FP6-NEST-PATH projet no: 29085

Closing the Loop of Sound Evaluation and Design (CLOSED)

Deliverable 4.1

Everyday sound classification: Sound perception, interaction and synthesis

Part I - State of the art

Olivier Houix, Guillaume Lemaitre,
Nicolas Misdariis, Patrick Susini (IRCAM)
Karmen Franinovic, Daniel Hug, Karmen Franinovic,
Jacqueline Otten, Jill Scott, Yon Visell (HGKZ)
Delphine Devallez, Federico Fontana, Stefano Papetti,
Pietro Polotti, Davide Rocchesso (UNIVERONA)

January 2007

Contents

Introduction

The CLOSED project aims at providing a functional-aesthetic sound measurement tool that can be profitably used by designers. To achieve this goal, the CLOSED consortium incorporates four different expertises, ranging from physics and signal processing, design, acoustics and psychology of perception, to computer science.

This document is the first scientific deliverable (4.1) of the CLOSED project. It takes part to the workpackage WP4 (sound reception), lead by IRCAM. HGKZ and UNIVERONA have as well contributed to the results of the deliverable.

The overall aim of this deliverable is to provide recommendations for sound design tools and for the development of proto sound products. It is divided into two parts. This first part (due on month 5) consists of a review of perceptual studies of everyday sound perception, design and synthesis. This review is complemented by an introduction to the influences of context and culture on the perception of product sound quality. The second part of the deliverable (due on month 11) will provide a model for the classification of everyday sounds, based on this prior review, developed and described according to different type of properties. Properties will include acoustical patterns conveying meaningful information and semantic attributes and underlying functional-aesthetic experiences. This work will complete and extend the theoretical assumptions on perception and classification of everyday sounds and challenge major theoretical contributions. These results will also provide useful guidelines for the definition of the naturalistic settings in WP2 for sound synthesis algorithms and for the development of proto sound products in WP3.

This document is made of three main parts. The first part consists of a state of the art of everyday sound perception. Through this review, we will give insights into how listeners perceive and cognitively represent sound events. Sound events are sounds which are perceived and represented as the identified event that produces the sound. We will review the properties of the event that listeners are able to auditorilly recover. The issue of sound identification will therefore lead us to the theories of categorization, in order to understand how listeners do categorize everyday sounds. We will also provide some examples of existing classifications of everyday sounds.

The second part of this document focuses on everyday sound interaction and design. Indeed sound quality and design have applied many experimental results found in studies of everyday sound perception. Particularly, we will focus on the influence of the contextual and cultural factors on the perception of product sound quality. Then basic interaction design will be thoroughly described, from a theoretical and historical perspective.

Finally, the third part of this document is a introductory and detailed guideline to the taxonomy of sound synthesis algorithms proposed by UNIVERONA. This collection was initiated in the Sounding Object project (2001-2003). New models are currently added to complete the taxonomy. This taxonomy is inspired by perceptual studies of everyday sound perception.

1 Everyday sound perception

The first five months of the workpackage 4 of the CLOSED project were dedicated to reviewing the scientific literature related to the perception of everyday sounds and to the theories of categories and categorization. The state of the art proposed below provides us with experimental results and theoretical assumptions which are the first bricks for a classification model of everyday sounds.

Introduction

The study of auditory perception has been historically based on synthetic simple stimuli, music, and speech. More recently, researchers have also began to focus on the perception of another category of sounds: environmental and every day sounds. One of the first findings was that listeners seemed to produce response patterns, when they listened to everyday sounds, that were different from those they provided when they listened to musical sounds. This results raised the following issues: do human listeners perceive differently environmental sounds and, for instance, musical sounds ? Can the results found with synthetic sounds be generalized to other kind of sounds ? These questions lead to develop new methodologies to study everyday sounds, and at the same, and shed new light on the perception of musical sounds. In the following sections, we propose a state of the art of these studies who have focused on everyday sound perception, from two standpoints.

In a first part (section 1.1), we will review the works that have focused on the perception of everyday sounds. We will in particularly show that some sounds are perceived and represented by the listeners as the producing source: we will label these sounds the sound events. We will then review the studies which have studied how listeners perceive the sound events and their properties. This will lead us to the more general issues of identification and classification.

Thus, we will focus in the second part (section 1.2) on how human listeners organize their knowledge. Particularly, we will work on the psychological categorization of everyday sounds, starting from the general theoretical notions of categorization. First we present definition of categorization and develop different theories on the structure of categories and its relation with similarity. We will then review theoretical approach to describe the relation between categories, illustrated by models of semantic network. And we give examples of everyday sound classification proposed in the literature to propose framework of classification model of everyday sounds based on theoretical assumptions we exposed.

1.1 Perception of everyday sounds

Introduction

Although many works reported in the literature deal with environmental or everyday sound perception, very few of them actually define which kind of sounds are labelled by these expressions. We found only the following definition, proposed by Vanderveer ([135] pp. 16-17) :

> "... any possible audible acoustic event which is caused by motions in the ordinary human environment. (...) Besides 1) having real events as their sources (...) 2) [they] are usually more "complex" than laboratory sinusoids, (...) 3) [they] are meaningful, in the sense that they specifiy events in the environment. (...) 4) The sounds to be considered are not part of a communication system, or communication sounds, they are taken in their literal rather than signal or symbolic interpretation."

In short, the "environmental sounds" used in the studies that we will review below are every sounds that may occur in the human environment, and which are neither music, speech, animal communication, or abstract synthetic sounds (altough we may cite some results concerning these sounds, for the sake of comparison). In this document, we will use the expressions equally "environmental" and "everyday" sounds.

We will first spend a few words on soundscape perception (section 1.1.1). These studies, although a little bit outside of our main interests, introduce nevertheless a very interesting distinction between the *amorphous* sound sequences and the sound sequences with *sound events*. Actually, only this latter category fits all of the Vanderveer's definition requirements, and is relevant to the issues addressed by this deliverable. We will then focus on this latter category of sounds (section 1.1.2). Afterwards (section 1.1.3) we will lend more attention on some works addressing the theoretical issues of identification and classification of environmental sounds. Finally (section 1.1.4), we will provide a synthesis of these reviews.

1.1.1 Perception of soundscapes

The term "soundscape" was introduced in the late 70's by the Canadian composer R. Murray Schafer [118], who defined soundscape as the auditory equivalent to landscape. Beside Schafer's project, the term soundscape perception is used in a scientific context to characterize how inhabitants perceive, experience and appraise their sonic environnement. Since these studies are often related to *urban* soundscape perception, they are thus as well closely related to the notion of annoyance (see Guski [58] for a discussion of the concept of annoyance). While some studies have studied annoyance caused by urban soundscapes in terms of acoustic level [116], or urban structure and traffic configuration [136], several studies have revealed that soundscape appraisal relies on a much wider variety of aspects, including semantic values [34], socio-cultural and biographical dimensions [120].

Because of this complexity, studying soundscape perception requires specific methods and experimental procedures that exceed the classical psychophysical paradigms. Following this idea, a series of experiments were realized in France, which are worth detailing (see Dubois [32] for a summary of the issues addressed by this series of papers, and Dubois and Schulte-Fortkamp for a review of recent advances [119]). These experiments were realized in situ (e.g. in streets, parcs), or in laboratory. In this latter case however, a great care was taken to reproduce the actual condition in which listeners normally experience the soundscapes: this is the issue of ecological reproduction [56, 101, 57]. Psycholinguistic methods were used, because, according to the authors, cognitive representations cannot be accessed directly [87, 56, 104, 103, 34, 57]. However, since linguistic devices used by the listeners reflect, through collective expressions, the individual cognitive representations and knowledge of the listeners, they provide an indirect way to study listeners' perception. They listened to rather long sequences of environmental sounds, and were asked to freely classify and describe what they heard. Among the rich insights provided by these studies, a conclusion is of major importance for the issue of classification of everyday sounds. The authors noticed quite systematically that listeners perceive differently *amorphous* sound sequences ("background noises") and sound sequences in which listeners were able to identify emerging *sound events* (see [32, 33] for a discussion). At a generic level, they split the sounds into these two categories, they used rather different linguistic devices to describe the sequences, and differences appeared also in the fine classification of the sound sequences and in the appraisal of the sequences. Listeners described the sounds belonging to the former category of amorphous sequences as rather abstract objects, external to the listener, mainly described in terms of apparent properties (timbre, structure, etc.), while they described the latter category by refering to the identified sound sources, and with reference to how the source affects themselves in their everyday lives. Sound events are not distinguished from their sources and listeners appraise the source and the values they associate to the source.

Therefore, common to these results and to the Vanderveer's definition arises the notion of *sound event*. Sound events are perceived by listeners as specifying an event occurring in the environment. Sound events are those kind of sounds that we are interested with.

1.1.2 Perception of sound events

When an event occurs in the environment and produces a sound (i.e. a *sound event*), the sound wave propagates and reaches the ears of the listener, who perceives an *auditory event*: we define here the auditory event as the cognitive representation of a sound event. The perception of an auditory event is thus mediated by an acoustic wave. However, as shown, among others, in the studies of soundscape perception reported above, what a listener perceives are not necessary the properties of the acoustic wave (or *proximal stimulus*) but the properties of the event (*distal stimulus*): he identifies the source of the sound, the object that makes the sound, its properties, how an agent may have interacted with this object to produce a sound, associates the sound with the semantic, hedonic values of its category, etc. (see also [135, 9]). To understand the formation of an auditory event, there are, according to Li et al. [79], three pairwise relationships to consider (see Figure 1.1):

- The relationships between the properties of the occurring sound event and the properties of the radiated acoustic wave (1),
- The relationships between the properties of the radiated acoustic wave and the perceived auditory event (2),
- The relationships between the properties of the occurring sound event and the perceived auditory event (3).

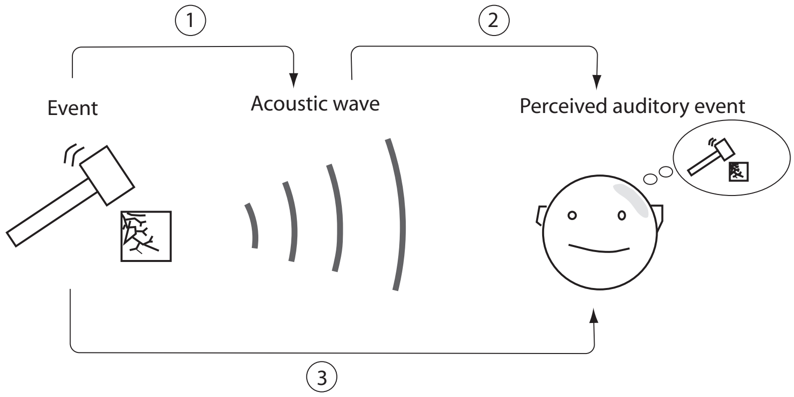


Figure 1.1: When an event occurs and produces a sound perceived by a listener, three relationships have to be considered: 1. between the occurring sound event and the radiated acoustic wave, 2. between the radiated acoustic wave and the perceived auditory event, 3. between the occurring sound event and the perceived auditory event (from Li et al. [79]).

Most of the studies reported in the literature choose to focus on one of the two latter relationships (2 or 3), while the first relationship is almost always implicitly addressed. These two issues define the two categories of experimental works that we review in the paragraphs below.

Occurring sound events and perceived auditory events

Several studies, often inspired by the Gibsonian ecological psychology, have focused on which characteristics of a sound event a listener is able to perceive (relation 3 in Figure 1.1). Gibson's idea of *direct perception* [51], according to which a listener perceives directly the environment (in our case: what makes the sound), shares similarities with the motor theory of speech perception (see for instance [80]). Indeed Fowler (see the controversy between Fowler and Diehl et al. in [45, 28, 46]) claims that speech perception is not special: just as a listener perceives the phonetic mechanisms that produce speech (what phonetic motion he would have to do to pronounce it), he perceives the mechanical properties of the event that makes the sound. In [45] for instance, she reported a series of experiments in which she let subjects listen to the sound of a steel ball rolling on different configurations of a two-phase track (muffled downward ramp + sounding flat track, or muffled downward ramp + sounding upward ramp) and asked them to assess the steepness of the first ramp. If listeners were only sensitive to low level auditory attributes (in this case: if their judgement of steepness would have been only based on the duration of the rolling along each part of the track, i.e. *durational contrast*), the different configurations of the second part of the track would have trapped them: indeed, when the second part of the track is bent upward, the ball takes more time to roll along it than when this track is flat, just as if the first part of the track were shallow. However, the results showed that the listeners were able to correctly classify the slope of the first phase of the tracks as steep or shallow, without being confused by the sounds resulting from the different configurations of the second part of the track, indicating that the listeners were not only sensitive to basic attributes (duration of each phase), but actually recover the speed of the ball.

Structural invariants Among the properties of a sound event that may be perceived, ecological psychologists distinguish between those which "specify the kind of object and its properties under change": *structural invariants*, and those which "specifies the style of change itself": *transformational invariants* [138]. Structural invariant properties of sound events (mostly impact) have been reported in many papers. Carello et al. [20] showed that listeners are able to report the *length* of wooden rods dropped on the floor relatively correctly. Lakatos et al. [77] let listeners listen to pairs of recorded sounds of struck bars made of wood or metal, and match them to a pair of pictures depicting the *sections* of the bars. Listeners succeeded relatively well to the task. Matching scores were transformed into similarities and analysed with a multidimensional scaling technique. They found two dimensions. The first one was correlated with the width/height ratio (roughly correlated to the torsion modes frequencies), and the second one with the spectral centroid. A cluster analysis showed that listeners discriminated mostly between thin plates and to thick blocks. Kunkler-Peck and Turvey [76] hung plates of different shapes (circular, squared, rectangular) and materials behind a curtain. Their results showed that listeners were able to report the *shape (squared, rectanguar or circular)* better than chance, and the *materials* almost perfectly. Grassi [55] showed that listeners were able to identify correctly the *shape* of a ball dropped on a plate. When the listeners had to draw the *size* of the ball, across different plate and ball sizes, he found a power law between the perceived and actual size of the ball, but exponent changed accross the thicknesses of the plates. He concluded that listeners are not able to judge independently the effect of the sounding object (the plate) and of the non-sounding object (the ball). Giordano and McAdams [52] studied the identification of the *material* of recorded struck plates. Listeners were able to discriminate roughly between metal and glass on the one hand, and from wooden and Plexiglas on the other hand. However, within each of these categories, they failed to identify the material correctly. Rather, listeners tended to associate small plates with glass or wood, and large plates with metal or Plexiglas. They found a predictive model based on loudness and frequency, and proposed also an ecological explanation to these results: there may be an ambiguity between the sound of a glass or metal bar. But as listeners are not used to manipulate big glass objects, they associate big objects with metal, and small objects with glass.

Transformational invariants Transformational invariants have been less often investigated. Warren and Verbrugge [138] studied the identification of *bouncing* or *breaking* events (glass objects falling). They showed that the invariants which characterize bouncing or breaking are based on the synchronicity of the individual impacts that form the sound of a bouncing or breaking event. Cabe and Pittenger [18] studied how listeners may use the sound that makes a vessel when it is *filled with a liquid* to control the pouring of the liquid. They showed that blindfolded subjects were able to fill the vessel to a normal drinking level, or to the brim. They concluded that the sound of pouring water "affords" filling, and suggested that there exists an invariant in the sound that allows to perceive the time remaining before overflowing.

Radiated acoustic wave and perceived auditory event

These results show that listeners are able to auditorilly recover some of the characteristics of a sound event. Even if a mere ecological (Gibsonian) psychologist would claim that this perception is direct, the classical standpoint of cognitive psychology (information processing) assumes that the recovery of the properties of a sound event is inferred from the perceptual attributes of the sound wave [89]. Therefore, a lot a studies reported in the literature aim at finding the acoustic basis of the perception of a sound event (some of the studies reported above sought as well the acoustical correlates of the perceived properties of the sound events): this is the relation 2 in Figure 1.1.

