties to the perception of musical instrument tones contradicts the definition of musical listening by Gaver (1988, 1993a), conceived as focused on the properties of the proximal stimulus, the sound, rather than on those of the distal stimulus, the source. Thus, either studies conducted on isolated musical tones, as the vast majority of those reviewed here, produce results not generalizable to musical listening, or this latter concept has to be redefined. It is this latter possibility that we would like to explore. Both studies conducted on environmental and musical sounds revealed a tie of perception with the properties of the distal stimulus. Nonetheless, it is undoubtable that when listening to a musical composition for marimbas the length of the struck bars, or the hardness of the mallets is far from our thoughts, while the same is not true when we hear the sound of a large, or small object falling on the floor. The distinction between everyday listening and musical listening might be better redefined on the basis of the properties of the linguistic content activated in the two contexts. Indeed, while in both cases our perceptions map properties of the source, as well as properties of the acoustical signals, the amount of source-related linguistic content would be much higher in everyday conditions than in musical listening contexts.

4.8 Conclusions

In Chapter 2, two theories of perception were presented: the ecological and the information-processing approaches. Several points of opposition between the two theories were found. The ecological approach postulated unique specification of the distal stimulus (the sound source) in invariant properties of the proximal stimulus (the acoustical signal), whereas the information-processing approach postulated such a relationship to be ambiguous. The ecological approach assumed perception to be unmediated by internal representations, and to be unaided by additional information provided by past experience, whereas the information-processing approach made the opposite assumptions. In other words, the ecological approach postulated perception to be direct, and the information-processing approach postulated perception to be mediated. Finally, the ecological approach assumed perception to be veridical and, ultimately, denied the appropriateness of labeling perception as in error, while such an assumption was not made by the information-processing approach, which contemplated errors of perception as useful to the understanding of the functioning of the perceptual system. Experimental evidence on source perception was reviewed in this chapter.

Only a few studies have attempted to test for the presence of invariants or acoustical parameters that uniquely specified the source property whose perception was being investigated. Among these Li et al. (1991) found linear combinations of different spectral descriptors to discriminate among actual walker genders, and Warren and Verbrugge (1984) pointed out the temporal patterning of impacts in bouncing and breaking sounds to discriminate among the two events. Notably with both studies the highlighted invariant acoustical information was used by participants to judge the investigated source properties. On the other hand, results from several studies on the acoustical determinants of judgments showed perception to rely on noninvariant acoustical properties. For example, geometry perception was explained in terms of the modal frequencies of the sounding object, influenced by source properties other than geometry, namely the material of the sounding object (see Section 4.2.1). These latter results might point toward a limited usefulness of the concept of invariant in explaining source perception. It should be noted, however, that experimental judgments may have been based on noninvariant properties of the proximal stimulus simply because invariant acoustical parameters were not available in the experimental context. Thus the notion of invariant can be rejected as empirically useless only if both listeners base their judgments on noninvariant acoustical properties and source-specific information is highlighted. As pointed out above, this test is seldom found in the literature on source perception.

Concerning the direct versus indirect perception debate, at the end of Chapter 2 it was stated that a valid empirical test for the correctness of the direct perception assumption was based on the demonstration of percept-percept couplings, i.e., the mutual influence of perceptual dimensions. Evidence in favor of the indirect approach was found, however, only for the visual modality. None of the studies reviewed in this chapter found percept-percept couplings in audition. This issue is thus still open.

Concerning veridicality, good source perception capabilities were found across studies. Nonetheless, several systematic biases revealed a misalignment between actual and perceived source properties. For example, the perceived geometry of struck plates was also influenced by their material (Gaver, 1988; Kunkler-Peck & Turvey, 2000; Tucker & Brown, 2003), or the perceived size of bouncing balls was influenced by the size of the object on which they bounced (Grassi, 2005). Following the considerations made in Section 2.1.4, these results cannot be interpreted as contradicting the veridicality assumption of the ecological approach, as they simply reveal that the dimensions of perception do not faithfully map those of physics. This issue is thus still open. It should be pointed out, however, that ecological theorists deny in ultimate stance, the appropriateness of applying the term error and the true and false labels to the facts of perception (see Section 2.1.4). For this reason, any study which might find nonveridical perception (i.e., false or in error) will never undermine the ecological approach.

Finally, empirical evidence on the perception of musical instrument tones was reviewed and reanalyzed in order to test for the relevance of source-based distinctions in explaining data. Independently of the experimental methodology used, source-based distinctions were found to be reflected in participant judgments. This result was used to argue against the hypothesis made by Gaver (1988, 1993a) according to which source properties are at best secondary to musical listening, this latter focusing primarily on the properties of the acoustical signal. As a consequence, independently of whether investigations focus on musical material or on environmental sounds, both source and acoustical properties should be conceived as equally useful to the understanding of audition (cf. Section 3.1).

Chapter 5

Plan of studies

Is perception veridical? Is perception direct? Whatever the truth, the perceiver makes use of the available sensory information to extract properties of the environment. Thus, ascertaining the environmental properties relevant to the perceiver and the properties of the sensory information upon which perception is based is a relevant issue, in and of itself, to the study of the perception of sound sources. This problem has motivated the empirical investigations to be presented in Chapters 6-8.

All studies focused on the hammer/sounding object impact sound source described in Section 3.2. The study in Chapter 6 complements a series of experiments conducted by Giordano (2003) on the identification of the material type of sounding objects, investigating the acoustical determinants for identification and testing for the presence of sufficient information for perfect performance. In Chapters 7 and 8, this work is referred to as Giordano and McAdams (submitted). The study in Chapter 7 tests for the relevance of the properties of the hammer/sounding object interaction in determining perceived hammer properties and for the perceptual independence of hammer and sounding object. These issues are investigated by asking listeners to rate the hardness of either the hammer or the sounding object. Finally, the study in Chapter 8 investigates the relative relevance to dissimilarity ratings of hammer properties, sounding object properties, and properties of their interaction.

Throughout these studies both the acoustical and physical or source-based determinants of perception were investigated. An effort was made to limit assumptions concerning the specific nature of the determinants, using the most complete possible characterizations (see Section 3.1). Thus, a large set of acoustical descriptors has been used to explain participants' judgment in all studies. Among these descriptors, the acoustical measure of damping ($tan\phi$) was included, as it had been hypothesized to mediate perception of the material of the sounding object by Wildes and Richards (1988). The need to avoid assumptions has also determined the choice to extend the analysis of physical determinants beyond the properties of the sounding object, where in both Chapters 7 and 8 the properties of the interaction between hammer and sounding object were also considered, and in Chapter 8 the properties of the hammer were also included among the variables potentially useful to the explanation of perception.

With all the investigations particular care has been taken to generate knowledge generalizable to everyday conditions or, in other words, to maximize the ecological validity of experimental results. Different methodological choices were made to this purpose. First, participants were given minimal nonauditory source-related information, and no feedback on response correctness was provided. In this way, participants were more likely to use response criteria closer to those used to perceive source properties under everyday conditions, rather than response criteria learned in the experimental setting through examples of source-acoustical signal pairings or through feedback on response correctness[1]. Second, stimulus sets were generated varying several different source properties. In this way the experimental context aimed to reproduce the source uncertainty that might characterize many listening conditions, where sources are not constrained to vary along a single property.

As outlined above, different experimental techniques were used throughout the studies. In Chapters 6 and 7, participants were instructed to judge explicitly a given property of the sound source. In both cases, the focus was on material-related properties: in Chapter 6 participants were asked to identify the material type of the sounding object, and in Chapter 7 they rated the hardness of either the hammer or the sounding object. As pointed out in Chapter 8, the use of techniques based on the explicit judgment of sound source properties might limit the ecological validity of the studies, because it might focus the attention of listeners on source and acoustical properties that are not of primary concern outside the experimental context. For this reason in Chapter 8 perception was tested using a technique that did not require the use of source-related linguistic labels: dissimilarity ratings. At any rate, whatever the task used to test perception in the laboratory, generalization to everyday conditions requires assuming that the experimental judgment is representative of everyday perception. Given the difficulties inherent in testing this assumption, more safely generalizable results shall emerge from the comparison of studies based on different experimental tasks. Accordingly, one can more safely conclude that a given source property is relevant to everyday perception if it explains data gathered with different experimental judgments. This final summary is presented in Chapter 9 where results concerning the acoustical and physical determinants of judgments gathered in the different studies are compared.

Finally, with all the studies the relationship between source and acoustical properties was investigated, ascertaining the presence of sufficient acoustical information for unbiased perception of the manipulated source properties, i.e. for an alignment of perception with the physical properties of the source. Following the considerations made in Section 4.8 such a test also allowed us to verify the usefulness of the concept of invariant to the explanation of source perception. In Chapters 7 and 8, this analysis also served the purpose of linking together the studies of source and acoustical determinants of the judgments. In other words, it allowed an understanding of why source property A and acoustical property B explained equally well the behavioral data on the basis of the fact that B was, above all the acoustical properties, the one most strongly associated with

[1]Feedback-based techniques model the so-called "best case scenario" which assumes highly practiced listeners (Lutfi & Oh, 1997).

source property A.

Chapter 6

Material identification of real impact sounds

Bruno L. Giordano† and Stephen McAdams††

Abstract

Listeners' performance in auditory material identification tasks was found to be perfect with respect to gross material categories, comprising materials of vastly different properties. Impaired performance was observed when materials with similar mechanical properties were involved, identification being based on the size of the objects with material type having no effect. Several acoustical criteria for identification, including an acoustical measure of damping, were tested concerning their ability to explain listener performance. The damping descriptor only accounted for the discrimination between materials belonging to different gross categories, while discrimination within the same gross category appeared to be based mainly on signal frequency. Sufficient acoustical information for perfect material identification was found. Procedural biases for the origin of the effects of size could be discarded, pointing toward their cognitive, rather than methodological, nature.

6.1 Introduction

A growing branch of research investigates the perceptual correlates of the properties of sound sources. This branch has been variously labelled *ecological acoustics* (Vanderveer,

†Dipartimento di Psicologia Generale; Università degli Studi di Padova; Via Venezia 8, I-35100 Padova, ITALY and Institut de Recherche et de Coordination Acoustique/Musique (STMS-IRCAM-CNRS);1 place Igor Stravinsky; F-75004 Paris, FRANCE. Electronic mail: bruno.giordano@unipd.it

††STMS-IRCAM-CNRS; 1 place Igor Stravinsky, F-75004 Paris, FRANCE and CIRMMT, Faculty of Music; McGill University; 555 Sherbrooke St. W.; Montréal, QC, Canada H3A 1E3. Electronic mail: smc@music.mcgill.ca

1979), *auditive kinetics* (Guski, 2000), *psychomecanics* (McAdams, 2000), or, generally, sound source recognition. The object of study in this field can be described at three different levels: the physical or mechanical level (the properties of the sound source), the acoustical level (the properties of the sound wave emitted by the source), and the perceptual level (the properties of the sound event as perceived by the listener). The research design in sound source recognition analyzes all the pairwise relationships among these levels (Li et al., 1991). This design was applied to one of the most investigated issues: identification of material type in impact sounds.

Several previous works investigated material identification with synthetic or real signals, focusing particularly on the effects of acoustical measures of damping. Wildes and Richards (1988) defined a shape invariant acoustical parameter for material type, the coefficient of internal friction $tan\phi$, which models material damping:

$$tan\phi = \frac{\alpha}{\pi f} \tag{6.1}$$

where α is the damping coefficient of the vibrational component, i.e. the inverse of the time required for the component to decay to $1/e$ of its original amplitude, and f is its frequency. The higher $tan\phi$ the greater the damping of the material, and the faster the decay time decreases with increasing frequency. Wildes and Richards (1988) proposed material type recognition to be based on the $tan\phi$ coefficient.

The effect of damping measures on auditory material judgments was tested in several studies. Klatzky et al. (2000) investigated stimuli synthesized according to a physical model of a struck bar (van den Doel & Pai, 1998), varying a parameter inversely proportional to the $tan\phi$ coefficient, as well as the frequency of the lowest vibrational mode, later referred to as frequency, which spanned over 3.3 octaves. Four response categories were used: rubber, wood, steel, and glass. Both experimental variables affected the judgments: rubber and wood were chosen for higher $tan\phi$ values than glass and steel; glass and wood were chosen for higher frequencies than steel and rubber. Avanzini and Rocchesso (2001a) investigated material identification with stimuli generated according to the physical model of a one-mode resonator, varying the $tan\phi$ coefficient, and frequency (range: 1 octave). The same response categories were used, and results were analogous to those of Klatzky et al. (2000), although frequency effects were less clear. Roussarie (1999) investigated stimuli synthesized according to a physical model of a struck plate (Lambourg, 1997), varying damping coefficients, elastic properties, and density of the simulated plates around those characterizing glass and aluminum. The properties of the simulated hammer were also manipulated, using parameters typical of either wood or rubber. Two response categories were adopted: glass and aluminum. Responses were found to be influenced only by the damping properties of the plates, strongly correlated with an acoustical parameter analogous to $tan\phi$ and with the average spectral center of gravity, respectively. Variations in density and elasticity, associated with a variation in frequency equivalent to a musical interval of a perfect fifth, had no effect. In summary all of the cited studies demonstrated the relevance of damping measures for material identification, while frequency effects were demonstrated only with experimental sets where this acoustical feature ranged over at least one octave.

Other studies focused on material identification performance. Gaver (1988) asked participants to judge whether variable-length struck bars were made of iron or wood. High recognition performance was observed, responses being uninfluenced by the geometrical properties of the objects. Similar results were obtained with stimuli synthesized according to a physical model of a struck bar. Kunkler-Peck and Turvey (2000) generated stimuli by striking triangular, rectangular, and square plates made of steel, wood or plexiglas. Material recognition was almost perfect, and only a secondary tendency to associate materials with shapes was found. Perfect material recognition performance was not confirmed, however, in a study conducted on synthetic signals (Lutfi & Oh, 1997). Stimuli were synthesized according to the wave equation of a struck clamped bar, with stimulus variability created by perturbing the density and elasticity terms. Participants were asked which of two stimuli was generated by striking a given target material (iron or glass), the alternatives being different metals, crystal or quartz. Signal frequency was given a disproportionate weight by listeners, resulting in poor performance.

Inconsistencies between results by Kunkler-Peck and Turvey (2000) and by Lutfi and Oh (1997) were explained by Carello, Wagman, and Turvey (2003) in terms of the lack of acoustical richness that might characterize synthetic signals, and thus of the absence of sufficient information for the task. Additional studies, however, showed that impaired performance is also found with real signals. Giordano (2003) reported preliminary results of three experiments conducted with real signals generated by striking rectangular plates made of steel, glass, wood, and plexiglas. Different stimulus sets were investigated, varying the height/width ratio of the plates and their area (both with freely vibrating and externally damped plates), or varying the area of the plates and the material of the hammer. The height/width ratio and hammer material variables had no significant effect. With freely vibrating plates steel and glass were almost perfectly discriminated from wood and plexiglas, and vice versa. In line with results by Gaver (1988), Kunkler-Peck and Turvey (2000), discrimination between these gross categories was not influenced by the geometrical properties of the objects. External damping caused glass-plate signals to be confused with wood and plexiglas sounds while categorization of steel plates was not influenced by this manipulation.

The strong confusion between wood and plexiglas was confirmed in another study conducted on real signals (Tucker & Brown, 2003). Stimuli were generated by striking wood, plexiglas, and aluminum plates of constant area and variable shape. Wood and plexiglas were strongly confused with one another and were almost perfectly discriminated from steel. A second experiment, performed on underwater recordings of the same struck plates, gave the same results. A large portion of the data (69% for the first, 62% for the second) was found to be explained by a parameter related to $tan\phi$. Studies by Lutfi and Oh (1997), Giordano (2003), Tucker and Brown (2003) thus point to the presence of limited material recognition capabilities, perfect only when discriminations among materials of vastly different properties (e.g., woods and metals) are involved.

Studies by Kunkler-Peck and Turvey (2000), and Tucker and Brown (2003) also found material identification to be independent of the geometrical properties of the objects when a discrimination between wood and plexiglas was involved. Such results, however, are not consistent with further effects reported by Giordano (2003). With experiments conducted on freely vibrating plates, discrimination between steel and glass, and between wood and plexiglas was in fact found to be based on the area of the plates: larger plates were more often labelled as steel and plexiglas than as glass and wood. Analogous tendencies were found with externally damped plates. These response patterns appeared strongly consistent across listeners, although a small percentage of participants showed an opposite effect of area on wood and plexiglas categorizations, the former material being associated with larger plates than the latter. Given the increase of frequency with a decrease in the size of an object the association of wood and glass to low-area plates and steel and plexiglas to large-area plates would seem to confirm results by Klatzky et al. (2000). The informal nature of the acoustical analyses presented in Giordano (2003), however, does not allow us to draw conclusions on this point. Indeed other acoustical parameters than frequency, affected by a variation in size, might have been used by listeners in selecting their responses. Also, one potential explanation for the emergence of size effects on material identification in Giordano (2003) was that no acoustical information was present in the stimulus sets to discriminate between steel and glass, on the one hand, and between wood and plexiglas signals, on the other. The absence of relevant information for these discriminations might indeed have led participants to focus on source properties irrelevant to the task, namely size, for which a variation in the acoustical features was present. However, no tests for the presence of sufficient acoustical information for perfect material identification were performed.

Despite all the work made on this topic, little is known about the acoustical criteria used in material identification. In fact with previous research behavioral data were explained with limited sets of descriptors that included, at best, an acoustical measure of the damping properties, frequency, and the average spectral center of gravity (Roussarie, 1999). Furthermore, the observed behavioral relevance of acoustical measures of damping like $tan\phi$ is not sufficient to conclude that they are used by listeners to identify the material of the objects. Indeed, other acoustical properties, associated with variations of $tan\phi$, might be used by participants to judge material type. For example signals characterized by lower $tan\phi$ values should also have a longer duration. Also, a change in $tan\phi$ should be associated with a change in the temporal variability of loudness, where the higher the $tan\phi$ value, and thus the faster the decay of amplitude for each spectral component, the faster the decrease in loudness. A recent dissimilarity rating study (McAdams et al., 2004) demonstrated the perceptual relevance of acoustical variables other than the damping descriptors for the judgment of impact sounds. With this study signals synthesized according to a physical model of a struck bar (Chaigne & Doutaut, 1997; Doutaut et al., 1998) were investigated. Participants were asked to rate the dissimilarity of pairwise presented sounds. Stimulus sets were generated varying a parameter related to damping, η, which models the frequency dependence of damping using a quadratic function, and either the density of the bar or its length. Two-dimensional spaces were found to account for the dissimilarity ratings in both experiments. One of the dimensions was related to the η coefficient, and the other dimension was related to either the density or the length of the bar.

Frequency explained the location of the stimuli along the density/length dimension. A linear combination of a level decay-rate descriptor and of the average spectral center of gravity explained the location of stimuli along the η-related dimension. Given the known relationship between identification and dissimilarity ratings (e.g., Grey, 1977; see also McAdams, 1993), one is supported in expecting similar acoustical parameters to influence auditory material identification.

A new study on material identification was performed, using a subset of the real signals studied in Giordano (2003). The complete research design in sound source recognition was adopted (Li et al., 1991). Analysis of the relationship between the physical, acoustical, and perceptual levels allowed for testing the physical and acoustical determinants of material identification. In particular, the relationship between auditory material identification and a wide set of acoustical descriptors was investigated, comparing their explanatory power with that of the acoustical measure of damping, $tan\phi$. An analysis of the relationship between the acoustical and physical levels allowed us to test for the presence of sufficient acoustical information for perfect material recognition. Given the interindividual differences for the wood and plexiglas recognition strategies reported by Giordano (2003), an analysis of the similarities among data from different participants was also performed.

6.2 Methods

6.2.1 Stimuli

Sounds were generated striking plates made of four different materials: plexiglas (polymethyl methacrylate), soda-lime glass, steel, and Tanganyka walnut. All plates were 2 mm thick. All plates were square. Five different values were used for the length of the sides of the plates: 8.66, 12.24, 17.32, 24.49 and 34.64 cm, yielding areas from 75 to 1200 cm^2. Each plate was drilled close to the right and left top corners and close to the left and right borders, at the middle of their height (diameter: 4 mm). The upper holes were used to suspend the plates; the lower ones to stabilize them after being struck, thus avoiding amplitude modulations due to an excessive movement of the plate relative to the microphone. Plates were struck with a steel pendulum (diameter: 2 cm; weight: 35.72 g).

The apparatus used to suspend the plates was similar to that used by Kunkler-Peck and Turvey (2000) (see Figure 6.1). The main structure was made of pine wood. Both the plates and the pendulum were hung from the top shelf with nylon lines (diameter: 1 mm.). The lateral holes of the plates were attached to two 150-g weights with nylon lines, passing through holes drilled in two horizontal planks attached to both sides of the structure. The pendulum was hung from the top shelf, 15 cm from the plane in which the plates lay, and was released from a fixed guide attached to the front of the top shelf so that its starting angle was kept constant. Plates were struck in their centers. No audible multiple impacts of the pendulum on the plate were observed.

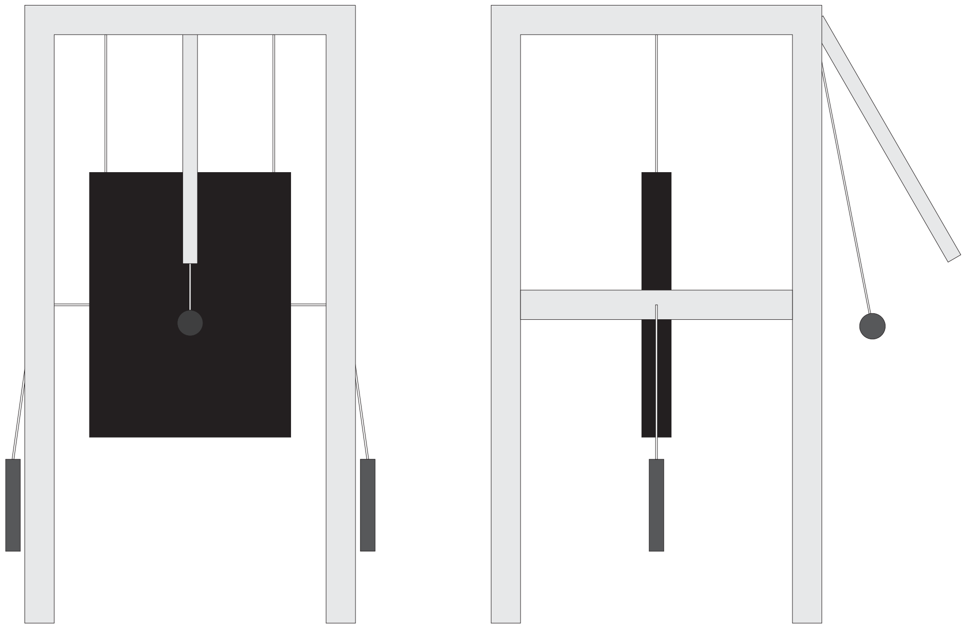

Figure 6.1: Sketch of the device used to suspend and strike the plates. In dark grey the pendulum and the stabilizing weights.

Sounds were generated in an acoustically isolated room with highly absorbing walls and were recorded using a TASCAM DA-P1 DAT recorder (48000-Hz sampling rate, 16-bit resolution) and Beyer Dynamic digital microphone (MCD101/MPD200) positioned 45 cm from the center of the plate, opposite to the struck surface. Recordings were transferred to a computer hard disk through the digital input of a Sound Blaster Live Platinum sound card. Signals longer than 1 sec were reduced to this duration by applying a 5-ms linear decay. Informal listening tests showed that material identification was not influenced by this sound wave editing process. Signals were not equalized in loudness. The presentation level was the maximum level which kept the background noise, constant across the samples, inaudible. The peak levels of the signals ranged from 54 to 72 dB SPL.

6.2.2 Procedure

Stimuli where presented through AKG K240 headphones, connected to a NIKKO NA-690 amplifier, which received the output of the Sound Blaster Live soundcard of the PC used to program the experiment. Participants sat inside a sound-proof booth. Stimulus presentation and data collection were programmed into the Mathworks Matlab environment, using the facilities provided by the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Participants were told that on each trial they would be presented a single sound generated by striking an object made of one of four different materials. In order to make instructions straightforward, it was decided to use generic linguistic labels for all materials: glass for soda-lime glass, metal for steel, plastic for plexiglas, and wood for Tanganyka walnut. Given that the stimulus set comprised only one material type per generic category, it was assumed that this linguistic choice did not affect participants' responses. No mention was made of the geometrical properties of the objects, minimizing the non-auditory information given to participants. After presentation of the stimulus, participants were asked to identify the material of the struck object. Before giving the response, participants were allowed to replay the stimulus as many times as needed. Responses were given by pressing appropriate keyboard keys. The 20 stimuli were presented in blocked-randomized order for each of seven repetitions, for a total of 140 trials.

6.2.3 Participants

Twenty-five listeners took part in the experiment on a voluntary basis (age: 22 - 49 years; 17 males, 8 females). All of them reported having normal hearing.

6.3 Acoustical descriptors

The analysis model used to extract the $tan\phi$ parameter, and the loudness and brightness descriptors was meant to simulate the output of the cochlea in response to the incoming acoustical signal. Outer and middle ear filtering were simulated by means of a cascade of two IIR and one FIR filter, in order to account for peak sensitivity at 2 kHz, and for loss of sensitivity at lower and higher frequencies. The transfer function was derived from measures of the minimum audible field (Killion, 1978). Processing of the signal inside the cochlea was simulated with a gammatone filter bank (Patterson, Allerhand, & Gigure, 1995), with center frequencies f_c uniformly spaced on an equal-resolution scale (Moore & Glasberg, 1983), between 30 and 16000 Hz. The power in output from the cochlear filters was then added to the power delayed by $1/4f_c$ (Marozeau et al., 2003).

$Tan\phi$ was extracted from this representation. Damping factors α for the signal output from each channel were computed using the regression model $log(P) = a + bT$, where P is power, T is time, and $b = -\alpha/2$. The regression model was applied to the signal from peak power to a fixed threshold power. Figure 6.2 shows the analysis of a harmonic complex given by the sum of six damped sinusoids with a fundamental frequency of 2000 Hz. Damping factors have been chosen to yield a $tan\phi$ of 0.01. Also shown is the upper limit for the damping factor of the signal in output from the cochlear channels, calculated analyzing an unitary amplitude impulse with this procedure.

Computation of $tan\phi$ from the extracted damping factors was based on the assumption that the perceptual relevance of the output of each cochlear filter was not constant. In particular it was assumed that the higher the total power in output from the filter, the higher the perceptual relevance of the signal in determining $tan\phi$. Thus $tan\phi$ was defined as:

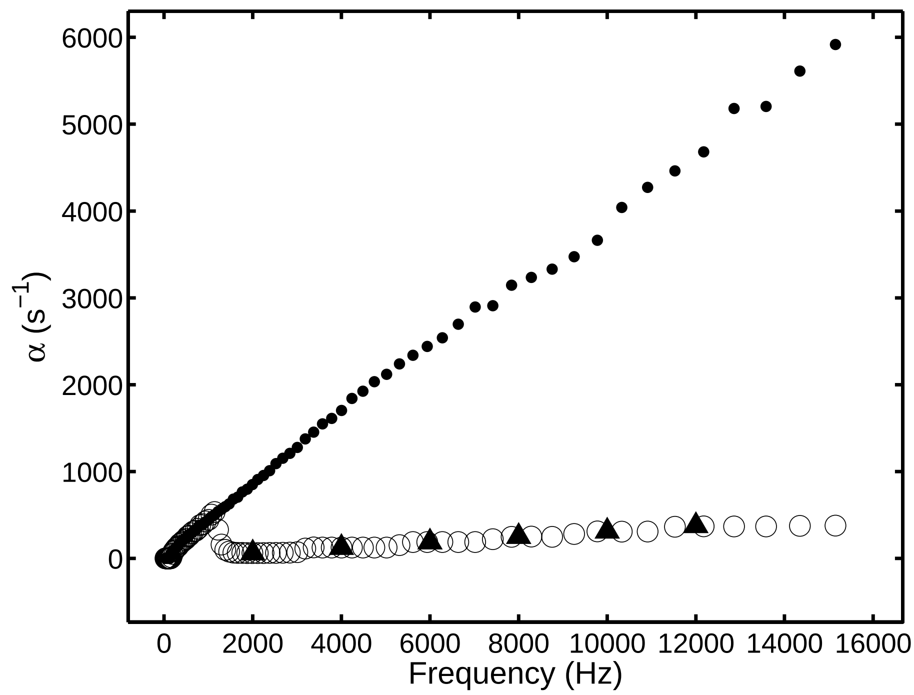

Figure 6.2: Extraction of the damping factors from the output of the analysis model. Filled circles show the upper limit for the damping factor, empty circles show the damping factors extracted from a six-component harmonic complex with fundamental frequency of 2000 Hz and $tan\phi$ of 0.01; filled triangles show the damping factors of the input signal.

$$\frac{\sum_{i=1}^{N} \frac{\alpha_i}{\pi f_{ci}} w_i}{\sum_{i=1}^{N} w_i} \tag{6.2}$$

where f_{ci} is the center frequency in Hertz, and w_i is the sum of power from peak value to threshold. This procedure yielded, for the signal shown in Figure 6.2, a $tan\phi$ value of 0.0101.

A second representation was used to extract loudness and brightness descriptors. The representation used to compute the $tan\phi$ parameter was downsampled, convolving it with a 10-ms square window, yielding a temporal resolution similar to that for loudness integration (Plack & Moore, 1990). The power in each channel was finally raised to the power of 0.25 to approximate partial loudness (Hartmann, 1997). For each temporal frame of the final representation, both loudness and brightness measures were derived. Loudness was defined as the sum of the partial loudnesses computed from the output of each cochlear channel (Zwicker & Fastl, 1999). Given that loudness was calculated on the sound files, without taking into account the actual presentation levels, its unit of measure was termed “pseudo-sone”. Brightness was measured by means of the spectral center of gravity (SCG), defined as the specific loudness weighted average of frequency on the ERB-rate scale. Finally a duration (Dur) measure was extracted, defining the offset as that instant where the signal reached the loudness of the background noise (about 0.2 pseudo-sones).

For both the loudness and SCG functions over time, an attack and an average measure were extracted (Lou_{att}, SCG_{att}; Lou_{mea}, SCG_{mea}). For 17 of the sounds the SCG_{att} measure corresponded to the maximum SCG value, while for the remaining three signals the peak was found in the third analysis frame (20-30 ms from onset). With loudness, the attack corresponded to maximum loudness in nine signals, while for the remaining 11 maximum loudness was found in the second analysis frame (10-20 ms from onset). Further descriptors characterized the temporal evolution of loudness and brightness and were extracted with linear regression procedures. The loudness function was characterized using two measures. The first one (Lou_{sl1}) measured the slope from the attack value to the point where loudness equalled half of the attack value. The second measured the slope from the point were loudness was double the final value up to the end (Lou_{sl2}). The SCG-over-time function was found to be non monotonic for 15 signals, for which an initial decrease was followed by a final increase. A slope measure was extracted considering the initial portion of the SCG-over-time function, from attack to the minimum value (SCG_{slo}). Figure 6.3 shows the loudness and SCG over time functions for the signal generated by striking the 150-cm^2 glass plate. Also shown are the linear regression functions used to extract the slope measures.