From pure tones to everyday sounds Traditionally, the perception of sounds have been studied in the psychoacoustical framework, and the perception simple sounds (pure tones, noise bursts) have been extensively inquired. In 1989, a panel of scientists published a review of the literature on the perception of *complex sounds* (spectrally complex, temporally complex, brief and embedded in noise, excluding speech and music) [141]. They proposed, among other recommendations for future research, to adapt and develop new theories and methods "to form a bridge between studies of simple stimuli and those that might be conducted using complex sounds" (p. 5). Particularly, they proposed to expand methods to account for the multidimensional nature of complex sounds.

The multidimensional nature of sound perception Indeed, multidimensional techniques have been widely used to study the perception of the timbre of musical sounds (see Hajda et al. [62] for a review). The multidimensional aspect of natural sounds (songs of birds, a purring cat, a crying baby, shouting people, etc.) has been studied by Björk [13] using the semantic differential method, although the results were mostly qualitative. More quantitative results were found by translating the multidimensional scaling technique used to study the timbre of musical sounds (see [91]) to environmental sounds. Susini et al. [130] applied the multidimensional scaling technique to air-conditioning noises. They found three perceptual dimensions characterizing the perceived differences between sounds produced by different brands and models of indoor air-conditionning units. These dimensions were correlated respectively with the ratio of the noisy part of the spectrum to the harmonic part, with the spectral center of gravity, and with the loudness. Using the same technique, Lemaitre et al. [78] studied the perception of car horn sounds. They found again three dimensions characterizing the perceived differences between the sounds, correlated with the roughness, the spectral center of gravity, and with a descriptor of the fine shape of the spectral envelope. They found as well that listeners were actually able to discriminate the different mechanisms involved to produce the sound on the basis of these three dimensions. These studies have single out the perceived dimensions underlying the perceived differences between the sounds of a set, and the acoustic descriptors related to these dimensions. However, as indicated by Susini et al. [129], this method is only valid for sounds that listeners are able to compare along continuous dimensions. When the sounds are identified as different sources, the perceived differences can not be characterized by the differences along a reduced set of continuous perceptual dimensions. Moreover the perceptual dimensions characterize the sounds within the set, but do not explain what characterize the set.

Acoustic correlates to material, shape, hardness ? While these works studied the perception of a set of different sources, another series of studies focused on the perception of some properties of an event, and tried to relate the perceived properties with the properties of the sound wave. Freed [47] studied the perceived hardness of mallets striking pans. Listeners had to rate the hardness of the mallet by listening to the sound. The results showed that they were able to rank correctly the mallets according to the hardness, even if the pans varied. The perceived hardness was however poorly correlated with the acoustic parameters they tested. Li et al. [79] studied the identification of the gender of walkers. Rather than a discrimination between male and female, they found a continuum of "maleness" (defined as the proportion of listener identifying the walker as a male). This maleness ratio was correlated with some descriptors of the spectrum (shape of spectral peak, and contribution of high frequencies). They showed as well that listeners identified women wearing male shoes as men and conversely. They concluded that the identification of the gender of the walker is based on inferences made by listeners on the basis of the perceived spectral properties of the acoustic signal, and on a stereotypical association between combinations of these properties and the assumed gender of the walker. Quite the same conclusions were obtained by Repp [107] who studied the perception of hand clapping. While listeners were able to associate the sound of the hand clapping with the picture of the configuration of the two hands, they failed to identify the clapper or even the gender of the clapper. Rather, he noticed that listeners associate stereotypically slow, loud and low frequency sounds with male clappers.

McAdams et al. [90] studied the perception of the material of impacted bars (synthesized sounds with physical models) with the multidimensional scaling technique. They intended to simulate sounds of bars from different materials by varying the damping coefficients and the density parameter of the model. The perceived dissemblances between the synthesized bars of different materials were perceived along two dimensions. The first dimension, on the one hand, was correlated with the damping parameter, and on the other hand with a combination of analytical parameters derived from the signal describing the decreasing slope of the temporal envelope. The second dimension was correlated on the one hand with the material density, and on the other hand with the first partial frequency.

Event perception as a decision Synthesized sounds were studied as well in a series of papers published by Lutfi et al. (they showed in [84] that synthesized sounds elicit the same response patterns as natural sounds do, and that results obtained with these sounds can thus be generalized to natural sounds). Their experimental paradigm was quite original. In [85], they studied the perception of the material of impacts. Carefully trained listeners had to choose which one of two presented sounds corresponded to a given material. They modelized the situation as a bivariate decision task (see [82] for a description of the model). They synthesized the sounds in order to create a theoretical optimal response criteria based on the perceived density and elasticity of the material. Their results showed that the criteria used by the listeners overweights the frequential aspects of the sounds, and thus was not the optimal criteria. This lead them to conclude that listeners are not able to recover all of the information available in an acoustic signal characterizing the sound event. In [83], they studied the perception of the hollowness of a struck bar, and found two groups of subject. While the first group was actually able to use the optimal criteria (based on the perception of a combination of partial amplitudes, frequency or decay) and thus was able to optimally discriminate the hollowness of the bars, the second group of subjects overweighed their perception of frequency and thus failed to precisely discriminate the hollowness of the struck bars.

Is there a specific spectral region containing the information ? Quite a different paradigm was used by Gygi et al. [60] who sought to identify if a privileged area in the spectrum of natural sounds would embed the information required to recognize the source of the sounds. They filtered environmental sounds and asked listeners to choose among four presented labels which one correspond to the source. Their results were highly dependant on the type of sound sources.

1.1.3 Identification and classification of environmental sounds

From the studies reported above arise systematically two important issues: identification of the source or of its properties, and classification of sounds into categories. These two issues have been more thoroughly studied by several authors. As these issues form actually the core of our problematic, we review these studies with much details in the following paragraphs.

Vanderveer's work The PhD thesis defended by Nancy Vanderveer in 1979 [135] was a milestone in environmental sound perception research, which raised some of the most important issues (and particularly on identification and classification) that are still studied nowadays. Her work was intended to be mostly exploratory (she considered that there was not any theory available that accounted for everyday sound perception), aiming at providing the research community with experimental data that would help forming assumptions on sound perception. She considered, along with the Gibsonian tradition, sound perception as a *pick-up* of information available within the environment. She classified information that may be picked up into three categories: information that allows orientation within an environment (less specific information), information that leads to orient oneself toward a specific event (active listening and/or visual scanning to localize a sound), and specific information about an event. In this latter category, she distinguishes *propriospecific* information (information about oneself and one's activities) and *exterospecific* information (about matters on the external worlds). Propriospecific information may be picked up by sounds generated by one's own activities, or imposed from external, independent sources[1].
She also described several experiments (which have to be however carefully interpreted, because of the lack of control of the experimental settings, as noted for instance by Ballas [9]). In an identification experiment, she let listeners listen to sounds recorded on a tape, and asked them to describe what they heard. Because of the very wide variety of responses she got, she had to define a heuristic aiming at defining what is a "correct" answer, by carefully studying the linguistic devices produced by listeners. Indeed, it is very difficult in any experiment where listeners describe what they hear, to define what is a "correct" answer. Conversely, studying the linguistic material provided by the listeners is often very informative. She also noticed that listeners very rarely spoke about the sound itself, but about the event that causes the sound. They described mostly 1) the action, 2) the object of the action 3) the place where the action took place. They described very rarely the agent that had made the action. The same kind of experiment was realized with children and revealed that they were also able to identify the source, at least roughly. With a recognition paradigm, she found that listeners were able to recognize sounds that they had previously heard with a good performance, even as late as three months later. The errors revealed that when listeners made errors (they pretended to have recognize a sound that they actually had not heard), it was because the sounds belonged to a common category. A set of classification experiments revealed that the sounds were grouped together, either because they had similar sound signals, or because they were produced by similar events.

Ballas' studies Following Vanderveer's work, the issues of identification of everyday sounds were explored thoroughly in a series of papers published by Ballas and Howard. The main idea of these authors is that the perception of environmental sounds shares some similarities with the perception of speech: since the sounds are perceived as events in the environment, they actually convey meanings, and thus may be thought as a language. Unlike Gaver however [49, 50] (see the sections 1.2.5 for a review of Gaver's work), they claimed that the information available in the acoustic signal is not sufficient to account for the identification of everyday sounds. Rather, identification of sounds results from both a bottom-up process (recovering of the information available in the sound and in the context) and a top-down process (using previous knowledge and expectations): "It is not only what we hear that tells us what we know; what we know tells us what we hear" [67][2]. In [67], they let listeners learn to categorize target sequences of brief sounds, in order to assess the influence of the syntactic (temporal organization of the sounds) and semantic factors on the performance. When sequences were organized following a set of grammatical rules (unknown by the listeners), they were more easily learned than when they were randomly organized, in the case of abstract sounds. When the sequences were

[1] It is worth mentioning that these distinctions help us to separate the studies that we have reviewed above. Indeed, most of these studies were focuses on pick-up of exterospecific information (with the exception of the vessel filling sounds [18]). Indeed, there is a category of sounds which are only rarely studied: those which guide action. They are however studied as *enaction*

[2] cited from R.A. Cole and J. Jakimik: Understanding speech: how words are heard, in G. Underwood (Ed.): Strategies of information processing, New York, Academic Press, 1978, p. 113

made of everyday sounds, sequences with a grammatical structure were more easily learned when the rules were interpretable (a usual water-pipe scenario), than when the rules were arbitrary. This lead them to conclude that, in interpreting sound events, human listeners rely on their knowledge (implicit knowledge of what is the syntactic sequence), as well as on the perceptual information available in the sound itself.

Based on a summary of some of their previous works, they further developed in [9] the parallels between speech and environmental sound perception. They reported for instance similar patterns of semantic interpretations of isolated speech and environmental sounds (some prompted stereotypical responses, some not). They reported as well homonym-type sounds (perceptually discriminable, but leading to confusing identification). In this case, the context helped listeners to choose among the alternative causes of the sounds. These homonymous sounds were further investigated in 1991 by Ballas and Mullins [10]. They presented to the listeners quasi-homonymous sounds, and the listeners had to assign the correct label. The sounds were embedded in sequences which were either coherent with the actual cause of the sound, or with the alternative cause. Incoherent sequences biased the subjects' responses toward the alternative cause of the sounds. However, coherent sequences did not lead to better performance of identification than in a neutral case. They argued that this latter result indicated that the parallels between speech and sound perception must be considered carefully.

Ballas studied extensively the influence acoustical, ecological and cognitive factors on identification of everyday sounds in a series of experiments reported in [8]. He showed that the performance of identification were related to different variables, including acoustic variables, ecological frequency (the frequency with which a listener encounters a specific sound event in his everyday life), causal uncertainty (measured as the amount of reported alternative causes for a sound) and sound typicality. Actually, acoustic variables could account only for about half of the variance in identification time and accuracy. Rating data (listeners had to judge the sounds along forty one cognitive scales) suggested that sound identifiability is related to the ease with which a mental picture is formed of the sound, context independence (when the sound can be identified easily without context), the familiarity with the sound, the similarity of the sound to a mental stereotype (the author indeed showed sounds that were more typical than others), the ease using words to describe the sound, and the clarity of the sound.

Perceptual and semantic encoding Categorization of sounds has a lot to do with how sounds are memorized. There is for instance a distinction between perceptual encoding of the auditory events and semantic encoding. Working with a priming paradigm, Chiu and Schacter [22] showed that identification (i.e. subjects have to describe what they hear and identify the source) of a sound is favored by perceptual encoding while recognition (the subjects have to determine if the sound was previously heard) is favored by semantic encoding. This distinction is coherent with Eustache et al.'s results [39]. Studying the performances of subjects with localized brain damages, they found a subject able to perform perceptual auditory tasks (compare, discriminate), but not to name the sounds. Another subject with others brain damages performed conversely.

1.1.4 Synthesis of the reviews

We can formulate several conclusions and assumptions from the results reviewed in the above paragraphs.

Two different listenings Experimental studies of soundscape perception (see section 1.1.1 and the special issue of Acta Acustica on soundscapes perception [119]) have revealed two different kinds of responses when listeners have to describe a sound sequence. On the one hand, when they recognize the sound events that form the sequence, they describe the events that makes the sound and its properties. On the other hand, when they are not able to identify the events, they describe the properties of the abstract (i.e. abstracted from any physical event) acoustic signal. Other authors have emphasized such a distinction between two different "listenings". Gaver [49, 50] defines *musical listening*, when listeners focus on the acoustic signal properties, and *everyday listening* when listeners focus on the event that makes the sound. The same distinction is also noticed by Hajda et al. [62] in the case of the perception of the timbre of musical instruments: they distinguish the *"source" mode of timbre perception*, when the listeners perceive the physical actions and mechanisms that makes the sound, and the *"interpretative" mode of timbre perception*, when the listeners interpret the low-level auditory attributes resulting from the time/frequency analysis provided by the peripheral auditory system. It should be emphasized that these are different *modes* of perception. Indeed, for a same sound, both these modes of perception may occur, and can be experimentally primed. Musical listening is for instance primed when an experimenter asks a subject to scale the dissemblances between the timbre of two sounds in a multidimensional scaling paradigm. Even if listeners are able to identify the two sounds of the pair (which can be investigated in an identification paradigm), they are also able to compare the sounds on the basis of low-level auditory attributes correlated to sound signal characteristics. Outside the laboratory, it seems that the modes of listening are triggered by several factors, including expectations and previous knowledge (top-down processing) and context (bottom-up processing).

Identifying sound events and their properties Therefore, and as demonstrated by the studies reviewed above, listeners spontaneously identify the event that makes the sound and recover its properties. However, this ability is far from being perfect. The properties are sometimes only roughly recovered: Lakatos et al. [77] found that listeners distinguish mainly between thin plates and thick blocks when they have to evaluate auditorilly the section of a struck bar. Grassi [55] showed that evaluated size of a ball dropped on a plate was not independent from the thickness of the plate. Giordano and McAdams found that listeners tend to associate the sound of struck plates of small dimensions with glass, and bigger plates with Plexiglas, independently of the actual material of the plate. Lutfi et al. [85, 83] showed that, in a task where listeners have to recognize the material, or rate the hollowness, of an impacted bar, some listeners are able to use the criteria that would allow them to perform the task optimally, while others based their judgements on an overweighing of the spectral information. Moreover, it seems that listeners tend to associate some sets of perceived properties with stereotypical responses. This is the case for instance when they have to judge gender of hand clappers [107] or of people walking on a stage [79]. This pleads in favor of the existence, parallel to a bottom-up processing, of a top-down processing: previous knowledge and expectations tend to select or emphasize specific information in the information available, and in order to associate them with the representation of expected events.

Influence of context, expectations and previous knowledge Ambiguous sounds (identical sound signals produced by different events) and quasi-ambiguous sounds (sounds which are discriminated by listeners, but which lead to confusing identification of the sound event) are a very interesting case that has been studied by Ballas et al. in [9, 10]. They showed for instance that embedding the sounds in a sound sequence coherent or not with the actual source of the sound modifies the identification performances. From these results, we can assume that listeners implicitly form expectations from the context (here: the previous sequence) which bias the identification toward one source or the other. More generally, Ballas' work (see [8]) clearly demonstrates the existence of a bottom-up processing of acoustic information combined with a top-down processing in the formation of auditory events. Information are incoded from the acoustic signal and from the context, but also selected on the basis of knowledge and expectations based on previous experience with the sounds (familiarity with the sounds, ecological frequency).

This latter point favors thus the view along which auditory events are formed from an iterative set of inferences based on perceived information picked up in the environment (the sound event and its context, including stimulation of other sensory modalities than auditory), and on knowledge, memory and expectations, rather than a direct perception of the events, as postulated by the original Gibsonian approach.

Acoustic cues to auditory perception It is therefore justified to seek what are the acoustic cues that a listener use to perceive an auditory event, for this auditory event is also formed partly from the integration of auditory cues. It must be however noticed that the results of such studies are not straightforward. It is of course always possible to find a combination of adequate acoustic properties that would correlate with the perceived properties of an event. The question is: how can these results be generalized to other sound events? Moreover, when researchers use the multidimensional scaling technique, they quite often correlate the perceptual dimension with a mixture of acoustic properties and mechanical properties of the event, because they did not succeed in finding any acoustic descriptor related with some of the perceptual dimensions (see McAdams et al. [90] for instance). One explanation may be that there is not relationship between the perceptual dimensions and the acoustic properties. But since the dissimilarity rating procedure is known to focus listeners' listening toward the properties of the acoustic signal, this explanation is doubtful. The other is that the authors have not looked for the right acoustic properties. Indeed, most of the time, acoustic descriptors are relatively simple descriptors of the traditional Fourier time/frequency representation. The approach of Cabe and Pittenger [18], or Warren and Verbrugge [138] offers an interesting alternative: rather than trying to describe the short-time spectrum with descriptors like spectral centroid, attack time, roughness, etc., they tried to define what are the invariants in the acoustic signal which characterizes the property under study (respectively time before overflowing or breaking/bouncing). This idea of invariant characterizing an action is also used by researchers studying *enaction* [137].

Auditory representation and categorization Finally, above these assumptions lies the notion of how knowledge is psychologically organized. Many of these results can indeed be explained if we assume that auditory representations are organized into categories: homonymous sounds may be thought as sounds belonging to a common category (discriminable, but sharing the same label), errors in identification may sometimes be explained by a response at an inappropriate level of categorization ("road noise" to describe a passing car), the evidence that some sounds are more typical of a category than others, etc. To explain more thoroughly this issue, we will focus in the following section on the theoretical concepts of categorization, and more specifically on the classification of everyday sounds.

1.2 Categories, categorization and models of classification

In the previous section were presented studies on the perception of everyday sounds. We introduce here different theories of category and categorization and why these concepts are related to our purpose of building a classification of everyday sounds. We will present example of classifications of everyday sounds in order to illustrate the relation between the review on perception of everyday sounds (section 1.1) and a more theoretical approach on categories and categorization.

Introduction

When we encounter objects, listen to sound events, or experience events, we need to form expectations in order to elaborate or anticipate action, or to produce thought [128]. But in our interaction with environment, experiences are never the same, all these variations make necessary the resort to our past experiences : i.e. the use of our memory, our knowledge.

We focus in this review on one aspect of our long term memory (L.T.M) : the semantic memory [131]. The semantic memory is our accumulated knowledge of basic meanings and facts, as general knowledge about our world, contrary to specific knowledge in episodic memory. The organization of our memory provides us with the opportunity to link information from past experiences to present experiences [38], and categorization is one of those major cognitive processes to achieve it.

For example, in our everyday sound perception, our semantic knowledge permits us to know if the *sound of a car* is known as *motor vehicle*, or as a *car* or as a *combustion sound*. We can recognize this sound from member of these different categories depending on the context of perception without necessarily identify it. When listening to a *sound event*, we experience it as an individual *auditory event*, but we remember or identify it as an instance of a categorie [38].

First we will introduce more deeply the concepts of *categories* and *categorization* (sec. 1.2.1) and link the different theoretical views of categories (sec. 1.2.2) and similarity in a general framework (sec. 1.2.3). Next we will present different approaches to relation between categories 1.2.4. In relation with the review of examples of classification of everyday sounds (sec. 1.2.5), we will then conclude by summarizing important points (sec. 1.2.6) with the aim of developing our classification framework of everyday sounds.

1.2.1 Categories and categorization

Categorization and *categories* are important concepts in cognitive science but controversial when people try to define the role and the structure of categories and the process of categorization [100]. Different theories explain sometimes specific set of experimental data, and can not always be transposable to other data. In this paragraph we will define the concepts of categorization and categories and enlighten the role of categorization and categories.

Introductory definitions

First we propose the definition of Sloutsky on categories and categorization :

> Categories are equivalence classes of different (i.e discriminable) entities and categorization is the ability to form such categories and treat discriminable entities as members of an equivalence class. [125] p. 246

This idea of *equivalence classes* is maybe too strong as the famous example of the concept of *game* from Wittgenstein concerning a fuzzy category regrouping very different entities as *Go*, *badmington*. For example, these two games do not shared a lot of properties in common, and it is hard to find a common definition for *game* : "activities offer pleasure", or "activity with rules" ?. This example point out a crucial question : i.e. what is *equivalence* ?, and is deeply related to how we structure categories, this point is developed in section 1.2.2.

Rosch, who made a seminal work on the study of *categories* (see [111] for a major introduction to her work), asserts a strong assumption on the formation of categories : our external world is structured, because of apparition of complexes of correlated attributes in the world [113], for example :

> Creatures with feathers are more likely also to have wings than creatures with fur, and objects with the visual appearance of chairs are more likely to have functional sit-on-ableness than objects with the appearance of cats. [113] p. 383.

In everyday live, we are confronted with these recurrent information set and are naturally sensible to it. For example some environmental sounds occur with a higher ecological frequency than others [8]. For instance bright, short and damped sounds are more associated with *metalic sound* than with *sound* [48]. From these correlated perceived structures emerge *natural categories*.

Nevertheless categorization has different meanings according to other authors. For Malt [88], categorization has two meanings, one related to the people's capacity to recognize objects having common properties with entities stored in memory. This cognitive faculty is more related to *recognition*. The other capacity relates to how people connect objects with words. This cognitive faculty is more related to *identification*. But Rosch does not separate these two meanings because categories are generally designated by names (e.g. *car*, *table*), but not always. Richard [108] states that we should not confuse the semantic categories with the words referring to them. Generally, lexicalized categories are those which capture stable properties across a lot of objects [108], so young children learn first nouns associated with these categories [113], as the noun *dog* for example. These categories are generally connected to a specific level of thought : the basic level (section 1.2.4).

Some authors make a distinction between *concept* and *category* [54]. *Concept* refers to the psychological state, as a cognitive representation of a category, for example when we think of the idea of a *car*, and *categories* refer to a set of entities in the real world, for example the external entities categorized as cars [54]. But generally category and concept are interchangeable [74] [38], as long as we do not focus on abstract concept or complex schemes. A more common view, is the distinction between *intension* and *extension*. Entities or objects recognized as X are the *extension* of the category and *intension* is related to the properties associated with this category or defining it. Intension may also be viewed as the mental representation of the concept/category. So a property is viewed as the basic constituent of a concept intension [100]. This distinction is useful when we will introduce the relation between categorization and inference in the next paragraph, because inference is a powerful cognitive process to provide new information.

In the next paragraph, we focus on the question : why are categories and categorization important?

Cognitive economy and informativeness

Funes, a hypothetical character fantizied by Jose Luis Borges[3], did memorize absolutely every situation, every entity, he experienced, but was therefore unable of any generalization, comparison, or inference of any kind. Conversely, when we interact with our environment, we can not memorize every sensitive experience: we must reduce information in a cognitively acceptable way. As Rosch wrote :

> ..., one purpose of categorization is to reduce the infinite differences among stimuli to behaviorally and cognitively usable proportions. It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand. [111] p29.

Nevertheless this cognitive economy should not mask an important cognitive process : relating experiences to each other in a productive way. We use categories and categorization to produce new information which is not necessarily directly accessible, in order for example to infer properties. We can distinct two types of inference related to categorization [106]. One, called classification or deduction, allows person to reason from general to specific situation. In this case when we categorize an entity, unobserved properties can be inferred on the basis of category membership. This entity categorized, have new inferred properties. The second process, called property induction, permit us to reason from specific to general, we project observed properties from an entity to the category to which it belongs and then to other category members.

Thus the process of categorization seems to be a central concept in cognitive psychology. For a large number of scholars, categories are the building blocks of thought. That is why we will introduce in the following paragraph different theories of categorization and expose important concepts related to categorization.

1.2.2 Theoretical views on structure of categories

General views on structure of categories

There are different approaches to explain how categories are structured. Mainly two distinct theories are presented in literature (see [74] and [54] for major introductions on the subject). These theories have distinctive characteristics in how an entity is associated to a specific category :

- *Rule based view* → People categorize an entity with respect to some necessary and sufficient properties, (or additional properties according to new development of this theory),
- *Similarity based view* → People compare an entity to all the instances of the category or with the prototype(s) representing the category.

Before giving a more complete account of each theory, it is important to introduce what are the main facts that these theories should account for. Theories evolve or oppose themselves through two major findings that theories should reconcile [108] [100] :

[3] Funes, the Memorious. By Jorge Luis Borges

- *Typicality of category members* → After Wittgenstein's observations, Rosch [112] [113] reported experimental observations of the typicality effect : category members have different status. People judge some members to be more typical than other members of the category. For example a *robin* is more typical of the category "bird" than a *canary.* This findings were observed in experiments like : production of members of categories, reaction times for category membership ...
- *Categories have conceptual core* → Stability of categories can not be explained only by perceptual similarity between members of category. The effect of context changes on which properties the similarity is based [132]. Murphy [97] explained that category are formed by people's theory on the world and not by similarity, that is why concepts have a stable core.

In next sections we present these theories across these two tendencies.

Rule based view

First of all, this approach was the first explanation of categories, the classical view, based on a logical statement, on a rule: the necessary and sufficient properties to be member of category. Rules can be defined in different forms as natural language rules, logical rules ... [124]. According to Sloman [124] "Rules take one or more mental representations as input, carry out a finite number of internal steps, and produce one or more representations as output". But in the classical view, members have equal status due to necessary and sufficient condition. Thus, this can not explain typicality. Following this limit, new extension of this theory [126] differentiate *defining features*, as essential and involved in the definition of membership of a category, and *characteristic features* as not necessary features related to the definition of the category but introducing typicality effect between members of a category. But this distinction is sometimes hard to explicit and that is why this model is problematic [40]. So one major problem with the rule based view is that membership is independent of the context because of the necessary and sufficient condition. Nevertheless one major advantage of rule based category are the well defined core of categories by definition, that is why a number of semantic network representing relation between categories in semantic memory are based on rules (section 1.2.4) and are often used for the development of classification model.

Similarity based views

In the resemblance based view, two different theories existed, the prototype theory and the exemplar theory.

The prototype theory Contrary to category formation based on rules, the coherence of a category is grounded on *family resemblance* between the members of a category, introduced by Wittgenstein [139]. In this case, members share common properties with close members, but all the properties are not distributed across all the members of this category. Categories are organized around a central tendency: the prototype [111]. The prototype is abstracted across members of the category and generally owns properties that are common to a majority of category members but owns different properties in contrast with other categories at the same level of abstraction. We find organization around a prototype in artificial categories, natural categories and artifacts categories [112] [94] [111] and for some abstract concepts [63]. Hampton has introduced [65] [64] an explicit version of prototype theory, with the resemblance cue calculated on the similarity between the properties of an entity to be categorized and the prototype. Typicality is associated with the weighting of these properties involved in similarity. A threshold of resemblance cue explain the membership to a category and the inter and intra-individual differences.

The exemplar theory The exemplar theory is a general approach. For example in prototype theory, only properties who which are not shared by the prototype, or which do not contribute to the family resemblance, are not included in the representation of a category. In exemplar theory, each member of the category can have its own properties. The context model [93] and its multidimensional extension

[98] postulates that all the members of a category are stored in memory. But when context changes, different properties are engaged in similarity between members of a category. Some theories assume an abstraction of the instances of a category [38] or the representation of the instances as points in a multidimensional space [3]. The first approach can converge with prototype theory if we consider that exemplar theory is a generalization of prototype theory with multiple prototypes. Typicality effect is explained by the similarity variation between members of a category and by the high frequencies or redundancy of specific members, contrary to less represented and so less typical members [74] [54], and the membership is based on probability measure [98] [93].

One interesting aspect of these theories based on similarity is their inherent flexibility. According to the context, similarity can change, and in this case, we form more variable categories. The core of the concept seems to be more difficult to explicit.

1.2.3 Categorization and similarity

Through the different theories we have presented previously, we propose in this paragraph a framework to explain relations between members in a category in terms of typicality and conceptual coherence. One major distinction between the theories presented in the previous section is the distinction between rule based categories and similarity based categories. Rule based categories are by definition well-defined categories. Similarity based categories are ill-defined categories. We propose and justify in this paragraph a single framework based on similarity.

We will not present here the different models of similarity (see Tversky [132] for the model of contrast based on traits, see Shepard [121] [122] and Krumhansl [75] for the multidimensional approach in geometric space).

Similarity as a single framework

Pothos [102] introduces a long discussion on the distinction between rules and similarity with commentaries from different authors. Generally rules and similarity are understood as separated [127] [124], but Pothos claims "that rules operations are simply a special case of similarity ones" [102] and even if there is a core distinction between rules and similarity, both can converge [61].

Pothos [102] argues that properties relevant for the membership of a category are not necessary perceptual properties, but may be as well even abstract properties if we consider abstract categories. He postulates that when an entity is categorized based on a small subset of the relevant entity properties, then categorization should be understood as a rule process. But when we categorize an entity using most of the relevant entity properties, with a similar weight, then categorization can be viewed as an overall similarity process. So there is a continuity of the similarity between two extremes, rules and overall similarity.

For example, Poitrenaud et al. [100] keep from the classical theory only the well-defined conceptual content as properties defining category, so the notion of rules is reduced to the definition of the type of properties, i.e. the distinction between traits and attributes.

Similarity in context

Assumption of conceptual core is not necessary accepted. Barsalou [11] shows that *ad hoc* categories can be formed in a specific goal (for example "things to bring for camping") and are for this reason dependent on the context of the task. Rosch [111] claims that categories do not have a conceptual core but members are related to each others through *family resemblance*. Poitrenaud et al. [100] explain differently this argument, people do not agree when they produce properties defining category, the apparent lack of conceptual core, because of pragmatic factors. People in this case will give more preferentially distinctive properties of the members of the category than shared properties between members of this category. They argue that people implicitly know that members have common properties and pragmatically do not use this information provided by the definition of the category. Nevertheless people agree when properties are provided to define a specific category. Faure [41] shows this result for sounds of musical instruments : when people describe them, she found a lot of variability. But when people use a set of properties describing a sound, people associated this set generally with the same sound. So category can have a set of properties defining a category but they are not necessary used during a specific task. Finally categorization is "how we use categories and properties in context" [100].

Dissociation between similarity and judgment of categorization has been sometimes observed [110, 109, 11] for particular cases. Nevertheless similarity plays a central role in categorization, (see Medin et al. [92] and [53, 125]).

1.2.4 Relation between categories

Introduction

In the first part of our review, we have focused on the nature and the structure of categories, based on rules or similarity. We focus here on the relation between categories, i.e. how categories are connected to each others, because it is an important point if we consider the creation of a classification of everyday sounds.

Hierarchical levels of abstraction

It is common idea that we think about entities at different level of generality. For example when we listen to a sound of "plane in the sky", we can categorize this sound as a sound of *a motor sound*, but at the same time the same sound can be heard as *an airplane*, and we can be more precise if we recognize a *supersonic airplane*. This idea has been studied in antrophological studies concerning zoological and biological categories [12] [4]. The main idea is that for example categories of plants used in everyday life by Mexican people correspond to a specific level of a scientific taxonomy (class inclusion in a hierarchical tree). This specific level is called *generic rank* [12] and is the best level to summarize categories in an informative way for an everyday use.

Privileged levels of abstraction Rosch in her different works [111] and with collaborators [113], showed experimentally that we use different privileged levels of abstraction when we are using categories. For example if we take the same previous example, three different levels are mainly exploited:

- *Superordinate level* → *a motor sound*,
- *Basic level* → *an airplane*,
- *Supraordinate level* → *a supersonic airplane*.

These different levels of abstraction are organized in a taxonomy. Rosch et al. have demonstrated that at the superordinate level, entities do not share many properties contrary to the specific level with many common properties. A the intermediate level, the basic level, members of a category have a lot of common properties but at the same time categories at this level are well delineated. For *natural categories* [111], at the basic level, categories seem to reflect the correlated structure of our environment. Basic level is well balanced between cognitive economy and informativeness. This level have specific cognitive properties as :

- we use the same motor control to manipulate members of a category
- members of a category have a similar form or many properties in common
- semantic label is associated spontaneously to an object at this level, first learned by children

It is important to say that basic level is modulated through expertise, a musician will use specific categories as basic level to describe musical instruments contrary to novice.