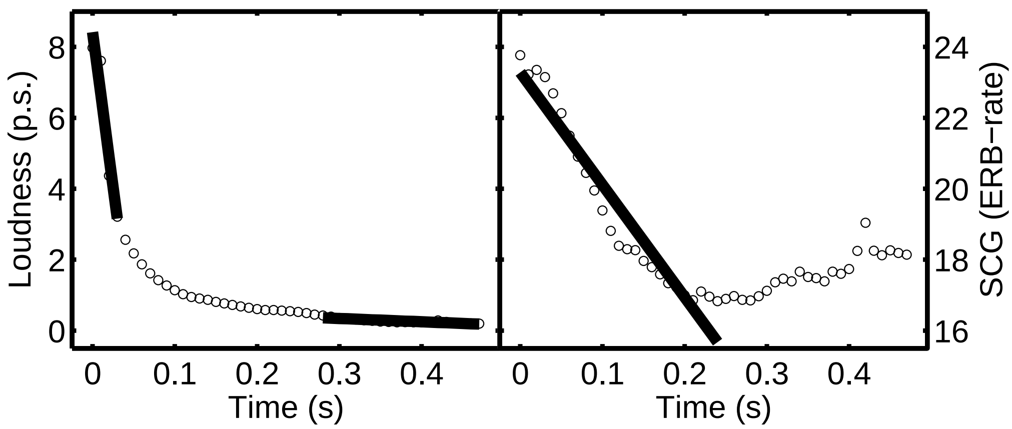

Figure 6.3: Temporal functions of loudness and SCG for the signal generated by striking the 150 cm^2 glass plate. Also shown are the linear regression functions used to extracted the slope measures.

Finally a measure of the frequency of the lowest spectral component F was extracted, on the basis of the fast Fourier transform of the first 4096 samples of the signals (Hanning window). F was defined as the frequency of the first amplitude peak exceeding a fixed threshold. Amplitude threshold was defined as the maximum amplitude of the low frequency background noise across the recorded samples. Table 6.1 shows for each signal the extracted acoustical indices. Approximate density measures for the investigated materials are also given. Notably, $tan\phi$ discriminated perfectly among material types, where an increase in this measure is found from plexiglas to wood to glass to steel.

Table 6.1: Acoustical descriptors extracted from each signal. Mat=material; S = steel; G = glass, W = wood; P = plexiglas; ρ=density; p.s.=pseudo-sones. See text for an explanation of the meaning of each acoustical descriptor.

Mat.	Area (cm^2)	ρ (kg/m^3)	$tan\phi \times 10^{-3}$	Dur (s)	F (Hz)	Lou_{att} (p.s.)	Lou_{mea} (p.s.)	Lou_{sl1} (p.s./s)	Lou_{sl2} (p.s./s)	SCG_{att} (ERB–rate)	SCG_{mea} (ERB–rate)	SCG_{slo} (ERB–rate/s)
S	75	7708.30	1.05	0.98	1535.15	7.24	0.95	-138.84	-0.62	25.98	21.41	-6.00
S	150	7708.30	0.86	0.98	773.44	6.89	1.45	-45.61	-0.96	24.41	19.90	-5.90
S	300	7708.30	0.90	0.98	386.72	6.17	1.81	-25.04	-1.24	23.27	20.64	-2.85
S	600	7708.30	0.37	0.98	187.50	6.80	3.15	-11.96	-2.50	24.10	21.69	-2.63
S	1200	7708.30	0.27	0.98	93.75	5.84	3.88	-4.78	-2.93	23.83	20.35	-2.97
G	75	2301.70	1.52	0.52	1406.25	8.09	1.25	-153.39	-1.17	25.38	22.48	-7.70
G	150	2301.70	4.46	0.47	750.00	7.97	1.09	-175.29	-0.93	23.76	18.59	-31.70
G	300	2301.70	2.59	0.63	386.72	8.58	1.50	-105.47	-1.12	23.07	19.33	-5.72
G	600	2301.70	1.68	0.98	187.50	7.06	1.47	-42.14	-0.69	22.96	16.86	-7.26
G	1200	2301.70	2.55	0.94	105.47	6.59	1.34	-38.41	-0.54	22.56	17.20	-5.56
W	75	718.33	19.29	0.17	527.34	5.19	0.93	-175.51	-1.40	23.51	18.18	-171.61
W	150	718.33	22.33	0.19	257.81	4.55	0.95	-131.95	-2.03	22.34	16.44	-102.38
W	300	718.33	19.03	0.30	128.91	4.56	0.83	-121.13	-1.12	21.40	15.75	-44.01
W	600	718.33	19.78	0.16	58.60	4.15	1.07	-104.69	-3.41	20.98	16.69	-52.39
W	1200	718.33	17.55	0.23	23.44	4.06	1.05	-64.57	-2.64	21.00	16.10	-36.04
P	75	1413.30	26.09	0.10	527.34	4.78	1.11	-176.74	-3.99	23.26	18.36	-153.04
P	150	1413.30	39.62	0.10	281.25	4.05	1.15	-127.72	-4.52	22.11	17.25	-148.12
P	300	1413.30	41.03	0.13	140.63	3.83	1.10	-110.59	-3.19	21.33	16.84	-114.28
P	600	1413.30	31.03	0.16	70.31	3.71	0.91	-99.08	-2.50	20.98	16.46	-123.87
P	1200	1413.30	24.50	0.17	35.16	3.79	0.91	-84.09	-2.46	20.80	15.38	-85.39

6.4 Results

Due to the repetitions, for each sound a distribution of responses across the four categories was possible for each listener. Analyses were conducted on the individual modes of these distributions, hereafter referred to as “modal responses”.

6.4.1 From Physics to Perception

Response profiles of small groups of participants presented macroscopic differences with respect to data pooled across all participants. Statistical criteria were applied to extract groups of homogeneous response profiles. Cluster analysis was used to this effect. Distances among individual response profiles were calculated using a general nominal dissimilarity measure, defined as the proportion of consistent categorizations among two participants (Gordon, 1999). An agglomerative hierarchical algorithm (average linkage) was used. The choice of the number of clusters to extract from the hierarchical solution was based on the analysis of the variation of a set of statistical indices across partitioning levels. These indices measure the goodness-of-fit between the input data and the resulting clustering partitions (Milligan, 1996). Among the available indices, a subset was chosen that had been found to have superior performance in recovering the correct number of clusters (Milligan, 1981; Milligan & Cooper, 1985): the c index (Hubert & Levin, 1976), the Goodman-Kruskal γ (Baker & Hubert, 1972), and the point biserial correlation (Milligan, 1980). For the first index, lower scores indicate higher goodness-of-fit, and better partitions, while for the latter two, better partitions are characterized by higher scores. Following the approach suggested by Gordon (1999), indications concerning the correct number of clusters were sought in local maxima/minima across partition levels, and the correct number of clusters was established on the basis of the concordance among indices. Figure 6.4 shows the value of the three indices as a function of the number of clusters. Local maxima or minima used to extract the correct number of clusters are also shown.

The final number of clusters was taken to be equal to three as this partitioning level was indicated by all the three indices. The three clusters contained, respectively, 21, 3, and 1 participant(s). Tables 6.2 and 6.3 show the frequency of the modal response for each stimulus for the first and second clusters of participants. Table 6.4 shows the modal response for the participant in the third cluster. Within the first group of participants, the responses metal and glass never occurred for wood or plexiglas sounds; the responses wood and plastic were only given once for the glass and steel sounds. Also a strong tendency to associate the responses glass and wood with smaller plates and the responses metal and plastic with larger plates was found. The same tendencies characterized the second cluster of participants, the only difference being the inversion of the effect of the area of the plates on the responses wood and plastic, associated, respectively, with larger and smaller wood, or plexiglas plates. The participant in the third cluster was the only one who responded metal for larger wood and plexiglas plates. Also the modal responses for this participant did not include the plastic category. Subsequent statistical modelings were performed only on data from the main cluster with 21 participants.

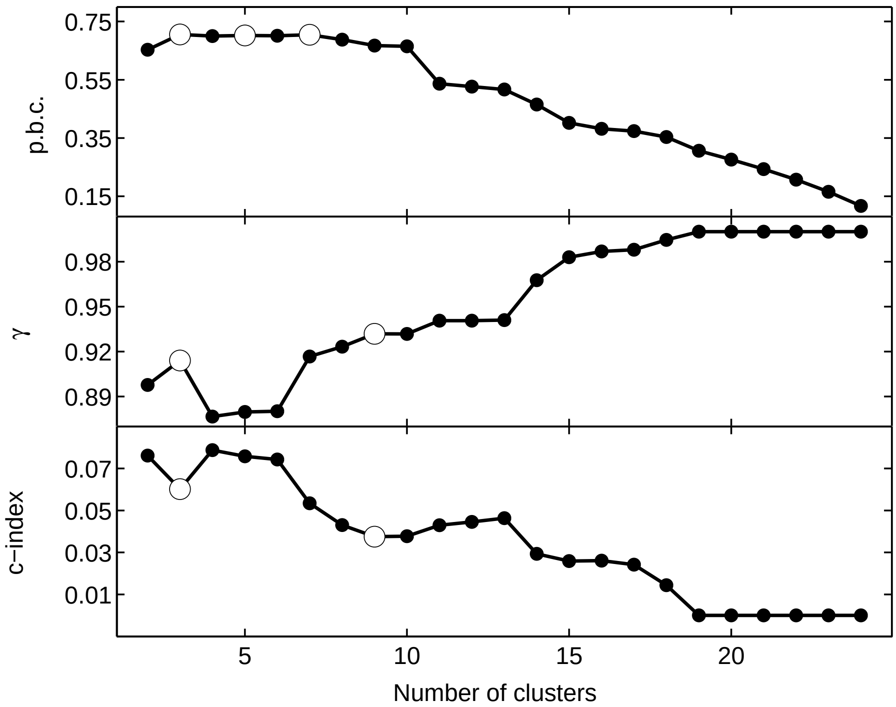

Figure 6.4: Statistical indices used to evaluate the number of clusters present in the dataset, across partitioning levels. White circles mark local maxima for the point biserial correlation, and for the Goodman-Kruskal γ index, and local minima for the c index.

Table 6.2: Contingency table for the modal response in the first cluster of participants (N=21). Material: S = Steel, G = Glass, W = Wood, P = plexiglas; Area: A1-A5 = 75-1200 cm^2. Response categories in italics.

	Metal					*Glass*				
S	1	3	17	21	21	20	18	4	0	0
G	2	2	18	21	21	19	18	3	0	0
W	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0
	Wood					*Plastic*				
S	0	0	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	0	0	0	0
W	19	13	10	13	2	2	8	11	8	19
P	21	21	16	6	3	0	0	5	15	18
	A1	A2	A3	A4	A5	A1	A2	A3	A4	A5

Table 6.3: Contingency table for the modal response in the second cluster of participants (N=3). Material: S = Steel, G = Glass, W = Wood, P = plexiglas; Area: A1-A5 = 75-1200 cm^2. Response categories in italics.

	Metal					*Glass*				
S	0	3	3	3	3	3	0	0	0	0
G	0	1	2	3	3	3	2	1	0	0
W	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0
	Wood					*Plastic*				
S	0	0	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	0	0	0	0
W	0	1	2	3	2	3	2	1	0	1
P	0	0	0	2	2	3	3	3	1	1
	A1	A2	A3	A4	A5	A1	A2	A3	A4	A5

Table 6.4: Modal response for the participant in the third cluster. Material: S = Steel, G = Glass, W = Wood, P = plexiglas; Area: A1-A5 = 75-1200 cm^2. Response categories in italics (M = metal).

S	*G*	*M*	*M*	*M*	*M*
G	*G*	*W*	*M*	*M*	*M*
W	*W*	*W*	*W*	*M*	*M*
P	*W*	*W*	*W*	*M*	*M*
	A1	A2	A3	A4	A5

Statistical models were then constructed in order to test for the relevance of the sound source properties in determining the judgments. Separate logistic regression models (Agresti, 1996) were constructed for the perceptual discriminations between metal and glass on the one hand, and between wood and plastic on the other. The variable selection approach suggested by Hosmer and Lemeshow (1989) was adopted in order to generate parsimonious models for observed data. Thus, before entering predictors into multivariate models, the significance of their effect was tested in the univariate models. Both the material, and the area of the plates, were coded as categorical variables. Data predicted by the regression models were compared with observed data using the deviance, and Hosmer-Lemeshow (Hosmer & Lemeshow, 1989) goodness-of-fit measures. Non-significant statistics indicated the equivalence of observed and predicted data, and thus the validity of inferences based on the regression models. Logistic regression could not be used to model the discrimination between wood and plastic on one hand, and glass and metal on the other, because of the almost perfect performance level observed. Simple χ^2 association tests were thus adopted.

Discrimination between metal/glass and wood/plastic was influenced by the material, but not by the area of the plates ($\chi^2(3) = 416.038$, $p < 0.001$, $\chi^2(4) = 0.038$, $p = 1.000$). Furthermore the frequency of choosing the responses metal or glass was equivalent for steel and glass plates ($\chi^2(1) = 1.005$, $p = 0.316$), and that of choosing the responses wood or plastic was equivalent for wood and plexiglas plates ($\chi^2(1) = 0$, $p = 1$). The probabilities of choosing the response metal over the response glass, and that of choosing the response wood over the response plastic, were not influenced by the material of the plates (Wald $\chi^2(1) = 0.052$, $p = 0.820$, Wald $\chi^2(1) = 1.962$, $p = 0.161$). The effect of area, on the other hand, was highly significant in both cases (Wald $\chi^2(4) = 47.386$, $p < 0.001$ for metal versus glass, Wald $\chi^2(4) = 48.52$, $p < 0.001$ for wood versus plastic). Finally goodness-of-fit statistics showed the observed data to be well accounted for by area effects alone: for both the metal-glass and wood-plastic models, the Hosmer-Lemeshow and deviance statistics were non-significant (glass-metal: deviance= 0, p= 1, Hosmer-Lemeshow $\chi^2(2) = 0.009$, $p = 0.996$; wood-plastic: deviance= 0, $p = 1$, Hosmer-Lemeshow $\chi^2(3) = 0$, $p = 1$).

6.4.1.1 Discussion

In line with results from previous research conducted on real signals (Gaver, 1988; Kunkler-Peck & Turvey, 2000; Giordano, 2003; Tucker & Brown, 2003), participants showed perfect recognition abilities when discrimination between gross material categories (metal-glass and wood-plastic) were involved, the geometrical properties of the objects having no effect. Analysis of response profiles showed that such an ability characterizes the near totality of listeners. Such gross material categories are characterized by vast differences in physical properties such as density (see Table 6.1), steel and glass being denser that wood and plexiglas, or elasticity, steel and glass being stiffer than wood and plexiglas (cf. Waterman & Ashby, 1997). In principle, then, both density, and elastic properties could explain the observed perfect identification performance.

Consistently with results by Giordano (2003) and Tucker and Brown (2003), but inconsistently with results by Kunkler-Peck and Turvey (2000), participants were not able to discriminate between wood and plexiglas. The same effect, consistently with the impaired discrimination performances reported by Lutfi (2001), was observed for the discrimination between harder materials, i.e., steel and glass. Indeed, for these discriminations physical materials had no effect, i.e. steel was perceptually equivalent to glass, and wood to plexiglas. Inconsistently with results by Kunkler-Peck and Turvey (2000) and Tucker and Brown (2003), the choice of the response categories appeared to be based on the geometrical properties of the objects. All participants, in fact, associated the response glass with smaller plates than metal, and the vast majority of them associated the response wood with smaller plates than plastic. The possible sources for the inconsistencies between the present results and those of Kunkler-Peck and Turvey (2000) and Tucker and Brown (2003) concerning geometry effects, will be addressed in Section 6.4.2.

In the introduction, it was hypothesized that the inability of listeners to discriminate between wood and plexiglas, and between steel and glass, were caused by the absence of acoustical differences between the investigated steel and glass signals, on the one hand, and between the wood and plexiglas signals, on the other. However, it must be pointed out that, even though no significant acoustical differences were present between these two sets of signals, the observed strong concordance among listeners in associating material types with object size points toward a cognitive origin for these effects, rather than to a procedural bias. Indeed, if these associations were caused by a general tendency in participants to focus on the only source property that carried significant acoustical variations, an equal number of participants associating given material types to opposite sizes would have been expected. This was not the case in the current data.

6.4.2 From Acoustics to Perception

The acoustical basis for the perceptual discrimination between metal and glass, and between wood and plastic was investigated, using the procedure outlined in Section 6.4.1. A different approach was used for the discrimination between wood and plastic on the one hand, and between metal and glass on the other, and will be presented in Section 6.4.2.

Before building the final regression models, a choice concerning the transform of each acoustical predictor was made. This problem is particularly relevant to regression models, because the transform of the predictors affects their association with the behavioral outcome. The investigated transforms were limited to linear (identity transform) and logarithmic for all indices, plus the ERB-rate transform for F. The univariate models with the different transforms of the same predictor were compared. The chosen model should be the one closer to the unknown true probability distribution from which observations were sampled (Golden, 2000). This distance is commonly estimated using the log-likelihood of the model, or penalized versions of this statistic which adjust for the number of parameters in the model (Akaike Information Criterion, AIC), and for the sample size (Bayes Information Criterion, BIC). In the case of the log-likelihood, the higher its value, the closer the model is to the unknown true probability distribution.

Thus, for each acoustical descriptor, the transform for which the regression model had the highest log-likelihood was chosen. For the Lou_{sl1}, Lou_{sl2}, and SCG_{slo} descriptors the logarithmic transform was evaluated using the absolute value of the measures. Table 6.5 reports, for each predictor, the log-likelihood for the linear and logarithmic models, and the log-likelihood for the ERB-rate(F) model. Table 6.6 shows the correlations among the acoustical indices, transformed according to this analysis, for both subsets of stimuli. The upper triangular matrix reports correlations for the metal-glass dataset; the lower triangular matrix reports correlations for the wood-plastic dataset.

Table 6.5: Log–likelihood (LL) of the models computed to select the transform for the acoustical predictors. MG = metal vs. glass dataset; WP = wood vs. plastic dataset. The LL of the models with the selected transform is shown in bold face.

Data set	Acoustical descriptor	LL linear model	LL logarithmic model	LL ERB model
MG	$tan\phi$	-135.71	**-133.10**	
	Dur	-124.33	**-122.66**	
	F	-57.50	**-48.53**	-50.25
	Lou_{att}	-123.61	**-122.02**	
	Lou_{mea}	-86.70	**-85.29**	
	Lou_{sl1}	-89.08	**-88.37**	
	Lou_{sl2}	**-124.79**	-131.52	
	SCG_{att}	-87.93	**-87.84**	
	SCG_{mea}	-127.49	**-127.24**	
	SCG_{slo}	-119.08	**-108.97**	
WP	$tan\phi$	**-135.23**	-135.55	
	Dur	-126.59	**-121.59**	
	F	-105.30	**-101.79**	-103.34
	Lou_{att}	**-127.87**	-127.97	
	Lou_{mea}	**-130.25**	-130.85	
	Lou_{sl1}	-103.58	**-103.02**	
	Lou_{sl2}	**-134.52**	-138.72	
	SCG_{att}	-110.08	**-109.87**	
	SCG_{mea}	**-103.86**	-104.03	
	SCG_{slo}	**-118.63**	-121.96	

Table 6.6: Correlation among acoustical predictors, transformed according to the analysis summarized in Table 6.5. The upper triangular matrix reports correlations for the wood–plastic dataset, and the lower triangular matrix reports correlations for the metal–glass dataset. Significant correlations (df=8, p< 0.05) are shown in bold face.

MG \ WP	$tan\phi$	Dur	F	Lou_{att}	Lou_{mea}	Lou_{sl1}	Lou_{sl2}	SCG_{att}	SCG_{mea}	SCG_{slo}
$tan\phi$		**-0.692**	0.219	-0.49	0.499	0.099	-0.603	-0.047	0.178	-0.476
Dur	**-0.637**		-0.461	0.086	**-0.752**	-0.41	**0.844**	-0.377	**-0.649**	**0.718**
F	0.315	-0.471		**0.722**	0.172	**0.968**	-0.092	**0.917**	**0.802**	**-0.783**
Lou_{att}	**0.693**	**-0.783**	0.596		-0.142	**0.761**	0.345	**0.857**	0.616	-0.355
Lou_{mea}	**-0.804**	0.416	**-0.694**	-0.619		0.074	**-0.902**	0.164	0.493	-0.239
Lou_{sl1}	**0.81**	**-0.669**	**0.786**	**0.847**	**-0.929**		-0.026	**0.884**	**0.784**	**-0.732**
Lou_{sl2}	**0.8**	-0.191	0.456	0.455	**-0.934**	**0.778**		-0.021	-0.357	0.27
SCG_{att}	-0.265	-0.125	**0.749**	0.203	-0.239	0.322	-0.038		**0.852**	**-0.745**
SCG_{mea}	-0.534	-0.081	0.487	0.022	0.213	-0.036	-0.45	**0.764**		**-0.784**
SCG_{slo}	**0.792**	**-0.74**	0.455	**0.637**	**-0.685**	**0.744**	0.54	0.065	-0.354	

Given the presence of strong correlations among predictors, in principle different regression models may account for the same data. This couldn't be highlighted using model selection techniques such as backward elimination, forward selection or stepwise selection. In fact these procedures generate a single model as output, and during the model creation process may discard effects which appear unrelated to the outcome because a strongly correlated predictor is already included. It was thus decided to compute all possible regression models, starting from the univariate ones and progressively increase the number of predictors until at least one of the models was associated with a non-significant goodness-of-fit statistic.

For the metal-glass dataset, the F predictor alone was sufficient to account for the observed data (deviance(8) $= 7.723$, $p = 0.259$; Hosmer-Lemeshow $\chi^2(8) = 9.700$, $p = 0.138$). The probability of choosing the metal response was found to increase with decreasing F.

For the wood-plastic dataset, none of the acoustical predictors alone could account sufficiently well for the data. Five of the two predictor models were associated with non-significant goodness-of-fit statistics (deviance(8) ≤ 13.533, $p \geq 0.060$; Hosmer-Lemeshow $\chi^2(8) \leq 11.948$, $p \geq 0.154$). For the first two models the most important predictor, i.e. that associated with the highest standardized parameter estimate, was F, the second predictor being either Lou_{mea} or Lou_{sl2}. The probability of choosing the response "wood" increased with increasing F and Lou_{mea} and with decreasing Lou_{sl2}, i.e. with faster loudness decays. For the other three models the most important predictor was Lou_{sl1}, the second predictor being either Lou_{mea}, Lou_{sl2} or Dur. The probability of choosing the response "wood" increased with decreasing Lou_{sl1}, and Lou_{sl2}, with decreasing Dur and increasing Lou_{mea}. It is worth nothing that the parameters F and Lou_{sl1} are highly correlated for the wood-plastic dataset. Figure 6.5 shows the selected regression model for the metal-glass dataset, and the F - Lou_{sl2} model for the wood-plastic dataset.

6.4.2.1 Discussion

Consistent with the results of Klatzky et al. (2000), discrimination between glass and metal was based on the frequency of the signals, glass being associated with higher frequencies than metal. The high relevance of frequency for the discrimination between hard materials is also consistent with results from Lutfi and Oh (1997). As pointed out in the introduction, impaired discrimination performance reported by Lutfi and Oh (1997) was due to an excessive weighting of signal frequency. A similar explanation for impaired performance might apply here.

The relevance of frequency for the discrimination between metal and glass is, however, not consistent with results by Roussarie (1999). As pointed out in the introduction, the simplest explanation for this inconsistency is based on the range of variation of frequency within the tested signals sets, one perfect fifth in Roussarie (1999), 4.05 octaves with the present study.

Frequency was also found to explain the perceptual discrimination between wood and plastic, where, consistently with Klatzky et al. (2000), the first category was chosen for

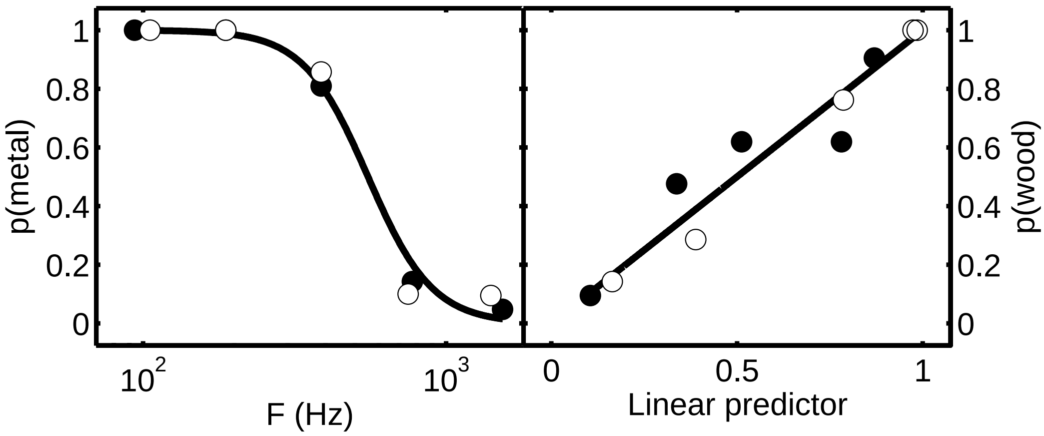

Figure 6.5: Left panel: observed and predicted proportions of choosing the response "metal" as a function of the F parameter. Filled circles = steel; empty circles = glass. Right panel: observed and predicted proportion of choosing the response "wood" as a function of the linear predictor in the F-Lou_{sl2} model. Solid circles = wood; empty circles = plexiglas.

higher frequencies than the second. However, this variable alone was not sufficient to account for observed data. In fact, average signal loudness, and the Lou_{sl2} parameter had a secondary but significant association with participants' responses. The same data were also explained in terms of Lou_{sl1}, strongly correlated with F, and either duration, average loudness, or Lou_{sl2} as secondary variables.

Following a principle of parsimony, a common acoustical explanation for both the glass-metal and wood-plastic data was sought. It can be concluded, then, that these discriminations were based mainly on signal frequency, an acoustical parameter that also explains the relevance of the size of the objects in determining judgments. These effects are also consistent with those of McAdams et al. (2004), where the location of the stimuli along the bar length/density dimensions was explained by signal frequency.

Finally, these analyses demonstrate that the acoustical measure of damping, $tan\phi$, does not account for several auditory material discriminations.

6.4.3 From Acoustics to Physics

It was finally tested whether any acoustical decision rule existed that would have allowed participants to achieve perfect identification performance. Given the perfect auditory discrimination of steel and glass, on the one hand, and wood and plexiglas on the other, this question is equivalent to asking which are the most plausible acoustical indices used by participants to make this judgment.

As anticipated in Section 6.3, $tan\phi$ perfectly discriminated among materials. However, in Section 6.4.2 it was found that this parameter could not explain behavioral data for the metal-glass, and wood-plastic discriminations, thus casting some doubt on the perceptual relevance of this variable for these judgments. Consequently, an analysis of perfect performance decision rules for these discriminations was carried out without considering the $tan\phi$ parameter.

Given two categories, and one acoustical parameter, the perfect decision rule can be assumed to correspond to a threshold value, above and below which objects belong to one and only one category. Given two acoustical parameters the threshold can be assumed to be a line in a bidimensional space. Logistic regression models were used, considering material type as dependent variable. Those models associated to perfect identification performance were sought. This outcome defines complete separation, i.e. perfect prediction of the dependent variable (Albert & Anderson, 1984). When none of the acoustical parameters alone yielded complete separation, bivariate models were considered. Given that the threshold, in the bivariate case, was assumed to correspond to a line in the acoustical space defined by a pair of acoustical descriptors, their ability to discriminate perfectly among materials might have been influenced by a change in their transform. Thus the same transform selected in Section 6.4.2 was adopted. Table 6.7 shows, for each of the investigated discriminations, the acoustical variables found to yield complete separation.

Four different response criteria lead to optimal discrimination between steel and glass on the one hand, and between wood and plexiglas on the other. The first of these gross categories was thus associated with higher values of the Lou_{att}, Dur, and SCG_{slo} descriptors, and with lower values of the $tan\phi$ parameter. It is highly likely that at least one of these acoustical parameters had been used by participants to make this discrimination. Several optimal criteria, based on pairs of acoustical descriptors, were found for the steel-glass (11 pairs) and wood-plexiglas discriminations (8 pairs). Two different criteria could have led to perfect material identification in all investigated discriminations, the first based on Dur and SCG_{mea}, the other based on Lou_{att} and F. Figure 6.6 shows the optimal discrimination criteria based on the Dur and SCG_{mea} parameters. Given the results of these analyses, it can be concluded that sufficient information for perfect material identification was present in the investigated signals, but apparently not used by listeners.

It is also interesting to compare optimal and behavioral weighting of F. For the steel-glass discrimination, three of the optimal criteria associated steel with higher F values. Participants weighted this variable in the opposite way, associating metal with lower frequencies. For the wood-plexiglas discrimination the optimal criterion based on F associated wood to lower F values than plexiglas, while participants weighted this variable in the opposite way.

6.4.3.1 Discussion

Several indices were found to account for the perceptual discrimination between the super-ordinate categories of metal-glass and wood-plastic. Among them, the $tan\phi$ parameter which, consistently with results by Klatzky et al. (2000), and Avanzini and Rocchesso (2001a), had a lower value for the glass-metal category than for the wood-plastic cate-

Table 6.7: Acoustical descriptors found to discriminate perfectly between the contrasted materials. For each acoustical descriptor the sign of the association with the bold face category is also shown (e.g., the model in the bottom row shows that wood is associated with a combination of higher Lou_{att}, and higher SCG_{slo} values).

Dataset	Acoustical descriptors
Steel or Glass	$Dur(+)$
vs.	$Lou_{att}(+)$
Wood or Plexiglas	$SCG_{slo}(+)$
	$Dur(+)\ SCG_{att}(+)$
	$Dur(+)\ SCG_{mea}(+)$
	$Dur(+)\ SCG_{slo}(+)$
	$F(+)\ Lou_{att}(-)$
Steel	$F(+)\ Lou_{sl1}(+)$
vs.	$F(+)\ SCG_{slo}(+)$
Glass	$Lou_{att}(-)\ SCG_{att}(+)$
	$Lou_{att}(-)\ SCG_{mea}(+)$
	$Lou_{sl1}(+)\ SCG_{att}(+)$
	$SCG_{att}(+)\ SCG_{slo}(+)$
	$SCG_{mea}(+)\ SCG_{slo}(+)$
	$F(-)\ Lou_{att}(+)$
	$Dur(+)\ Lou_{att}(+)$
Wood	$Dur(+)\ Lou_{mea}(+)$
vs.	$Dur(+)\ Lou_{sl1}(-)$
Plexiglas	$Dur(+)\ SCG_{mea}(+)$
	$Lou_{att}(+)\ Lou_{sl1}(+)$
	$Lou_{att}(+)\ SCG_{att}(-)$
	$Lou_{att}(+)\ SCG_{slo}(+)$

gory. Other parameters, however, were able to explain the same discrimination, Dur, Lou_{att}, and SCG_{slo}. It should be noted that, in two of the three experiments reported by McAdams et al. (2004), SCG_{slo} was found to explain the perceptual relevance of material damping, a criterion used to judge dissimilarity of impact sounds. Given the fact that multiple indices explain this discrimination equally well, no conclusions can be drawn concerning which of them is used to discriminate between these super-ordinate material categories. Thus it cannot be excluded that for this judgment $tan\phi$ is attended to by listeners.

This conclusion, however, can't be drawn for the discriminations within the gross categories. Indeed, although $tan\phi$ was found, for the investigated stimuli, to discriminate perfectly among all materials, the observed absence of perfect identification revealed this acoustical parameter not to be used to discriminate between wood and plexiglas, and

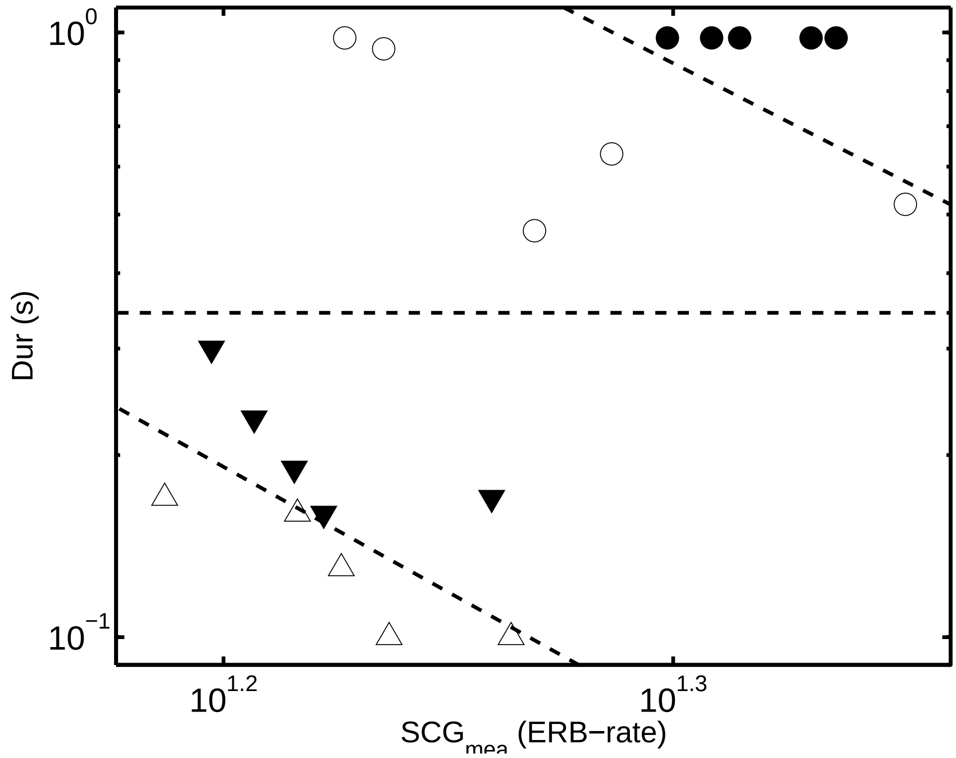

Figure 6.6: Optimal criteria for material categorization. Dashed lines show the equal probability boundaries (thresholds) for the optimal criteria. Black circles = steel, white circles = glass; black triangles = wood, white triangles = plexiglas.

between steel and glass. It has been shown that multiple decision rules, based on pairs of acoustical indices, would have allowed perfect discrimination within the gross categories. Thus acoustical information for perfect identification performance was present, so that we can rule out the hypothesis that participants' focusing on plate size was caused by the absence of acoustical differences between materials in the same gross category.

Among the optimal decision rules for the discrimination between steel and glass, three made use of frequency, where the optimal criterion, contrary to the behavioral criterion, associated glass signals with lower frequencies than steel. Also one of the optimal criteria for the discrimination between wood and plexiglas was based on frequency, and, similarly to what was found for the steel-glass discrimination, the behavioral weighting was opposite with respect to the optimal one. Indeed, participants associated wood with higher frequencies, where the optimal criterion associated this category with lower frequencies. Thus two causes for impaired performance might be found. First, the wrong weighting of signal frequency, second the absence of focus on the other acoustical parameters necessary for perfect identification, such as attack loudness.

Why the impaired wood-plexiglas discrimination was not observed by Kunkler-Peck and Turvey (2000), in contradistinction to the current results and those of Giordano (2003) and Tucker and Brown (2003) remains unclear. Two explanations seem reasonable. Kunkler-Peck and Turvey (2000) generated stimuli live, while in the other studies recorded stimuli were used. Live generation might have provided additional information for material type. Firstly, the repetitions provided to the participants were not acoustically identical, thus favoring the extraction of invariant acoustical information specific to material type. Secondly, manipulation of the plates (hanging them on the supporting device after each trial) might have generated additional acoustical signals (e.g. scraping sounds) potentially informative with respect to the object's material.

Another point which remains to be explained is the fact that with our data geometrical properties of the objects were found to significantly affect responses, while this result was not observed by Kunkler-Peck and Turvey (2000) or Tucker and Brown (2003). A range explanation might be used, where the shape factor manipulated in Kunkler-Peck and Turvey (2000) and Tucker and Brown (2003) caused less acoustical variations than the area variation for the signals investigated in the current research. This would explain also why Giordano (2003) found area, but not shape, to influence the wood-plexiglas and steel-glass discriminations.

Several hypotheses can be advanced concerning the origin of the observed response profiles, based on the properties of the impacted sound sources experienced on an everyday basis, i.e. on the regularities of everyday listening. Available measures of the mechanical properties of a wide range of engineering materials show that plastics (polymers) and woods are characterized by large differences when compared to metals and glasses (Waterman & Ashby, 1997). Given these large differences among material properties, and their influence on signal properties (Fletcher & Rossing, 1991), it follows that, despite variations in the geometrical properties of the source, signals originating from wood and plastic objects would always be differentiated from those originating from metal and glass objects. For this reason, both the absence of effects of geometrical properties on auditory discrimination among the gross categories, and the perfect performance observed with the almost totality of participants, do not appear surprising.

Several speculative hypotheses can be advanced for the origin of the material-size associations for the discriminations within the gross categories. The simplest are based on the geometrical properties. The glass impact sounds we may experience everyday are probably generated by smaller objects than the metal ones (e.g., klinking glasses): large, freely vibrating glass objects, such as those used to generate part of the investigated stimuli, in fact would probably be too fragile to be of any ordinary use. The validity of an explanation based on object size might be questioned because we also experience signals generated by striking small metallic objects such as coins or keys in everyday life. These sources, however, generate more complex acoustical events than those investigated with the current research, being characterized by multiple, rather than single impacts, interleaved with signals generated by more complex interactions among objects (e.g., friction). However, generalization of source recognition effects demonstrated with impact sounds to signals generated by different interactions among solid objects does not appear legitimate. Consequently the size explanation for the glass-metal identification still seems valid. However, an explanation based on object size does not appear convincing in explaining the wood-plastic discrimination.

Given that the wood-plastic discrimination was also based mainly on frequency, in principle any source property that significantly affects this signal property might be a potential candidate to explain the observed effects.

An increase in the frequency of the vibrational modes of an object is determined, for example, by a decrease in the size of the objects, an increase in thickness, a decrease in density, and an increase in Young's modulus or simply in stiffness (cf. Fletcher & Rossing, 1991). One might therefore hypothesize that the observed association of wood with higher frequencies than plastic might be due to the fact that the wood sounds we experience everyday are generated by thicker objects than for plastic sounds. This might be true, given that with plastic materials production of thin layers should be easier than with woods. The same explanation does not seem reasonable for the metal-glass case. Concerning Young's modulus, one should expect woods and glasses to be characterized by higher stiffness than, respectively, plastics and metals. Such a clear difference is not however found in measures of engineering materials (Waterman & Ashby, 1997). Alternatively, one might explain these effects in terms of density, where one would expect glass to be typically characterized by a lower density than metals, and plastic to be characterized by a higher density than woods. This is indeed what is found with engineering materials, where woods are characterized by a lower density than plastics, and metals are characterized, on average, by a higher density than glasses (Waterman & Ashby, 1997). The use of an identical explanation for both the glass-metal, and wood-plastic discriminations, makes this hypothesis particularly attractive.

6.5 Conclusions

Material identification from impact sounds was investigated. All the pairwise relationships between source features, signal properties, and recognized source properties were studied.

Analysis of the relations between source properties and recognition performance highlighted perfect discrimination of the gross material categories metal-glass and wood-plastic, as well as impaired discrimination of materials within the same gross categories. Within these categories participants were found to identify object materials on the basis of size. Only secondary discrepancies among participants responses were observed.

Acoustical criteria for material categorization were also investigated. An acoustical measure of damping, $tan\phi$, was contrasted, in its ability to explain the behavioral data with that of a large set of signal descriptors. This damping measure was found to account only for the discrimination between the gross metal-glass and wood-plastic categories. The same discrimination was equally well accounted for by other signal properties: duration, peak loudness and spectral center of gravity decay rate. Consequently only partial support for the perceptual relevance of $tan\phi$ was found.

Discrimination within the gross categories was found to be based mainly on signal frequency, although the wood-plastic discrimination was found to be equally well accounted for by loudness decay descriptors, and, as secondary variables, signal duration, and average loudness.

Analysis of the relationship between signal and source properties highlighted the presence of sufficient information for perfect discrimination among materials. Observed material recognition biases were hypothesized to be caused by regularities in the sound sources frequently encountered in everyday listening. Several alternatives were discussed.

Acknowledgments

Bruno L. Giordano was supported by the European Commission's Future & Emergent Technologies collaborative R&D programme, Project no IST-2000 25287 (Sob - The Sounding Object), and by a Marie Curie Training Site Fellowship hosted by KTH Music Acoustics in Stockholm, FP5 contract number HPMT-CT-2000-00119.

The authors wish to thank Alain de Cheveigné for providing the code implementing the simulation of the peripheral auditory system processing, and for suggestions concerning the $tan\phi$ extraction procedure.

Chapter 7

Hardness scaling of synthetic impact sounds

Bruno L. Giordano[†], Stephen McAdams[††] and Davide Rocchesso[§]

Abstract

The perceptual independence of hammer and sounding object properties was investigated with reference to the perceptual continuum of hardness using synthetic signals simulating an impact sound source. Perceived hammer hardness was strongly influenced by a parameter specifying the properties of its interaction with the sounding object, but also by variations in the properties of the latter. Perceived sounding object hardness was influenced by sounding object properties and to a secondary extent by the interaction parameter. Results showed limited but consistent abilities supporting the perceptual independence between hammer and sounding object, in spite of the fact that participants were given minimal non-auditory, source-related information. Analogous conclusions were drawn from the study of the acoustical criteria for hardness estimation. Similarities were found between the estimation of sounding object hardness and identification of sounding object material type.

[†]Dipartimento di Psicologia Generale; Università degli Studi di Padova; Via Venezia 8, I-35100 Padova, ITALY and Institut de Recherche et de Coordination Acoustique/Musique (STMS-IRCAM-CNRS);1 place Igor Stravinsky; F-75004 Paris, FRANCE. Electronic mail: bruno.giordano@unipd.it

[††]STMS-IRCAM-CNRS; 1 place Igor Stravinsky, F-75004 Paris, FRANCE and CIRMMT, Faculty of Music; McGill University; 555 Sherbrooke St. W.; Montréal, QC, Canada H3A 1E3. Electronic mail: smc@music.mcgill.ca

[§]Dipartimento di Informatica; Università di Verona; Strada Le Grazie 15, I-37134 Verona, ITALY. Electronic mail: rocchesso@sci.univr.it

7.1 Introduction

The study of sound source perception highlights the mapping between source properties and percepts on the one hand, and acoustical properties and percepts on the other (Li et al., 1991). The aim of this analysis is to give a complete account of the interaction of listeners with their acoustical environment, highlighting abilities and eventual biases in source recognition, as well as the perceptually relevant acoustical variables. Ballas (1993) carried out a survey on the frequency with which sounds other than speech or music are encountered in everyday life, finding impact sounds (e.g., typing, footsteps, car door closing), generated by a brief interaction between two solid objects, to be among the most frequently encountered. Not surprisingly most of the studies in this field have focused on this class of sounds (Gaver, 1988; Freed, 1990; Lakatos et al., 1997; Lutfi & Oh, 1997; McAdams et al., 1998; Roussarie, 1999; Guski, 2000; Klatzky et al., 2000; Kunkler-Peck & Turvey, 2000; Avanzini & Rocchesso, 2001a; Lutfi, 2001; Giordano, 2003; Houix, 2003; Tucker & Brown, 2003; McAdams et al., 2004; Giordano & McAdams, submitted; Grassi, 2005). In all of this research, the acoustical signal was generated by striking a vibrating object, the sounding object, with a highly damped object, the hammer.

An interesting issue emerging from this literature concerns the perceptual independence of hammer and sounding object, i.e. the absence of influence of sounding object properties on hammer perception and vice versa. Several previous studies conducted on isolated impact sounds highlighted perceptual independence, revealing a surprisingly fine tuning of the auditory system to the properties of the impacted sound sources. Roussarie (1999) and Giordano (2003) investigated identification of the material of the sounding object upon variations in the simulated/real hammer[1]. In both studies, participants performance was uninfluenced by hammer variations. Freed (1990) investigated participants' abilities to estimate the hardness of variable-material hammers from the sound they generated striking metallic pans of different sizes. Ratings were scaled with actual hardness and were independent of the size of the pans. McAdams et al. (1998) investigated the perceptual independence of hammer and sounding object with sounds generated by striking variable hardness xylophone bars and variable tension/size tympani membranes with variable hardness hammers. Also, in two of the experiments sounding objects were struck with variable force. Participants were asked to identify either the hammer or the sounding object. When temporal constraints on the response were given (speeded judgment), hammer and sounding object were found not to be perceptually independent. However, when no time limits on the response were given (unspeeded judgment), a condition closer to those investigated in the above-mentioned research, results supported perceptual independence. Results in contrast with the notion of perceptual independence of hammer and sounding object were found by Grassi (2005) in a study conducted on sequences of impact sounds, i.e. bouncing sounds. Participants were presented with sounds of variable size balls bouncing over variable size ceramic dishes, and were asked to estimate the size of the balls. Size estimates were scaled with the actual size of the balls and affected by

[1] Roussarie (1999) modeled the hammer by varying an interaction property, the force stiffness coefficient.

a variation in the size of the sounding object.

A tentative explanation for the disagreement among previous studies involves the amount of non-auditory, source-related information available to participants. It might indeed be hypothesized that everyday listeners do not discriminate hammers from sounding objects, and that they can learn to do so by means of proper training. Consistently, McAdams et al. (1998) provided participants extensive feedback on the correctness of the response, Roussarie (1999) demonstrated to participants the acoustical effects of hammer variations asking them to avoid considering hammer differences for their responses, and Freed (1990) showed participants the hardest and softest mallets striking the largest and smallest pans, while Grassi (2005) did not give feedback on response correctness, and did not inform participants concerning the variation in the size of the sounding object. This training-based hypothesis might also explain why McAdams et al. (1998) did not find perceptual independence in the speeded condition, where the imposed time constraints increased the likelihood of participants basing their judgments on over-learned response strategies used in everyday life, which, according to our hypothesis, do not discriminate between hammer and sounding object. However, this explanation is not able to account for all experimental data, as Giordano (2003) found independence even in the absence of training or feedback.

Additional factors might then be taken into account to explain inconsistencies: 1) the extent of the variations in the acoustical structure associated with changes in the hammer or in the sounding object, e.g., with the study by Giordano (2003) the acoustical changes associated with variations in the hammer material might have been insufficient to influence significantly participants responses; 2) the nature of the investigated sounds, e.g., Grassi (2005) investigated bouncing sounds, while all the other studies investigated sets of isolated impact sounds; 3) the type of judgment requested of participants (identification vs. scaling); 4) the source property under estimation (hardness, size, material type). Choosing among these different factors is however not possible with the available data. A good choice, then, would be to test for perceptual independence in untrained participants using the same set of stimuli, the same type of judgment, and the same source property when investigating perception of both objects. These indications were followed in this study, testing perception of the hardness of both the hammer and the sounding object impacted by the hammer.

Two additional relevant issues concern the acoustical and source properties upon which independent hammer and sounding object perception might be based and, more specifically, the source and acoustical properties for hardness estimation. From the acoustical point of view, the high damping of the vibrations of the hammer makes the signal it radiates likely to go undetected by the auditory system, being at best masked by that radiating from the sounding object. Consequently, perception of the properties of both objects might be assumed to rely on the signal radiating from the sounding object.

Following these considerations, hammer properties seem a bad candidate to explain hammer perception, as their influence on the sounding object signal is not direct, but mediated by the properties of its interaction with the latter. Consequently, hammer perception should not rely directly on hammer properties, but on interaction properties.

Among these an interesting candidate is found in the duration τ of the contact between the two objects during the stroke, which increases with increasing hammer stiffness and weight, and with decreasing hammer striking velocity (Landau & Lifshitz, 1981; Chaigne & Doutaut, 1997). The influence of τ on the structure of the radiated signal stands in the fact that those sounding object vibrational modes with a period higher than τ are not excited by the blow (Benade, 1979). Consequently, a decrease in hammer stiffness, and an increase in τ, determine a decrease in the loudness of the radiated signal, and in the amount of energy at high frequencies, or, conversely, in the spectral center of gravity (SCG). Not surprisingly, loudness and SCG-related descriptors (average value and temporal variability of the initial portion of the signals) were reported by Freed (1990) to be strongly associated with hammer hardness estimates. Similar acoustical variables are expected to account for hammer hardness perception in the current study. It should be noted that τ variations are likely to explain at least in part data from Freed (1990) and Grassi (2005), where increasing hammer hardness and weight were associated with decreasing τ values. However, none of the previous studies investigated systematically the influence of interaction properties on hammer perception.

Studies focusing on perception of the properties of the sounding object can be roughly divided into those that investigated geometry perception (Gaver, 1988; Lakatos et al., 1997; Kunkler-Peck & Turvey, 2000; Lutfi, 2001; Houix, 2003; Tucker & Brown, 2003) and those interested in material type identification (Gaver, 1988; Lutfi & Oh, 1997; Roussarie, 1999; Klatzky et al., 2000; Kunkler-Peck & Turvey, 2000; Avanzini & Rocchesso, 2001a; Giordano, 2003; Tucker & Brown, 2003; Giordano & McAdams, submitted). Trivially, in both cases sounding object perception was based on the properties of the sounding object and on the properties of the signal it radiates, namely the frequency and decay time of the spectral components, as well as on the time-variant distribution of energy across the spectrum. In particular, material identification studies found harder materials (metal and glass vs. plastic and wood) to be recognized in sounding objects characterized by lower damping and higher density and elasticity. Also an influence of the size of the objects on material identification was observed, smaller objects being more often categorizzed as glass and wood than as metal and plastic (Giordano & McAdams, submitted). From the acoustical point of view, signals characterized by a longer duration, a higher attack loudness, a lower rate of decay of the SCG, and by lower values of an acoustical measure of damping, later referred to as $tan\phi_{aud}$ (Giordano & McAdams, submitted) were more often identified as harder materials. Also, identification was influenced by the frequency of the lowest spectral component F (Klatzky et al., 2000; Giordano & McAdams, submitted). Finally, the absence of effects of hammer variations observed by Roussarie (1999) and Giordano (2003), points toward a negligible relevance of interaction properties to material identification. Logically, estimation of sounding object hardness should be based on similar source and acoustical properties as for material type identification. This hypothesis is supported by a semantic differential study conducted by Ohta et al. (1999) on impact sounds were stimuli perceived as metallic were rated high on a soft-hard scale. It was further tested in the present study.

Finally, the above analysis would require a reformulation of the concept of independence of the perception of hammer and sounding object, the former, as hypothesized, based on interaction properties as τ, the latter on the properties of the sounding object. It should however be noted that τ is also influenced by the properties of the sounding object, namely its weight and elasticity (Landau & Lifshitz, 1981; Chaigne & Doutaut, 1997). As a consequence, the outcome of perceptual independence, associated with that of a strong link with hammer perception and properties of the hammer/sounding object interaction, would pose a conceptual problem. Indeed, it remains to be explained why a source property influenced by the properties of the sounding object would be ignored when judging its properties.

With this study the perception of hammer and sounding object properties was investigated with reference to the perceptual continuum of hardness using synthetic signals. This allowed us to address the problem of the perceptual independence of hammer and sounding object, testing for the relevance of interaction properties to hammer perception, and investigating the relationship between perception of sounding object hardness and identification of material type.

7.2 Methods

7.2.1 Physically-based impact model

Stimuli were generated using a model of inertial mass hitting a resonating object. An implementation of the model originally proposed by Hunt and Crossley (1975), widely used in applied mechanics and robotics (Marhefka & Orin, 1999) and recently proposed for sound synthesis (Avanzini & Rocchesso, 2001b), was used. In the model, the contact force is expressed as

$$f(x(t), v(t)) = \begin{cases} Kx(t)^\alpha + \lambda x(t)^\alpha \cdot v(t) = Kx(t)^\alpha \, (1 + \mu v(t)) & x > 0, \\ 0 & x \leq 0, \end{cases} \tag{7.1}$$

where $v(t) = \dot{x}(t)$ is the compression velocity, and K and α are the force stiffness coefficient and a geometry-dependent exponent, respectively. The parameter λ is a damping weight, and $\mu = \lambda/K$ is a mathematically convenient term called the "viscoelastic characteristic" by Marhefka and Orin (1999). In the software realization of the model, the parameters K, α, and λ can be directly manipulated, together with the mass of the impacting object (Avanzini, Rath, & Rocchesso, 2002).

The impact model (Eq. 7.1) can be used as a coupling mechanism between two modal resonators. For our purposes, we used the simplified configuration where only one of the two objects actually resonates, the other being just an inertial mass whose displacement is indicated by $x^{(h)}$. According to modal analysis, the resonator is described through equations in which the variables $x_l^{(r)}$ are referred to as the *modal displacements*. Each mode obeys a second-order oscillator equation. Assuming the resonating object has $N^{(r)}$ modes, its displacement at a given point x is given by a linear combination of the modal displacements: $\sum_{j=1}^{N^{(r)}} t_{kj}^{(r)} x_l^{(r)}$. Assuming that the interaction occurs at point

$m = 1 \ldots N^{(r)}$ of the resonator, the continuous-time equations of the coupled system are given by:

$$\begin{cases} \ddot{x}^{(h)} = \dfrac{1}{m^{(h)}}(f_e^{(h)} + f) \\ \ddot{x}_j^{(r)} + g_j^{(r)}\dot{x}_j^{(r)} + \left[\omega_j^{(r)}\right]^2 x_j^{(r)} = \dfrac{1}{m_{jm}^{(r)}}(f_e^{(r)} - f) \quad (\text{for} \quad j = 1 \ldots N^{(r)}) \\ x = x_m = \sum_{j=1}^{N^{(r)}} t_{mj}^{(r)} x_j^{(r)} - t^{(h)} x^{(h)}, \\ v = v_m = \sum_{j=1}^{N^{(r)}} t_{mj}^{(r)} \dot{x}_j^{(r)} - t^{(h)} \dot{x}^{(h)}, \\ f(x, v) = \begin{cases} kx(t)^\alpha + \lambda x(t)^\alpha \cdot v(t), & x > 0, \\ 0, & x \leq 0, \end{cases} \end{cases} \tag{7.2}$$

where the parameters $\omega^{(r)}$ and $g^{(r)}$ are the oscillator center frequencies and damping coefficients, respectively. The parameters $1/m^{(r)}$ control the "inertial" properties of the modal oscillators ($m^{(r)}$ has the dimension of a mass). The terms $f_e^{(h)}, f_e^{(r)}$ represent external forces.

The continuous-time system (7.2) is discretized using the one-step Adams-Moulton method (Lambert, 1993), also known as bilinear transformation. The resulting discrete-time system appears as a parallel bank of second-order low-pass resonant filters, each accounting for one specific mode of the resonator. Details about the discrete-time system have been discussed elsewhere (Rocchesso & Fontana, 2003) and will not be addressed in this paper.

For the purpose of the experiments described in this paper, the resonator was set to have $N^{(r)} = 5$ modes, tuned according to the most prominent resonances of a clamped bar (Fletcher & Rossing, 1991). In the impact model, each mode is also characterized by its decay time t_e (time to reduce the amplitude by a factor e), computed according to the notions of internal and external damping (van den Doel & Pai, 1998) as

$$\frac{1}{t_e} = \frac{f_i}{\tau_d} + \frac{1}{\tau_{ext}} = \pi f_i \tan\phi + \frac{1}{\tau_{ext}} \ , \tag{7.3}$$

where f_i is the modal center frequency, $\tan\phi$ is the internal friction parameter (Wildes & Richards, 1988; Klatzky et al., 2000), and $1/\tau_{ext}$ is the external friction parameter.

Avanzini and Rocchesso (2001b) derived an equation that relates the contact time τ to the physical parameters of the contact model, in the special case where the resonator is a rigid wall (i.e., it does not resonate at all):

$$\tau = \left(\frac{m^{(h)}}{K}\right)^{\frac{1}{\alpha+1}} \cdot \left(\frac{\mu^2}{\alpha+1}\right)^{\frac{\alpha}{\alpha+1}} \cdot \int_{v_{out}}^{v_{in}} \frac{dv}{(1+\mu v)\left[-\mu(v - v_{in}) + \log\left|\frac{1+\mu v}{1+\mu v_{in}}\right|\right]^{\frac{\alpha}{\alpha+1}}}. \tag{7.4}$$

It can be easily shown that the power-law dependence $t_0 \sim (m^{(h)}/K)^{1/(\alpha+1)}$ holds. In order to see how the contact time varies when the resonator is not perfectly rigid, numerical simulations were performed (Avanzini & Rocchesso, 2001b), and it emerged that τ is always higher than the value predicted by equation (7.4), due to the compliance of the struck object.

7.2.2 Synthesis parameters

The model of section 7.2.1 has been implemented as indicated in Avanzini et al. (2002). Table 7.1 reports the model parameters kept constant for all the experimental stimuli. It should be noted that the geometry-dependent exponent α was set as for contacting spheres in Hertz's theory, and that the value chosen for the interaction damping parameter λ characterizes a lossless interaction.

Table 7.1: Parameters of the impact model kept constant for all experimental stimuli.

Parameter	Symbol	Value	Units
Hammer mass	$m^{(h)}$	0.5	kg
Geometry-dependent exponent	α	1.5	
Interaction damping	λ	0	kg /m$^{\alpha}$s
Strike velocity	$\dot{x}^{(h)}(t=0)$	-5	m/s
External friction	τ_{ext}	0.5	s

Modal frequencies $2\pi\omega_j^{(r)}$ were set by multiplying the lowest modal frequency F by $\{1, 6.26, 17.54, 34.37, 56.82\}$. Three values were used for F (50, 200, 800 Hz), interpreted as modeling variations in bar length and/or density and/or elastic properties. Three values were used for the internal friction parameter $tan\phi$ (31.83, 7.96, 1.99 $\times$ 10^{-}3), interpreted as modeling variations in the material of the bar (Wildes & Richards, 1988; van den Doel & Pai, 1998). Three values were used for the force stiffness coefficient K (5e+006, 2.24e+008, 1e+010 N/m$^{\alpha}$), corresponding, according to equation 7.4, to three different values for the contact time τ (2.55 0.56 0.12 ms), and interpreted as modeling the properties of the hammer-bar interaction. In particular increasing values of K, and decreasing values of τ could be interpreted as modeling an increase in the stiffness of both the hammer and the bar. A set of 27 stimuli was synthesized, combining the chosen values for the model parameters F, K, and $tan\phi$.

7.2.3 Acoustical descriptors

Signals were analyzed using the same procedure adopted by Giordano and McAdams (submitted), meant to simulate the processing that takes place in the peripheral auditory system. The first descriptor can be conceived as a global damping measure, which, above all, takes into account the frequency resolution of the auditory system ($tan\phi_{aud}$). Four loudness descriptors and a duration measure Dur were extracted from the temporal function of signals' loudness, in pseudo-sones[2]. Dur was defined as the temporal extent of the signal for which loudness was above a fixed threshold (0.2 pseudo-sones). Loudness descriptors were an attack measure, Lou_{att}, the loudness of the first 10 ms of the signal, an average measure, Lou_{mea}, and the slope of the initial and final portions of the temporal

[2]The unit of measure for loudness is termed pseudo-sone as it is calculated directly on the sound file, without taking into account the actual presentation levels.

function of loudness, respectively Lou_{sl1} and Lou_{sl2}. From the peripheral auditory system model a time-variant measure of brightness was also extracted, the spectral center of gravity SCG, defined as the specific-loudness-weighted average of frequency, measured on the ERB-rate scale (Moore & Glasberg, 1983). Three descriptors were extracted from this measure: an attack value, SCG_{att}, the SCG of the first 10 ms of the signal, an average measure SCG_{mea}, and the slope of the initial portion of the function, SCG_{slo}. Both the SCG and loudness slope measures were extracted by means of linear regression (see Giordano & McAdams, submitted, for a detailed description of the procedure). Finally, the frequency of the lowest spectral component F was also considered, specified by the value of this parameter in the synthesis model. Table 7.2 reports for each stimulus the value of the acoustical descriptors.

Table 7.2: Acoustical descriptors extracted from each signal. p.s.=pseudo-sones. See text for an explanation of the meaning of each acoustical descriptor.

F (Hz)	$tan\phi \times 10^{-}3$	$K \times 10^8$ ($Nm^{-\alpha}$)	$tan\phi_{aud} \times 10^{-}3$	Dur (s)	Lou_{att} (p.s.)	Lou_{mea} (p.s.)	Lou_{sl1} (p.s./s)	Lou_{sl2} (p.s./s)	SCG_{att} (ERB-rate)	SCG_{mea} (ERB-rate)	SCG_{slo} (ERB-rate/s)
50	31.83	0.05	47.92	0.26	1.43	0.48	-16.86	-1.26	13.08	9.25	-43.09
50	31.83	2.24	42.75	0.21	5.42	1.15	-155.47	-2.36	16.20	10.88	-48.48
50	31.83	100	47.11	0.17	8.45	1.62	-307.63	-2.15	20.99	12.21	-89.58
50	7.96	0.05	17.16	0.64	1.50	0.51	-5.83	-0.61	13.18	9.50	-9.92
50	7.96	2.24	12.25	0.61	6.03	1.05	-79.62	-0.89	16.32	11.12	-12.00
50	7.96	100	11.76	0.56	10.59	1.31	-208.16	-0.84	21.03	11.87	-30.77
50	1.99	0.05	8.10	1.20	1.55	0.57	-2.32	-0.38	13.22	10.01	-2.67
50	1.99	2.24	4.42	1.38	6.27	1.14	-23.52	-0.45	16.36	11.90	-4.59
50	1.99	100	3.78	1.31	10.74	1.59	-54.01	-0.46	21.04	13.22	-9.47
200	31.83	0.05	41.78	0.20	1.24	0.53	-11.09	-2.36	13.70	10.39	-157.26
200	31.83	2.24	44.75	0.23	4.08	0.90	-123.86	-2.33	16.15	10.66	-90.57
200	31.83	100	54.28	0.19	7.32	1.18	-258.58	-2.38	21.65	11.65	-204.05
200	7.96	0.05	12.37	0.60	1.31	0.53	-3.73	-0.74	13.85	10.34	-18.16
200	7.96	2.24	12.20	0.69	4.65	0.85	-57.34	-0.76	16.39	10.69	-21.03
200	7.96	100	14.00	0.57	8.40	0.98	-247.30	-0.77	22.10	11.85	-45.75
200	1.99	0.05	5.54	1.26	1.37	0.54	-1.81	-0.35	13.92	10.35	-4.67
200	1.99	2.24	4.65	1.48	4.89	0.92	-16.27	-0.35	16.47	11.24	-6.37
200	1.99	100	4.53	1.22	9.57	1.15	-88.15	-0.38	22.25	13.09	-17.84
800	31.83	0.05	52.87	0.04	0.63	0.38	-18.25	-14.15	15.02	14.46	-110.05
800	31.83	2.24	48.65	0.08	3.81	1.37	-122.72	-12.44	16.29	15.9	-180.42
800	31.83	100	42.65	0.09	6.13	1.86	-176.43	-9.76	20.62	16.76	-274.37
800	7.96	0.05	17.42	0.09	0.66	0.34	-11.81	-3.13	15.22	16.00	-95.58
800	7.96	2.24	12.00	0.29	4.25	1.16	-43.96	-2.48	16.46	16.32	-144.78
800	7.96	100	10.13	0.31	6.83	1.55	-119.04	-2.43	21.08	16.74	-41.91
800	1.99	0.05	5.76	0.25	0.67	0.31	-5.63	-0.91	15.28	16.75	-179.01
800	1.99	2.24	3.55	0.88	4.41	1.14	-13.70	-0.83	16.52	16.35	-127.88
800	1.99	100	2.97	0.94	7.69	1.52	-39.42	-0.83	21.26	16.80	-11.78

The relationship between synthesis parameters and the structure of the generated signals was studied with the experimental stimuli. The purpose of this analysis was to test for the presence of sufficient acoustical information for the perceptual independence of hammer and sounding object, i.e. for the presence of acoustical parameters specifying uniquely the investigated parameters of the synthesis model. Univariate ANOVA models were computed, with the synthesis parameters as independent variables, and the acoustical descriptors as dependent variables. The F measure of signal frequency was not considered among the dependent variables. The η_p^2 measure of effect size was adopted (Cohen, 1973) to highlight which acoustical descriptor had the strongest association with each of the synthesis parameters. With univariate ANOVA models η_p^2 is equivalent to R^2, the proportion of variance in the dependent variable explained by the independent variable. The results of this analysis are shown in Table 7.3.