Hierarchical categories and inference An important aspect of possible hierarchical organization of our categories is its powerful informativeness, young children seem to use hierarchical relations to get semantic information [30]. If we consider a category (for example a *car*) and its super-ordinate category (more general like *transport*), we consider that entities from the category *car* inherit properties from the category *transport*. Based on this idea, Collins and Quillian [25] proposed a hierarchical semantic network representing relation of inclusion between categories. This model is related to the classical view of categories (section 1.2.2). An assumption of this model is the inference process (section 1.2.1) related to categorical inheritance of properties, a controversial point defended by authors [140] [30] or discussed [123]. In our conclusion we will discuss the problem of purely hierarchical organization of categories (section 1.2.6).

Semantic network

In this paragraph we present a restricted but representative view of contrasted models of an organization of semantic knowledge, a network of spreading activation of Collins and Lotfus [24] and an electronic lexical database WORDNET [42]. This view is not exhaustive but these models have been a starting point for new models and widely used and we will use it to propose a model of classification of everyday sounds.

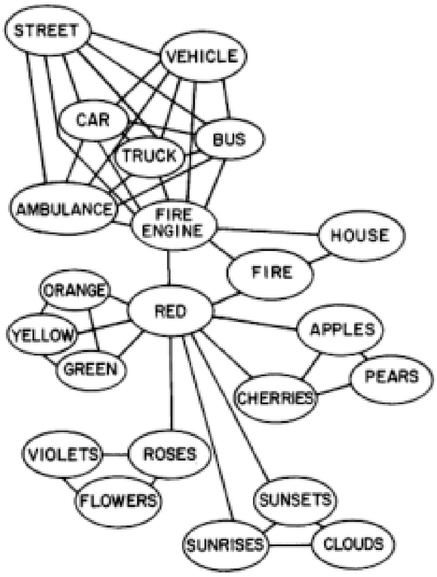


Figure 1.2: Semantic network of Collins and Lotfus. Taken from [24].

Spreading-activation network Collins and Lotfus have developed a theory after which the meaning is represented by hypothetic network. Semantic memory is organized in a network of concepts like a grid with many interconnections (see figure 1.2 for an example). When we search an information in memory, an activation is spread inside the network, each concept being represented by a node, as the concept of *red* in the figure 1.2. When the concept *red* is activated also all the concept associated with *red* are activated at the same time, as *apples* or *fire engine*. So if someone asks you if a *fire engine* is an *apple*, there is no direct intersection, the response will be no, contrary to the question "is a fire engine a vehicle". The networks that are often activated are reinforced and this reinforcement of specific connections explains typicality (section 1.2.2). Even if this theory is elegant and has been generalized to provide new accounts, one major problem of those theories, is the complexity of the network. This network is maybe too general to provide a powerful classification of everyday sounds.

The network WORDNET This approach results from studies in psycholinguist and psycholexicology, with the aim of determining lexical components of our language. Miller [95] develops a lexical database based on these works with approximatively 90000 forms of words and 70000 meaning of words. This lexicon is classified in five different lexical categories : nouns, adjectives, verbs, adverbs and functional

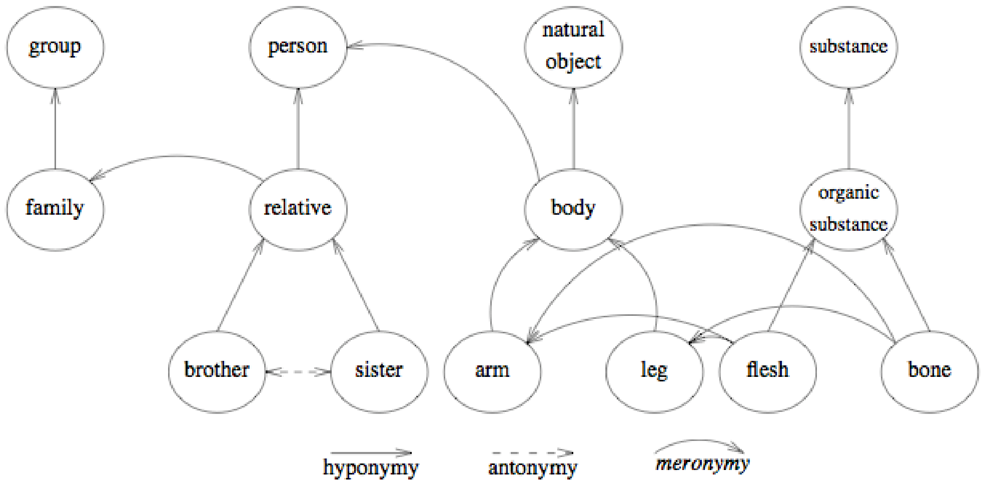


Figure 1.3: Semantic network and its lexical and semantic relations in WORDNET. Taken from [42] .

words. These categories were obtained from results on the organization of semantic memory. When we ask people to associate a word with an other, people systematically give a word from the same category. Each category has different semantic organization :

- *Nouns* → Hierarchical thematic organization
- *Verbs* → Themes with implication relations
- *Adjectives* → Multidimensional organization in a space

One crucial point is the conventional relation between a concept and its representation, i.e. in this case the relation between the label (form of the word) and the meaning (concept expressed by a word). Wordnet [42] uses a special way to represent concepts. Concepts are represented by a definition, a meaning of a word M_1 is represented by a list of word forms $F_1, F_2, F_3...$, this list is called *synset* a set of synonyms. So a synset represents a concept. Semantic relations and lexical relations extend the relation of synonymy in a complex relational and hierarchical network. We summarize here the principal relations used in WORDNET :

- *Lexical relations* → Lexical relations are relations between word forms.
  - *Synonym* → In a sentence two synonym words can indifferently be interchangeable without changing the meaning of this sentence.
  - *Antonym* → A pair of words between which there is an associative link built up by co-occurrences. For example wet and dry are antonyms.
- *Semantic relations* → are relations between *synset*, i.e. between concepts.
  - *Hypernym, hyponym* → X is an hypernym of Y, if Y is a sort of X in terms of meaning, for example X : *plant* and Y : *tree.*
  - *Meronym* → is the relation that holds between a part and the whole.

All these relations are the backbone of the network. If we focus on the organization of nouns, concepts represented by synset are structured through a hierarchical network with the hyponym relation from general to specific synset. Moreover a specific category inherits properties from general categories as the classical view of categories (section 1.2.2). The network starts with 25 semantic themes :

> {act, action, activity}, {animal, fauna}, {artifact}, {attribute, property}, {body, corpus}, {cognition, knowledge}, {communication}, {event, happening},{feeling, emotion}, {food}, {group, collection}, {location, place}, {motive}, {natural object}, {natural phenomena}, {person, human being}, {plant, flora}, {possession}, {process}, {quantity, amount}, {relation}, {shape}, {state, condition}, {substance}, {time}

Each semantic theme is organized with its own hierarchical network and has a maximal level of inclusion 10. This network is close to the Collins and Quillian's semantic network, but a major improvement of

Wordnet is the use of *meronym* relation that permits connexions between concepts a the same level or or different levels of inclusion. The figure 1.3 illustrates the different relations : for example *body* is a sort of *natural object* and *arm* is part of *body* and the word sister and brother are antonyms.

WORDNET is an interesting example of network because of its rich relations, in spite of its hierarchical form, and because of its implementation in a lot of different computer languages, that is why it is used in many classifications `http://wordnet.princeton.edu/links`. We present in the next section a use of WORDNET in a classification of environmental sounds and discuss it.

1.2.5 Classification of everyday sounds

After presenting theoretical approaches of categories and categorization, we introduce here a set of examples of classification of everyday sounds. These classifications are empirical (Gaver [50]), experimental (Guyot [59]), and related to database with examples of sound databases.

Taxonomy of interactions sound proposed by Gaver

Original taxonomy proposed Gaver According to the ecological approach of Gaver [48] to everyday listening, sound provides information about an interaction of materials at a location in an environment, and therefore what we hear is suggested by physics. In order to understand what people hear in the world, the author has proposed a hierarchical structure of sound-producing events, as shown in Fig. 1.4.

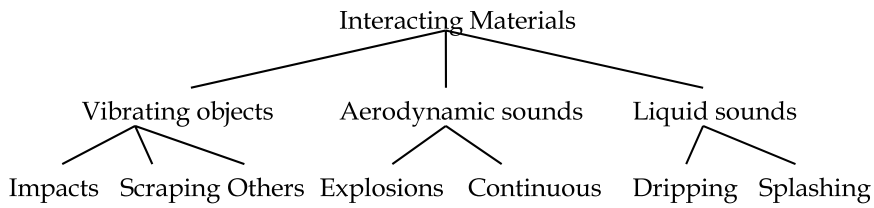


Figure 1.4: Gaver's proposal of a hierarchical description for simple sound events. Taken from [50].

First a sound event occurs because of the interaction between two materials. The pressure waves are indeed produced by the resulting vibrations of the objects and depend on the force, duration, and changes over time of their interaction. The resulting sound depends as well on the size, shape, material and textures of the objects themselves. Therefore a second level in the hierarchical structure of sound-producing events describes three categories related to the nature of the interacting materials (solid, liquid or gas), which are unlikely to be confused by people. The first category consists of vibrating solids which are a very common source of sounds, such as footsteps and closing a door. The second one describes aerodynamic events. The resulting sound may be created by solid objects due to its interaction with the atmosphere (like explosion), or when changes in pressure themselves transmit energy to objects, causing them to vibrate (like wind passing through a wire). Finally the third category of sound-producing events involves liquids, in that case sound results from the formation and change of resonant cavities in the surface of the liquid (e.g. splashing). The next level consists of *basic level* sound-producing events within each former category. These are simple interactions, such as impacts for solids, and explosions for gas. From these events, more complex ones may be derived. Figure 1.5 shows a more complete classification of sound-producing events, where complexity grows towards the centre.

For solids, the author proposes four different basic events which are deformation, impacts, scrapping and rolling. For each basic event, a list of physical attributes that may also be perceptually relevant is provided. For the example of impact, it means that the vibrating object's material(s), size and configuration, the surface hardness and the force of the impact may influence what people hear.

Following these events in the hierarchy, the author distinguishes three types of more complex events. The first one, *temporal patterning*, involves events whose patterns are made of simpler low-level events. For instance, breaking involves a patterned impact and crumpling a patterned deformation. The second complex level is produced by more than one type of basic level event and is called *compound* events. One example is bowling which involves rolling followed by impact. Finally a third degree of complexity describes hybrid events which involve more than one type of material, such as water dripping on a solid surface.

According to Gaver, this framework is far from being exhaustive and may be organized differently, but it already describes a large range of sound events and provides a way to understand a multitude of complex sound-producing events and their physical attributes that may be relevant for the human perception of everyday sounds.

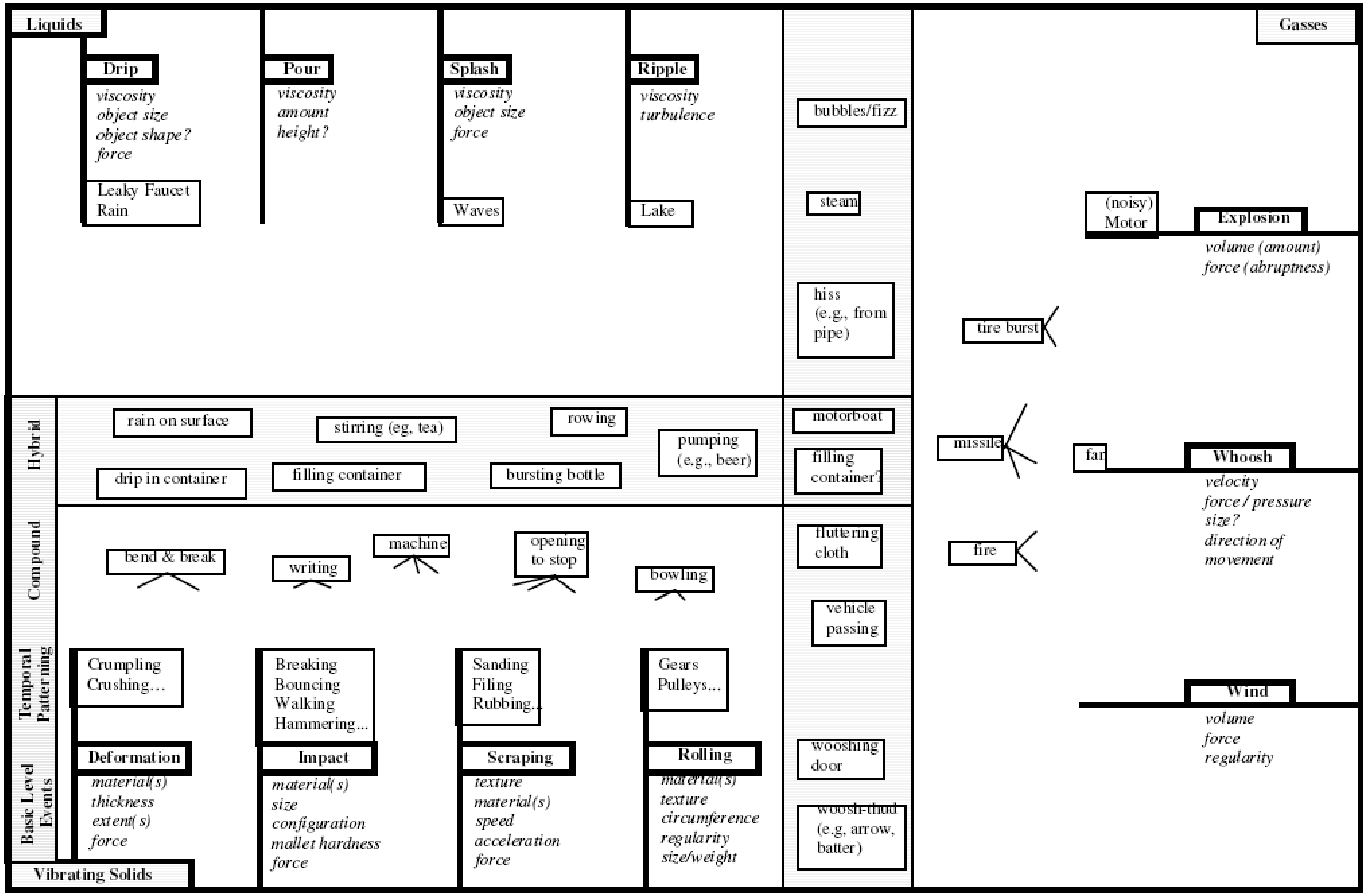


Figure 1.5: Gaver's hierarchical description of everyday sound events. Three fundamental sources (vibrating solids, liquids and aerodynamics) are shown in the three overlapping sections of the figure. Within each section, basic soundproducing events are shown in bold, and their relevant attributes next to them in italics. Complexity grows towards the centre of the figure, with examples showing temporally patterned, compound, and hybrid sounds. Taken from [50].

Extension of the classification proposed by Univerona Univerona propose here a taxonomy (without being exhaustive) for *everyday sounds* with related sound models based on the taxonomy of Gaver [50] presented previously. This taxonomy have been developed by Univerona with the aim of proposing a framework to their sound synthesis algorithms (see section 3 for an exploratory guide). The taxonomy graph (see Fig. 1.6) develops bottom-up starting with low-level sound models. The second level contains *basic events* and sound *textures* founded on (i.e. straightly derivable from) the low-level models. At the third level are *processes* based (e.g. implying temporal patterns) on one or more basic events and textures. Lastly, the highest level contains several examples of implemented simulations. These make use of results from lower levels, and further involve resonating objects models (*resonators*). The modeled resonators can be of different shapes (e.g. a bar, a wheel, a string, a tube, etc.) and materials (e.g. metal, wood, glass, etc.).

The catalog of physics-based models based on this taxonomy and presented in section 3 is widely based on Gaver's study, however some differences exist between the two taxonomies used to describe sound-producing events.