Table 7.3: Univariate ANOVA models computed to study the relationship between synthesis parameters and acoustical structure. For each descriptor, the F statistic and the associated p-value are reported. For all models the degrees of freedom of the F statistic are (2,24). For each model, the η_p^2 measure of effect size is also reported. Particularly large effects ($\eta_p^2 > 0.7$) are shown in bold face.

	Synthesis parameters					
Acoustical	F		$tan\phi$		K	
descriptor	F/p-value	η_p^2	F/p-value	η_p^2	F/p-value	η_p^2
$tan\phi_{aud}$	0.00/1.000	0.000	470.41/<0.001	**0.975**	0.04/0.957	0.004
Dur	2.20/0.132	0.155	32.98/<0.001	**0.733**	0.21/0.810	0.017
Lou_{att}	0.76/0.476	0.060	0.20/0.822	0.016	101.27/<0.001	**0.894**
Lou_{mea}	0.72/0.498	0.057	0.18/0.837	0.015	54.81/<0.001	**0.820**
Lou_{sl1}	0.35/0.710	0.028	3.74/0.039	0.238	14.19/<0.001	0.542
Lou_{sl2}	5.08/0.014	0.297	6.84/0.004	0.363	0.03/0.969	0.003
SCG_{att}	0.12/0.890	0.010	0.02/0.981	0.002	335.63/<0.001	**0.965**
SCG_{mea}	73.69/<0.001	**0.860**	0.22/0.801	0.018	1.19/0.323	0.090
SCG_{slo}	5.87/0.008	0.329	5.92/0.008	0.330	0.06/0.943	0.005

Variations in F were associated with significant variations in Lou_{sl2}, SCG_{mea}, and SCG_{slo}. Among these, SCG_{mea} had the strongest association with F, and was not significantly affected by variations in the other synthesis parameters. In particular SCG_{mea} was increased with increasing F. Variations in $tan\phi$ were associated with significant variations in $tan\phi_{aud}$, Dur, Lou_{sl1}, Lou_{sl2}, and SCG_{slo}. Among these $tan\phi_{aud}$ and Dur had a particularly strong association with $tan\phi$, where the first increased and the second decreased with increasing $tan\phi$. Also, these two parameters were not significantly affected by F and K. Finally K significantly affected Lou_{att}, Lou_{mea}, Lou_{sl1}, and SCG_{att}. Lou_{att}, Lou_{mea}, and SCG_{att} had the strongest association with K, all of them increasing with increasing K. None of these acoustical descriptors were significantly affected by F and $tan\phi$. In conclusion, sufficient acoustical information for independent perception of hammer and sounding object was highlighted, each of the synthesis parameters affecting strongly different signal properties.

7.2.4 Procedure

Stimuli were presented through Sennheiser HE60 headphones, connected to a Sennheiser HEV70 amplifier, which received the output of the sound card of the PC used to program the experiment. Stimulus presentation and data collection were programmed into the Mathworks Matlab environment. Participants sat inside a silent room. Signal peak level ranged from 41 to 88 dB SPL.

Participants were told they had to judge sounds generated by the interaction of two objects, a hammer, which does not vibrate after the impact, and a sounding object. They were then described verbally a few hammer/sounding object impacted sounds sources (e.g., finger tapping on a glass), and they were asked which of the two objects was the hammer and which was the sounding object (e.g., the finger is the hammer, the glass is the sounding object). All of them responded correctly with all the examples. Two conditions were investigated, participants being asked to estimate either the hardness of the hammer or that of the sounding object. Correct understanding of the dimension of judgment was tested by asking participants to identify between two sounds the one generated using the hardest hammer or the hardest sounding object. Two hammers (felt and wood) and two sounding objects (metallic and plastic bowls) were used. Sounds were generated live, out of participant sight. Two pairs of sounds were presented, generated, depending on the condition, with the same sounding object or with the same hammer, and with each of the hammers or sounding objects. No feedback on response correctness was given. Finally, participants were asked to estimate the hardness of the hammer/sounding object on a 1 - 100 scale (from really soft to really hard), typing a numerical estimate with the keyboard after presentation of each sound. Before giving the response, participants were allowed to replay the stimulus as many times as needed. The twenty-seven stimuli were presented in blocked-randomized order for each of ten repetitions for a total of 270 trials.

7.2.5 Participants

Fifty-one listeners took part in the experiment on a voluntary basis (age: 19 - 53; 33 females, 18 males). Twenty-four were assigned to the hammer condition, twenty-seven to the sounding-object condition. All of them reported having normal hearing.

7.3 Results

Three of the participants assigned to the sounding object condition failed with one or both of the hardness discrimination trials during the instruction phase. Their data were not considered further. Analyses were conducted on individual estimates averaged across repetitions. Data from the first block of trials, meant to familiarize participants with the task, were not considered.

Inspection of individual data revealed strong differences concerning the effects of the F parameter. Consequently, statistical criteria were used to isolate groups of participants with homogeneous response profiles. A hierarchical cluster analysis (average linkage) was performed on a Euclidean measure of the dissimilarity among individual response profiles. The choice of the number of clusters to be extracted from the hierarchical solution was based on the analysis of the variation of a set of statistical indices across partitioning levels, measuring the goodness-of-fit between the input data and the resulting clustering partitions (Milligan, 1996). Among the available indices, a subset was chosen that had been found to have superior performance in recovering the correct number of clusters (Milligan, 1981; Milligan & Cooper, 1985): the Calinski-Harabasz index (Calinski & Harabasz, 1974), the Goodman-Kruskal γ (Baker & Hubert, 1972), and the point biserial correlation (Milligan, 1980). For all the indices better partitions are characterized by higher scores. Following the approach suggested by Gordon (1999), indications concerning the correct number of clusters were sought in local maxima across partition levels, and the correct number of clusters was established on the basis of the concordance among indices. Figure 7.1 shows the value of the three indices as a function of the number of clusters. Local maxima or minima used to extract the final number of clusters are also shown.

The final number of clusters was taken to be equal to two as this partitioning level was indicated by all three indices. The two clusters contained 12 and 36 participants, respectively. Of the participants in the first cluster, seven had been assigned to the hammer condition, five to the sounding-object condition. The distribution of participants in the two clusters was independent of the experimental condition ($\chi^2(1)=0.444$, p=0.505).

A repeated-measures ANOVA model was computed with the synthesis variables as within-subjects factors and the condition and cluster belongingness as between-subjects factors. The interaction between condition and cluster belongingness as well as the one-way effect of condition were not significant ($p \geq 0.061$), while the effect of the cluster belongingness factor was significant ($F(1,44) = 20.696$, $p < 0.001$), indicating the uninteresting tendency of participants in the second cluster to emit lower estimates than those in the first. All the interactions between within- and between-subjects factors,

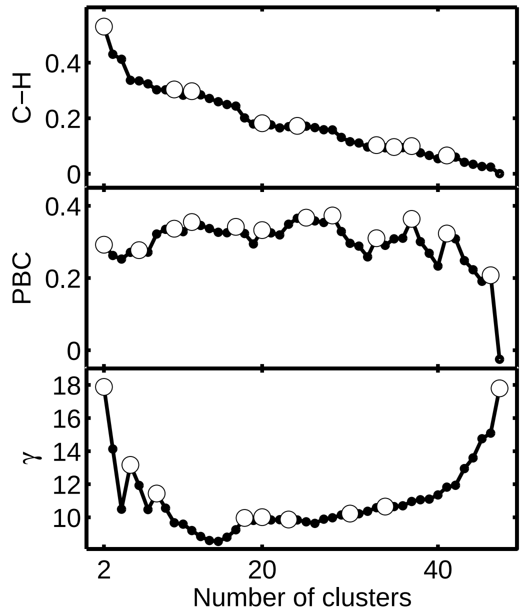

Figure 7.1: Statistical indices used to evaluate the number of clusters present in the dataset across partitioning levels. White circles mark local maxima. C-H = Calinski-Harabasz index; p.b.c = point biserial correlation; γ = Goodman-Kruskhal γ.

which included both between-subjects variables, were not significant ($p \geq 0.200$). These results point toward the orthogonality of the effects of interindividual differences and of instructions. Table 7.4 reports the significance of all the other effects in the model, as well as the η_p^2 measure of the size of each effect (Cohen, 1973).

Table 7.4: General ANOVA model computed to investigate the effect of synthesis parameters, of the experimental condition, and of the cluster belongingness factor on hardness estimates. Significant p-values are shown bold.

			Interaction with condition		Interaction with cluster	
Effect	F(d.f.)	p-value/η_p^2	F(d.f.)	p-value/η_p^2	F(d.f.)	p-value/η_p^2
F	6.925(2,88)	**0.003/0.125**	1.681(2,88)	0.192/0.037	49.728(2,88)	<**0.001/0.531**
$tan\phi$	36.098(2,88)	<**0.001/0.451**	0.667(2,88)	0.516/0.015	15.834(2,88)	<**0.001/0.265**
K	75.937(2,88)	<**0.001/0.633**	7.708(2,88)	**0.001/0.149**	1.297(2,88)	0.279/0.029
$F \times tan\phi$	0.282(4,176)	0.889/0.006	0.495(4,176)	0.740/0.011	2.283(4,176)	0.062/0.049
$F \times K$	4.550(4,176)	**0.002/0.094**	0.941(4,176)	0.442/0.021	21.940(4,176)	<**0.001/0.333**
$tan\phi \times K$	19.170(4,176)	<**0.001/0.303**	0.595(4,176)	0.667/0.013	2.545(4,176)	**0.041/0.055**
$F \times tan\phi \times K$	0.980(8,352)	0.451/0.022	0.460(8,532)	0.884/0.010	3.382(8,352)	**0.001/0.071**

In general all synthesis variables affected hardness estimates. However their effect was modulated by the condition and by idiosyncratic response tendencies. Two additional repeated-measures ANOVA models were computed to study the significant interactions of within-subjects factors with synthesis parameters, analyzing separately the effects of condition and cluster belongingness. The results of these analyses are shown in Tables 7.6 and 7.5.

Table 7.5: ANOVA models created to investigate the interaction between condition, and synthesis parameters. Significant p-values are shown bold.

	Hammer			Sounding object		
Effect	F(d.f.)	p-value	η_p^2	F(d.f.)	p-value	η_p^2
F	2.713(2,46)	0.077	0.106	1.993(2,46)	0.148	0.080
$tan\phi$	29.937(2,46)	**<0.001**	0.566	38.617(2,46)	**<0.001**	0.627
K	56.202(2,46)	**<0.001**	0.710	37.731(2,46)	**<0.001**	0.621
$F \times tan\phi$	0.432(4,92)	0.785	0.018	0.988(4,92)	0.418	0.041
$F \times K$	7.814(4,92)	**<0.001**	0.254	5.242(4,92)	**0.001**	0.186
$tan\phi \times K$	21.266(4,92)	**<0.001**	0.480	11.843(4,92)	**<0.001**	0.340
$F \times tan\phi \times K$	1.159(8,184)	0.326	0.048	1.649(8,184)	0.114	0.067

Independently of the condition, hardness estimates increased with increasing K and with decreasing $tan\phi$ (see Figure 7.2). However, the weight of K in determining estimates differed across conditions, as highlighted by the significance of the interaction between condition and K. This interaction was caused by a change in the weight of K in determining hardness estimates (see Table 7.5). Indeed, as highlighted by the η_p^2 statistic, while $tan\phi$ had a slightly higher relevance than K in determining judgments in the sounding object condition, K was much more relevant than $tan\phi$ in the hammer condition.

Table 7.6: ANOVA models created to investigate the interaction between cluster belongingness, and synthesis parameters. Significant p-values are shown bold.

	Cluster 1			Cluster 2		
Effect	F(d.f.)	p-value	η_p^2	F(d.f.)	p-value	η_p^2
F	24.069(2,22)	**<0.001**	0.686	32.933(2,70)	**<0.001**	0.485
$tan\phi$	1.508(2,22)	0.243	0.121	99.755(2,70)	**<0.001**	0.740
K	41.085(2,22)	**<0.001**	0.789	43.558(2,70)	**<0.001**	0.554
$F \times tan\phi$	1.108(4,44)	0.365	0.091	1.865(4,140)	0.120	0.051
$F \times K$	10.095(4,44)	**<0.001**	0.479	30.968(4,140)	**<0.001**	0.469
$tan\phi \times K$	5.145(4,44)	**0.002**	0.319	28.942(4,140)	**<0.001**	0.453
$F \times tan\phi \times K$	1.550(8,88)	0.152	0.124	4.377(4,140)	**<0.001**	0.111

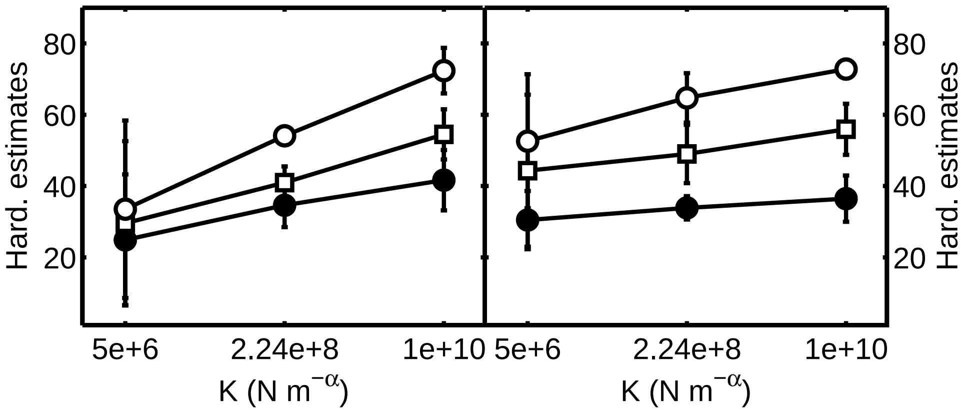

Figure 7.2: Hardness estimates as a function of K with $tan\phi$ as factor. Left panel: hammer condition; right panel: sounding-object condition. Factor level increases from white circles to white squares to black circles. Error bars bracket 95% confidence intervals about the mean.

Separate analyses of data from participants belonging to the two clusters revealed responses to be centered on different synthesis parameters in the two cases. As highlighted by the η_p^2 statistic, participants in the first cluster focused on K, and secondarily on F; participants in the second cluster focused on $tan\phi$, and secondarily on K. The origin of the significance of the interaction of the cluster belongingness factor with the synthesis parameters differs across cases. The rather secondary interaction with the three-way interaction among synthesis parameters is due to the significance of this effect for participants in the first cluster, but not for those in the second cluster. Similarly, the interaction with the $tan\phi$ parameter is due to the significance of the $tan\phi$ effect for participants in the second cluster, where hardness estimates increased with decreasing $tan\phi$, but not for participants in the first cluster. The remaining significant interactions are explained by a change in the shape of effects across clusters. Figure 7.3 shows the interactions between F and K and between $tan\phi$ and K for both clusters.

The origin of the interaction between cluster belongingness and the K-$tan\phi$ interaction stands in the fact that the effect of $tan\phi$ in modulating the effect of K is much stronger for participants in the second cluster. The interaction with the F-K interaction is caused by the fact that stronger effects of K are found for the lowest F level in the first cluster, and for the highest F level in the second cluster. The interaction with F is instead due to the fact that while for participants in the first cluster hardness estimates increased with increasing F, the opposite weighting was given by participants in the second cluster. Finally, independently of cluster belongingness, hardness estimates increased with increasing K.

In summary, the relevance of the synthesis parameters to hardness ratings was influenced by two orthogonal tendencies: idiosyncratic response tendencies and, to a much lesser extent, the object of the estimation (hammer vs. sounding object).

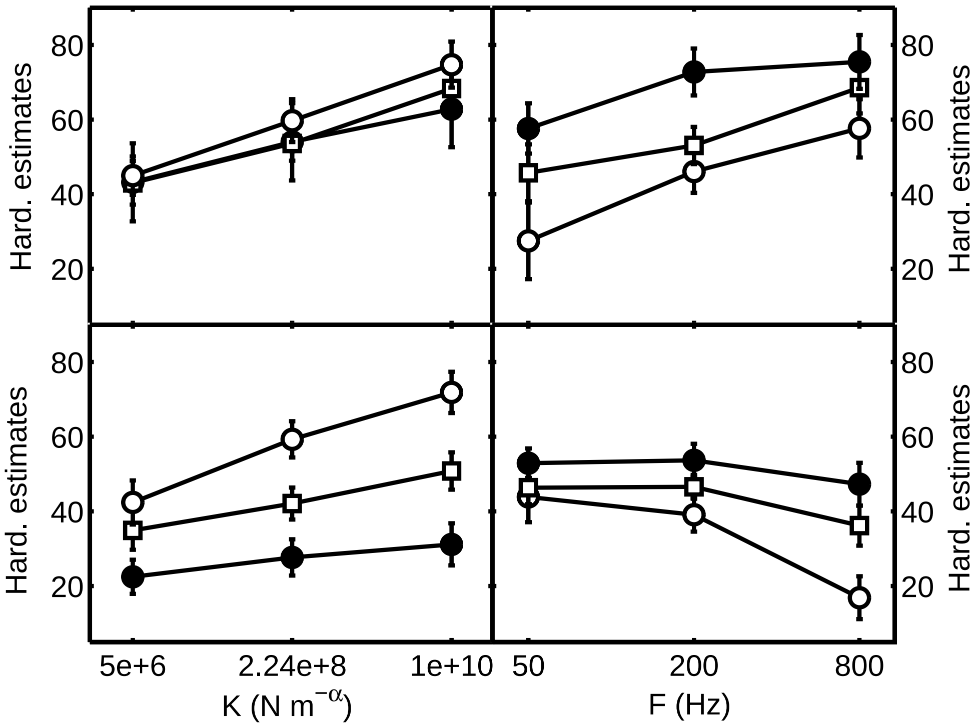

Figure 7.3: Left panels: hardness estimates as a function of K with $tan\phi$ as factor. Right panels: hardness estimates as a function of F with K as factor. Upper panels: cluster 1; lower panels: cluster 2. Factor level increases from white circles to white squares to black circles. Error bars bracket 95% confidence intervals about the mean.

7.3.1 Discussion

Perceived hammer hardness was strongly influenced by variations in the interaction parameter K where, consistently with its physical interpretation, increasing K values were associated with increasing hardness estimates. This effect supports the hypothesized relevance of interaction parameters to the perception of hammer properties, and supports also the hypothesized perceptual relevance of the contact time τ, directly influenced by K. In contrast with data from Freed (1990), hammer hardness was influenced by F. The origin of this inconsistency might well be the fact that Freed (1990) explicitly instructed participants to ignore variations in the size of the sounding object, while participants in the current study did not receive specific instructions in this sense. The perceptual relevance of $tan\phi$ to hammer hardness is instead a novel finding, given that the only study on the perception of this source attribute (Freed, 1990) did not include variations in the material of the sounding object.

Similarities were found between the criteria for sounding object hardness estimation and those for material identification. The $tan\phi$ parameter was indeed relevant in determining sounding object hardness for the majority of participants, increasing hardness estimates being given for decreasing $tan\phi$ values, while with identification studies higher

$tan\phi$ values were associated with the recognition of harder materials (metal and glass, Klatzky et al., 2000; Giordano & McAdams, submitted). The relevance of F in determining sounding object hardness estimates is also consistent with the relevance of F to material identification responses (Klatzky et al., 2000; Giordano & McAdams, submitted). This result is further analyzed in Section 7.4.1. A point of departure between hardness perception and material identification is instead found in the observed relevance of K to sounding object hardness, inconsistent with material identification results by Roussarie (1999) and Giordano (2003).

Limited support for the perceptual independence of hammer and sounding object was found. Indeed, independently of idiosyncratic response tendencies, hammer hardness was found influenced by variations in the sounding object parameters F, and $tan\phi$, and sounding object hardness was found influenced by variations in the interaction parameter K. Then, consistently with results by Grassi (2005) and with the hypotheses outlined in Section 7.1, in absence of a specific training listeners were found unable to tell apart the two objects involved in the generation of the impact sounds. This conclusion is however mitigated by the increased perceptual relevance of K in the hammer condition, which revealed a partial tendency of listeners to focus on different source properties, depending on the object under judgment.

7.4 Acoustical criteria for hardness scaling

The acoustical criteria used by participants to estimate hardness from sound were investigated using regression techniques. Four datasets were modeled separately: participants belonging to the two clusters assigned to the sounding-object and hammer conditions. Regression models were built using average hardness estimates for each of the four datasets. The following procedure was used to build and select the regression models:

1. for each of the α acoustical descriptors, estimate the monotone transform relating α with the hardness ratings ψ, estimating the parameters of the non-linear regression model $\psi = a + b\,|\alpha|^c$;

2. test the significance of the association of each transformed acoustical descriptor with the behavioral outcome, and discard from further modeling those not significantly associated (Hosmer & Lemeshow, 1989);

3. compute all the possible univariate and multivariate regression models in a linear regression framework, raising each of the predictors to the exponent c computed at point 1.;

4. select as final models those with the fewest number of predictors whose adjusted R^2 value is equal to or higher than a threshold value of 0.85, i.e., select the most economical models that fit the observed data well.

The results of this analysis are summarized in Table 7.7. For each predictor in the final regression models the standardized parameter estimate is also reported, where the higher the absolute value of this statistic, the stronger its weight in determining the modeled behavioral response. Figure 7.4 plots the best regression model for each considered dataset.

Table 7.7: Regression models built to study the acoustical criteria for hardness estimation in each of the four considered datasets. Cl. = cluster. For each dataset, the exponent and the standardized parameter estimate for each of the predictors in the model are reported. Also reported is the adjusted R^2 goodness-of-fit measure.

Dataset	Pred.$^{exp.}$	Stand. estim.	Pred.$^{exp.}$	Stand. estim.	$Adj.R^2$
Hammer/Cl.1	$Lou_{att}^{1.01}$	0.871	$F^{2.70\times10^3}$	0.700	0.945
	$SCG_{att}^{-2.34}$	-0.828	$F^{2.70\times10^3}$	0.344	0.893
	$SCG_{att}^{-2.34}$	-0.707	$SCG_{mea}^{6.38\times10^3}$	0.362	0.876
Hammer/Cl.2	$Dur^{9.06\times10^3}$	0.643	$Lou_{att}^{-0.41}$	-0.497	0.883
	$Dur^{9.06\times10^3}$	0.767	$SCG_{att}^{9.46}$	0.460	0.877
	$Lou_{att}^{-0.41}$	-0.632	$\|Lou_{sl2}\|^{-0.24}$	0.590	0.861
Sound.Obj./Cl.1	$F^{18.38\times10^3}$	0.649	$SCG_{att}^{-18.06}$	-0.457	0.852
Sound.Obj./Cl.2	$Dur^{0.31}$	0.966	—	—	0.930

7.4.1 Discussion

In line with results discussed in Section 7.3.1, analysis of the acoustical correlates highlighted limited support for the perceptual independence of hammer and sounding object. Indeed, for the vast majority of participants both acoustical parameters specifying the sounding object, and the acoustical parameters specifying the interaction parameter K were used for hardness estimation. This conclusion is however weakened by two findings. Firstly, the fact that for the majority of participants in the sounding object condition, judgments were based exclusively on signal duration and were uninfluenced by the interaction parameter K (see Section 7.2.3). Secondly, in line with results summarized in Section 7.3.1, two of the acoustical correlates of K, Lou_{att} and SCG_{att}, were much more relevant to hammer hardness rather than to sounding object hardness. Thus, even when only minimal non-auditory, source-related information was available, listeners showed a somewhat limited ability to discriminate acoustically between hammer and sounding object.

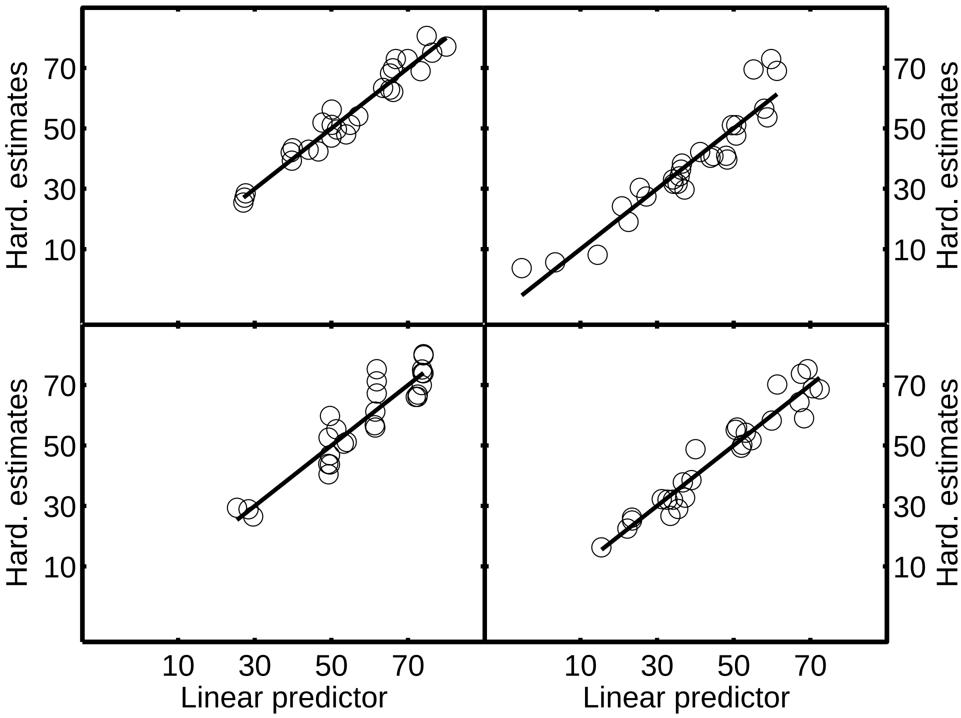

Figure 7.4: Best regression model for each of the analyzed datasets. Observed hardness estimates are plotted as a function of the linear predictor. Solid lines show the regression functions. Top-left panel: hammer condition, cluster 1; top-right panel: hammer condition, cluster 2; bottom-left panel: sounding object condition, cluster 1; bottom-right panel: sounding object condition, cluster 2.

Several results concerning the acoustical criteria for sounding object hardness estimation support the hypothesized link of this perceptual dimension with material identification. In particular, both in the current study and in the study by Giordano and McAdams (submitted), stiffer materials were associated with longer signals. Also consistent is the effect of signal frequency, highlighted by Klatzky et al. (2000) and by Giordano and McAdams (submitted). It should however be noted that, while in the current study frequency and duration were relevant to different groups of participants, Giordano and McAdams (submitted) showed that both of these variables are taken into account by the same listener. The source of this inconsistency is not clear. Another point of departure is found in the relevance of SCG_{att} to sounding object hardness, found to be unrelated to material identification by Giordano and McAdams (submitted). It should however be noted that this acoustical variable was significant for a smaller part of the participants, and had a smaller perceptual relevance than the sounding object parameter F.

Concerning hammer hardness, several inconsistencies are found with respect to the data of Freed (1990). With this study, both signal frequency and duration influenced hammer hardness, while none of these measures seem to explain the data of Freed (1990). In line with discussion of Section 7.3.1, this inconsistency might be explained on the basis of the difference in instructions between the two studies: in Freed (1990) listeners were aided in developing response strategies independently of the acoustical correlates of sounding object variations. Instead, the relevance of loudness and SCG-related descriptors, as well as the relevance of the initial portion of the signals observed by Freed (1990), is consistent with the results of the current study. A more rigorous comparison was then carried out by extracting the SCG_{att} measure from the time-varying analyses published by Freed (1990), and testing its power in explaining hammer hardness estimates. There were insufficient data to test for the relevance of Lou_{att} in explaining Freed's (1990) data[3]. Linear regression was used, with the average hardness ratings for the different mallet/pan-size combinations as dependent variable and the SCG_{att} parameter as predictor. This analysis highlighted a strong association of hardness estimates with $SCGatt$ ($Adj.R^2 = 0.848$), where, consistently with the current study, hardness estimates increased with increasing SCG_{att}. The effects of pan size, and of mallet hardness on SCG_{att} were studied with a two-way ANOVA. In line with results reported in Section 7.2.3, the effect of pan size was just below the critical p-value ($F(3,15) = 2.04$, $p = 0.025$), and had a quite small effect size ($\eta_p^2 = 0.453$) as compared to that of hammer hardness ($F(5,15) = 26.61$, $p < 0.001$, $\eta_p^2 = 0.947$).

7.5 Conclusions

Auditory perception of hardness was investigated for both the objects involved in impact sound generation: the hammer and the sounding object. Perceived hammer hardness was strongly influenced by a property of its interaction with the sounding object, the force stiffness coefficient K, which influences the duration τ of the contact between the two objects during the impact (Landau & Lifshitz, 1981; Chaigne & Doutaut, 1997). This result gives support to the hypothesis according to which hammer perception does not rely directly on the properties of the hammer, but on the properties of its interaction with the sounding object. From the acoustical point of view, hammer hardness estimates were influenced by two of the parameters specifying K, attack loudness and spectral center of gravity SCG, and, depending on idiosyncratic response tendencies, on one of two acoustical parameters specifying the properties of the sounding object: signal duration, strongly related with the $tan\phi$ parameter used to model the material of the sounding object, and frequency. The relevance of the attack SCG was consistent with previous data from Freed (1990), while the relevance of duration and frequency were not. This inconsistency was explained with the difference in experimental conditions which, in the study by Freed (1990), aided participants in avoiding basing their judgments on acoustical correlates of sounding object properties.

[3]The time-varying acoustical analyses published in Freed (1990) were not computed for the experimental stimuli, equalized in loudness, but for the unequalized signals. Even though loudness equalization was not likely to influence strongly SCG_{att}, the likelihood of an influence on Lou_{att} was much higher. Consequently this descriptor was not taken into account.

Several similarities were outlined between judgments of sounding object hardness and identification of sounding object material type. In particular, consistently with results by Giordano and McAdams (submitted), hardness estimates were influenced by the material of the sounding object, as modeled with the $tan\phi$ parameter, and, from the acoustical point of view, by signal duration or frequency. However, while for material identification both signal duration and frequency were used by the same participant for material identification (Giordano & McAdams, submitted), different participants based hardness estimation on one of these two acoustical parameters in isolation. Inconsistently with previous findings on material identification, perceived hardness was influenced by the interaction property K and, for a minority of participants, by the attack SCG. These results, in summary, pointed toward the presence of secondary but relevant differences between judgments of hardness and of material type.

The perceptual independence of hammer and sounding object was finally addressed. Results from the current study support the notion according to which independent perception was observed in previous studies (Freed, 1990; McAdams et al., 1998; Roussarie, 1999; Giordano, 2003) thanks to the focus on different source properties: interaction properties for hammer perception, sounding object properties for sounding object perception. Analysis of the previous literature highlighted the amount of non-auditory, source-related information available to listeners as potentially determining the outcome of perceptual independence. Thus, consistently with results by Grassi (2005), with minimal non-auditory, source-related information participants in the current study did not perceive these properties independently. Nonetheless, the perceptual relevance of K and of the related acoustical parameters increased when participants were asked to estimate hammer hardness rather than sounding object hardness. Thus, partial abilities of independent perception were observed even in naive listeners, abilities that could easily be improved with proper training. It finally remains to be explained why untrained participants chose to rely secondarily on interaction parameters when judging the hardness of the sounding object. Indeed, as pointed out in Section 7.1, both K and τ are influenced by variations in the properties of the sounding object, and, in the specific, in its stiffness. The most plausible reason for this strategy is that listeners chose to focus on source parameters and acoustical variables, which specify unambiguously the sounding object, interaction variables also being influenced by the properties of the hammer.

Chapter 8

Dissimilarity ratings of real impact sounds

Bruno L. Giordano[†] and Stephen McAdams[††]

Abstract

Dissimilarity ratings of impacted sound sources were studied, with reference to both mechanical and acoustical parameters. The perceptual relevance of the properties of the vibrating object, the sounding object, was compared with that of the properties of the striking object, the hammer, and with that of the properties of the interaction between the hammer and the sounding object. Results showed judgments to rely on the properties of the sounding object, and, for the minority of participants, also on interaction properties. From the acoustical point of view, judgments were found to be based on signal duration, on an acoustical measure of damping, on signal frequency and on the attack spectral center of gravity. No evidence for the perceptual relevance of the hammer was found. It was then concluded that in everyday conditions auditory perception relies mainly on the properties of the sounding object. A study of the relationship between acoustical structure and source parameters was also performed, highlighting the acoustical parameters most likely mediating perception of the considered source properties.

8.1 Introduction

The vast majority of sounds we encounter everyday are generated by real objects in interaction. The task of the perceptual system is that of extracting information on

[†]Dipartimento di Psicologia Generale; Università degli Studi di Padova; Via Venezia 8, I-35100 Padova, ITALY and Institut de Recherche et de Coordination Acoustique/Musique (STMS-IRCAM-CNRS);1 place Igor Stravinsky; F-75004 Paris, FRANCE. Electronic mail: bruno.giordano@unipd.it

[††]STMS-IRCAM-CNRS; 1 place Igor Stravinsky, F-75004 Paris, FRANCE and CIRMMT, Faculty of Music; McGill University; 555 Sherbrooke St. W.; Montréal, QC, Canada H3A 1E3. Electronic mail: smc@music.mcgill.ca

the relevant objects in the environment, the sound sources, starting from the properties of the proximal stimulus, the acoustical signal. The problem of source perception has been mainly studied with reference to a particular source, that of isolated impact sounds, composed of two objects: a highly damped object, the hammer, which strikes a vibrating object, the sounding object. Previous studies tested the ability to recover the properties of each of these objects, as well as the properties of the acoustical signals underlying performance. Gaver (1988), Lakatos et al. (1997), Kunkler-Peck and Turvey (2000), Lutfi (2001), Houix (2003), Tucker and Brown (2003) studied perception of the geometrical properties of the sounding object; Lutfi and Oh (1997), Roussarie (1999), Klatzky et al. (2000), Kunkler-Peck and Turvey (2000), Avanzini and Rocchesso (2001a), Giordano (2003), Giordano and McAdams (submitted) studied identification of the material type of the sounding object. Freed (1990) investigated perception of the hardness of the hammer. Finally, the perceptual correlates of the properties of the interaction between hammer and sounding object were studied in Chapter 7.

With the vast majority of these studies, participants were asked to judge explicitly a physical property of the sound source. Such an approach is not free of drawbacks. Firstly, it restricts the tests of perception to linguistic categories immediately understood by the participant, (e.g., material type, size), where perception of source properties whose definition is unclear to the participant can be tested only through alternative linguistic labels. For example, in a study conducted on sequences of impact sounds, i.e. bouncing sounds, Guski (2000) tested perception of the work done by a ball falling onto a drum asking participants to estimate the "impact force". Such a choice might bias judgments in uncontrolled ways, and might undermine the **internal validity** of results, i.e. the correspondence of the measuring variable (e.g., judged impact force) to the measured variable of interest (e.g., the source property work). Secondly, and most importantly, tests based on direct judgment come with a problem of ecological validity, i.e. the possibility of generalizing results to everyday conditions. When explicitly asking for the judgment of a given source property the experimenter is likely to direct the attention of listeners toward that source property. Generalization of these results to everyday conditions is possible only if it is assumed that the source property is attended to even in the absence of these experimental constraints, or if we assume that the overall context of everyday listening has the same effects as the experimental one. Such assumptions might however be inappropriate. Experimental techniques based on judgments of dissimilarity are not affected by these problems, where perception of source properties can be tested without facing issues concerning their linguistic definition, and without directing listeners' attention toward the source property of interest. Typically participants are asked to estimate the dissimilarity of pairwise presented stimuli (e.g. McAdams et al., 2004). Data are then analyzed with multidimensional scaling (MDS) models, that map observed dissimilarities to the distance of stimuli within a geometric representation, usually a Euclidean space with a given number of dimensions (see Borg & Groenen, 1997, for a good introduction to MDS). Dimensions, i.e. the coordinates of stimuli along the axes of the MDS space, are then interpreted with reference to known properties of the stimuli, i.e., source and acoustical properties. The criteria used by participants to rate the dissimilarity are finally assumed to be based on those stimuli properties that explain the dimensions of the spaces.

The technique of dissimilarity rating has been previously used for the study of isolated impact sounds by Gaver (1988), Roussarie (1999) and McAdams et al. (2004). In two separate experiments, Gaver (1988) investigated signals generated striking either real or simulated wood and iron bars of variable length. Two-dimensional MDS spaces were derived in both cases, the first dimension categorically partitioning wood sounds from iron sounds, the second being explained by the length of the bars. Roussarie (1999) investigated signals generated with the simulation of a struck plate. The properties of the plates (damping coefficients, density, elastic coefficients) were varied around those characterizing aluminum and glass. Two sets of signals were created, varying across sets the value of an interaction parameter, the force stiffness coefficient, interpreted as reflecting the properties of the hammer. The lowest and highest values were interpreted as modeling rubber and wood hammers, respectively. Two and three-dimensional solutions, respectively, were derived for the wood hammer, and rubber hammer datasets. In both cases the first dimension strongly correlated with the damping of the plate. Two acoustical parameters were found to explain the first dimension. First, the spectral center of gravity (SCG). Second, a parameter similar to the acoustical measure of damping $tan\phi$ found to be associated with auditory identification of material type (Klatzky et al., 2000; Giordano & McAdams, submitted): the slope of the function relating the damping factors of the spectral components to their frequencies. The second dimension was found to be related to the flexural and torsional wave velocities, as modulated by the elasticity and density of the plates. The second dimension was explained acoustically by two measures of pitch, the frequency of the spectral peak, and the frequency of the lowest resonant mode. Notably the MDS spaces for the two stimulus sets were found to be highly correlated, result interpreted as supporting the absence of perceptual relevance of the force stiffness coefficient or, in the interpretation of the author, of hammer material. McAdams et al. (2004) investigated simulated struck bars. In a first experiment signals were synthesized manipulating the damping and density of the bar. A two-dimensional space was found, the first dimension being related to the damping parameter, and to the SCG temporal decay, the second dimension being related to bar density and signal frequency. Two stimulus sets, judged in separate sessions, were investigated in the second experiment, varying bar damping and length, one of the sets being characterized by lower frequencies than the other. Two-dimensional spaces were found with both datasets. In both cases, one of the dimensions was related to bar length and signal frequency. The other dimension was related to bar damping and to level decay descriptors for the higher frequency set, or to SCG-related descriptors in the lower frequency set. Data from all experiments were explained in terms of signal frequency, related to either bar density or length, and in terms of a linear combination of the average SCG and the level decay in the last portion of the signals, parameters related to bar material.

In summary across studies dissimilarity ratings were based on two independent criteria, related to different source properties. The first criterion was associated with the material type of the sounding object (Gaver, 1988), and to a material type-related property, damping (Roussarie, 1999; McAdams et al., 2004). The second criterion was related to the size of the sounding object (Gaver, 1988; McAdams et al., 2004) or, alternatively, with other sounding object properties that discriminate between material types: density and elasticity (Roussarie, 1999; McAdams et al., 2004). From the acoustical point of view, the damping-related criterion was associated with different acoustical variables: a parameter similar to the acoustical measure of damping $tan\phi$, the overall SCG (Roussarie, 1999; McAdams et al., 2004), the SCG decay rate (McAdams et al., 2004), and the level decay rate (McAdams et al., 2004). The size-related criterion was, to the contrary, invariably related to a frequency component of the signal (Roussarie, 1999; McAdams et al., 2004).

None of these studies compared the perceptual relevance of hammer and sounding object properties. Also the perceptual relevance of the parameters characterizing the interaction between the two objects was only secondarily addressed in the study by Roussarie (1999). Indeed, as the force stiffness coefficient was varied across experimental sessions, participants were never given the possibility to compare different values of this parameter, forcing judgments to be unrelated to this source property. Finally, following the same arguments made by Giordano and McAdams (submitted) concerning the acoustical criteria for material identification, the perceptual relevance of mechanical damping, and of acoustical measures of damping, point to the perceptual relevance of a simple property of the signals, duration. Indeed, the higher the damping of a sounding object, the shorter the signal. However, none of the previous studies related dissimilarity ratings with measures of signal duration.

With this study perception of impact sound sources was investigated, using the experimental technique of dissimilarity rating. Properties of the sounding objects, of the hammers, and of the interaction between hammers and sounding objects were compared in their perceptual relevance, providing a more complete framework to understand perception of impact sounds in everyday conditions. A large set of descriptors to investigate the acoustical criteria for dissimilarity rating was adopted. Finally, a study on the relationship between the properties of the impact sound source and the structure of the acoustical signals was performed in order to highlight the acoustical parameters more likely to mediate perception of the investigated source properties.

8.2 Physical and acoustical characterization of impacted sound sources

A large set of impact sound sources, and impact sounds, was created. The same paradigm adopted in previous research on the perception of impact sounds was used for sound generation (e.g. Lakatos et al., 1997; Kunkler-Peck & Turvey, 2000). Accordingly, the sound source was composed of two objects, a vibrating object, a plate, struck by a highly damped object, referred to as the hammer. All sources were characterized both physically and acoustically. Physical characterization was carried out by measuring the properties of plates and hammers, as well as the parameters characterizing their interaction. Acoustical characterization was carried out by analyzing signals with the same methodology outlined in Giordano and McAdams (submitted). A study of the acoustical correlates of the source properties was finally performed.

8.2.1 Plate properties

Plates were made of seven different materials: aluminum (Alu.), alumina ceramic (Cer.), soda-lime glass (Gla.), oak, pine (Pin.), polymethyl methacrylate or plexiglas (Ple.), and steel (Ste.). All plates were approximately square in shape, and had approximately the same thickness (1 cm). For each material three areas were used, approximately 225, 450, and 900 cm^2. A different steel was used for the largest of the steel plates. Each plate was drilled at four locations with a 3 mm $\emptyset$ screw, for mounting in the device used to strike them (see Section 8.2.3). Two top holes were located 1.5 cm from the top border, and from the left/right borders. Two lateral holes where located at the middle height, 1.5 cm from the left/right borders. Oak and pine plates were cut along the grain, i.e. the grain was parallel to the length of the plates. The largest wood plates were prepared by gluing side by side three planks of the same width. Plate weight was measured with an accuracy of $\pm$ 2 g for weights up to 2500 g, of $\pm$ 5 g for weights from 2500 to 5000 g. Plate area was calculated assuming an angle of $\pi/2$ between the upper and right sides, and between the left and lower sides. Density estimation took into account the volume of the drills. Table 8.1 reports this initial set of plate properties.

8.2.1.1 Estimation of the elastic properties

McIntyre and Woodhouse (1988) and Chaigne and Lambourg (2001) outline a methodology for the rough estimation of the elastic properties of thin plates, based on the measurement of the frequency of a limited number of their vibrational modes. This methodology was used here to estimate the rigidities D, or, equivalently, the Young moduli E, Poisson's ratios ν and the shear modulus G, relating the stresses operating on the plate to the resulting strains or deformations (cf. Lambourg, 1997).

With orthotropic materials, as for pine and oak, elastic properties are symmetric with respect to three orthogonal axes. In the special case of orthotropic thin plates with axes of symmetry parallel to the sides, the case for the investigated wood plates, elastic properties are characterized by four independent elastic constants, the rigidities $D_{1,\ldots,4}$. These can be estimated on the basis of Equations (8.1).

$$D_1 \simeq 0.0789\frac{\rho f_{(2,0)}^2 l_x^4}{h^2}; D_2 \simeq 0.6\sqrt{D_1 D_3}; D_3 \simeq 0.0789\frac{\rho f_{(0,2)}^2 l_y^4}{h^2}; D_4 = f_{(1,1)}^2\frac{4\pi^2}{144}\frac{\rho\left(l_x l_y\right)^2}{h^2} \tag{8.1}$$

where $f_{(1,1)}$ is the frequency of the first torsional mode; $f_{(2,0)}$ is the frequency of the first bending mode along the length l_x of the plate, i.e. along the dimension parallel to the grain; $f_{(0,2)}$ is the frequency of the first bending mode along the width l_y of the plate; h is the thickness of the plate; ρ is the density.

Table 8.1: Properties of the investigated plates.

Material	Area (cm^2)	Thickness (mm)	Left h. (mm)	Right h. (mm)	Upper w. (mm)	Lower w. (mm)	Weight (g)	Density ρ (kg/m^3)
Alu.	224.62	10.0	149.5	150.0	150.0	150.0	598	2665.57
Alu.	448.91	10.0	211.5	212.0	212.0	212.0	1200	2674.83
Alu.	899.55	10.1	299.7	299.5	300.0	300.5	2394	2635.81
Cer.	224.25	9.0	149.0	150.0	150.0	150.0	436	2163.01
Cer.	449.44	9.1	212.0	212.0	212.0	212.0	872	2133.42
Cer.	900.00	9.2	300.0	300.0	300.0	300.0	1720	2077.95
Gla.	225.00	10.0	150.0	150.0	150.0	150.0	562	2500.92
Gla.	449.65	10.0	212.2	212.0	212.0	212.0	1116	2483.48
Gla.	901.50	10.0	300.0	301.0	300.0	300.0	2228	2472.21
Oak	229.50	10.2	150.0	150.0	153.0	153.0	170	727.11
Oak	449.44	10.0	212.0	212.0	212.0	212.0	356	792.60
Oak	894.00	10.2	300.0	300.0	301.0	295.0	788	864.42
Pin.	223.46	10.0	150.0	150.0	149.2	148.75	148	663.14
Pin.	447.85	10.0	212.0	211.0	212.0	211.5	312	697.10
Pin.	886.50	10.0	300.0	300.0	295.0	296.0	592	668.01
Ple.	225.00	10.0	150.0	150.0	150.0	150.0	264	1174.81
Ple.	447.85	10.2	212.0	211.0	212.0	211.5	540	1182.87
Ple.	901.50	10.2	300.0	301.0	300.0	300.0	1104	1200.99
Rub.	226.13	9.0	150.0	150.0	150.5	151.0	312	1534.99
Rub.	449.44	9.0	212.0	212.0	212.0	212.0	644	1593.11
Rub.	903.15	9.5	300.5	300.9	300.5	300.2	1278	1489.99
Ste.	225.00	10.0	150.0	150.0	150.2	149.8	1736	7725.26
Ste.	445.10	10.0	210.9	210.0	211.5	211.5	3460	7778.44
Ste.	900.00	10.1	300.0	300.0	300.0	300.0	7110	7824.24

The Young moduli $E_{x,y}$, Poisson's ratios $\nu_{xy,yx}$, and the shear modulus G_{xy}, are related to the rigidities $D_{1,\ldots,4}$ by Equations (8.2).

$$E_x = \frac{12D_1D_3 - 3D_2^2}{D_3}; E_y = \frac{12D_1D_3 - 3D_2^2}{D_1}; \nu_{xy} = \frac{D_2}{2D_3}; \nu_{yx} = \frac{D_2}{2D_1}; G_{xy} = 3D_4 \text{ with } \frac{E_x}{E_y} = \frac{\nu_{xy}}{\nu_{yx}} \tag{8.2}$$

where it is assumed that $\nu_{xy}\nu_{yx} = 0.3^2$.

With isotropic materials (aluminum, ceramic, glass, plexiglas, steel) physical properties are constant along every direction, and only two independent coefficients are needed to characterize their elastic behavior. In particular Young's modulus E and Poisson's ratio ν can be estimated with Equations (8.3).

$$\nu \simeq 1.48 \left[\frac{f_O^2 - f_X^2}{f_O^2 + f_X^2} \right]; E \simeq 0.46 \left(1 - \nu^2\right) \left(f_O^2 + f_X^2\right) \rho l^4 / h^2; G = \frac{E}{2(1+\nu)}; \quad (8.3)$$

where where f_O is the frequency of the ring vibrational mode given by the in-phase combination of the $(2, 0)$ and $(0, 2)$ bending modes; f_X is the frequency of the X vibrational mode, given by the out-phase combination of the $(2, 0)$ and $(0, 2)$ bending modes; l is the length of the plate. It should be noted that in the isotropic case $E_x = E_y = E$ and $\nu_{xy} = \nu_{yx} = \nu$.

For the isotropic case, the rigidities $D_{1,\ldots,4}$ are related to E and ν by Equations (8.4).

$$D_1 = D_3 = E/12 \left(1 - \nu^2\right); D_2 = 2D_1 - D_4; D_4 = E/6 (1 + \nu); \quad (8.4)$$

8.2.1.1.1 Measurement of the modal frequencies Measurement of the frequencies of the vibrational modes of interest was based on the observation of the patterns of nodal lines. For both isotropic and orthotropic materials the nodal lines of the first torsional mode $(1, 1)$ form a cross with arms parallel to the width and length of the plate, crossed at its center. For orthotropic materials, the (2,0) mode has nodes with, approximately, the shape of two lines parallel to the height, while the two nodal lines for the (0,2) mode are parallel to the width. For isotropic plates the nodal pattern for the X mode is a cross whose arms connect opposite corners of the plate, while the nodal pattern for the O, or ring, mode is a circle centered on the center of the plate.

The Chladni technique was used. Plates were placed on four small cork supports, placed at nodal positions for the modes of interest, covered with particles of a light material, and excited with a sinusoidal signal. Two different setups were used for the different materials. Acoustical excitation was used for aluminum, glass, oak, pine, sillimanite, and steel, where the signal emanated from a loudspeaker placed above the plate. Plexiglas plates were excited mechanically with an LDS model V203 shaker, placed below the plate, and attached to an antinodal location. Both the loudspeaker and the shaker received as input a sinusoidal signal generated with a Leader LAG125 audio generator, amplified with a KH-MB140 amplifier. A Racal Instruments 9911 frequency meter, connected to the tone generator, was used to measure the frequency of the signal. Table 8.2 reports the frequencies of the vibrational modes of interest. Although not used for the calculation of the elastic coefficients, the frequency of the (1,1) mode is also reported for the isotropic plates.

8.2.1.1.2 Estimated elastic coefficients The elastic properties of the different plates were estimated by applying equations 8.1-8.3. For orthotropic plates, the average of the upper and lower width was used as l_y, and the average of the right and left length was used as l_x. For the isotropic plates, the average of the length of all sides was used as l. The resulting measures are reported in Tables 8.3 and 8.4.

Table 8.2: Modal frequencies used to estimate the elastic coefficients of the plates.

Material	Area (cm^2)	$f_{(1,1)}$ (Hz)	f_X (Hz)	f_O (Hz)
Alum.	225	1413.38	2120.47	2690.68
	450	715.96	1073.03	1361.49
	900	358.2	534.21	672.18
Cer.	225	872.62	1252.61	1320.74
	450	468.62	676.97	706.37
	900	215.57	308.73	320.32
Gla.	225	1525.82	2238.15	2623.28
	450	780.38	1141.25	1340.06
	900	393.3	571.89	668.42
Ple.	225	605.12	863.95	1101.5
	450	242.23	439.12	542.05
	900	135.44	198.51	264.44
Ste.	225	1455.76	2153.65	2607.93
	450	751.96	1106.43	1319.65
	900	370.2	556.81	672.88

Material	Area (cm^2)	$f_{(1,1)}$ (Hz)	$f_{(2,0)}$ (Hz)	$f_{(0,2)}$ (Hz)
Oak	225	604.69	1234.74	2411.5
	450	262.88	595.3	1031.65
	900	153.99	333.75	517.38
Pin.	225	630.3	1245.98	2475.24
	450	310.5	635.72	1185.28
	900	147.78	325.97	539.63

Table 8.3: Elastic coefficients estimated for the isotropic plates.

Material	Area (cm^2)	D_1 (GPa)	D_4 (GPa)	E (GPa)	ν
Alu.	225	6.051	7.915	63.919	0.346
Alu.	450	6.209	8.123	65.598	0.346
Alu.	900	5.909	7.869	62.987	0.334
Cer.	225	1.706	3.144	20.342	0.078
Cer.	450	1.91	3.579	22.824	0.063
Cer.	900	1.509	2.853	18.051	0.055
Gla.	225	5.771	8.852	65.492	0.233
Gla.	450	5.964	9.116	67.588	0.236
Gla.	900	5.96	9.191	67.77	0.229
Ple.	225	0.447	0.578	4.695	0.353
Ple.	450	0.425	0.589	4.623	0.307
Ple.	900	0.393	0.461	3.913	0.413
Ste.	225	17.15	24.701	189.681	0.28
Ste.	450	17.519	25.993	196.221	0.258
Ste.	900	18.167	26.272	201.282	0.277

Table 8.4: Elastic coefficients estimated for the orthotropic plates.

Material	Area (cm^2)	D_1 (GPa)	D_2 (GPa)	D_3 (GPa)	D_4 (GPa)	E_x (GPa)	E_y (GPa)	ν_{xy}	ν_{yx}
Oak	225	1.623	0.519	0.461	0.369	17.727	5.031	0.563	0.16
Oak	450	1.344	0.465	0.448	0.303	14.681	4.888	0.52	0.173
Oak	900	1.421	0.543	0.576	0.432	15.521	6.288	0.471	0.191
Pin.	225	1.623	0.483	0.4	0.361	17.722	4.369	0.604	0.149
Pin.	450	1.546	0.499	0.447	0.37	16.884	4.88	0.558	0.161
Pin.	900	1.243	0.437	0.427	0.314	13.576	4.663	0.512	0.176

For each material except steel, further calculations based on the elastic coefficients were made averaging estimates across area levels. For steel, instead, separate estimates of the elastic coefficients were considered for the larger, and for the two smallest plates. In this latter case, estimates were averaged across areas.

8.2.2 Hammer properties

Seven hammers were created of the same material types used for the plates (aluminum, ceramic, glass, oak, pine, plexiglas, steel). The steel hammer was created using the same steel as for the two smallest plates. Their shape was approximately semi-spherical (radius = 1 cm). Weight was measured using a high precision Ohaus galaxy 400D scale (precision: ± 1 ng). The radius was calculated assuming the shape of a spherical cap. For each material the elastic coefficients and density of the hammers were assumed to correspond to those measured on the plates, averaged across areas. For the steel hammer, elastic coefficients and density were averaged across the measures for the two smallest plates. Table 8.5 reports the properties of the hammers. Only the elastic coefficients $D_{1,\ldots,4}$ are shown.

Table 8.5: Properties of the investigated hammers.

Material	Radius (cm)	Weight (g)	Density (kg/m^3)	D_1 GPa	D_2 GPa	D_3 GPa	D_4 GPa
Alu.	1.025	4.213	2658.736	6.056	4.144	6.056	7.969
Cer.	1.063	4.402	2124.795	1.708	0.224	1.708	3.192
Gla.	1.005	5.801	2485.538	5.898	2.743	5.898	9.053
Oak	1.056	1.844	794.71	1.463	0.509	0.495	0.368
Pin.	1.056	1.543	676.084	1.471	0.473	0.425	0.348
Ple.	1.007	1.844	1186.222	0.422	0.3	0.422	0.543
Ste.	1.041	10.697	7751.854	17.335	9.322	17.335	25.347

8.2.3 Recording session

The apparatus used to suspend the plates was similar to that used by Kunkler-Peck and Turvey (2000) and Giordano and McAdams (submitted) (see Figure 6.1). The main structure was made of pine wood. Plates were hung from the top shelf with nylon lines, attached to the top holes. The lateral holes of the plates were attached to two 132 g weights with nylon lines, passing through holes drilled in two horizontal planks attached to both sides of the structure. Hammers were mounted on the bottom end of an aluminum guide, using a small amount of wax. The guide was damped with a heavy piece of garment, in order to prevent the generation of audible signals after the blow of the hammers on the plates. The guide weighed 1.8 kg, garment included. The guide was anchored to the top shelf, 20 cm from the plane of the plates. Plates were struck in their centers, releasing the guide from a fixed angle of 22.5°. No audible multiple impacts of the hammers on the plates were observed during the recording phase.

Sounds were generated in an acoustically isolated room. A Brüel & Kjær type 4003 condenser microphone was positioned 25cm from the center of the plate opposite the struck surface. The signal captured by the microphone was delivered to a Symetrix SX202 microphone preamplifier, connected to a Loughborough Sound Images PC/C32 DSP board. The signal was acquired through the DSP board with a sampling rate of 44100 Hz, and a resolution of 16 bits. One signal for each of the plate/hammer pairs was recorded, for a total of 147 samples.

For each of the recorded signals, additional measures were collected to characterize the hammer/plate interaction. An Endevco model 22 Picomin light accelerometer (weight: 0.14 g) was attached to the back side of the guide and located on the opposite side of the striking surface of the hammers. The acceleration signal was amplified with a Brüel & Kjær type 2635 charge amplifier and delivered to a Tektronix TDS-210 two-channel digital oscilloscope (sampling rate: 1 GHz).

From the acceleration signal, two measures were extracted: the duration τ of the contact between hammer and plate, and the maximum hammer acceleration during the stroke $acc_{\max}$. The beginning of the contact between hammer and plate was clearly marked by a sudden rise of the acceleration signal above the DC value. The end of the contact time corresponded to the instant where acceleration went below the DC value. Figure 8.1 shows the acceleration of the plexiglas hammer striking the 450 cm^2 steel plate, and the beginning and end of the contact between hammer and plate, defined as above.

8.2.4 Interaction properties

A methodology close to that outlined by Chaigne and Doutaut (1997) was used to measure additional parameters characterizing the interaction between hammers and plates, the dynamic mass of the hammers, and the force stiffness coefficient K. The dynamic mass of the hammer is defined as the ratio of the hammer acceleration to the hammer striking force and measures the mass of the guide/hammer system while striking the plate. The relevance of the dynamic mass of the hammer stands on its influence on the amount of energy introduced into the plate. The force stiffness coefficient K, according to Hertz's law of contact, relates the striking force F to the compression δ of the mallet during the contact ($F = K\delta_h^{3/2}$ in the analysis developed by Chaigne and Doutaut, 1997). The relevance of K, dependent on the elastic properties of both the hammer and the plate (cf. Chaigne & Doutaut, 1997), stands on its influence on an acoustically relevant parameter of the hammer/plate interaction, the contact time τ between the hammer and the plate during the impact (Benade, 1979).

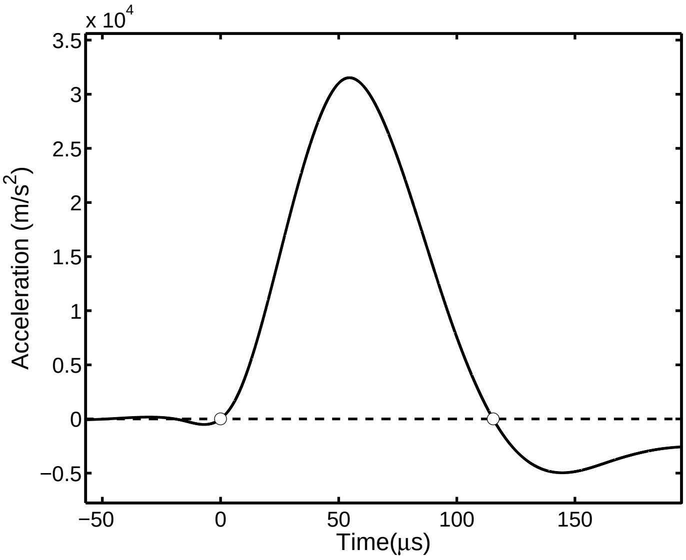

Figure 8.1: Acceleration of the plexiglas hammer striking the 450 cm^2 steel plate. The dashed line shows the DC value of the acceleration signal. White circles mark the beginning and the end of the contact between hammer and plate. Time starts with the beginning of the contact.

8.2.4.1 Dynamic mass of the hammers

The dynamic mass of the hammers was estimated from two signals: hammer acceleration and striking force. An Endevco model 22 Picomin light accelerometer (weight: 0.14 g) was mounted on the back side of the aluminium guide using a small amount of wax. The accelerometer was located exactly on the opposite side of the striking surface of the hammers. The hammer struck a rigidly fixed Brüel & Kjær type 8001 impedance head, which delivered a signal proportional to the striking force. The impedance head was mounted in a position corresponding to the striking location of the plates during the recording session. The acceleration and force signals where delivered to a Brüel & Kjær type 2635 charge amplifier. The amplified signals were then delivered to a Tektronix TDS-210 two-channel digital oscilloscope (sampling rate: 1 GHz) for further measurement.

For each hammer, thirty measures were collected, releasing the guide from a starting angle of 17.5°. The same procedure was also followed with the guide without any hammer mounted. In this case, the impedance head was struck in a slightly lower position than when a hammer was mounted. The hammer dynamic mass was defined as the ratio of the peak striking force to the peak hammer acceleration. Figure 8.2 shows, for each of the hammers and for the guide without a hammer, the peak force as a function of the peak acceleration.

Linear regression was used to study the relationship between the dynamic mass and static mass of the hammers. Dynamic mass measures averaged across the thirty repetitions were considered. The regression equation $\textit{Dynamic mass} = 8.542\text{ g} + 0.99767\ \textit{Static mass}$

accounted for 99.7% of the variance of the average dynamic mass estimates. The hammer dynamic mass was thus found to be given by the static mass of the hammers plus 8.542 grams, i.e., the dynamic mass of the guide. This quantity overestimates the direct estimate of the dynamic mass of the guide (5.699 g). The misalignment between these measures is due to the fact that in this latter case the guide struck the impedance head in a lower location than the hammers, i.e., farther from its fulcrum of rotation. Figure 8.3 shows the average dynamic mass for the different hammers and for the guide without hammers. Average dynamic mass estimates are reported in Table 8.6 along with the regression estimates. Further calculations were based on the regression estimates of the dynamic mass.

Table 8.6: Hammer dynamic mass measures.

Material	Average estimates (g)	Regression estimates (g)
Alu.	12.792	12.745
Cer.	12.755	12.934
Gla.	14.529	14.330
Oak	10.355	10.382
Pin.	10.291	10.082
Ple.	10.175	10.382
Ste.	19.172	19.214
Guide	5.699	8.542

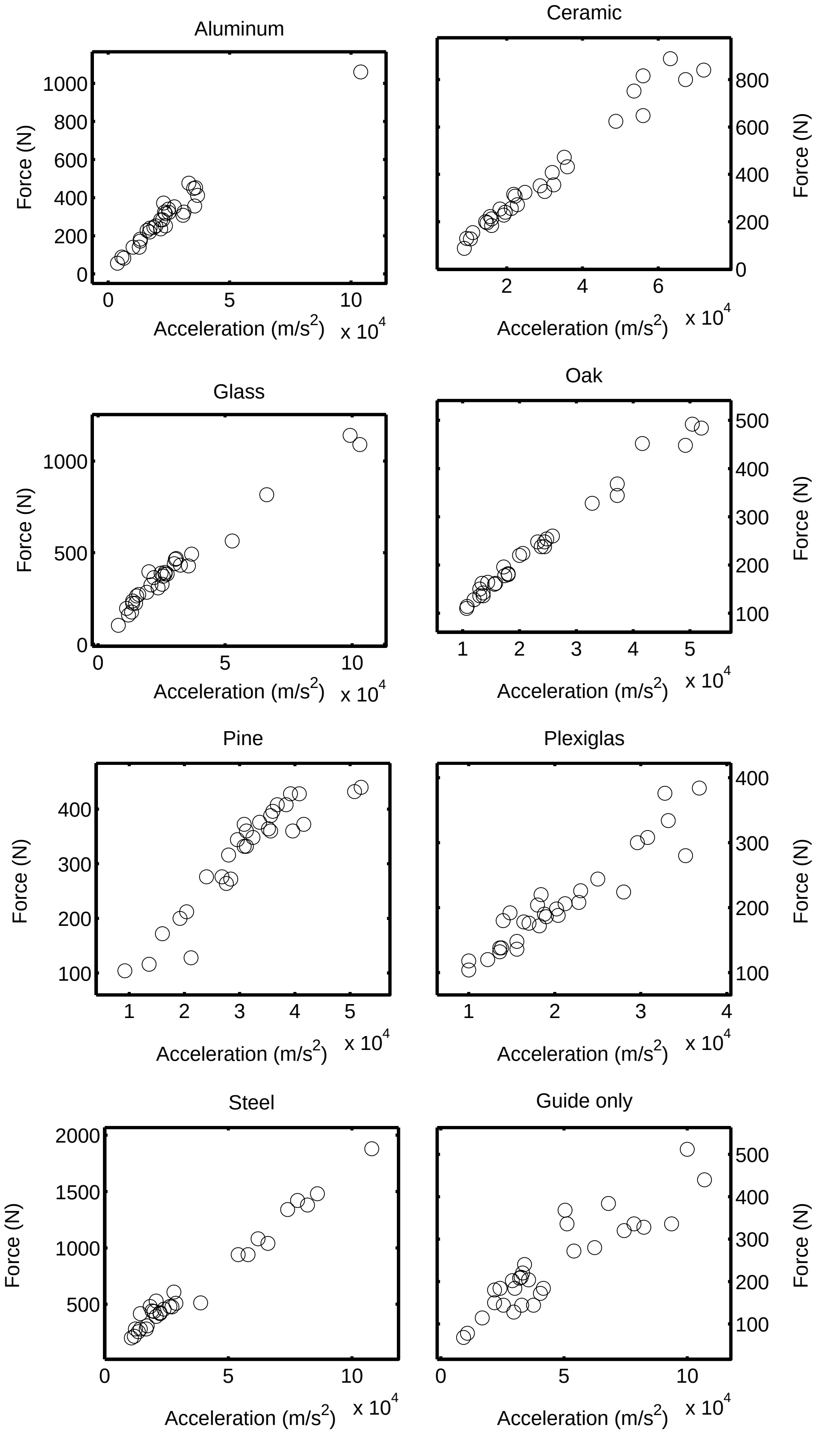

Figure 8.2: Data for the measurement of the dynamic mass of the hammers. For each of the hammers, and for the guide without a hammer, peak force is shown as a function of peak acceleration during the contact time period.