Conversely to Gaver's taxonomy (section 1.2.5) whose attempt is to decompose a specific sound-producing event into simpler ones to understand its underlying perceptual attributes, the methodology of the CLOSED project consists of creating physical-based models more and more sophisticated in order to model any sound-producing event. Therefore the hierarchy is established from low levels sound events to more complex ones. This explains the differences between Gaver taxonomy and ours (which may be seen in Fig. 1.6). Differences refer to the naming of sound events, their position in the hierarchy, and also some events present in one taxonomy may not be present in the other because none of them is exhaustive.

The first difference lies in the first level for each class of material. While Gaver directly starts with simple sound events, we refer to *low level models* used to create basic sound events. Therefore our first level holds fewer models than the first level of Gaver's taxonomy. For solids, these models are fracture, impact and friction, which would correspond to impact and scraping in Gaver's taxonomy. Deformation, which is used in his hierarchy to create the sound events of crumpling and crushing also refers to our model of impact (together with the fracture model for the case of crushing). Besides in our taxonomy rolling lies on the second level of the hierarchy because this event is derived from the impact model.

Our next level in the hierarchy refers to *basic events and textures* that correspond to the basic level events of Gaver's taxonomy and some of its temporal patterning events, which for us relate to textures (e.g. crumpling). Unlike Gaver, we indeed make the distinction between textures, i.e. sound events that have a periodic temporal pattern and are reversible, from other temporal patterning events which temporal structure is unperiodic and irreversible. The latter events lie within the *derived processes* level and include what Gaver calls *compound* events, i.e. those which are derived from more than one basic event.

Finally, our taxonomy describes on the higher level, called *simulation examples*, real sound-producing events that may be simulated from the aforementioned types of interaction between objects, while Gaver proposes hybrid events derived from diverse classes of material. In our taxonomy, this type of events appears at the basic event level. For example, dripping, which is primarily derived from the bubble model of liquids, also includes the impact model.

Referring to liquids, our low-level models of bubble and fluid flow enable us to create the basic events of burbling, dripping and flowing, and more complex events such as splashing, filling and pouring, whereas Gaver considers splash and pour as basic level events.

About sound events produced by gasses, we propose the models of turbulence and explosion to create basic events such as whooshing, burning, exploding and popping, whereas Gaver's proposition of explosion, whoosh and wind as basic events are deduced from the three specific complex events of tire burst, missile and fire.

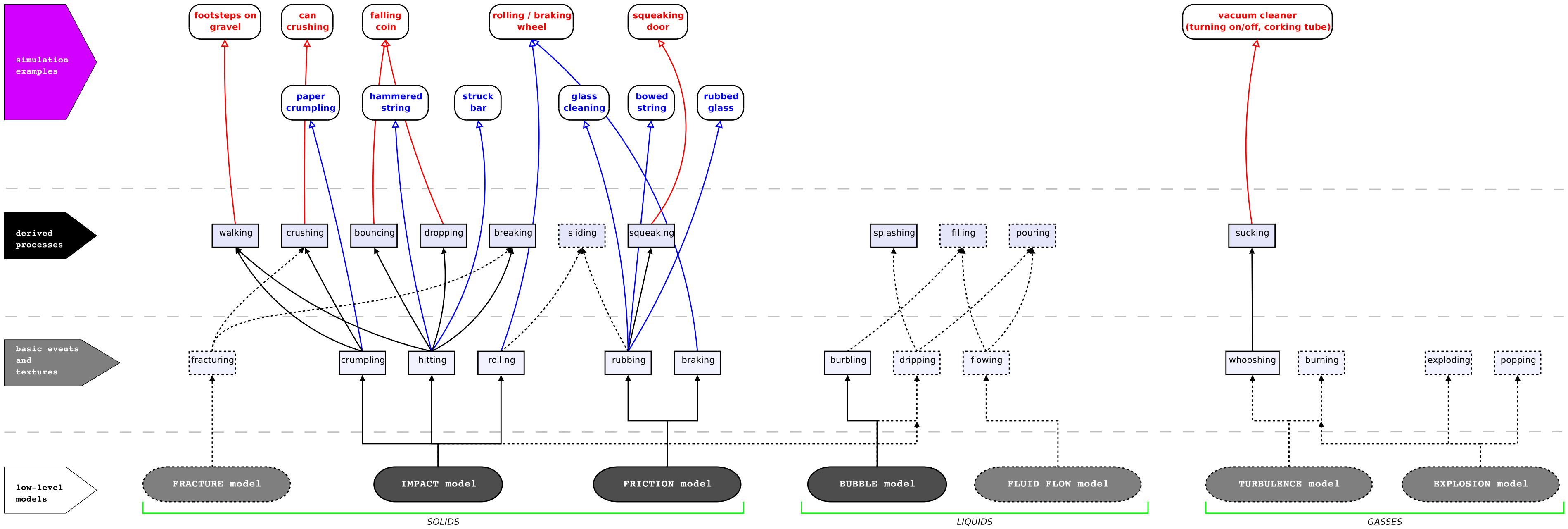


Figure 1.6: Overview of the proposed taxonomy for *everyday sounds* with parent sound models. Elements outlined by dashed lines are here proposed and not yet implemented in sound synthesis algorithms; dashed connections represent expected dependencies. Proposed by Univerona.

Classification related to sound databases

We focus here only on databases of environmental sounds. These examples are particular because the goal of these databases is to find and to classify sounds. We focus here only on database of environmental sounds.

Engine search Some internet web sites as `http://www.findsounds.com` or `http://www.soundfisher.com/` propose a search engine to find sounds. For example "SoundFisher" develops some tools in order to search sounds on the basis of similarity (acoustic descriptors calculated on signal) and on some basic processes to classifiy sounds with a semantic description of sound (as descriptions given by the sound database Sound Ideas `http://www.sound-ideas.com/`). But this classification is totally hierarchical and just based on the description of the sounds.

Freesound Freesound `http://freesound.iua.upf.edu/` is an interesting project using the semantic network WORDNET. They use the different relations of this network to permit fuzzy queries according to synonyms and to relate different categories of sounds with their semantic descriptions [19]. For example *piano* has two meanings, the musical attribute or the musical instrument, and a piano has a keyboard, a pedal ... Wordnet is used to label new sounds, when a new sound is imported, sound is compared with close sounds and depending on similarity to cluster of these sounds, their description is associated. So the new sound belong now to the taxonomy. This is an automatic classification not necessarily related directly to the perception.

Classification based on perceptual studies

We present here a summary of the work of Guyot [59] on the classification of domestic sounds, or everyday sounds. Her work is in the theoretical framework of prototype theory (see section 1.2.2). For the categorization of domestic sounds, she used a set of recorded sounds in a natural condition. The set of sounds was chosen to be in an ecological condition, i.e. a realistic sound environment in everyday life without presupposed of particular expertise.

An important aspect of this study is to try to find the basic level of categorization in order to understand relation between categories of domestic sounds. She asked to people to listen to 25 five sounds and classify sounds in different classes according to their perceptual similarity. After this classification people explicited each category with verbal production.

An interesting result reveals that people used two different strategies to classify sounds. The first strategy was based on psychoacoustical criteria like pitch, temporal evolution and the other one on type of excitation of the source (mechanical, electronic, ...). Her explication is a categorization with logical criteria, because an object belongs only to one category. The second strategy is based on the representation of how the sound has been produced (identification of the source or movement). The statistical analysis of the classification data is done with an additive tree [117] having the property to reflect gradient of typicality (section 1.2.2).

For the second strategy, the results showed two modalities of categorization, one modality is related to similar sources and similar functions and the other related to sounds generating by movements or same gestural (friction, ...).

The categorization of objects based on function has been demonstrated for categorization of artifacts, most authors agree that people's categorization of artifacts is influenced by two forms of similarity: functional similarity, and low physical similarity [29]. Results are similar for child.

Guyot explains these results with two different cognitive processes, for the first modality, it is a level of identification of sources contrary to the second modality more on abstract level movement. Guyot proposed an hypothetical classification of everyday sounds with three level of abstraction as Rosch (section 1.2.4). The basic level corresponds to the identification of the production of sound, at the supraordinate, people identify sources, and a the superordinate level people identify abstract production like mechanical sounds, electronic sounds ...

1.2.6 Synthesis for a classification of everyday sounds

The review of perceptual studies (section 1.1) has demonstrated that everyday sound perception is based on inferences from context and knowledge, and, particularly, on a categorical organization of knowledge. To explore thoroughly this issue we have thus exposed in this section the major findings on theories of categories and categorization.

In section 1.2.1, we have provided a basic overview of categories and categorization in the general framework of cognition. In order to build a classification we must understand how people organize their knowledge with categories. Thus, we have focused in section 1.2.2 on how categories can be structured. With well defined categories based on basic rules, we argue that it is difficult to account for typicality between members of a category. The better explanation accounting for typicality is provided by categories constructed on similarity. In section 1.2.3, we have shown that similarity seems to be a potential framework to catch these different views. We furthermore assume that it is crucial to know on which potential properties similarity is based. For example the review of perception of everyday sounds presented in section 1.1 gives some insights on this potential properties that people use in auditory perception. Indeed, when people identify the source of the sound, they spontaneously describe the event that makes the sound and its properties. When they are not able to identify the event, they describe the properties of the meaningless acoustic signal. Moreover, these results depend as well on the context of listening. Thus, to build a model for the classification of everyday sounds, we propose to structure the classification with a well defined core as a general framework, based on similarity build from properties as well as selective weightings allowing to emphasize or suppress certain properties depending on the context. The type of properties is not limited, and includes perceptual similarity (like timbre, pitch, ...) or higher-level similarity (semantic properties, properties of sources, type of interaction, emotion, etc.).

Another challenge is to understand how categories are connected together, and more specifically, how listeners organize categories of sounds. We have reviewed in section 1.2.4 different theoretical internal organization of categorization. Finally, in section 1.2.5, we have presented some examples of categorization and classification of everyday sounds. As we have reported it from Guyot [59], people can categorize sounds differently depending on the focus of properties. This result, known as *cross classification*, may also be found in other modality, for example for food [114]. To account for the results from studies of everyday sound perception, the structure of the organization of categories of sounds do not have to be restricted to a hierarchical structure. For example, Wordnet/Freesound, and the taxonomy proposed by Univerona, introduced relations within a same level, and between different level of abstraction. This view is a compromise between a strict hierarchy (not very flexible) and a semantic network, as Collins and Loftus (too much flexible to be implemented in a computational model).

These results and reflexions are intended to be the first guidelines to build a theoretical a classification of everyday sounds. The second part of the deliverable will consist of experimentally testing these assumptions and propositions, in order to provide NIPG with data aiming at implementing a computational model of this everyday sound classification.

2 Everyday sound interaction and design

Introduction

The studies that we have reviewed above have provided us with experimental and theoretical results concerning the perception of environmental or everyday sounds. Since many of these everyday sounds are produced by everyday manufactured products, these results have naturally found applications in the field of sound quality and sound design, which are the core of the CLOSED project. The idea of sound quality has emerged relatively recently. It refers to the fact that the sounds produced by an object or a product are not, or not only, annoying or unpleasant, but are also required for people to interact with an object.

The following paragraphs review the ideas founding the domains of sound quality and sound design. In section 2.1, we will focus on the influences of context and culture on the perception of product quality. Then, in section 2.2, we will more precisely detail basic interaction design and its theoretical background.

2.1 On sound reception, design for context, and inter-cultural differences

One recurring issue of import for the stated aims of the project, especially for those of developing tools linked to human reception of everyday sound, concerns the treatment of inter-cultural differences between potential users. Intuitively, cultural differences may play a factor in determining the way that the sound associated to a product is received, so it is desirable to take this into account from the outset.

One way we might address the implications of inter-cultural variability in CLOSED is to consider it as a particular instance of a context-dependent effect. Context may be thought of as the set of actual factors that can influence the user's assessment of the design, including her environment, mental state, task, cultural preconditions, and so forth.

A related perspective on this question is provided by research in product sound quality. The definition of product sound quality suggested by Blauert and Jekosch [15] has been widely cited:

> Product-sound quality is a descriptor of the adequacy of the sound attached to the product. It results from judgments upon the totality of auditory characteristics of the said sound, the judgments being performed with reference to the set of those desired features of the product which are apparent to the users in their actual cognitive, actional and emotional situation. (Blauert, Jekosch, 1997, [15]).

The user's situation, including her annoyance level, or the perceived noisiness of a sound, depends heavily on contextual features such as the contents of the ambient soundscape, the user's task, operating expectations, cultural preconditions, and so forth. Indeed, it is possible to construct sound examples that differ dramatically in subjective measures such as annoyance level but which register equally on traditional ear-based acoustic measures [142], due, for example, to semiotic or other context-dependent effects (the sound of the mosquito is a classic example). Blauert [14] has argued already in 1986 that an accurate assessment of sound quality demands a close connection to and examination of prevailing factors within the real operating environment of the subject. One should, in general, also consider non-acoustic aspects of the interactive appearance of a product, including its tactile and visual appearance, and mechanical or informational function, as intimately involved in the user's quality judgment.

Culture might loosely be taken to mean the set of cognitive, physical, social, and environmental preconditions that may be associated to a particular collection of people. Research in sound perception has underlined the importance of cultural factors : timbre of musical instruments, although seen as an integration of auditory attributes, may also be thought as a cultural construction [21]. Soundscape are another example of cultural constructions [35] (see section 1.1.1). Thus, quality assessment can be seen as well to be an intrinsically culture-dependent process. Due to the complexity which this implies, a general purpose tool for assessing product sound quality that will function in all cultures is a practical impossibility. It is common in sound quality assessment to employ a user group or jury drawn from the target audience to lend meaning to hearing-based psychoacoustic measures [86], but field investigations in which other contextual factors, such as task, interaction, environment, etc., are accounted for are less common [16]. This is the case despite many examples from perceptual quality research that have revealed cases in which context dominates laboratory measures; for example, in the work on mobile video quality assessment of Sasse et al [115].

More direct consideration has been given to cross-cultural effects by researchers in the social sciences. Within the field of anthropology, sensual qualities are, as a rule, always considered relative to a cultural setting, place, and people. Sensory cues have often been seen as distinguishing markers of cultural territories, as for example in Alain Corbin's study of bells as acoustic markers of place in 19th century France [26]. It is the norm that anthropological knowledge of sensory events be articulated by case studies that situate those events within their context. David Howes [68] has cited the work of Classen [23] in concluding that sensory perception may be seen as a cultural as well as a physical act, channeling cultural values.

> When we examine the meanings associated with various sensory faculties and sensations in different cultures we find a cornucopia of potent sensory symbolism. Sight may be linked to reason or to witchcraft, taste may be used as a metaphor for aesthetic discrimination or for sexual experience, an odour may signify sanctity or sin, political power or social exclusion. Together, these sensory meanings and values form the sensory model espoused by a society, according to which the members of that society 'make sense' of the world, or translate sensory perceptions and concepts into a particular 'worldview.' There will likely be challenges to this model from within the society, persons and groups who differ on certain sensory values, yet this model will provide the basic perceptual paradigm to be followed or resisted. (Howes, 2005, [68])

A greater challenge to sensory anthropology lies in accounting for cross-cultural effects, which are capable of impacting various situations in which a cultural artifact is presented outside of its original context; Museum exhibit design is a familiar example [36]. A number of studies of sonically-based inter-cultural collisions and re-interpretations can also be cited [37].

There has existed for some time a connection between anthropology and diverse design disciplines, based on the common aim of understanding the behavior and situation of a group of interest. Context-based methods from anthropology, such as ethnographic inquiry, have played an important and remarkably diverse role in design and in human-computer interaction design research [31]. As Dourish notes, over the last centuries, ethnography in anthropology has grown from its roots (partly) in colonial administration, of "objective, instrumental, and actionable" accounts, toward situated encounters [31]. He goes on to remark that the parallel process in design has meant a migration from a narrow vision of design ethnography, as the description of practices and habits of potential user groups or product consumers, to a broader framing of situated encounters and partnerships blending both sides of the designer-producer exchange. This migration has been particularly evidenced in the Participatory Design movement, an approach partly adopted, in abstracted form, in CLOSED. Dourish directs sharp criticism at the human-computer interaction design research community for clinging to a narrower, and now dated, interpretation of the role for and manifestation of ethnographic inquiry in design.