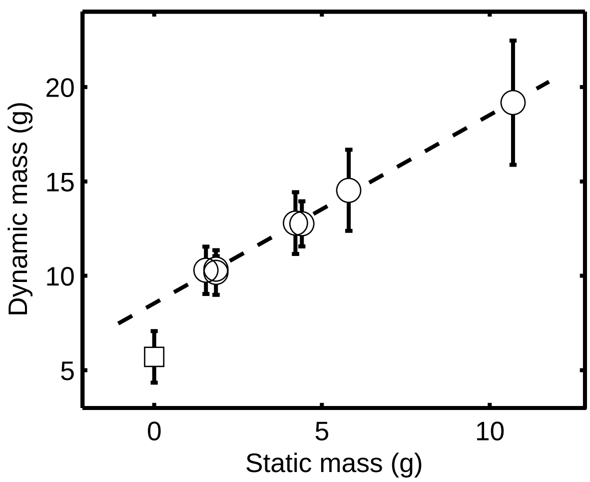

Figure 8.3: Average dynamic mass as a function of the static mass of the hammers (white circles), and of the guide (white square). Error bars = ±1 SD. The dotted line shows the regression function calculated using average measures from all hammers.

8.2.4.2 Force stiffness coefficient

Chaigne and Doutaut (1997) estimated the force stiffness coefficient K from measures of the contact time τ between hammer and plate, of the mass of the plate m_P, of the dynamic mass of the hammer m_h, and of the maximum impact force $F_{\max}$ (Chaigne & Doutaut, 1997). These quantities are related as in Equation (8.5).

$$K = 35.4\frac{1}{\tau^3}\sqrt{\frac{\mu^3}{F_{\max}}} \text{ with } \mu = \frac{m_h m_P}{m_h + m_P} \tag{8.5}$$

For each plate/hammer pair, thirteen τ and $F_{\max}$ measures were collected, releasing the mounting guide for the hammers from thirteen different starting angles, from 15° to 30° in 1.25° steps. Plates were mounted as during the recording phase and struck at their centers. $F_{\max}$ and τ were extracted from the hammer acceleration signal, measured with an Endevco model 22 Picomin light accelerometer (weight: 0.14 g) mounted on the back side of the guide on the opposite side of the surface struck by the hammer. The acceleration signal was amplified with a Brüel & Kjær type 2635 charge amplifier and delivered to a Tektronix TDS-210 two-channel digital oscilloscope (sampling rate: 1 GHz). Measures of the contact time τ were extracted from the acceleration signal as outlined in Section 8.2.3. $F_{\max}$ was computed multiplying $acc_{\max}$, the maximum hammer acceleration during the impact, with the dynamic mass of the hammers. Figure 8.4 shows the contact time τ and $F_{\max}$ measures collected for the 900 cm^2 steel plate struck with the different hammers.

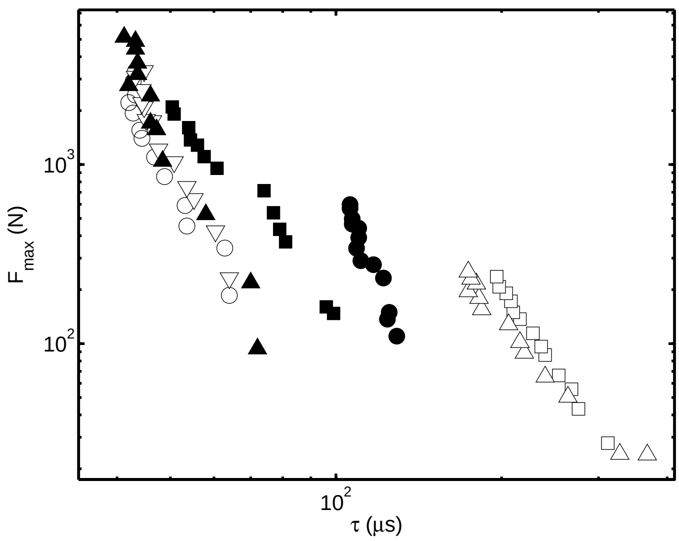

Figure 8.4: τ and $F_{\max}$ measures collected for the 900 cm^2 steel plate struck with the different hammers. White circle: aluminum; black square: ceramic; white triangle-down: glass; white square: oak; white triangle-up: pine; black circle: plexiglas; black triangle-up: steel. Decreasing $F_{\max}$ values are associated with decreasing starting angles of the guide.

Equation (3.10) was applied to calculate the coefficient K for each pair of τ and $F_{\max}$ measures, i.e. from data for each plate/hammer/guide starting angle triplet. The final K measures for each of the plate/hammer pairs were computed averaging estimates across starting angles of the guide. Out of the 2184 K estimates, 75% of them deviated less than $\pm$ 23.43% from the respective average estimate. Table 8.7 gives the final K estimates for all the plate/hammer pairs.

Table 8.7: K measures for all the plate/hammer pairs ($\mathrm{N/m}^{3/2} \times 10^9$). Mat_h = hammer material; Mat_P = plate material.

		Mat_h						
Mat_P	Area (cm^2)	Alu.	Cer.	Gla.	Oak	Pin.	Ple.	Ste.
Alu.	225	8.374	2.815	13.425	0.321	0.368	1.339	10.2
Alu.	450	8.417	3.015	16.232	0.466	0.267	1.226	12.571
Alu.	900	8.096	2.814	11.695	0.466	0.254	1.213	12.258
Cer.	225	2.528	1.517	1.94	0.381	0.302	0.707	2.467
Cer.	450	4.281	1.919	4.675	0.368	0.269	0.892	5.623
Cer.	900	3.835	1.683	2.911	0.499	0.377	1.041	3.372
Gla.	225	6.867	4.05	10.64	0.367	0.279	1.538	9.407
Gla.	450	8.242	4.289	8.48	0.451	0.274	1.44	9.037
Gla.	900	9.101	2.573	7.445	0.398	0.231	1.594	9.936
Oak	225	0.247	0.292	0.328	0.144	0.145	0.209	0.189
Oak	450	0.557	0.273	0.55	0.165	0.172	0.347	0.468
Oak	900	0.523	0.616	0.351	0.227	0.216	0.381	0.417
Pin.	225	0.162	0.202	0.25	0.135	0.099	0.145	0.197
Pin.	450	0.4	0.224	0.406	0.172	0.151	0.184	0.188
Pin.	900	0.368	0.209	0.54	0.163	0.137	0.272	0.279
Ple.	225	1.175	0.89	1.177	0.273	0.242	0.648	0.839
Ple.	450	1.288	0.579	0.917	0.258	0.426	0.616	1.059
Ple.	900	1.151	0.897	0.962	0.3	0.266	0.683	0.963
Ste.	225	13.609	3.732	11.223	0.282	0.339	1.352	22.76
Ste.	450	12.406	3.461	14.242	0.496	0.541	1.417	19.596
Ste.	900	13.857	6.742	14.288	0.297	0.36	1.458	20.946

As pointed out at the beginning of this Section, K can also be extracted from measures of the elastic properties of the hammer and plate and from the radius of the hammer. This relationship is outlined in Equation (8.6).

$$K = \sqrt{R_h}/D \text{ with } D = \frac{3}{4}\left(\frac{1-\nu_P^2}{E_P} + \frac{1-\nu_h^2}{E_h}\right) \tag{8.6}$$

where R_h is the radius of the hammer, E is Young's modulus, ν is Poisson's ratio, and the subscripts h and P refer to the hammer and plate, respectively.

The K estimates extracted from the τ-$F_{\max}$ measures were compared to those predicted by Equation 8.6), in order to check for eventual biases. Only isotropic materials were considered. Figure 8.5 shows the results of this comparison. The K estimates derived from the elastic properties of hammer and plates were found to underestimate those extracted from τ and $F_{\max}$ measures. The reason for this is unknown. However, a linear function accounts well for the relationship among the two sets of measures ($R^2 = 0.907$) as tested with linear regression. Given that the perceptual relevance and the acoustical correlates of K were tested with regression procedures (see Section 8.4), for which linearly related variables are equivalent, whether the first or second set of measures is that closer to the true K value is irrelevant. The K estimates derived from τ and $F_{\max}$ measures were thus considered in the following analyses.

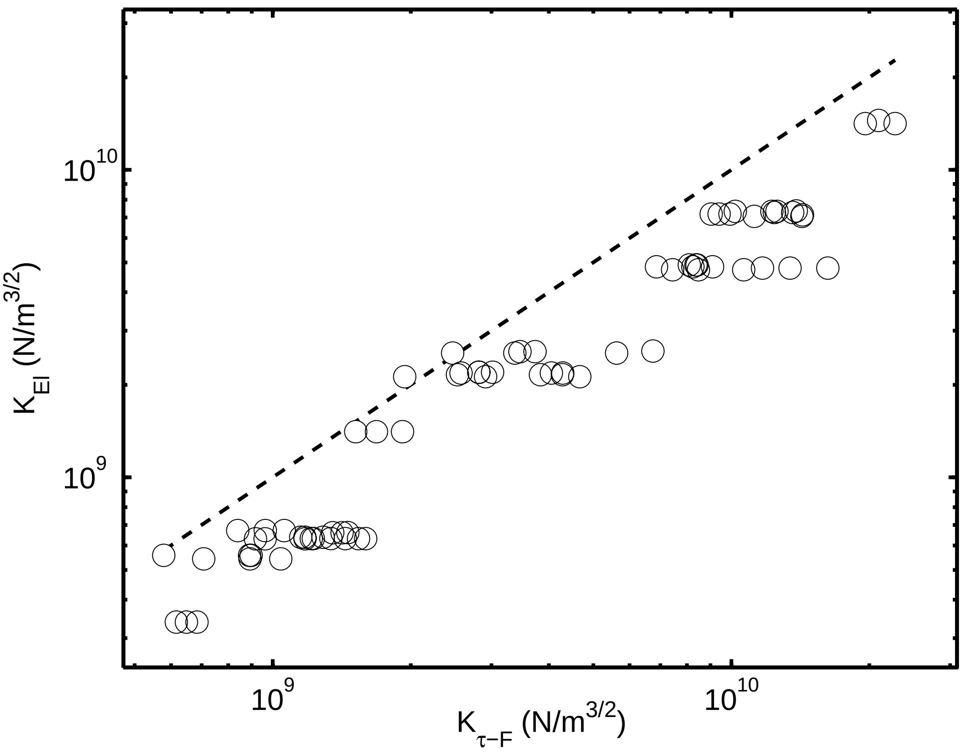

Figure 8.5: Comparison of the K measures derived from measurement of τ and $F_{\max}$ ($K_{\tau-F}$), with the K measures predicted on the basis of the elastic properties of hammers and plates (K_{El}). The dashed line highlights the condition of perfect correspondence of the two sets of measures. Only data from isotropic materials were considered.

8.2.5 Selected source parameters

Only a subset of source parameters was considered in the rest of this study. First, the source parameters with little or no variation in the database were discarded (e.g., plate thickness, hammer radius). Further criteria for source parameter selection were based on considerations of the statistical methods used to determine their perceptual relevance (see Section 8.4). The association of the source parameter ϕ with the behavioral outcome ψ was thus tested with regression models of the form $\psi = a + b\phi^c$. In this framework source parameters ϕ_1 and ϕ_2 related by a function of the form $phi_1 = d + e\phi^f$ are statistically equivalent. For example, plate volume, given by the product of plate's area and an almost constant term, plate's thickness, was discarded. Also plate height and width were not considered, as they were almost equal to the square root of plate area. Finally, among the elastic coefficients, only the rigidities $D_{1P,...,4P}$ were considered. The following source parameters were thus selected: plates rigidities $D_{1P,...,4P}$, plates density ρ_P, plate area, hammer rigidities $D_{1h,...4h}$, hammer density ρ_h, maximum hammer acceleration $acc_{\max}$, maximum impact force $F_{\max}$, contact time τ, and stiffness coefficient K.

A preliminary investigation on the correlation among selected source parameters was performed. Cluster analysis was used for this purpose. A measure of the association between source properties was defined as $d_{x,y} = 1 - r_{(x,y)}$, where $r_{(x,y)}$ is the Pearson correlation between the log physical parameters x and y, and $0 \leq d_{x,y} \leq 1$. Hierarchical cluster analysis (average linkage) was used to analyze this distance measure. Figure 8.6 shows the resulting dendrogram.

Setting $d_{x,y} = 0.25$ as a threshold distance, four groups of highly correlated parameters are found: [1] rigidities and density of the hammer, where increasing rigidities are associated with increasing densities; [2] rigidities and density of the plates; [3] plate area; [4] interaction parameters, where increasing values of τ are associated with decreasing values of K, $acc_{\max}$, and $F_{\max}$. The strong correlations observed among rigidities and densities are consistent with measures on engineering materials published by Waterman and Ashby (1997), where stiffer materials are, in general, characterized by higher densities. The strong association among interaction parameters is consistent with Equation (3.10).

8.2.6 Acoustical descriptors

Signals were analyzed using the same procedure adopted by Giordano and McAdams (submitted), meant to simulate the processing that takes place in the peripheral auditory system. The first descriptor can be conceived as a global damping measure, which, above all, takes into account the frequency resolution of the auditory system ($tan\phi_{aud}$). Extraction of $tan\phi_{aud}$ was performed on the first 1000 ms of the signals. Four loudness descriptors and a duration measure Dur were extracted from the temporal function of signal loudness, in pseudo-sones[1]. Dur was defined as the temporal extent of the signal for

[1] The unit of measure for loudness is termed pseudo-sone as it is calculated directly on the sound file, without taking into account the actual presentation levels.

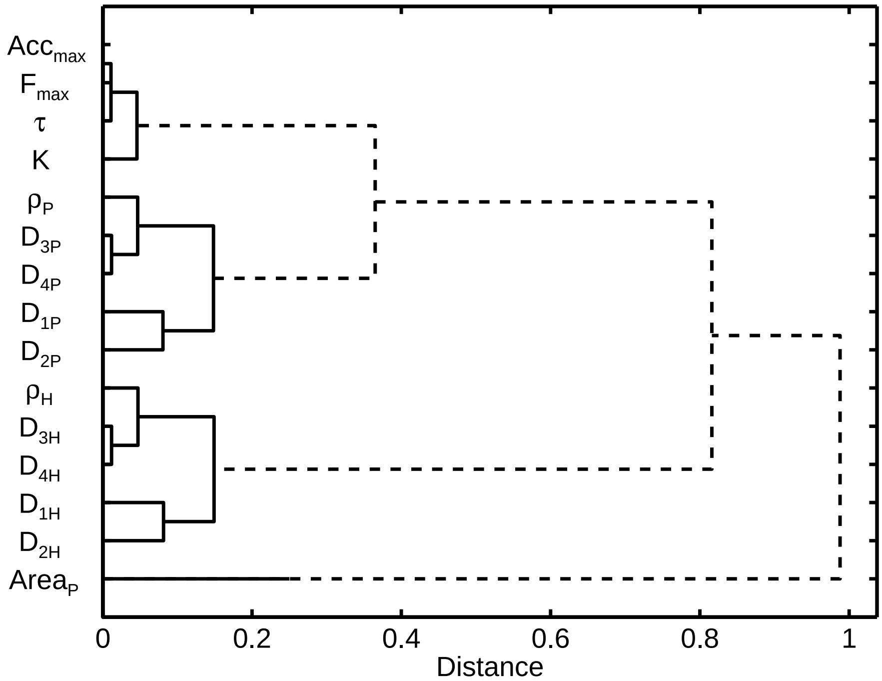

Figure 8.6: Analysis of the correlations among source parameters. Solid lines highlight the clusters of strongly correlated parameters.

which loudness was above a fixed threshold. The maximum loudness of the background noise in the database was 0.148 pseudo-sones; the loudness threshold for Dur estimation was fixed at 0.15 pseudo-sones. The loudness-related descriptors were an attack measure, Lou_{att}, the loudness of the first 10 ms of the signal, an average measure, Lou_{mea}, and the slope of the initial and final portions of the temporal function of loudness, respectively Lou_{sl1} and Lou_{sl2}. From the peripheral auditory system model, a time-variant measure of brightness was also extracted, the spectral center of gravity SCG, defined as the specific loudness-weighted average of frequency, measured on the $ERB - rate$ scale (Moore & Glasberg, 1983). Three descriptors were extracted from this measure: an attack value, SCG_{att}, the SCG of the first 10 ms of the signal, the average value SCG_{mea}, and the slope of the initial segment of the function, SCG_{slo}. Both the SCG and loudness slope measures were extracted by means of linear regression (see Giordano & McAdams, submitted, for a detailed description of the procedure). The last descriptor considered was the frequency of the lowest spectral component F. This measure was extracted on the basis of the fast Fourier transform of the first 512 samples of the signal (Hanning window)[2]. F was defined as the frequency of the first amplitude peak exceeding a fixed threshold. Level threshold was defined on the basis of the analysis of the background noise, as measured

[2]Such a short analysis window was needed for the short length and fast damping of many signals, particularly those generated by striking wood and plexiglas plates. If longer windows were chosen no significant spectral peaks emerged above the level of the background noise.

from the recording of the 250 ms silence preceding the signals. A threshold level for F extraction was then defined on the basis of the maximum spectral level of the background noise across the recorded samples[3]. A final remark should be made on the measurement of F. With isotropic plates, 89% of the measures were equal to the frequency of the ring mode ± 10%; 1% of the measures were equal to the frequency of the X mode ± 10% (see Figure 8.7); none of the measures were close to the frequency of the (1,1) mode. This is not surprising, since during the recording phase, plates were struck in their centers, i.e., in a position that corresponded to an antinode of the ring mode and a node of the X and (1,1) modes. With orthotropic plates, 21% of the measures corresponded to the frequency of the (0,2) mode ± 10%; 5% of the measures corresponded to the frequency of the (2,0) mode ± 10%; 2% of the measures corresponded to the frequency of the (1,1) mode ± 10%. The vast majority of the remaining measures had a frequency intermediate between that of the (1,1), and (0,2) modes.

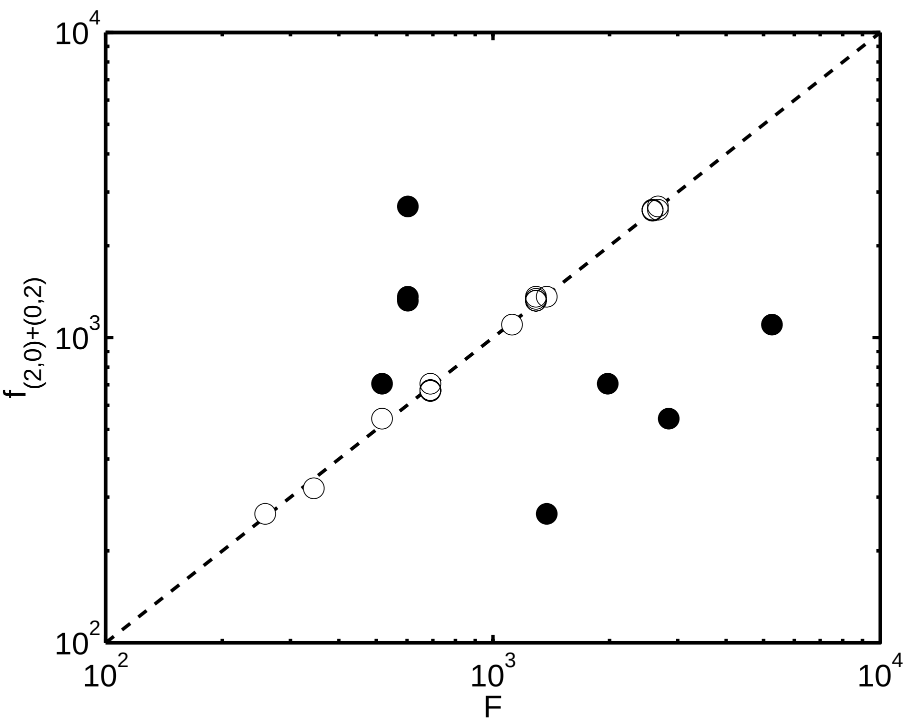

Figure 8.7: Comparison of the F measure with the frequency f_O of the ring vibrational mode. White circles show the F measures that differed by less than 10% from f_O (89% of all measures). The dashed line highlights the condition of perfect correspondence between the two sets of measures.

The correlation among acoustical descriptors was studied using the same methodology as for the source parameters. A slightly different distance measure $d_{x,y}$ was used: $d_{x,y} =$

[3]Across all samples the maximum spectral level of the background noise was -48 dB from the spectral amplitude of the unitary amplitude sinusoid. The level threshold for F extraction was fixed at -45 dB.

$1 - r_{(|x|,|y|)}$, where $r_{(|x|,|y|)}$ is the Pearson correlation between the log absolute value of the acoustical descriptors x and y. Figure 8.8 shows the resulting dendrogram.

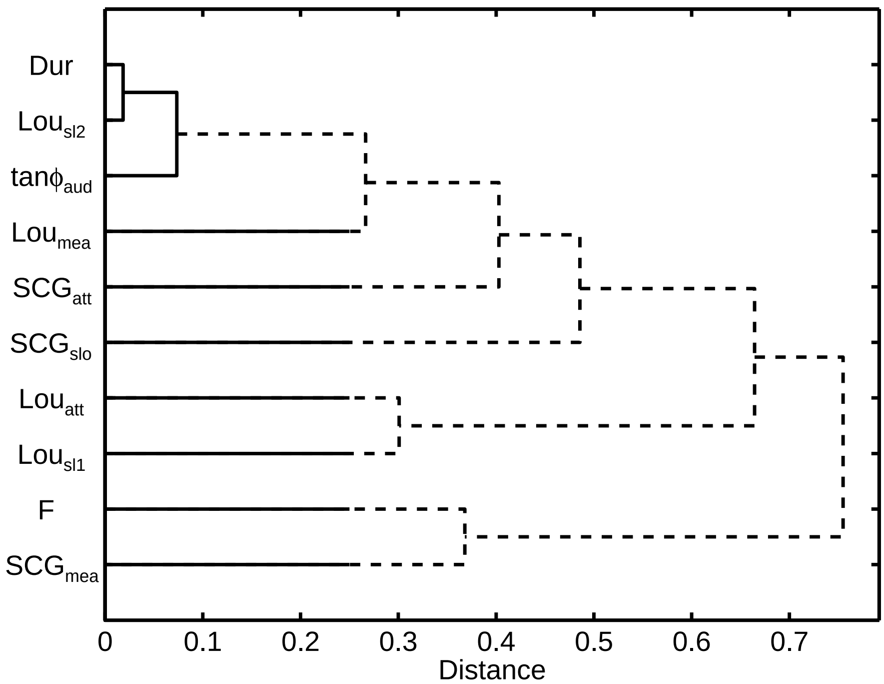

Figure 8.8: Analysis of the correlation among acoustical parameters. Solid lines highlight the clusters of strongly correlated parameters.

Setting $d_{x,y} = 0.25$ as a threshold distance, only one group of highly correlated acoustical parameters is found: Dur, $Lousl2$, $tan\phi_{aud}$, where increasing values of $tan\phi_{aud}$ are associated with decreasing values of Dur and Lou_{sl2}. All the other parameters are, instead, found belonging to separate clusters.

8.2.7 Acoustical correlates of source parameters

The relationship between source parameters and acoustical structure was studied. The goal of this analysis was to highlight, for each source parameter ϕ, the most strongly associated acoustical descriptor α, and thus the most likely acoustical correlates for the perception of each of the source parameters. The following regression model was used: $\alpha = a + b\phi^c$, where the exponent c was estimated using an iterative least squares procedure. Table 8.8 reports the results of this analysis, showing, for each of the source parameters, the two most strongly associated acoustical descriptors. Also, for each of the acoustical parameters the R^2 statistic is reported as a measure of their association with the source parameters.

Several interesting points emerge from this analysis. First, as could be expected, the area of the plates significantly affects F, F decreasing with plate area. Also, F is weakly

Table 8.8: Analysis of the acoustical correlates of the investigated source parameters. For each of the considered physical properties, the two most strongly associated acoustical descriptors are shown. Also reported is the statistic R^2, conceived as a measure of the strength of the association of source and acoustical properties.

Source property	Acoustical descriptor	R^2	Acoustical descriptor	R^2
$Area_P$	F	0.476	SCG_{mea}	0.292
ρ_P	$tan\phi_{aud}$	0.789	SCG_{att}	0.540
D_{1P}	$tan\phi_{aud}$	0.698	Lou_{sl2}	0.671
D_{2P}	Dur	0.579	$tan\phi_{aud}$	0.547
D_{3P}	$tan\phi_{aud}$	0.910	SCG_{att}	0.592
D_{4P}	$tan\phi_{aud}$	0.888	SCG_{att}	0.622
ρ_h	Lou_{att}	0.359	Lou_{sl1}	0.247
D_{1h}	Lou_{att}	0.223	Lou_{sl1}	0.166
D_{2h}	Lou_{att}	0.218	Lou_{sl1}	0.162
D_{3h}	Lou_{att}	0.302	Lou_{sl1}	0.224
D_{4h}	Lou_{att}	0.334	Lou_{sl1}	0.243
K	SCG_{att}	0.870	$tan\phi_{aud}$	0.427
τ	SCG_{att}	0.923	$tan\phi_{aud}$	0.471
$Acc_{\max}$	SCG_{att}	0.935	$tan\phi_{aud}$	0.543
$F_{\max}$	SCG_{att}	0.893	$tan\phi_{aud}$	0.444

affected by variations in the other source parameters (average R^2 for the other source parameters: 0.073; maximum R^2 for the other source parameters: 0.109). The density, and rigidities of the plates, instead, strongly affect $tan\phi_{aud}$, this acoustical parameter decreasing with increasing plate density and rigidity. Although $tan\phi_{aud}$ is weakly to moderately affected by variations in other source properties (average R^2: 0.466; maximum R^2: 0.698), the association with plate density and rigidity is the strongest. It should be noted that the observed association of the rigidity of the plate with the acoustical measure of damping $tan\phi_{aud}$ is consistent with measurements conducted on engineering materials (Waterman & Ashby, 1997), where the damping parameter η, termed loss coefficient, decreases with increasing stiffness of the materials, as measured with Young's modulus. Concerning the hammer parameters, the most strongly affected acoustical property was Lou_{att}. Although the association of hammer properties with Lou_{att} was rather moderate, the association of this acoustical parameter with other source properties was even lower (average R^2: 0.054; maximum R^2: 0.148). Finally, interaction parameters strongly affected SCG_{att}, increasing values of SCG_{att} being found for increasing values of K, $acc_{\max}$, $F_{\max}$, and for decreasing values of τ. Although other source parameters affected this acoustical property, the strength of their association was in any case lower than that found for interaction parameters (average R^2: 0.341; maximum R^2: 0.540). In conclusion, acoustical parameters selectively specifying each of the source parameters, and thus those most likely mediating perception of these latter, were highlighted.

8.3 Methods

8.3.1 Stimuli

A stimulus set was extracted from the database of recordings. The selection criteria were designed to extract a set of sound sources representative of the entire database, i.e. a set of stimuli for which source properties had the same range and a similar inter-correlation structure as that characterizing the database. Selection of the stimulus set was made by applying the following constraints: all the plate and hammer materials had to be included; for each physical variable the maximum and the minimum value had to be included, as well as the value at the center of the log range. Small deviations of the area of the plates from 225, 450, and 900 cm^2 were not considered in this selection procedure.

Given these constraints a set of eighteen stimuli was selected randomly from the database. Signals longer than 1 sec (one out of eighteen) were reduced to this duration by applying a 5-ms linear decay. Acoustical descriptors were extracted using the same methodology outlined in Section 8.2.6. Tables 8.9 and 8.10 report the source and acoustical parameters, respectively, for each of the selected stimuli.

Table 8.9: Source properties for the stimuli in the experimental set. The subscripts P and h stand for plate and for hammer, respectively.

Mat_P	$Area_P$ cm^2	ρ_P kg/m^3	D_{1P} GPa	D_{2P} GPa	D_{3P} GPa	D_{4P} GPa	Mat_h	ρ_h kg/m^3	D_{1h} GPa	D_{2h} GPa	D_{3h} GPa	D_{4h} GPa	$K \times 10^9$ N/m$^{3/2}$	τ μs	Acc_{max} m/s$^2 \times 10^3$	F_{max} N
Alu.	224.62	2658.74	6.06	4.14	6.06	7.97	Cer.	2124.79	1.71	0.22	1.71	3.19	2.82	94	26.6	339.28
Gla.	225	2485.54	5.9	2.74	5.9	9.05	Ple.	1186.22	0.42	0.3	0.42	0.54	1.54	111	24.4	248.28
Oak	449.44	794.71	1.46	0.51	0.49	0.37	Pin.	676.08	1.47	0.47	0.42	0.35	0.17	246	8.64	88.91
Oak	894	794.71	1.46	0.51	0.49	0.37	Alu.	2658.74	6.06	4.14	6.06	7.97	0.52	210	11.7	149.66
Pin.	223.46	676.08	1.47	0.47	0.42	0.35	Oak	794.71	1.46	0.51	0.49	0.37	0.14	292	7.36	76.21
Pin.	223.46	676.08	1.47	0.47	0.42	0.35	Pin.	676.08	1.47	0.47	0.42	0.35	0.1	260	7.52	77.39
Pin.	886.5	676.08	1.47	0.47	0.42	0.35	Oak	794.71	1.46	0.51	0.49	0.37	0.16	348	5.76	59.64
Pin.	886.5	676.08	1.47	0.47	0.42	0.35	Ste.	7751.85	17.33	9.32	17.33	25.35	0.28	392	5.24	100.46
Ple.	225	1186.22	0.42	0.3	0.42	0.54	Gla.	2485.54	5.9	2.74	5.9	9.05	1.18	160	16.6	241.17
Ple.	901.5	1186.22	0.42	0.3	0.42	0.54	Ple.	1186.22	0.42	0.3	0.42	0.54	0.68	152	14.4	146.52
Cer.	224.25	2124.79	1.71	0.22	1.71	3.19	Cer.	2124.79	1.71	0.22	1.71	3.19	1.52	131	19.6	250
Cer.	449.44	2124.79	1.71	0.22	1.71	3.19	Alu.	2658.74	6.06	4.14	6.06	7.97	4.28	92	28.4	363.28
Cer.	449.44	2124.79	1.71	0.22	1.71	3.19	Ple.	1186.22	0.42	0.3	0.42	0.54	0.89	130	23	234.03
Cer.	900	2124.79	1.71	0.22	1.71	3.19	Ple.	1186.22	0.42	0.3	0.42	0.54	1.04	143	20.8	211.65
Ste.	225	7751.85	17.33	9.32	17.33	25.35	Ste.	7751.85	17.33	9.32	17.33	25.35	22.76	50.4	88	1687.09
Ste.	445.1	7751.85	17.33	9.32	17.33	25.35	Alu.	2658.74	6.06	4.14	6.06	7.97	12.41	46.4	128	1637.34
Ste.	445.1	7751.85	17.33	9.32	17.33	25.35	Ste.	7751.85	17.33	9.32	17.33	25.35	19.6	42	126	2415.61
Ste.	900	7824.24	18.17	10.06	18.17	26.27	Oak	794.71	1.46	0.51	0.49	0.37	0.3	218	12.6	130.47

Table 8.10: Acoustical properties of the experimental stimuli. p.s.=pseudo-sones. See text for an explanation of the meaning of each acoustical descriptor.