Broadly stated, good design is that which is appropriate to the user group it addresses, and to the various contextual factors that are significant to that group. Some would argue that certain markers of product quality may be identified as universal, but this idea is problematic. As Blauert has noted, even so comfortable an example as the sonic signature of the build quality of a luxury automobile's door closing mechanism can be perceived as undesireable, when heard out of context [14]. In light of what may be assumed to be generically strong inter-group differences, including inter-cultural differences, it is important to clarify that the goodness of a design must, as noted above, be judged relative to the user group it services and the role it plays in their context. Some have seen this idea as closely associated with the tendency in product design toward increasing variation and specialization within product categories. A reasonable expectation is that a tool set for assisting design quality assessment should be capable of a similar type and degree of specialization if it is to be of any use.

Of course, many successful products find uses outside the context for which they are designed, some even attracting global markets. The fact that certain globally marketed products manage to elicit emotional associations that entirely transcend context, on an even global scale may be explained, in part, by the observation that the design of these products has led them to local specialization by means of heavily contextually-targeted marketing campaigns. The construction of localized models for the relation of a group to a given product, and of scenarios that illuminate these relations, form an intimate part of the design process. At the same time, the effects of globalization over the past decades have been such that certain products – for example, luxury brands – have paradoxically found it possible to establish apparently seamless international identities, forging consistent associations across very different cultures. While this has been seen by some as reflecting a general erosion of cultural differences, new cultural and social groupings continue to be generated, mixing old and new, and facilitated in part by the same modes of exchange and electronic communication that have been instrumental in globalization. Thus, the need for contextually sensitive design and design tools is likely to persist for the forseeable future.

To predict the reception that an extra-contextual or extra-cultural use of design may meet, some degree of context-based investigation to address or validate the response in complementary user groups or subgroups will be invaluable. This kind of investigation may be integrated within the CLOSED project by examining subgroup differences in both experimental psychology and participatory design components, and by carrying out small-scale comparative investigations with groups outside the target context. Finally, alternative design scenarios may be built to explore how a product may accomodate the reception of a group of users outsize of the target context; such a reframing of the design does not necessarily require a material accomodation of the added group, but may often rely on the suggestion of an appropriate model for its role in the shifted context and in the minds of its users. It will be valuable to consider such alternate scenarios in CLOSED, if only to shape discussion around the question of context displacement. This stated, based on the remarks above, the most appropriate perspective on CLOSED as design research may be to consider its outcome as a new method for developing certain computational tools for sound design, and for interpreting them, rather than as a universal tool which might apply to all contexts. As noted above, the latter aim is probably impossible.

2.2 Basic Sonic Interaction Design

A key activity in the project is to provide new tools for sound design, in the context of new interactions with sonically augmented or otherwise sonically-designed products. HGKZ is carrying out design research in CLOSED with the aim of developing basic knowledge applicable to product interactions incorporating sound. One intended outcome is to arrive at a repertoire of primitive product interactions that can serve as building blocks for composition with interactive sound, in a way that is organized by the categorization activities of D4.1, and which can exploit the sound synthesis modules that are being developed in WP2.

Basic Design is a method combining educational practice with the theoretical and methodological foundation of design [2]; Figure 2.1 provides a diagram of the Basic Course, through which Basic Design was introduced in the Bauhaus curriculum. Recently, various authors have stressed the importance of Basic Design as central to the discipline of design [2] [43] . Although its focus has been on visual and formal aspects of artefacts, its methods, such as reduction, translation and morphological analysis,

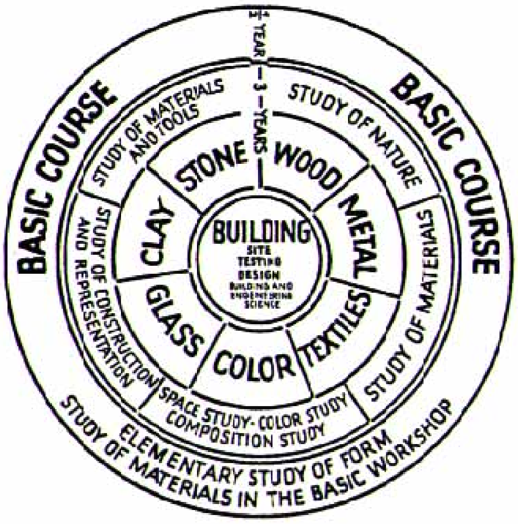


Figure 2.1: The Bauhaus Basic Course was conceived by several significant artists of the last century: J. Itten, W. Kandinsky, L. Moholy-Nagy and P. Klee.

make it particularly suitable for design research in CLOSED. The latter is focused on understanding the relations between sonic, formal and interaction proprieties of sounding artefacts.

Basic Design originates in the kindergarden movement and with the Bauhaus school, and is based on the analysis of visual experience in terms of simple, abstract properties, such as forms, patterns, or colors. Reduction from real world experiences (see Fig 2.5) is used to determine abstract elements. We will use this method of reduction to define sonic and interactive object properties, abstracting them from everyday experiences, and the ways in which sound and gesture are coupled in them. The Bauhaus' notion of an abstract element was always linked to its dynamic properties: "line is a the track made by the moving point: that is, its product. It is created by movement..." (Kandinsky, [71]). Similarly basic sound interaction design refers to couplings of sonic and interactive, or movement-based, affordances of an object, rather than upon an understanding of these as separate elements. Sound can often be described in terms of physical interaction: the sound of pouring water, the sound of walking, of cutting, typing, and so forth. In the design process, it also proves useful to analyze and decompose these interactions in order to use them to explore and generate new ideas, as discussed further below.

The Bauhaus explored design abstraction in relation to human perception, with the aim of uncovering a universal visual language, independent from such cultural limitations as are present in alphabetical writing. In these experiments, researchers were not interested in individual preferences, but in intuitive, biological responses and in the most frequently occurring perceptual relations between abstract properties: graphics, color, texture and so on. In Kandinsky's "psychological test" in 1923, he asked participants to fill in basic shapes with the basic colors, in order to identify a perceptual link between the two (Fig 2.2). Kandinsky called this correspondence between different modalities "translation", which he hoped to be a method for unifying all perceptual experiences through visual language [81].

Later, the Bauhaus disassociated its aims from this ideal, and employed the manipulation of basic elements primarily for generating new design ideas (Fig 2.3). The focus was not on the vision only, but included tangible qualities of objects. Tactile charts and structures (Fig 2.4) explored sensations of pressure, temperature, vibration bringing relational complexity into abstract design:

> If the same methodology was used generally in all fields we would have the key to our age - *seeing everything in relationship.* (Moholy-Nagy, [96])

In the same spirit as this more exploratory approach suggests, CLOSED researchers will use the strategy of analyzing, transforming and re-arranging formal, sonic and interactive properties of an

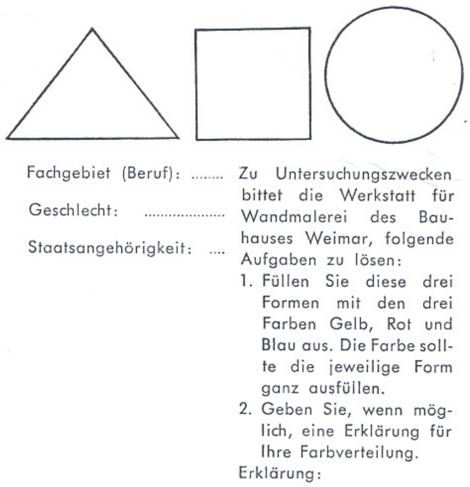


Figure 2.2: W. Kandinsky. 1923. Human perception and abstraction. The instruction asked to fill the shapes with yellow, red, and blue color, and to provide an explanation for the choice.

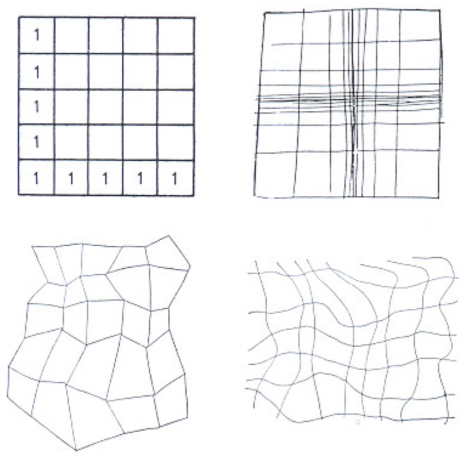


Figure 2.3: Drawings from the “Thinking Eye” [73]. Transformation of abstract element, the grid : from a static reference system to dynamic field.

artefact, with the aim of uncovering future interactions and products.

Relations between potential abstract elements or properties of a design are more complex where interactive sonic artifacts are concerned than in the case of purely graphic design. Partly as a result of this, we will rely on other analytic design tools in addition to the methods of the Bauhaus that have been described above, including the *morphological matrix.* The latter represents a structure that may be used to understand and to organize the multi-dimensional qualities (sonic, formal, interactive) of a design – in our case, those of an interactive sound-based experience. Fritz Zwicky developed the morphological matrix formalism in order to decompose otherwise seemingly non-reducible complexity [143]. He described the method of the Morphological Box in five steps:

> *First Step:* The problem which is to be solved must be exactly formulated.
> *Second Step:* All of the parameters which might enter into the solution of the given problem must be localized and characterized.
> *Third Step:* The morphological box or multidimensional matrix which contains all of the solutions of the given problem is constructed.
> *Fourth Step:* All of the solutions which are contained in the morphological box are closely

Figure 2.4: L. Moholy-Nagy, Hand sculptures. Designing for the eyes and hands.

> analyzed and evaluated with respect to the purposes which are to be achieved.
> *Fifth Step:* The best solutions are being selected and are carried out, provided that the necessary means are available. This practical application requires an additional morphological study. (Zwicky, [143])

The morphological matrix is used predominantly in design conceptualization rather than in analysis [143] [99]. We will use it in both ways. In the analysis and abstraction of case studies (as outlined above), we will employ it to organize basic properties from existing sound design solutions or interactions. Later we will apply it in multidimensional form as a means to generate novel design concepts accounting for sonic, formal and interactive qualities, alongside other design ideation methods such as bodystorming and interaction relabeling.

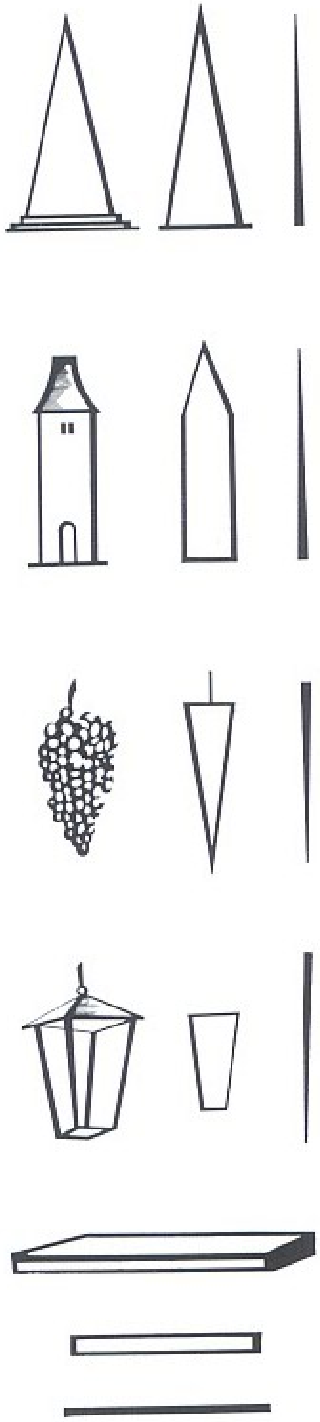


Figure 2.5: J.Ramsauer. Drawing Tutor. 1821. Reduction of standing and hanging objects to graphical signs.

3 Everyday sound synthesis algorithms

Several sound synthesis algorithms were proposed as part of the Sounding Object project (2001-2003). In this chapter, these and more recently-developed models and algorithms are presented, based on a taxonomy of everyday sounds (see section 1.2.5). Such taxonomy has been inspired by the literature on everyday sound perception (see section 1.1) as well as by observations driven by our modeling practice.

3.1 Introduction

In 1969, Risset published a ground-breaking catalog of computer-synthesized sounds [69], which served the purpose of illustrating the emerging techniques of sound analysis and synthesis. Those examples and studies are still influential for composers and sound scientists, especially those working with signal-processing tools and in the context of musical sounds. The discipline of psychoacoustics has provided, over the years, a solid support to connect signal processing to human perception.

A new stream of studies started in the early eighties from the observation that everyday listening is different from musical listening [48]. Both new psychoacoustic and sound modeling methods and results are needed for this new framework. On the perceptual side, the viewpoint of ecological psychology is very useful [49]. On the modeling side, the physically-based modeling paradigm seems to be the best sound production strategy to address everyday listening in interactive applications.

The EU-funded project "the Sounding Object" (SOb) [1] was launched in 2001 to provide a corpus of knowledge in everyday sound perception, accompanied by suitable new methods and tools for physics-based sound modeling and for high-level control of these models.

The SOb project aimed at "sounding objects" that incorporate a (possibly) complex responsive acoustic behavior, expressive in the sense of ecological hearing, rather than the (re-)production of fixed isolated signals. Although "real" sounds hereby serve as an orientation, realistic simulation is not necessarily the perfect goal: simplifications which preserve and possibly exaggerate certain acoustic aspects, while losing others considered less important, are often preferred. Besides being more effective in conveying certain information, such "*cartoonifications*" are often cheaper to implement, just like graphical icons are both, more clear and easier to draw than photo-realistic pictures.

The main idea behind the modeling approach is that elementary physical phenomena, such as impact, friction, etc., are accurately simulated, while more structured processes, such as bouncing or breaking, are modeled via geometrical or statistical simplifications.

Contacts of solid bodies form a large class of sound-emitting processes in every-day surroundings. Many typical forms of contact interaction can be successfully modeled on the basis of a physically founded but "abstracted", flexible and efficient one-dimensional impact or friction algorithm. Specific characteristics of the macroscopic scenarios which are of high perceptual relevance are modeled explicitly, for instance as macro-temporal distributions of micro-impacts.

Following this example of impacts sounds, several other models have been developed (friction, bubble), or will be developed (fracture, fluid flow, turbulence, explosion), that allow to form subsequent basic events, derived processes and simulation examples. All this collection actually defines a taxonomy of everyday sounds which is detailed in section 1.2.5.

These notes are intended to be an explanatory guide to our collection of sound models and examples, as they are available nowadays from the SOb project website[2] as `pd` and Max/MSP *externals* and

[1] `http://www.soundobject.org/SObBook/SObBook_JUL03.pdf`
[2] `http://www.soundobject.org`

patches. Detailed explanations of the inner structure and the development of some models can be found in [27]. Basic interactions (resonator, impact, friction, bubbles) are described in section 3.2. Higher-level models describing phenomena with complex temporal patterns are presented in section 3.3. Finally, section 3.4 briefly gives some examples of how the sound models can be associated to everyday objects, thus providing their typical sonic behavior in an interactive, real-time fashion.

3.2 The low-level physics-based models

Most of the here considered models for solid objects comply with the following framework: two resonating objects interact by means of a contact model. This way, as opposed to wave-table techniques, the corresponding algorithms can be instantiated with an infinite variety of interaction- and resonator-properties, which can be easily tuned to attributes of ecological hearing (material, shape, impact position or surface properties a.s. - see also section 1.1.2). Also, the exact form of each interaction depends on the actual state of the involved resonators.

Models for sounds from liquids are also being explored. To date, models based on single resonating cavities in liquids (bubbles) are adopted to reproduce dripping-like sounds, and mixtures of bubble models and textures are adopted to reproduce composite liquid sounds such as splashes.

3.2.1 Resonators

The two interacting objects are built under the premises of modal [1] or digital waveguide [70] synthesis[3].