Mat_p	Mat_m	$Area_P$ cm^2	$tan\phi \times 10^{-}3$	Dur (s)	F (Hz)	Lou_{att} (p.s.)	Lou_{mea} (p.s.)	Lou_{sl1} (p.s./s)	Lou_{sl2} (p.s./s)	SCG_{att} (ERB-rate)	SCG_{mea} (ERB-rate)	SCG_{slo} (ERB-rate/s)
Alu.	Cer.	224.62	7.39	0.20	2670.12	2.34	0.60	-52.41	-1.82	25.22	25.39	-167.46
Gla.	Ple.	225.00	7.83	0.12	2583.98	2.57	0.84	-57.92	-4.39	24.58	23.25	-96.14
Oak.	Pin.	449.44	24.22	0.09	689.06	3.45	1.16	-86.66	-6.78	20.74	17.66	-69.26
Oak.	Alu.	894.00	15.50	0.17	516.80	4.04	0.99	-89.59	-2.46	22.06	15.95	-93.14
Pin.	Oak.	223.46	18.40	0.08	1119.73	2.84	1.05	-74.25	-6.75	20.82	19.10	-47.55
Pin.	Pin.	223.46	21.48	0.07	1119.73	2.90	1.17	-78.45	-9.85	21.31	19.14	-53.16
Pin.	Oak.	886.50	18.63	0.15	602.93	2.42	0.76	-54.56	-2.39	19.70	16.10	-54.83
Pin.	Ste.	886.50	22.57	0.12	689.06	3.43	1.10	-82.60	-4.14	19.97	16.62	-49.34
Ple.	Gla.	225.00	19.61	0.07	1119.73	3.47	1.15	-108.04	-8.12	23.40	20.39	-64.28
Ple.	Ple.	901.50	18.67	0.07	258.40	3.61	1.34	-97.58	-10.50	22.52	19.18	-60.03
Cer.	Cer.	224.25	7.83	0.24	602.93	3.66	0.71	-85.68	-1.10	23.18	19.85	-48.62
Cer.	Alu.	449.44	6.03	0.28	516.80	5.44	0.93	-128.06	-1.57	24.84	19.68	-42.97
Cer.	Ple.	449.44	5.49	0.24	689.06	4.10	0.85	-93.02	-1.62	23.82	19.42	-26.50
Cer.	Ple.	900.00	5.98	0.30	344.53	3.38	0.74	-68.12	-1.33	23.47	17.16	-37.05
Ste.	Ste.	225.00	1.34	0.58	2583.98	3.14	0.65	-53.40	-0.98	26.96	27.80	-3.92
Ste.	Alu.	445.10	1.64	0.53	1291.99	2.66	0.72	-23.88	-0.99	27.26	26.91	-127.54
Ste.	Ste.	445.10	1.60	0.62	1291.99	3.10	0.86	-20.82	-1.02	27.23	26.21	-137.56
Ste.	Oak.	900.00	0.67	0.98	689.06	2.06	0.81	-4.84	-0.63	23.52	19.22	-7.86

The same methodology used in Sections 8.2.5 and 8.2.6 was used to analyze the correlation structure among source parameters and among acoustical parameters. The dendrograms resulting from these analyses are shown in Figure 8.9. Comparison of the dendrograms reported in Figure 8.9 with those reported in Figures 8.6 and 8.8 shows that the structure of the correlation among source parameters and among acoustical descriptors in the experimental set, bears strong similarities with that in the database. Indeed, using $d_{x,y} = 0.25$ as a distance threshold to isolate separate groups of highly correlated parameters, the same clusters as for the database are found for the source parameters, while the only difference concerning the clusters of acoustical descriptors stands in the fact that while in the database SCG_{mea} and SCG_{att} belonged to separate clusters, for the experimental set they belong to the same cluster. In general these analyses show that the experimental set is a representative sample of the entire database.

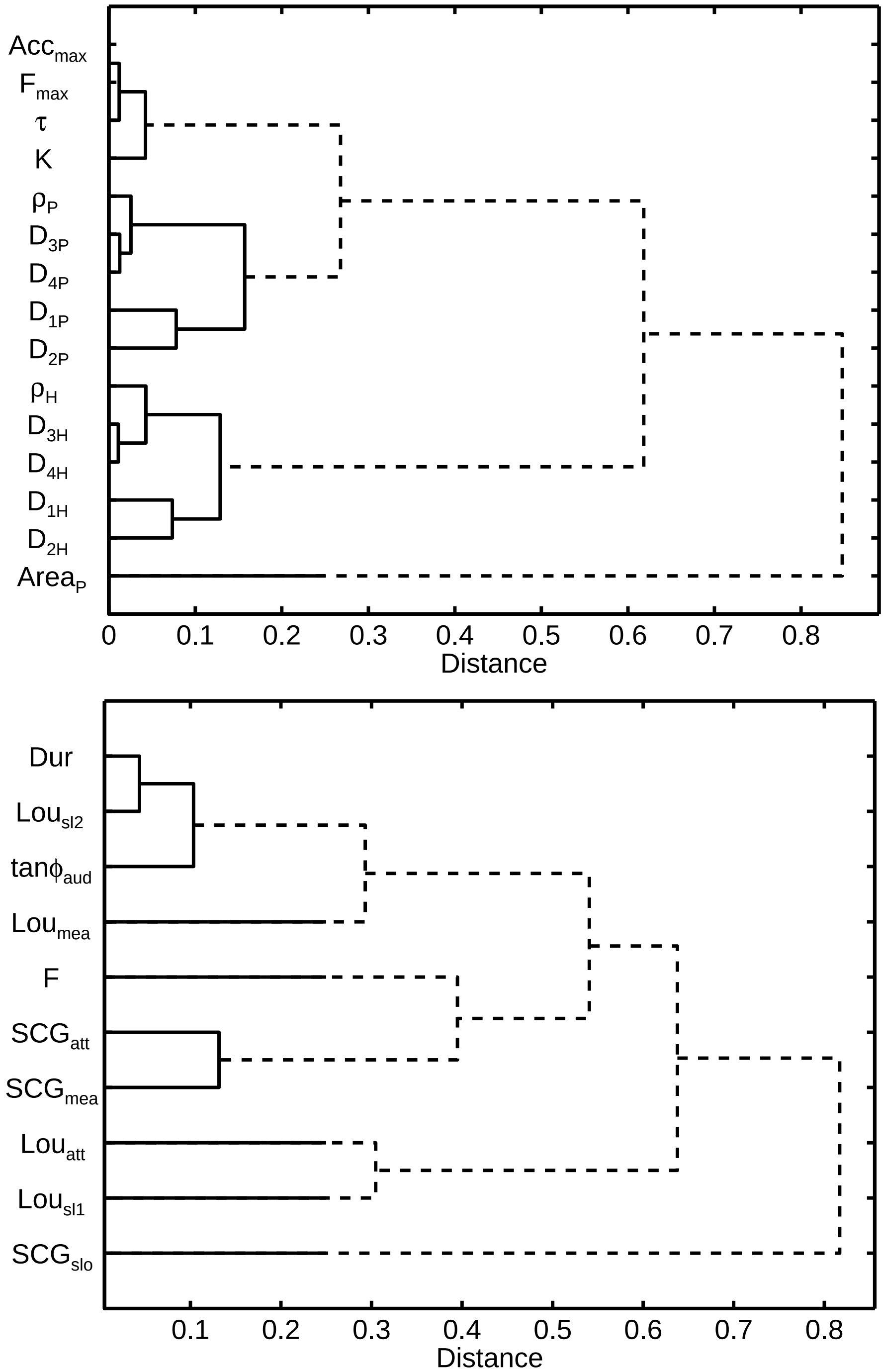

Figure 8.9: Dendrograms computed for the analysis of the correlation structure among source parameters (top panel) and among acoustical descriptors (bottom panel) in the experimental set.

8.3.2 Procedure

Stimuli were stored in the hard disk of a Linux Workstation, equipped with a RME Hammerfall 9652 sound card. Audio signals were converted with the RME Analog/Digital Interface ADI-8 PRO, equipped with anti-aliasing filters, amplified with a Yamaha power amplifier P2075, and presented through Sennheiser HD250 linear II headphones. Participants sat inside an audiometric booth. Stimulus presentation and data collection were programmed into the Mathworks Matlab environment.

Stimuli were presented pairwise. Participants were asked to rate the dissimilarity between stimuli in the pair using any salient criteria. Responses were given by moving a slider on a scale whose extremes were marked very-similar and very-dissimilar. Before giving the response, participants were allowed to replay the pair as many times as needed. Once participants were satisfied with their estimate, they could move to the next pair by clicking with the mouse on the appropriate on-screen button. At the beginning of the experimental procedure, participants were presented with all the stimuli in random order for a minimum of three times and were asked to get a rough idea of the maximal similarity and dissimilarity among the stimuli in the set. This entire procedure was first practiced with a set of four stimuli not presented during the main experiment, comprising all the area levels and two hard and two soft materials for both the hammers and the plates (aluminum, glass, oak and plexiglas, for the plates; steel, ceramic, pine and oak for the hammers). In the experimental phase each non-identical pair was presented only once for a total of 153 trials. The order of the pairs and the order of the stimuli within the pairs were chosen randomly for each participant.

8.3.3 Participants

Twenty-five listeners took part in the experiment (age: 21 - 50 years; 14 females, 11 males). All of them reported having normal hearing. They were paid for their participation.

8.4 Results

Data were analyzed with multidimensional scaling. An extended version of the CLASCAL model (Winsberg & De Soete, 1993; McAdams et al., 1995) was used, which maps observed dissimilarities to the distance of stimuli in a spatial model, comprising a Euclidean space common to all the stimuli, whose dimensions are weighted differently by different classes of participants, and a set of dimensions specific to each of the stimuli, referred to as "specificities". A latent class approach is used to model differences among groups of individuals. It is assumed that each of the $k = \{1, ..., N\}$ participants belongs to one of $T \ll N$ classes, with a probability λ_k, where $\sum_{t=1}^{T} \lambda_{kt} = 1$. Model distances are given by Equation (8.7).

$$d_{ijt} = \left[\sum_{r=1}^{R} w_{tr} \left(x_{ir} - x_{jr} \right)^2 + v_t \left(s_i + s_j \right) \right]^{1/2} \tag{8.7}$$

where d_{ijt} is the model distance between stimulus i and j for participants in the latent class t; x_{ir} is the coordinate of stimulus i along dimension r; s_i is the coordinate of stimulus i along its specific dimension; w_{tr} and v_t are the weights of the Euclidean dimension r and of the specificities for latent class t.

Parameters of the CLASCAL model are estimated using maximum likelihood procedures. Therefore model selection can be based on information criteria. In particular, given the number of latent classes T, Bayes' Information Criterion BIC (Schwarz, 1978) is used to make a decision concerning the number of dimensions R and whether the model should include specificities or not. Models with lower BIC values are preferred. T is chosen on the basis of a Monte Carlo procedure proposed by Hope (1968), which, conditional on a given spatial model, tests for a significantly better fit of the model with $T+1$ latent classes, over the model with T classes. The same procedure is also applied to test for significant differences between competing spatial models. Model selection is based on the following steps (Winsberg & De Soete, 1993; McAdams et al., 2004; Caclin, 2004):

1. Select T applying Hope's procedure to the null model, i.e. the matrix of average dissimilarities among all stimuli, where $1 \leq T \leq 6$.

2. Compute the spatial models with one to eight dimensions, both with and without specificities. Retain the three models with the lowest BIC.

3. Apply Hope's procedure to choose among the three competing models.

4. Use Hope's procedure to test for T on the spatial model selected at step three.

5. If T at step four equals T used at step two the model selection procedure is terminated. Otherwise repeat steps two to four until T converges.

The initial Monte Carlo test on the null model highlighted two possibilities for T: one and three latent classes. Application of the above outlined model selection procedure with these T values led to two different models, referred to as model A, and model B. Neither model not included specificities.

8.4.1 Analysis of model A

Model A had one latent class and two dimensions. It explained 62% of the variance in the individuals' ratings, and 97% of the variance in the average ratings. The dimensions were not weighted. Therefore the model was rotationally invariant. It was rotated to the $T = N$ classes solution, orienting the axes along psychologically meaningful dimensions. The rotated model explained 65% of the variance in the individuals' ratings, and 97% of the variance in the average ratings. A low correlation among the coordinates of stimuli along the two dimension was found (Pearson $r = 0.182$, p=0.470, df=16).

The relative perceptual salience of the two dimensions of the model was compared through their range of variation, where the higher the range, the higher the perceptual salience. The range of dimension 1 in the common space (i.e., in the unweighted space) was 1.46 times that of dimension 2, indicating a higher salience of the first of the two dimensions. This was confirmed inspecting the individual ranges (i.e., the ranges multiplied by the individual weights), were the range of dimension 1 was the larger for the vast majority of participants, while for the minority the two dimensions were approximately equally salient (see Figure 8.10).

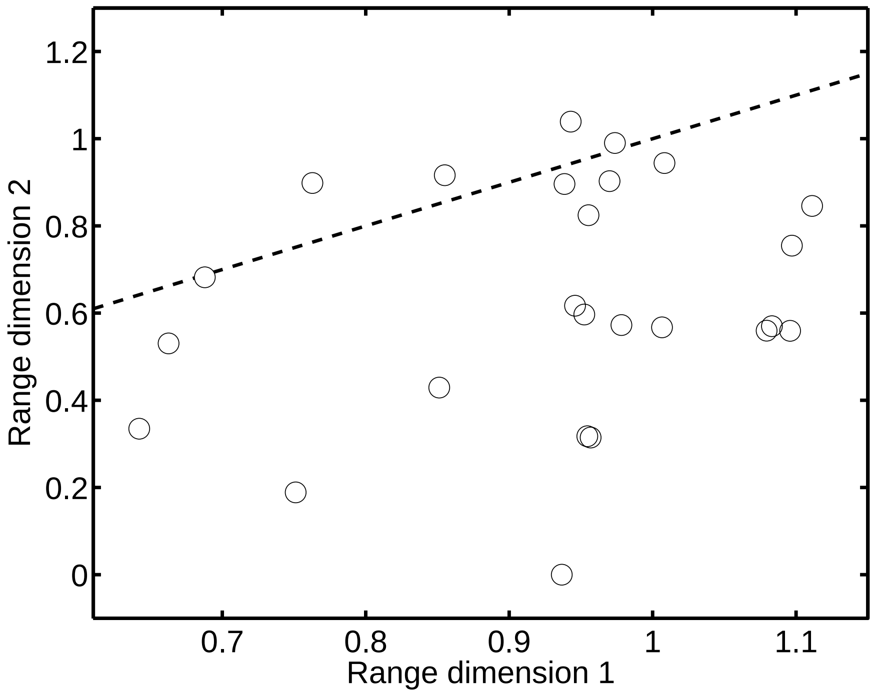

Figure 8.10: Range of variation of the weighted dimensions 1 and 2 in model A for the different individuals. The dashed line highlights the condition of equal range of variation for the both dimensions.

The spatial model was then interpreted with reference to the properties of the investigated stimuli. Regression models were used, where source descriptors ϕ, and acoustical descriptors α were tested separately in their power to explain the location of stimuli along the dimensions of the MDS space. The following procedure was used for each of the dimensions:

1. For each of the α/ϕ descriptors, estimate the monotone transform relating α/ϕ with the coordinate ψ of the stimuli along the dimension, estimating the parameters of the non-linear regression model $\psi = a + b\,|\alpha|^c$ / $\psi = a + b\phi^c$.

2. Test the significance of the association of each transformed descriptor with the

behavioral outcome, and discard from further modeling those not significantly associated (Hosmer & Lemeshow, 1989).

3. Compute all the possible univariate and multivariate regression models in a linear regression framework, raising each of the predictors to the exponent c computed at step one.

4. Select as final models those with the fewest number of predictors whose adjusted R^2 value is equal to or higher than a threshold value of 0.75, i.e., select the most economical models that fit well the observed data. If none of the models exceeds the threshold adjusted R^2, choose the one with the highest value of this statistic.

Tables 8.11 and 8.12 report the main statistics of the final regression models based on source and acoustical parameters, respectively. Figure 8.11 plots the best source and acoustical-based models for each of the dimensions.

Table 8.11: Source criteria for dissimilarity ratings emerging from the analysis of model A. For each dimension, the exponent and the standardized parameter estimate for each of the predictors in the regression model are reported. Also reported is the adjusted R^2 goodness-of-fit measure.

	Pred.$^{exp.}$	Stand. estim.	$Adj.R^2$
Dimension 1	$D_{4P}^{0.24}$	-0.990	0.979
	$D_{3P}^{0.05}$	-0.986	0.976
	$\rho_P^{0.12}$	-0.978	0.953
	$D_{1P}^{0.38}$	-0.928	0.853
	$D_{2P}^{0.82}$	-0.903	0.803
Dimension 2	$Area_P^{3.37\times10^3}$	-0.639	0.371

Finally, the best regression model for each dimension were combined to check their power in explaining the observed dissimilarity ratings. The best source-based model explained 82% of the variance in the average ratings, while the best acoustical model explained 72% of the variance in the average ratings.

Table 8.12: Acoustical criteria for dissimilarity ratings emerging from the analysis of model A. For each dimension, the exponent and the standardized parameter estimate for each of the predictors in the regression model are reported. Also reported is the adjusted R^2 goodness-of-fit measure.

	Pred.$^{exp.}$	Stand. estim.	Pred.$^{exp.}$	Stand. estim.	$Adj.R^2$
Dimension 1	$tan\phi_{aud}^{0.93}$	0.949	—	—	0.894
	$SCG_{att}^{-2.33}$	-0.875	—	—	0.751
Dimension 2	$\lvert Lou_{sl2}\rvert^{0.45}$	0.750	$F^{0.49}$	0.595	0.788
	$F^{0.49}$	0.811	$Lou_{mea}^{5.44}$	0.791	0.764

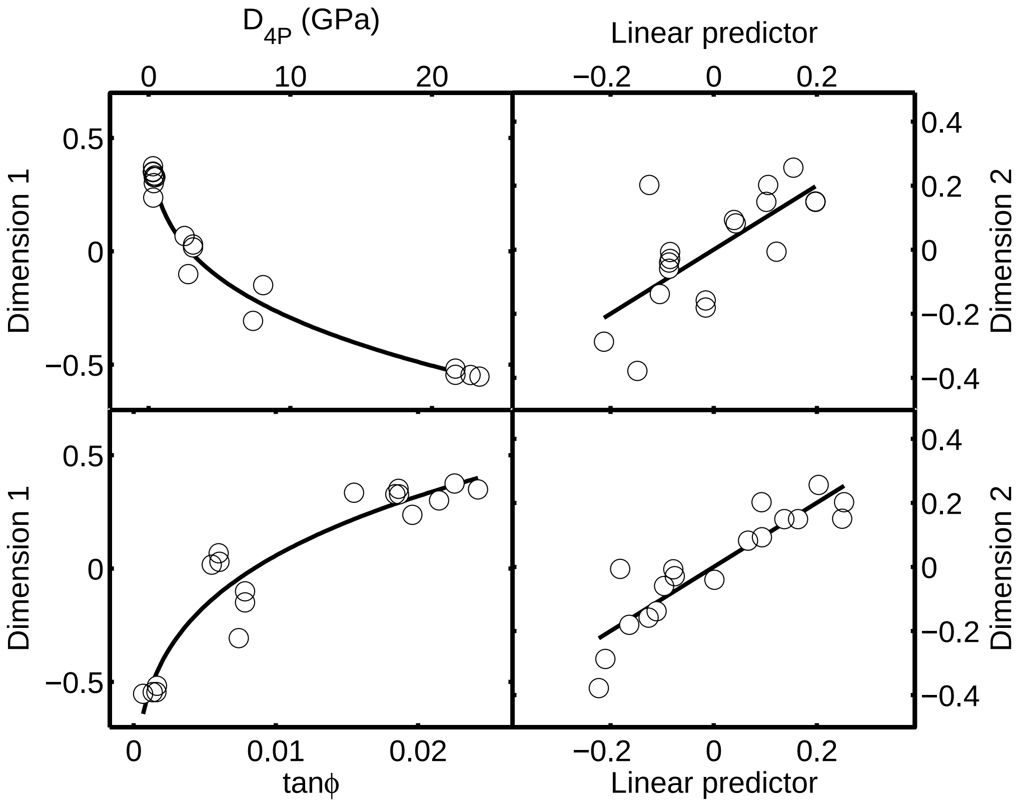

Figure 8.11: Analysis of model A: best regression models based on source properties (top panels) and acoustical properties (bottom panels) of the investigated stimuli. Continuous lines show the regression models.

8.4.2 Analysis of model B

Model B had two latent classes and four dimensions. It explained 66% of the variance in individuals' ratings, and 97% of the variance in the ratings averaged across participants belonging to the same latent class. Belongingness of participants to the latent classes was very clear ($\lambda > 0.8$) except for one participant, for whom the posterior probability of belonging to the first cluster equalled 0.58. Quite strong correlations were found among the coordinates of stimuli along the different dimensions. In particular, dimension 1 is significantly correlated with dimensions 3 and 4 ($r \geq 0.663$, $p \leq 0.003$, df=16), and dimensions 3 and 4 are significantly correlated each other ($r = 0.605$, $p = 0.008$, df=16).

The relative salience of the four dimensions differed across classes. In particular, for the first latent class the most salient dimensions were the first (range = 0.974) and second (range = 0.551), the range of the third and fourth dimensions being 0.238 and 0.146, respectively. For the second latent class, instead, the most relevant dimensions were the third and the fourth (range = 0.714, 0.682), the range of the first and second dimensions being 0.660 and 0.551, respectively.

The same methodology outlined in Section 8.4.1 was adopted to investigate the source and acoustical parameters underlying dissimilarity judgment. Tables 8.13 and 8.14 show the relevant parameters for the regression models selected for each dimension. Figures 8.12 and 8.13 plot part of the selected regression models. In both Figures 8.12 and 8.13, the regression models shown for dimensions 1 and 3 are based on the same parameters, thus facilitating the comparison of the transform of the properties of the stimuli to the MDS coordinates.

When combined the best source-based regression models explained 77% of the variance in the average observed dissimilarities. The best acoustical-based regression models explained 82% of the variance in the average observed dissimilarities.

8.4.3 Discussion

Two similar MDS models were accounted equally well for the observed data. Strong commonalities were found among these two models, with respect to both the source-based explanation and to the acoustical-based explanation of the criteria for dissimilarity rating. However, in the following only model B will be discussed, for three reasons. Firstly, the second dimension of model A was much more poorly accounted for by source properties than the second dimension of model B. Secondly, the comparison of source-based and acoustical-based regression models is completely unambiguous for model B, ambiguous with model A. In particular, for model B the acoustical explanations for the same dimension are based on descriptors that specify the same group of source parameters. For model A, instead, the first dimension is equally well explained by parameters specifying different groups of source properties (rigidities/density of the plate and interaction parameter), and the best regression model for the second dimension is based on two acoustical parameters specifying each two separate groups of source parameters (the area of the plate and the rigidity of the plate). Thirdly, model B is more economical than model A, as

Table 8.13: Source criteria for dissimilarity ratings emerging from the analysis of model B. For each dimension the exponent and the standardized parameter estimate for each of the predictors in the regression model are reported. Also reported is the adjusted R^2 goodness-of-fit measure.

	Pred.$^{exp.}$	Stand. estim.	$Adj.R^2$
Dimension 1	$D_{4P}^{0.52}$	-0.985	0.969
	$D_{3P}^{0.39}$	-0.985	0.968
	$D_{2P}^{0.50}$	-0.979	0.956
	$D_{1P}^{0.59}$	-0.959	0.916
	$\rho_P^{0.98}$	-0.950	0.896
Dimension 2	$Area_P^{0.73}$	-0.776	0.578
Dimension 3	$D_{3P}^{-2.11}$	0.926	0.850
Dimension 4	$D_{4P}^{-1.26}$	0.904	0.805
	$\tau^{1.18}$	0.888	0.775
	$\rho_P^{-1.10}$	0.884	0.768
	$acc_{\max}^{-0.80}$	0.879	0.759

with this latter removal of the rotational invariance comes with the price of introducing many more additional parameters (one weight per dimension for each listener).

Two latent classes of participants were found using different criteria when judging the dissimilarity of sounds. Participants in the first class based their ratings mainly on the rigidity or density of the plates. From an acoustical point of view, the ratings were based on two parameters strongly associated with the rigidity and density of the plates, the acoustical measure of damping $tan\phi_{aud}$ or signal duration. Thus the most relevant perceptual dimension for this class reflected differences in the material of the sounding objects, and from the acoustical point of view, in the temporal properties of the acoustical signals. Similar explanations were found for the first dimension of participants in latent class two, explained by one rigidity coefficient of the plates, and by three acoustical parameters specifying the temporal properties of the signals: Dur, $tan\phi_{aud}$ and Lou_{sl2}. Also for these participants the most relevant dimension reflected the variation in the material of the sounding object. However the transform relating the stimulus descriptors to the coordinates of stimuli along these dimensions was found to differ across latent classes. For example, as shown in Figure 8.13, the transform relating the parameter $tan\phi_{aud}$ to the most relevant dimensions of each class was much more compressive for

Table 8.14: Acoustical criteria for dissimilarity ratings emerging from the analysis of model B. For each dimension the exponent and the standardized parameter estimate for each of the predictors in the regression model are reported. Also reported is the adjusted R^2 goodness-of-fit measure.

	Pred.$^{exp.}$	Stand. estim.	Pred.$^{exp.}$	Stand. estim.	$Adj.R^2$
Dimension 1	$tan\phi_{aud}^{-0.06}$	-0.950	—	—	0.897
	$Dur^{0.59}$	-0.906	—	—	0.809
Dimension 2	$F^{0.25}$	0.643	$SCG_{mea}^{-1.29}$	-0.286	0.740
Dimension 3	$Dur^{-1.01}$	0.940	—	—	0.877
	$tan\phi_{aud}^{1.13}$	0.928	—	—	0.853
	$\lvert Lou_{sl2}\rvert^{0.35}$	0.918	—	—	0.832
Dimension 4	$SCG_{att}^{-1.99}$	0.880	—	—	0.761

participants in the first latent class than for those in the second. The second dimension for participants in latent class one, which is the third most salient dimension for participants in latent class two, was related to the area of the plates, and to two acoustical parameters strongly associated with it: F and SCG_{mea}.

Thus, independently of the inderindividual differences, and consistently with results from Roussarie (1999) and McAdams et al. (2004), and in line with results by Gaver (1988), dissimilarity ratings of impact sounds relied strongly on two criteria, the first of them related to the acoustical measure of damping $tan\phi_{aud}$ and, in line with the hypothesis formulated in Section 8.1, with signal duration, while the second criterion was based on signal frequency. However, while in the current study the damping-related dimension was associated with the density and rigidity of the sounding objects, in the study by Roussarie (1999) and in the first of the experiments performed by McAdams et al. (2004) these source properties were found associated with the frequency-related dimension. The reason for this inconsistency might be that the extent of the size variation in the current study was such as to mask the influence of the density and elasticity coefficients on the modal frequencies of the sounding object. Consistently, in the cited studies by Roussarie (1999) and McAdams et al. (2004) the size of the sounding objects was kept constant, thus favoring the emergence of a strong link between rigidities or densities, and signal frequency. Finally, consistently with results by Gaver (1988) and with those from the second experiment by McAdams et al. (2004), the frequency-related dimension was strongly correlated with the size of the sounding object.

The second dimension for participants in the second latent class was equally well explained by the rigidity coefficient D_{4P} and by the density of the sounding object, as

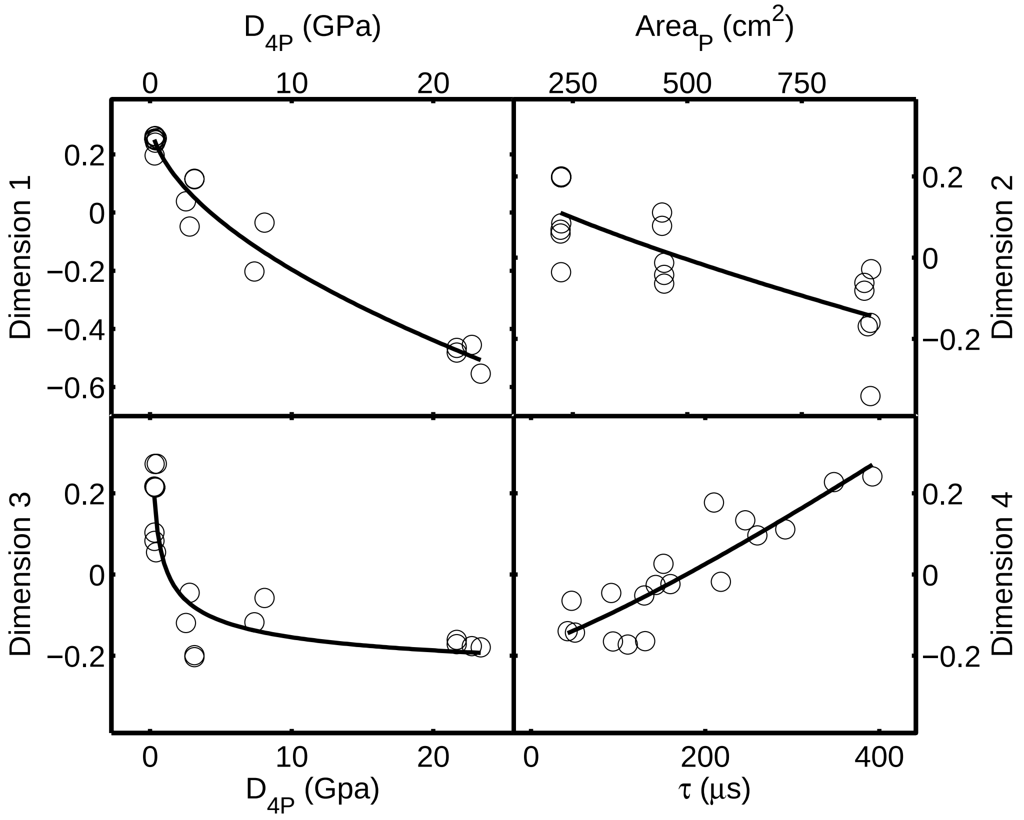

Figure 8.12: Source-based criteria for dissimilarity rating emerging from the analysis of model B. Continuous lines show the regression models.

well as by the interaction parameters τ and $acc_{\max}$. This ambiguity in the source-based explanation can be rectified considering that the only acoustical parameter that explains well this dimension was SCG_{att}, influenced by D_{4P} and ρ_P, but most strongly related to the interaction parameters τ and $acc_{\max}$ (see Table 8.8). For this reason this dimension is better interpreted as reflecting the perceptual salience of the properties of the interaction between hammer and sounding object, and of the related acoustical parameter SCG_{att}. It should be noted that the perceptual salience of this dimension for participants in the first latent class was almost null. Thus data from this experiment support the perceptual salience of interaction parameters, but only for a minority of participants.

Finally, independently of interindividual differences, hammer-related parameters were associated with none of the dimensions, thus revealing the absence of their perceptual relevance to dissimilarity rating.

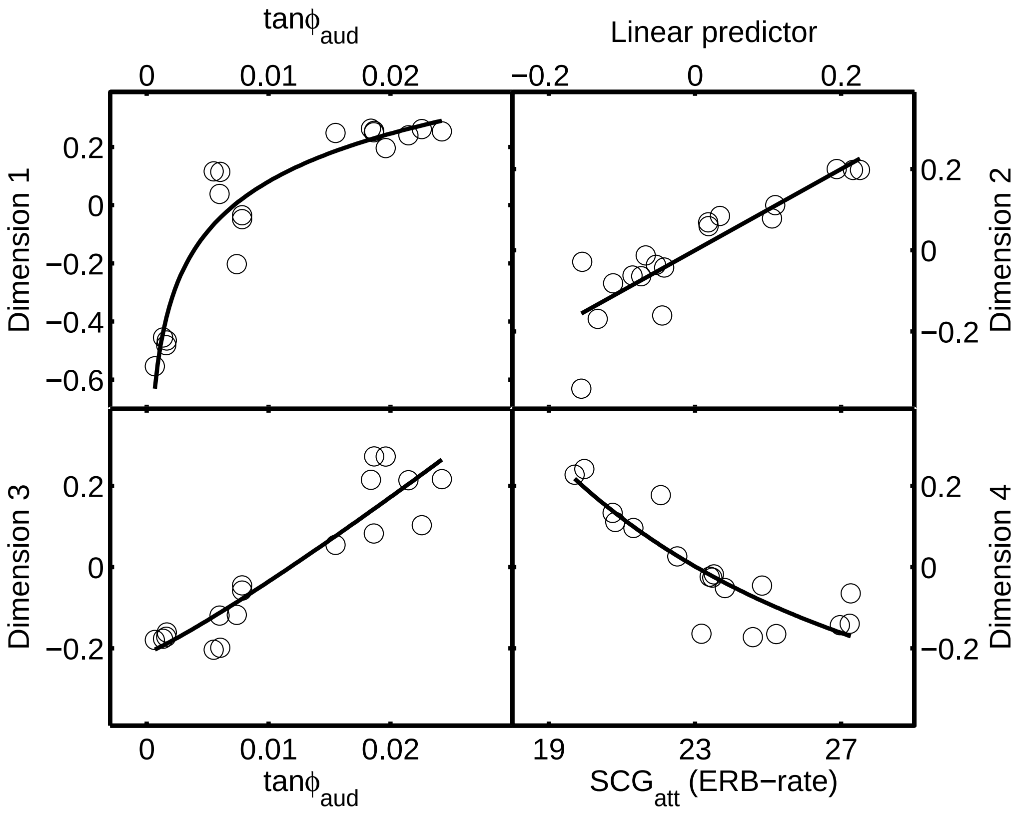

Figure 8.13: Acoustical-based criteria for dissimilarity rating emerging from the analysis of model B. Continuous lines show the regression models.

8.5 Conclusions

Dissimilarity ratings of impact sound sources were studied with reference to both the source and acoustical properties. A large database of sound sources was built, using several different materials for both the hammers and the plates, and varying the size of the plates. All sources were characterized in their physical properties, i.e. measuring the properties of the hammers, sounding objects, as well as those of the interaction among the two objects. The relationship between the acoustical structure of the impact sounds and the source parameters was studied, highlighting those acoustical properties more likely to mediate perception of each of the considered physical parameters.

A set of stimuli representative of the database was investigated. Results highlighted, consistently with previous dissimilarity rating studies, two main perceptual dimensions, the first related to the damping of the sounding object and to signal duration, the second related to the size of the sounding objects. Interaction parameters, and the related acoustical parameter attack SCG, were perceptually relevant, but only for a minority of participants. Independently of interindividual differences no evidence for the perceptual relevance of hammer properties was found. Some ambiguities in the source-based and acoustical-based explanations for the judgmental criteria were found. For example sounding objects' rigidities and densities explained equally well the main dimension for the largest group of participants. Future studies will need to decorrelate these source properties, which are strongly associated in real materials (cf. Waterman & Ashby, 1997).

Finally, the generalization of these results supports the notion according to which everyday perception of impact sounds is based mainly on the properties of the sounding object, and on two simple acoustical parameters: signal frequency and duration.

Chapter 9

Conclusions

The studies presented in Chapters 6-8 investigated perception of impact sound sources focusing on both source and acoustical properties as determinants of perception.

9.1 Everyday perception of impact sounds

As outlined in Chapter 5, a relevant issue concerns the determinants of perception in everyday conditions. Following this goal, several methodological choices were made in order to improve the ecological validity of the results.

Across studies differences in the physical and acoustical determinants of experimental judgments were observed. Five different sources may be at the origin of these variations: the experimental task, the nature of the stimulus set, the statistical models used to test for the relevance of stimuli properties in determining experimental judgments, interindividual differences and cultural differences (participants in the study in Chapter 6 were Swedish, Italian in Chapter 7, and French in Chapter 8). In Chapter 5, it was stated that the generalization of experimental results to everyday conditions requires assuming that the experimental task is representative of everyday perception. It was also stated that, given the difficulties in testing this assumption, a good option for the gathering of ecologically valid knowledge on the determinants of source perception would have been to compare results from studies based on different experimental tasks. Indeed, those acoustical and source properties emerging as relevant despite variations in experimental task would be characterized by a higher likelihood of also being relevant under everyday conditions. A similar line of reasoning can be followed when comparing results collected with different groups of participants, different stimulus sets, and based on the use of different statistical models. Thus, what remains constant despite variations in all these factors is highly likely to be relevant to everyday perception.

Concerning the physical determinants, the study in Chapter 6 found material identification of the sounding object to be influenced by the size and material of the sounding object, associated with changes in the density and, on the basis of measures published in Waterman and Ashby (1997), with changes in the elastic properties. Also, the acoustical measure of damping, $tan\phi$, was found to be relevant to the explanation of experimental judgments. Furthermore, it should be noted that results collected by Giordano (2003) highlighted material identification to be independent of variations in the material of the hammer. Consequently, and consistently with results by Roussarie (1999), material identification of the sounding object can also be conceived as independent of variations in the properties of the interaction between the hammer and the sounding object, such as the force stiffness K and the duration τ of the contact between these two objects during the impact. Overall, the study in Chapter 7 found perceived hardness of the hammer and of the sounding object to be influenced by variations in the damping of the sounding object, and by variations in the properties of the hammer/sounding object interaction, as modeled by the K parameter. Also, the relevance of signal frequency F pointed toward an influence of those properties of the impact sound source on this acoustical parameter, namely density, elasticity, size and shape of the sounding object (see Section 3.2). Finally, the study presented in Chapter 8 did not find dissimilarity ratings to be influenced by the properties of the hammer, but to be influenced by the size of the sounding object, the material-related properties of the sounding object (density, elasticity, damping), and, for a minority of participants, by the interaction property τ. In conclusion, following the considerations made above, it can be hypothesized that everyday perception of impact sounds relies on variations in the size and material-related properties of the sounding object (density, elasticity, damping as measured with the acoustical parameter $tan\phi_{aud}$) and, to a limited extent, on variations in the properties of the interaction among these two objects, but not on the properties of the hammer. A final remark on the generalization of these results should be made. With all the studies presented in this thesis, variations in the shape of the sounding object were not directly investigated. Given the perceptual relevance of shape already outlined in the experimental conditions (see Section 4.2.1), it is plausible to hypothesize everyday perception of impact sounds to be also influenced by this source property. Concerning the absence of relevance of the properties of the hammer, it should be noted that with the study presented in Chapter 8, and also in the study carried out by Giordano (2003), the range of variation of the density and elastic properties of the hammer was, to some extent, limited, and not completely representative of the values encountered in everyday conditions. Thus it cannot be excluded that experimental sets that also included highly soft materials (e.g., soft rubbers; felts) would have allowed an observation of relevant effects of hammer properties, i.e., would have allowed us to conclude that hammer properties have relevance in everyday listening.

Results concerning the acoustical determinants of experimental judgments can be found in Section 6.4.2 and Table 6.7 of Chapter 6, in Table 7.7 of Chapter 7 and in Table 8.14 of Chapter 8. As can be noted, in several cases the same dataset was equally well explained by different acoustical descriptors, or by different combinations of acoustical descriptors. Despite these ambiguities, several regularities emerged across studies. In particular, across experiments participants' judgment were associated with two simple acoustical variables: signal frequency and duration. An exception to this regularity is found only in the study presented in Chapter 7, where signal frequency and duration were relevant, but instead of both being used by the same participant, different participants based their judgment on only one of them. Leaving aside this difference, it can be concluded that in everyday conditions, source perception in impact sounds is based mainly on signal frequency and duration. Two other results are nonetheless worth mentioning. Firstly, across all studies the perceptual relevance of the acoustical measure of damping $tan\phi_{aud}$ (labeled $tan\phi$ in Chapter 6) was tested. This variable explained material identification and dissimilarity ratings, but not hardness ratings. Given this result, it might be concluded that $tan\phi_{aud}$ is not a good candidate as an acoustical parameter for the explanation of everyday perception of impact sound sources. This conclusion is also supported by the fact that a much simpler variable strongly associated with $tan\phi_{aud}$ explained judgments in all experimental investigations: signal duration. Secondly, one attack property, namely the spectral center of gravity of the first 10 ms of the signals, was associated with behavioral outcome in the hardness- and dissimilarity-rating studies. Thus it might be hypothesized that such a variable might play also a role in everyday perception of impact sound sources, although the appropriateness of this generalization is less firm than when signal duration and frequency are considered.

Finally, it should be stressed that the regularities observed across studies give some information concerning the source/acoustical properties relevant to the perception of impact sound sources, but do not give indications concerning which stimuli properties define the perceptual class of impact sounds and sources.

9.1.1 Generalization of experimental results

A few remarks should be made concerning the everyday conditions to which generalizations of the experimental results of this thesis should aim.

It is first important to consider the nature and availability of source-related information in everyday conditions, where acoustical information does not necessarily reach the perceiver in isolation. Indeed the acoustical signal might come along with information from other sensory modalities (one might see a small plastic object hitting the floor and hear the impact sound; one might touch, smell and tap on a melon to check if it is ripe), and with nonsensory source-related information (one might be in a restaurant, where clinking glasses are likely to be heard, rather than inside a garage, where impacts of metallic objects as tools are more likely to occur). All these additional types of information are likely to determine which acoustical and source properties are relevant to the perceiver. Concerning the influence of other sensory modalities, for example, one will almost surely recognize the material glass in a large struck glass plate when visual information is available, thus limiting the relevance of signal frequency observed in several different studies (see Chapter 6). However, it is often the case that the sound reaches the perceiver in the absence of additional source-related sensory information, i.e. it may be passively imposed on us and it might come from out-of-sight locations. Thus, experimental knowledge gathered with purely auditory stimuli might be more properly generalized to this latter condition, which should describe a rather relevant portion of everyday auditory experience. The influence of nonsensory source-related information on source perception is supported both by the experiments of Ballas and Mullins (1991), and by the partial relevance of this concept to the explanation of differences among studies related to the sounding object/hammer perceptual independence (see Section 7.1). It should be recognized that such nonsensory sources of information are hardly absent in everyday conditions. Nonetheless little or no experimental evidence is available concerning the relevance of this factor in determining the perceptually relevant source and acoustical properties. For this reason such issues may be reasonably ignored for the moment, although it is strongly recommended that future experimentation take them under consideration.

Additional factors that should be taken into account when discussing the ecological validity of experimental results concern the perceivers, i.e. their expertise with specific classes of acoustical signals and their goals when perception is taking place. It is highly likely that both these factors influence the source and acoustical properties relevant to the perceiver. Concerning expertise, it is not unrealistic to hypothesize that, when exposed to the sound of a vase shattering on the floor, an artisan that manufactures baked-clay, ceramic and porcelain objects will not only recognize that something broke, but will also be able to recognize of which of these three materials the object was made. On the other hand, a less expert listener will probably not be able to make such a discrimination, focusing mainly on the temporal patterning of the impacts that allow recognition of a breaking event. The results outlined in the preceding section would therefore be more properly generalized to a population of listeners without expertise for specific subsets of impact sounds. Concerning the goals of the perceiver, a mechanic might pay attention to different properties of the sound of the engine of a car under diagnosis, depending on the suspected problem. A similar ability to focus on different source and acoustical properties, depending on the requested judgment, was outlined in Chapter 7 with naive listeners. Such a result might then point toward the need to take into account the motivations of the perceiver when generalization of experimental results is of concern, but, given the paucity of studies, this problem will have to await future studies on this issue.

9.2 Theoretical issues

A relevant issue concerns the usefulness of the concept of invariant for the explanation of source perception. Testing this issue requires firstly ascertaining the presence of an acoustical structure that uniquely specifies the source properties under investigation, secondly to point out whether participants' judgments rely on such properties. Empirical evidence to these purposes was found in Chapters 6-7. In Chapter 6, acoustical parameters specifying the material of the sounding object independently of size variations were found, and auditory material identification was pointed out to be based on acoustical criteria other than those perfectly identifying the actual material. In Chapter 7, sufficient acoustical information for the independent perception of hammer and sounding object was present. In particular, acoustical parameters uniquely related to the interaction parameter K and to the sounding object parameters were found. Nonetheless, the acoustical criteria for the estimation of the hardness of the hammer and of the sounding object did not discriminate between these groups of source properties. Finally, the presence of acoustical parameters specifying selectively the different properties of the impact sound source was ascertained also with real sounds in Chapter 8, although the adopted experimental judgment did not allow for the testing of the presence of a related invariant (i.e., dissimilarity is purely psychological and, linguistically, has no physical referent). On the basis of the above-mentioned results it might then be concluded that the notion of invariant is not useful in explaining the perception of impact sound sources.

A final interesting issue, already pointed out in the study on material identification (see Section 6.4.3.1), concerns the origin of the response criteria when perception is tested with an explicit judgment of the properties of the sound source. In Chapter 6, the hypothesis was made that these originated from regularities in the everyday acoustical environment. The identification criteria for glass, for example, were hypothesized to reflect the size of freely resonating glass objects encountered in everyday life. Then, source perception might be conceived as reflecting the statistical properties of the acoustical environment, where judgment of source property A would rely also on property B if, in our acoustical environment, levels of B were unevenly distributed over levels or categories of A. In other words, glass identification would also rely on the size of the objects because in our everyday acoustical environment glass objects are much more frequently small than large. Despite the problems connected to ascertaining the statistical properties of everyday acoustical environment, it is highly probable that such an approach will be fertile with respect to the development of a theory of source perception and will be effective in explaining future and past experimental evidence.

References

Grey, J. M., & Gordon, J. W. Perceptual effects of spectral modifications on musical timbres. *J. Acoust. Soc. Am.*

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

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, *71*(1), 1-10.

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

Anderson, J. R. (1980). *Cognitive psychology and its implications.* New York: W. H. Freeman and Company.

Arabie, P., Hubert, L. J., & De Soete, G. (1996). *Clustering and classification.* River Edge, NJ: World Scientific Publ.

Avanzini, F., Rath, M., & Rocchesso, D. (2002). Physically-based audio rendering of contact. In *Proc. ieee int. conf. on multimedia & expo* (p. 445-448). Lausanne.

Avanzini, F., & Rocchesso, D. (2001a). Controlling material properties in physical models of sounding objects. In *Proceedings of the international computer music conference 2001* (p. 91-94). La Habana, Cuba.

Avanzini, F., & Rocchesso, D. (2001b). Modeling collision sounds: Non-linear contact force. In *Proc. COST-G6 Conf. on Digital Audio Effects (DAFx01)* (p. 61-66). Limerick, Ireland.

Baker, F. B., & Hubert, L. J. (1972). Measuring the power of hierarchical cluster analysis. *Journal of The American Statistical Association*, *70*, 31-38.

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.

Ballas, J. A., & Mullins, T. (1991). Effects of context on the identification of everyday sounds. *Human Performance*, *4*(3), 199-219.

Balzano, G. J. (1986). What are musical pitch and timbre? *Music Perception*, *3*, 297-314.

Benade, A. H. (1979). *Fundamentals of musical acoustics.* London: Oxford University Press.

Berger, K. W. (1964). Some factors in the recognition of timbre. *J. Acoust. Soc. Am.*, *36*(10), 1888-1891.

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

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*(10), 433-436.

Bregman, A. (1990). *Auditory scene analysis.* MIT press.

Cabe, P. A., & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filing. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 313-324.

Caclin, A. (2004). *Interactions et indepéndances entre dimensions du timbre des sons complexes.* Unpublished doctoral dissertation, Université Paris 6.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*, 1-27.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science*, *9*(3), 211-214.

Carello, C., Wagman, J. B., & Turvey, M. T. (2003). Acoustical specification of object properties. In J. Anderson & B. Anderson (Eds.), *Moving image theory: Ecological considerations.* Carbondale, IL: Southern Illinois University Press.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, *35*, 283-319.

Chaigne, A., & Doutaut, V. (1997). Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars. *J. Acoust. Soc. Am.*, *101*(1), 539-557.

Chaigne, A., & Lambourg, C. (2001). Time-domain simulation of damped impacted plates: I. Theory and experiments. *J. Acoust. Soc. Am.*, *109*(4), 1422-1432.

Clark, M. J., Luce, D., Abrams, R., Schlossberg, H., & Rome, J. (1963). Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones. *Journal of the Audio Engineering Society*, *11*(1), 45-54.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor anova designs. *Educational and Psychological Measurement*, *33*, 107-112.

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

de Silva, C. W. (2000). *Vibration: fundamentals and practice.* CRC Press LLC.

Doutaut, V., Matignon, D., & Chaigne, A. (1998). Numerical simulations of xylophones. II. Time-domain modeling of the resonator and of the radiated sound pressure. *J. Acoust. Soc. Am.*, *104*(3), 1633-1647.

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

Epstein, W. (1965). Nonrelational judgments of size and distance. *American Journal of Psychology*, *78*, 120-123.

Epstein, W. (1982). Percept-percept couplings. *Perception*, *11*, 75-83.

Eysenck, M. W. (2001). *Principles of cognitive psychology.* Hove, East Sussex, UK: Psychology Press Ltd. (Second Edition)

Fletcher, N. H., & Rossing, T. D. (1991). *The Physics of Musical Instruments.* New York: Springer-Verlag.

Fowler, C. A. (1990). Sound-producing sources as objects of perception: rate normalization and nonspeech perception. *J. Acoust. Soc. Am.*, *88*(3), 1236-1249.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *J. Acoust. Soc. Am.*, *87*(1), 311-322.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Gaver, W. W. (1988). *Everyday listening and auditory icons.* Unpublished doctoral dissertation, University of California, San Diego.

Gaver, W. W. (1993a). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, *5*(4), 285-313.

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

Gibson, J. (1966). *The senses considered as a perceptual system.* Boston: Houghton Mifflin.

Gibson, J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, *195*, 185-187.

Giordano, B. L. (2003). Material categorization and hardness scaling in real and synthetic impact sounds. In D. Rocchesso & F. Fontana (Eds.), *The sounding object* (p. 73-93). Firenze: Mondo Estremo.

Giordano, B. L., & McAdams, S. (submitted). Material identification in real impact sounds.

Golden, R. M. (2000). Statistical tests for comparing possibly misspecified and nonnested models. *Journal of Mathematical Psychology*, *44*, 153-170.

Gordon, A. D. (1999). *Classification.* Chapman & Hall /CRC.

Gordon, I. E. (1989). *Theories of visual perception.* New York: John Wiley & Sons.

Gower, J. C. (1966). Some distance properties of latent root and vector methods using multivariate analysis. *Biometrika*, *53*, 325-338.

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* (p. 392-397).

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

Grassi, M., & Burro, R. (2003). Impact sounds. In D. Rocchesso & F. Fontana (Eds.), *The sounding object* (p. 47-72). Firenze: Mondo Estremo.

Gregory, R. L. (1970). *The intelligent eye.* New York: McGraw-Hill.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, *61*(5), 1270-1277.

Guski, R. (2000). Studies in auditive kinetics. In M. M. A. Schick & C. Reckhardt (Eds.), *Contributions to psychological acoustics. results of the 8th oldenburg symposium on psychological acoustics* (p. 383-401). BIS.

Gygi, B. (2001). *Factors in the identification of environmental sounds.* Unpublished doctoral dissertation, Indiana University, Department of Psychology.

Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *J. Acoust. Soc. Am.*, *115*(3), 1252-1265.

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

Hajda, J. M. (1996). The effect of reverse playback and signal partitioning on the identification of percussive and nonpercussive musical tones. In B. Pennycook & E. Costagiomi (Eds.), *Proceedings of the 4th international conference on music perception and cognition* (p. 25-30). Liége, Belgium: European Society for the Cognitive Sciences of Music.

Hajda, J. M. (1999). *The effect of time-variant acoustical properties on orchestral instruments timbre.* Unpublished doctoral dissertation, University of Californa, Los Angeles.

Hajda, J. M., Kendall, R. A., & Carterette, E. C. (1994). Perceptual and acoustical analyses of impulse tones. In I. Deliége (Ed.), *Proceedings of the 3rd international conference on music perception and cognition* (p. 43-51). Liége, Belgium: European Society for the Cognitive Sciences of Music.

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* (p. 253-306). London: L. Erlbaum.

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

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

Handel, S., & Erickson, M. L. (2001). A rule of thumb: the bandwidth for timbre invariance is one octave. *Music Perception*, *19*(1), 121-126.

Handel, S., & Erickson, M. L. (2004). Sound source identification: the possible role of timbre transformations. *Music Perception*, *21*(4), 587-610.

Harsberger, M. L., Kendall, R. A., & Carterette, E. C. (1994). Comparative psychoacoustics and acoustics of indonesian gong tones. In I. Deliége (Ed.), *Proceedings of the 3rd international conference on music perception and cognition* (p. 313-314). Liége, Belgium: European Society for the Cognitive Sciences of Music.

Hartmann, W. M. (1997). *Signals, sound and sensation.* Woodbury, NY: AIP Press.

Hope, A. C. (1968). A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society, Series B*, *30*, 582-598.

Hosmer, D., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Houben, M. M. J. (2002). *The sound of rolling objects. perception of size and speed.* Unpublished doctoral dissertation, Technische Universiteit, Eindhoven.

Houben, M. M. J., Kohlrausch, A., & Hermes, D. J. (2004). Perception of the size and speed or ralling balls by sound. *Speech Communication*, *43*, 331-345.

Houix, O. (2003). *Categorisation auditive des sources sonores.* Unpublished doctoral dissertation, Université du Maine, France.

Houix, O., McAdams, S., & Caussé, R. (1999). Auditory categorization of sound sources. In M. A. Grealy & J. A. Thomson (Eds.), *Studies in perception and action v* (p. 47-51). Lawrence Erlbaum Associates, Inc.

Hubert, L. J., & Levin, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Psychological Bullettin*, *83*, 1072-1080.

Hunt, K. H., & Crossley, F. R. E. (1975). Coefficient of Restitution Interpreted as Damping in Vibroimpact. *ASME J. Applied Mech.*, 440-445.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *J. Acoust. Soc. Am.*, *94*(5), 2595-2603.

Jansson, G. (1993). Perception of the amount of fluid in a vessel shaken by hand. In S. Valenti & J. Pittinger (Eds.), *Studies in perception and action II* (p. 263-267). (Posters presented at the VIIth International Conference on Event Perception and Action,)

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201-211.

Kendall, R. A. (1986). The role of acoustical signal partitions in listener categorization of musical phrases. *Music Perception*, *4*(2), 185-214.

Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, *16*(3), 327-364.

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

Killion, M. C. (1978). Revised estimate of minimum audible pressure: where is the "missing 6 db"? *J. Acoust. Soc. Am.*, *63*(5), 1501-1508.

Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environment*, *9*(4), 399-410.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caracterisation du timbre de sons complexes: II. analyses acoustiques et quantification psychophysique. *Journal de Physique*, *4*, 625-628.

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* (p. 43-51). Amsterdam, Netherlands: Excerpta Medica.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 279-294.

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

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: simple geometric form. *Perception & Psychophysics*, *59*(8), 1180-1190.

Lambert, J. D. (1993). *Numerical Methods for Ordinary Differential Systems.* Chichester, UK: John Wiley & Sons.

Lambourg, C. (1997). *Modèle temporel pour la simulation numérique de plaques vibrantes - application à la synthèse sonore.* Unpublished doctoral dissertation, Université du Maine, France.

Lambourg, C., Chaigne, A., & Matignon, D. (2001). Time-domain simulation of damped impacted plates: II. Numerical models and results. *J. Acoust. Soc. Am.*, *109*(4), 1433-1447.

Landau, L., & Lifshitz, E. M. (1981). *Theory of elasticity.* Pergamon Press.

Lass, N. J., Eastham, S. K., Parrish, W. C., Schebrick, K. A., & Ralph, D. M. (1982). Listeners' identification of environmental sounds. *Perceptual and Motor Skills*, *55*, 75-78.

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

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *J. Acoust. Soc. Am.*, *90*(6), 3036-3049.

Lindsay, P. H., & Norman, D. A. (1977). *Human information processing - an introduction to psychology.* New York: Academic Press.

Lutfi, R. (2001). Auditory detection of hollowness. *J. Acoust. Soc. Am.*, *110*(2), 1010-1019.

Lutfi, R. A., & Oh, E. L. (1997). Auditory discrimination of material changes in a struck-clamped bar. *J. Acoust. Soc. Am.*, *102*(6), 3647-3656.

Marhefka, D. W., & Orin, D. E. (1999). A Compliant Contact Model with Nonlinear Damping for Simulation of Robotic Systems. *IEEE Trans. Systems, Man and Cybernetics-Part A*, *29*(6), 566-572.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *J. Acoust. Soc. Am.*, *114*(5), 2946-2957.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman.

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

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: the cognitive psychology of human audition* (p. 146-198). Oxford University Press.

McAdams, S. (2000). The psychomechanics of real and simulated sound sources. *J. Acoust. Soc. Am.*, *107*(5), 2792.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomecanics of simulated sound sources: Material properties of impacted bars. *J. Acoust. Soc. Am.*, *115*(3), 1306-1320.

McAdams, S., Kudo, K., & Kirchner, H. (1998). Perceptual interaction of excitor and resonator properties in percussive instrument sounds. In *Ica: International congress on acoustics* (p. 1885-1886). Seattle Washington, USA.

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.

McIntyre, M. E., & Woodhouse, J. (1988). On measuring the elastic and damping constants of orthotropic sheet materials. *Acta metall.*, *36*(6), 1397-1416.

Michaels, C. F., & Carello, C. (1981). *Direct perception.* Englewood Cliffs, New Jersey: Prentice-Hall.

Miller, J. R., & Carterette, E. C. (1975). Perceptual space for musical structures. *J. Acoust. Soc. Am.*, *58*(3), 711-720.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325-342.

Milligan, G. W. (1981). A monte carlo study in thirty internal criterion measures for cluster analysis. *Psychometrika*, *46*(2), 187-199.

Milligan, G. W. (1996). Clustering validation: results and implications for applied analyses. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), *Clustering and classification* (p. 341-375). River Edge, NJ: World Scientific Publ.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159-179.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.*, *74*(3), 750-753.

Ohta, K., Kuwano, S., & Namba, S. (1999). Sound quality of impulsive sounds in relaiton to their physical properties. *Technology reports of the Osaka University*, *49*, 189-199.

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

Palmer, S. E. (1975). Visual perception and world knowledge: notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (p. 279-307). San Francisco: Freeman.

Patterson, R. D., Allerhand, M., & Gigure, C. (1995). Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am.*, *98*(4), 1890-1894.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: transoforming numbers into movies. *Spatial Vision*(10), 437-442.

Plack, C. J., & Moore, B. C. J. (1990). Temporal window shape as a function of frequency and level. *J. Acoust. Soc. Am.*, *87*(5), 2178-2187.

Repp, B. H. (1987). The sound of two hands clapping: an exploratory study. *J. Acoust. Soc. Am.*, *81*(4), 1100-1109.

Richards, W. (1988). Sound interpretation. In W. Richards (Ed.), *Natural computation* (p. 301-308). Cambridge, MA: MIT Press.

Rocchesso, D., & Fontana, F. (Eds.). (2003). *The sounding object.* Firenze, Italy: Mondo Estremo. (<http://www.soundobject.org>)

Rocchesso, D., Ottaviani, L., Fontana, F., & Avanzini, F. (2003). Size, shape, and material properties of sound models. In D. Rocchesso & F. Fontana (Eds.), *The sounding object* (p. 73-93). Firenze: Mondo Estremo.

Rossing, T. D., Moore, F. R., & Wheeler, P. A. (2002). *The science of sound.* San Francisco, CA: Addison Wesley. (Third Edition)

Roussarie, M. V. (1999). *Analyse perceptive de structures vibrantes simulees par modele physique.* Unpublished doctoral dissertation, Université du Maine, France.

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *J. Acoust. Soc. Am.*, *36*(10), 2021-2026.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Serafini, S. (1993). *Timbre perception of cultural insiders: a case study with Javanese gamelan instruments.* Unpublished master's thesis, University of British Columbia, Vancouver.

Srinivasan, A., Sullivan, D., & Fujinaga, I. (2002). Recognition of isolated instrument tones by conservatory students. In *Proocedings of the 7th international conference on music perception and cognition* (p. 720-723).

Strong, W., & Clark, M. (1967). Synthesis of wind-instrument tones. *J. Acoust. Soc. Am.*, *41*(1), 39-52.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Tousman, A., Pastore, R. E., & Schwartz, S. (1989). Source characteristics: a study of hand clapping. *J. Acoust. Soc. Am.*, *85*, S53.

Tucker, S., & Brown, G. J. (2003). Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds. In *Aes 2003.*

van den Doel, K., & Pai, D. K. (1998). The sounds of physical shapes. *Presence*, *7*(4), 382-395.

Vanderveer, N. J. (1979). *Ecological acoustics: Human perception of environmental sounds.* Unpublished doctoral dissertation, Cornell University. (Dissertation Abstracts International, 40, 4543B. (University Microfilms No. 80-04-002))

von Helmholtz, H. (1877/1954). *On the sensation of tone.* New York: Dover.

von Hornbostel, E. M., & Sachs, C. (1914). Systematik der musikinstrumente. Ein versuch. *Zeitschrift für Ethnologie*, *4-5*, 553-590.

von Hornbostel, E. M., & Sachs, C. (1961). Classification of musical instruments (translated by Baines A., and Wachsmann K. P.). *The Galpin Society Journal*, *14*, 3-29.

Ward, W. D. (1970). Musical perception. In J. Tobias (Ed.), *Foudations of modern auditory theory, volume I* (p. 405-447). New York: Academic Press.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392-393.

Warren, W. H., Kim, E. E., & Husney, R. (1987). The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. *Perception*, *16*, 309-336.

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.

Waterman, N. A., & Ashby, M. F. (1997). *The materials selector.* London: Chapman & Hall. (Second Edition)

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

Wessel, D. L. (1973). *Psychoacoustics and music: a report from Michigan State University* (Vol. 30; Tech. Rep.). PACE: Bullettin of the Computer Arts Society.

Wildes, R., & Richards, W. (1988). Recovering material properties from sound. In W. Richards (Ed.), *Natural computation* (p. 356-363). Cambridge, MA: MIT Press.

Winsberg, S., & Carroll, J. D. (1989a). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika*, *54*, 217-229.

Winsberg, S., & Carroll, J. D. (1989b). A quasi-nonmetric method for multidimensional scaling of multiway data via a restricted case of an extended INDSCAL model. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (p. 405-414). Amsterdam, Holland: Elsevier Science.

Winsberg, S., & De Soete, G. (1993). A latent class approach to fitting the weighted euclidean model, CLASCAL. *Psychometrika*, *58*(2), 315-330.

Zener, C. (1948). *Elasticity and anelasticity in metals.* Chicago, Illinois: University of Chicago Press.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: facts and models* (2 ed.). New York, NY: Springer-Verlag.