The modal formulation supports particularly well our main design approach for its physical generality and, at the same time, for its intuitive acoustic meaning. A modal `resonator` is here characterized by the number of its modes and for each mode the three modal parameters: mode-frequency, exponential-decay time and level ("weight") of the mode at the point of interaction. Further, for each modal resonator, an arbitrary number of "pickups" can be defined, which are characterized by weighting-factors (for all modes).

The waveguide `resonator` here considered simulates an ideal vibrating string with propagation losses, and is characterized by the length, mass and tension of the string. In addition it is possible to vary the contact position along the string, which also corresponds to the pickup point.

3.2.2 Impact

In contrast to several studies of contact sounds of solid bodies that focus on the resonance behavior of interacting objects and widely ignore the transient part of the event, our approach is based on a physical description of impact interaction processes [6]. This physical model involves a degree of simplification and abstraction that implies efficient implementation as well as adaption to a broad range of impact events. The `pd` and Max/MSP *externals* are called "`impact_2modalb`", "`impact_modalb_wg`", "`impact_inertialb_modalb`", and "`impact_inertialb_wg`".

We consider two resonating objects and assume that their interaction depends on the difference x of two (1-dimensional) variables connected to each object. In the standard case of examined movements in one spatial direction, x is the *compression* (the inverse of the *distance*) variable in that direction ("positive compression" - or "negative distance" - corresponds to "contact"). Possible simultaneous interactions along other dimensions are excluded at this stage. This leads to a compact efficient algorithm that strikes the main interaction properties. The impact force f is stated as a nonlinear term in x (and $\dot{x}$):

$$f(x,\dot{x}) = \begin{cases} kx(t)^{\alpha} + \lambda x(t)^{\alpha} \cdot \dot{x}(t), & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{3.1}$$

[3] In fact, the *externals* are realized in a modular structure that enables the connection of numerous different `resonators` as well as `interactors`.

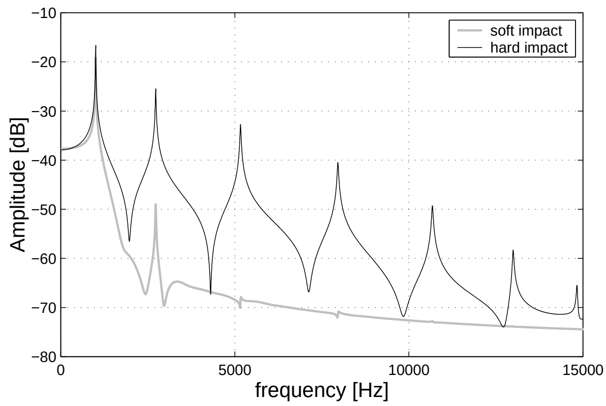


Figure 3.1: Sound spectra obtained when hitting a resonator with a soft mallet (low m_h/k) and with a hard hammer (high m_h/k).

where k is the *elasticity constant* (corresponding to the hardness of the impact), α, the exponent of the non-linear terms, shapes the dynamic behavior of the interaction (i.e. the influence of initial velocity), while λ controls the dissipation of energy during contact, accounting for friction loss.

Alternative versions, "`linpact_...`" exist with a simpler linear force term, that trade richness in detail for reduced computational cost.

It is often satisfactory and more convenient to use the modules "`impact_inertialb_...`" and "`linpact_inertialb_...`", where the first `resonator` is reduced to an inertial (point-)mass[4] and characterized only by one (mass-)parameter. This practical and computational simplification parallels the notion that in many practical contact scenarios the vibration of one involved object is hardly or not perceived.

Finally, all modules have three audio inlets for the input of signals representing external forces on both resonators (again at the point of interaction) and an additional positional offset, used mainly for surface profiles in rolling-/sliding-models.

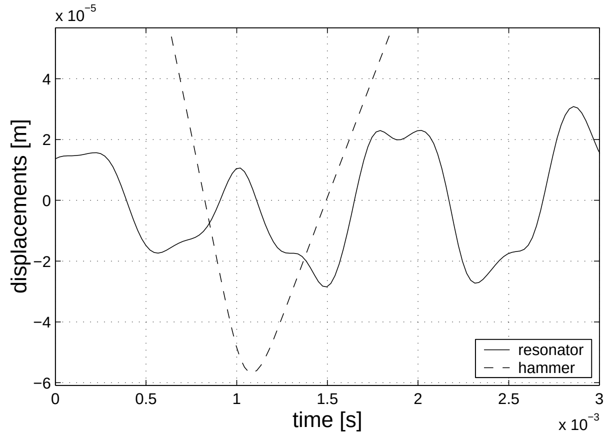
The impact model has been tested in order to assess its ability to convey perceptually relevant information to a listener. A study on materials [6] has shown that the decay time is the most salient cue for material perception. This is very much in accordance with previous results [72]; however, the physical model used here is advantageous over using a signal-based sound model, in that more realistic attack transients are obtained. The decay times of the resonator modes can therefore be used to "tune" the perceived material of the resonator in a collision with a hammer. See also Chapter 4 in [27] for more detailed discussion on material perception from recorded and synthesized sounds.

A study on hammer hardness [5] has shown that the *contact time* t_0 (i.e. the time after which the hammer separates from the resonator) can be controlled using the physical parameters. This is a relevant result, since t_0 has a major role in defining the spectral characteristics of the initial transient. Qualitatively, a short t_0 corresponds to an impulse-like transient with a rich spectrum, and thus provides a bright attack. Similarly, a long t_0 corresponds to a smoother transient with little energy in the high frequency region. Therefore t_0 influences the spectral centroid of the attack transient, and it is known that this acoustic parameter determines to a large extent the perceived quality of the impact [47].

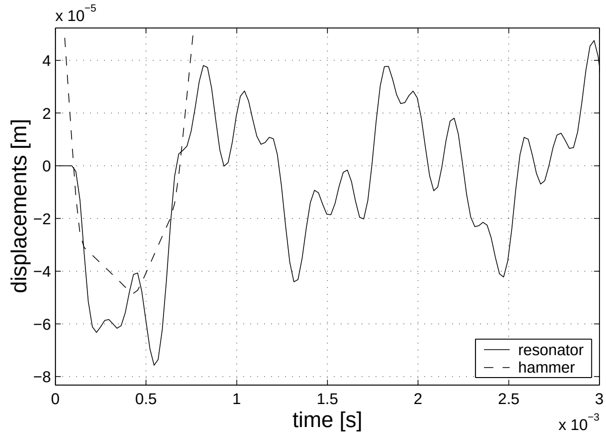
In [5], an expression for t_0 has been derived, and the ratio $m^{(h)}/k$ is found to be the most relevant parameter in controlling contact time and consequently the perceived hardness of the impact. Numerical simulations have shown excellent accordance between contact times computed using and those observed in the simulations. Fig. 3.1 shows an example of soft and hard impacts, obtained by varying m_h/k.

Due to the physical description of the contact force, realistic effects can be obtained from the model

[4]This is the special case of a modal resonator with only one resonant mode of frequency 0 and infinite decay time (undamped).



(a)



(b)

Figure 3.2: Numerical simulations; (a) impact on an oscillating resonator; (b) micro-impacts in a hard collision. Intersections between the solid and the dashed lines denote start/release of contact.

by properly adjusting the physical parameters. Fig. 3.2a shows an example output from the model, in which the impact occurs when the resonator is already oscillating: the interaction, and consequently the contact force profile, differs from the case when the resonator is not in motion before collision. This effect can not be simulated using pre-stored contact force profiles (as e.g. in [134]). Fig. 3.2b shows an example of "hard collision", obtained by giving a very high value to the stiffness k, while the other model parameters have the same values as in Fig. 3.2a. It can be noticed that several micro-collisions take place during a single impact. This is qualitatively in accordance with the remarks about hard collisions by van den Doel et al. [134].

3.2.3 Friction

For friction modeling, we use a computational structure very similar to the one used for impacts. The `pd` and Max/MSP *externals* are called "`friction_2modalb`" and "`friction_modalb_wg`".

The underlying model describes the average behavior of a multitude of micro-contacts made by hypothetical bristles extending from each of two sliding surfaces. When a modal decomposition is adopted for both interacting objects, the equations are:

$$\left\{ \dot{z}(z,v) = v(t)\left[1-\alpha(z,v)\frac{z(t)}{z_{ss}(v)}\right] \; f(z,\dot{z},v,w) = \sigma_0 z(t) + \sigma_1 \dot{z}(t) + \sigma_2 v(t) + \sigma_3 w(t) \right. \tag{3.2}$$

where v represents the *relative velocity* between the two rubbing objects, z is the *mean bristle displacement*, while w is a generic pseudo-random function which introduces a *noise component*. As far as the form of functions α and z_{ss} is concerned, we adopt a couple of previously proposed functions[7]. High-level interactions rely mainly upon three interaction parameters: the *external forces* f_{e1} and f_{e2} acting on (dragging) each of the two objects, which are tangential to the sliding direction, and the *normal force* f_N between the two objects. The remaining parameters belong to a lower level control layer, as they are less likely to be touched by the user and have to be tuned at the sound design level. Such low-level parameters can be grouped into two subsets, depending on whether they are related to the resonators' internal properties or to the interaction mechanism. A second subset of low-level parameters relates to the interaction force specification.

The phenomenological role of the low-level physical parameters of the friction model has been studied. The description given in Table 3.1 can be a helpful starting point for the sound designer.

Sym.	Physical Description	Phenomenological Description
σ_0	bristle stiffness	affects the evolution of mode lock-in
σ_1	bristle dissipation	affects the sound bandwidth
σ_2	viscous friction	affects the speed of timbre evolution and pitch
σ_3	noise coefficient	affects the perceived surface roughness
μ_d	dynamic friction coeff.	high values reduce the sound bandwidth
μ_s	static friction coeff.	affects the smoothness of sound attack
v_s	Stribeck velocity	affects the smoothness of sound attack
f_N	normal force	high values give rougher and louder sounds

Table 3.1: A phenomenological guide to the friction model parameters.

The triple $(\sigma_0, \sigma_1, \sigma_2)$ (see Eq. (3.2)) defines the *bristle stiffness*, the *bristle internal dissipation*, and the *viscous friction*, and therefore affects the characteristics of signal transients as well as the ease in establishing stick-slip motion. The triple (f_c, f_s, v_s) (see [7]) specifies the shape of the steady state Stribeck curve: specifically, the *Coulomb force* and the *stiction force* are related to the *normal force* through the equations $f_c = \mu_d f_N$ and $f_s = \mu_s f_N$, where μ_s and μ_d are the *static* and *dynamic friction coefficients*. Finally, the *breakaway displacement* z_{ba} is also influenced by the normal force [7].

3.2.4 Bubbles in liquids (dripping)

Dripping, the falling of an object into a liquid, can be considered one of the most basic events involving liquids. The dripping sound occurring in the occasion of such an event is due to the formation of radially oscillating bubbles just under the surface of the liquid [49, 133]. When assuming that the bubble cavity acts as a simple Helmoltz resonator, the equation of its impulse response is

$$p(t) = a \sin(2\pi f(t)t)e^{-dt} \tag{3.3}$$

where $f(t)$ is the time-varying resonance frequency, d the damping factor, a is the amplitude, and t is time. For the simulation of a single bubble sound, we used the following relations reported in [133]: the initial frequency $f_0 = 3/r$ (r being the bubble radius), the damping factor $d = 0.043 f_0 + 0.0014 f_0^{3/2}$, and instantaneous frequency $f(t) = f_0(1 + \sigma t)$ (σ being the slope of the frequency rise). Table 3.2 shows the link between the two principal parameters used for the dripping model and their acoustic counterparts.

Sym.	Physical Description	Phenomenological Description
r	bubble radius	affects the initial pitch f_0
σ	density of the liquid	affects the speed of frequency rise

Table 3.2: A phenomenological guide to the dripping model parameters.

The more the cavity of the bubble becomes larger, the more the simple single bubble sound model becomes inadequate to represent the dripping event. This is most probably due to the fact that large objects or drops falling into a resting liquid generate a large cavity (single resonance) as well as many

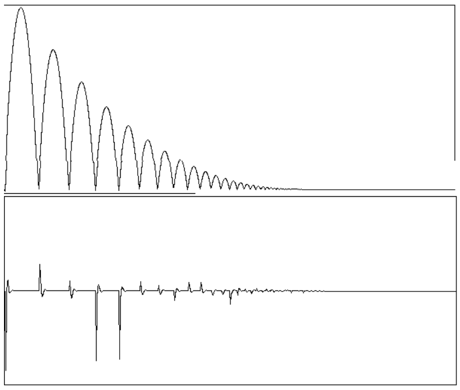


Figure 3.3: Temporal movement of an inertial mass (above) “bouncing” on a two-mode resonator (at `pickup-point`, below).

secondary bubbles and droplets events due to the mass displaced by the principal impact event. As a preliminary attempt to reproduce the temporal structure of a splashing sound, we choose to design this event as a sequence of three distinct events: 1. a short samples noisy impact sound, 2. a bubble sound modeled as detailed above, and 3. a secondary droplets event texture.

3.3 Higher-level structures

3.3.1 Bouncing and Breaking

Short acoustic events like impacts can strongly gain or change in expressive content when set in an appropriate temporal sequence. Examples are the grouping of impacts in “bouncing” and “breaking” patterns. Warren and Verbrugge [138] study on the perception of breaking- and bouncing-scenarios is a starting point for our related modelling efforts. They showed, that sound artefacts, created through layering of recorded collision sounds, were identified as bouncing or breaking scenarios depending on their homogeneity and the regularity and density of their temporal distribution. These results can be effectively exploited and expanded by higher-level sound models, making use of the “impact” module previously described.

Bouncing is the result of a constant external (gravity-)force term, as depicted in Fig. 3.3. The one-dimensionality of the impact algorithm only allows the immediate simulation of symmetrical, basically spherical, bouncing objects, by coupling an accelerating tempo to the impact parameters. A strict physical simulation of irregular bouncing objects would be highly complex to control and computationally too demanding. Instead, a high-level modelling of typical bouncing-patterns leads to *cartoonifications*, that are efficient to implement and able to express ecological attributes like regularity/irregularity of the bouncing object.

In the CLOSED sound-tools the temporal patterns are generated and controlled by a `pd` patch (implemented also in `Max/MASP`), called the “`dropper`”, which form the core element of both the bouncing and breaking models.

The first basic principle behind the bouncing process is the loss of macro-kinetic energy of the global vertical, horizontal and rotational movement, due to friction and acoustic vibrations. Under the assumption that the loss of energy at each bounce is proportional to the remaining kinetic energy, one obtains a global energy term that decays exponentially with the number of reflections. Therefore, in the case of simple spherical bouncing objects, the kinetic energy at floor level is proportional to the duration of the following bounce. We thus arrive at exponentially decaying terms for impact velocities and temporal intervals in a regular bouncing movement. The implementation of this basic scheme in fact proved to be convincing in comparison with recordings of bouncing (round) wooden balls.

For irregular objects, energy transfers occur between the vertical, horizontal and rotational terms, of which only the vertical velocity (and therefore the maximum height) contributes with a simple term to the impact intervals and velocities. On the other hand, the contribution of the rotational movement is not expressible in a simple form, while the contribution of the horizontal movement is basically zero. Energy transfers, however, can be approximately modelled by means of deviations of both inter-impact intervals and velocities from a strict exponentially decaying behavior. Furthermore, the effective relative masses and the weighting-factors of resonant modes are modulated through the rotation (and therefore changing contact points) of an irregular object. In general, the exact movement in the non-spherical case can only be simulated through a detailed solution of the underlying differential equations. However, it seems questionable how precisely shapes of bouncing objects (except for sphericity) can be recognized acoustically. Conversely, controlled-random patterns of impact parameters can generate expressive cartoonifications.

An important extension of the spherical case includes, for example, shapes with certain symmetries (e.g. disks or cubes). In these cases the transfer of energy between the vertical, horizontal and rotational terms can take place in regular patterns, closely related to those of spherical objects. This phenomenon is exploited in some modelling examples. Often however, such movements include rolling aspects, suggesting a potential improvement through integration of rolling models. A very prominent sound example with an initial "random"- and a final regular stage is that of a falling coin.

¿From an implementation point of view, the "`dropper`" patch generates temporal patterns of impact velocities, according to the previous considerations. Control parameters are:

1. the time between the first two reflections, representing the initial falling-height.
2. the acceleration factor, i.e. the quotient of two following maximal "bounce-intervals", describing the amount of microscopic energy loss/transfer at each reflection, thus the speed of the exponential time sequence. From a phenomenological point of view, it represents the object elasticity.
3. a parameter controlling the energy of the impact, related to the sensation of softness of the material.
4. another parameter specifying the range of random deviation of the impact velocities. The irregularity/sphericity of an object's shape is modelled in this way.
5. the initial impact velocity.
6. a threshold parameter that controls when the accelerating pattern is stopped and produces a "terminating bang", that can possibly trigger a following stage of the bouncing process.

The main ideas behind the structure of the breaking-model are now shortly sketched. Typical fragments of rupture present highly irregular form and are usually rather anelastic. This means that they perform a decelerating rather than an accelerating progression. It is important to keep in mind that emitted fragments mutually collide, and that the number of such mutual collisions rapidly decreases, starting with a massive initial density. These collisions cannot be described as bouncing patterns at all. According to these considerations the breaking-model was realized by means of the `dropper` with high values of "randomness", and a quickly **decreasing** temporal density, i.e. a behavior opposite to that of the bouncing movements. Following Warren and Verbrugge's results, a short noise impulse added to the attack portion of the pattern underlines the breaking character.

As another insight during the modeling process, several sound attributes, describable as structural invariants [138], showed to be important. Impacts with identical temporal structure seem to be less identifiable as a breaking event, when tuned to a metallic character in their modal settings; this may correspond to the fact that fractures of metal objects are rather rare in everyday experience. Also, extreme mass relations of "striker" and struck resonator in the impact settings led to more convincing results. Again, this is in correspondence with typical situations of breakage: a concrete floor has a practically infinite inertia in comparison to a bottle of glass.

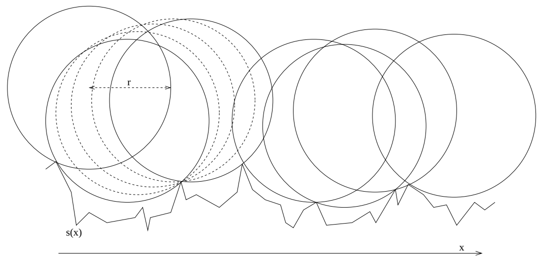


Figure 3.4: Sketch of the fictional movement of a ball, perfectly following a surface profile s(x). Relative dimensions are highly exaggerated for a clearer view. Note that this is **not** the de-facto movement; this idealization is used to derive the offset-curve to be used by the impact-model.

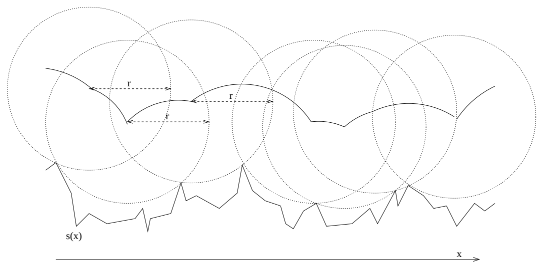


Figure 3.5: Sketch of the effective offset-curve, resulting from the surface $s(x)$.

3.3.2 Rolling

Among the various common mechanical interactions between solid objects, "rolling" scenarios form a category that seems to be characteristic also from the auditory viewpoint: Everyday experience tells that the sound produced by a rolling object is often recognizable as such, and in general clearly distinct from sounds of slipping, sliding or scratching interactions, even of the same objects. This may be due to the nature of rolling as a prominent **continuous** interaction process, where the mutual force on the involved objects is described as an impact without additional surface-tangential friction forces.

Consequently, the impact-algorithm has been embedded in a complex higher-level structure to reach an efficient cartoonification, that can express various ecological attributes of rolling-scenarios: material, size and shape of the objects, as well as velocity or acceleration/deceleration (transformational attributes [48]).

Rolling contact between two objects is restricted to distinct points: the supporting surface is not continuously traced. Fig. 3.4 sketches the idea; the rolling object is here assumed to be locally spherical without "microscopic" surface details. Any micro details in the surface of the rolling object can simply be added to the supporting surface. For deviations from sphericity, the radius of the remaining "smoothed macroscopic" curve could be varied.

The actual movement of the rolling object differs from the idealization of Fig. 3.4 due to inertia and elasticity. In fact, it's exactly the consequences of these physical properties that substantiate the use of the impact-model equations. It is important to notice that, in contrast to slipping-, sliding- or scratching-actions, the interaction force on the two objects involved in a simple rolling-scenario is approximately perpendicular to the contact surface, pointing along the line between the instantaneous contact point and the center of mass of the rolling object. This fact is not reflected in the sketches, since here relative dimensions are highly unrealistic, exaggerated for purposes of display. Summing up, the final vertical movement of the center of the ball can be approximated by use of the one-dimensional impact-model with the offset-curve shown in Fig. 3.5.

In a naive approach, the calculation of contact points is computationally highly demanding. To facilitate an efficient implementation, a "smart" algorithm had to be developed, that reduces the number of calculations and comparisons by factors up to 1000 [105]. The ideal offset-curve for the impact model is then calculated from the coordinates of the contact points.

The surface-signal which is processed by a "rolling-filter" as above might be derived through scanning of real surfaces. A flexible statistics-based generation though is preferable in our context over static storage of fixed profiles. One such approach is *fractal noise*, i.e. noise with a $1/f^\beta$ power spectrum, the real parameter β reflecting the fractal dimension or roughness. However, practical results became much more convincing, when the bandwidth of the surface-signal was strongly limited. This does not come as a surprise, when one keeps in mind that typical surfaces of objects involved in rolling scenarios, are generally smoothed to high degree. Cutting and arranging pieces of stone for a stone floor corresponds to high-pass-filtering, while smoothing on a microscopic level, e.g. polishing the stones, can approximately be seen as low-pass-filtering. Band-pass filtered white noise thus was chosen as a cheap and efficient solution for surface representation. It can eventually be enhanced by an additional second-order filter, whose steepness finally represents a "microscopic" degree of roughness as a very coarse approximation of the fractal spectrum.

The parameters of the impact itself, in particular the elasticity constant k, must also be carefully adjusted, as material properties strongly contribute to the expressiveness of the model.

Typical scenarios of rolling tend to show characteristic macroscopic acoustic features, that appear to be of high perceptual relevance, especially for velocity-expression. Macro-temporal periodicities result from typical patterns of more or less regular nature as found on many ground surfaces (such as paved floors, the periodic textures of textiles, or the pseudo-periodic furrows in wooden boards). Moreover, for rolling objects that are not perfectly spherical the velocity of the point of contact and the effective force pressing the rolling object to the ground vary periodically. In order to model such deviations from perfect sphericity, these two parameters must be modulated with narrow-band modulation signals. Of course all quasi-periodic modulations have to reflect the rolling-velocity in their frequency.

Finally it is to be noted that, like in everyday listening, acoustic rolling scenarios are recognized and accepted more easily with typical dynamics. As an example, consider the sound of a falling marble, that bounces until constant contact with the ground is reached, and the rolling action takes over.

3.3.3 Crumpling

Like most of the other sounds presented in this catalog, the `crumpling pd` patch results from providing the impact model with a control layer. Since `crumpling` does not model physical contacts between solid objects but, rather, time sequences of crumpling events, the use of closed-form formulas expliciting hammer/object collision mechanisms can be avoided.

The temporal distribution of crumples, as well as their own power, follow stochastic laws which are derived from physics [66]. Such laws govern the energy consumption of an impact and the temporal idle time between adjacent events. Both the temporal sequence of events and the dissipation law expose a characteristic parameter, on top of which we can design control maps affecting the average idle time between events and the average power of impacts, respectively.

By providing the crumpling process with an initial "potential" energy that will be progressively burned out by every single event, proportionally to its own power, then we can vary the average time length of the process once the parameters of the temporal and power stochastic laws have been set.

The physics-based approach to crumpling sound synthesis allows to design a control layer whose knobs straightforwardly map on top of the idle time, power, and potential energy process parameters. In this way the user interface presents a physically meaningful control panel, without the need of designing complex intermediate layers between the model and the interface.

Besides crumpling sounds, by means of this model we were able in particular to synthesize sounds of crushing cans and footsteps.

The user interface so far puts four sliders available to the user, along with three presets that give an initial direction to the user about the allowed parameter ranges and the potential of the model. These presets respectively initialize the model to the synthesis of crumpling, can crushing, and footstep sounds, meanwhile providing the corresponding parameter settings.

As to the sliders, their meaning is shown in Table 3.3.

- The `Metal` slider passes by the temporal/power control layer, and connects directly to the decay time parameter of the impact model thus providing a control of the "metallicity" of the crumpled or chrushed material.
- The `Object_HEIGHT` slider exposes the control of the initial potential energy, and can be visualized as a parameter of "size" (i.e., area, height, amount and so on) of the crumpled material.
- The `Crushing_FORCE_(against_Obj)` slider maps to the power of the indivdual crumpling events.
- The `Object_SOFTNESS` slider jointly affects the statistics of time and power, and refers to the property of the crumpled object to resist against the crumpling/crushing action coming from the external world.

Sym.	Physical Description	Phenomenological Description
`Metal`	decay time	metallic timbre of individual crumples
`Object_HEIGHT`	potential energy	average process time length
`Crushing_FORCE_(against_Obj)`	average power of individual crumples	dynamics of individual crumples
`Object_SOFTNESS`	object resistance against crumpling action	smoothness of the crumpling process

Table 3.3: A guide to the tuning of sliders of the crumpling model.

Yet, aiming at modeling higher level scenarios, the "user" might be an upper level control structure which triggers events according to some high-level process. Rules governing the temporal evolution of "walking" and "running" exist [17]. Those rules drive the crumpling model parameters directly, so to obtain interesting walking and running sounds. Crushing, walking and running are extensively described in [44].

3.3.4 Textures

In the taxonomy proposed in Fig. 1.6, layer II also includes sound *textures*. Textures are useful to reproduce sounds with complex temporal patterns that can be effectively represented as statistical distributions of short basic audio segments, or grains. The sounds of rain, waterfall, wind, fire, or crowds, are examples for which a sound texture model is usually the preferred choice.

In the examples illustrated so far, we used sound textures, for example, in the modeling of splash sounds. As already said, splash sounds are characterized by a main dripping event caused by the falling mass creating a large cavity (single resonance), followed by many secondary bubbles and droplets events. We found that using a texture to represent the secondary event was more simple and effective than using a huge number of dripping events based on the bubble model.

3.4 Familiar (Sounding) Objects

The expressiveness of the sound models is best recognized when parameters are set to values adequate for scenarios familiar from every-day experience. Such demonstrations often involve combinations of several models; we have chosen some items, partly accompanied with basic visualizations, from rather simple to complex ones:

- The sole impact-model can be tuned to *struck bars* of different sizes and materials,
- as the low-level friction-model can realize *squeaking doors* and *rubbed glasses.*
- The rolling-model with its strong ecological potential (velocity, direction, size . . .) "sonifies" different interactive "games" with *rolling balls.*
- Rolling and friction are two states of an interactive *wheel–brake* construction.
- The dropper-object delivers convincing *bouncing balls* as well as *dropping plastic bottles* and *metallic coins* or *breaking glasses.*
- Natural is the combination of *dropping and subsequently rolling balls.*
- Typical scenes of crumpling are *crushing cans* and the sound of *walking on gravel.*

Conclusion

The overall goal of the deliverable 4.1 (including this current deliverable - state of the art - and the following one - recommendations) is to provide recommendations for the development of sound design tools and proto sound products. This first part of this document is a state of the art of "everyday sound classification". Indeed, everyday sound classification is the first goal to reach to fulfil the requirements of this deliverable, for the first issue to address is to define everyday sounds and their cognitive representations, prior to start building design tools for everyday sounds.

This report is threefold. In the first part, we have provided a review of the literature on **perception of everyday sounds**. This review was split up into two main sections. The first section focused on experimental works studying the perception of everyday sounds. The studies have shown that there are two kinds of listening. Listeners may focus on the acoustical properties (meaningless) of the sound signal. They may as well focus on the sound event that has caused the sound: this is the most spontaneous reaction while listening. Sound events are actually the sounds with which the CLOSED project is interested. In this latter case, listeners identify the sound event and recover its properties. From the many studies that have analyzed how listeners perceive a sound event, we have concluded that sound event perception is based on dual bottom-up and top-down processing. This processing is made of inferences based, on the one hand, on the auditory attributes perceived from the sound and from the context, and, on the other hand, on the expectations that the listener forms, and on its previous knowledge. Thus, the issues of identification and recovery of the sound event properties have oriented us toward the more general problematic of cognitive categorization. In the second section of this part, we have reviewed the different theories of categories and classification. More specifically, we have emphasized the different notions of similarity. We have then reviewed some existing classifications and categorizations of everyday sounds, including the taxonomy of everyday sounds proposed by UNIVERONA, thoroughly explored in the third part of this document. This review has provided us with the first assumptions that will allow us to define the recommendations for sound design tools, awaited by NIPG.

The results of the experimental and theoretical studies of everyday sound perception have also found applications to the domains of sound quality and sound design. Thus, the second part of this document has proposed an overview of **everyday sound interaction and design**. Particularly, the effects of context, and intercultural differences have been reviewed on the basis of ethnographical studies. The effect of context had indeed already been acknowledged in experimental studies of sound perception. In this second part of the document, the effect of context was related more generally to the cultural factors that might influence the reception and the appraisal of a sounding artefact. From a theoretical and historical perspective, we have then defined the concepts of basic interaction design, which actually form the core the design part of the CLOSED project.

Finally, we have proposed in the third part of this document a thorough exploratory guide to the taxonomy of **everyday sound synthesis** algorithms that has been developed by UNIVERONA. This taxonomy was previously summarized and was related to the theories of categories and categorization. This taxonomy, originating from the Sounding Project, currently includes sounds caused by interactions of solids, and is extending to liquid and aerodynamics sounds. It has an overall structure made of three levels: basic interactions, higher level structures and familiar sounds build upon these elements. Each element of these levels were detailed in this last part of the document, as well as their implementation in Max or Pure Data, and the phenomenological description of the model parameters. Moreover, the rationale (including perceptual validations) of the simplifications that has allowed to reach *cartoonification* has been described.

The confrontation of these three standpoints over everyday sound perception reveals common issues. The perception of basic interaction properties - shape, size, material -, the need to account for contextual and intercultural factors, the need of a categorization framework accounting for complex and dynamic relations between the categories of sounds, are examples of these problematics shared by the different sensibilities that form the CLOSED consortium. It will be the goal of the following part of this deliverable - awaited at the end of month 11 - to synthetize these results, to propose a model for the classification of everyday sounds, and thus to propose the recommendations for the sound design tools and the proto sound products.

Bibliography

[1] J. M. Adrien. The Missing Link: Modal Synthesis. In G. De Poli, A. Piccialli, and C. Roads, editors, *Representations of Musical Signals*, pages 269–297. MIT Press, 1991.

[2] G. Aneschi. Basic Design, fondamenta del design. In M. A. Garito, G. Anceschi, and M. Botta, editors, *L'Ambiente dell'Appredimento - Web Design e Processi Cognitivi*, pages 57–67. McGraw-Hill, Milano, 2006.

[3] F. Ashby. *Multidimensional Models of Perception and Cognition.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1992.

[4] S. Atran. Scott folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral & Brain Sciences*, 21(4):547–569, 1998.

[5] F. Avanzini and D. Rocchesso. Controlling material properties in physical models of sounding objects. In *Proc. Int. Computer Music Conf.*, La Habana, Cuba, September 2001.

[6] F. Avanzini and D. Rocchesso. Modeling Collision Sounds: Non-linear Contact Force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 61–66, Limerick, Dec. 2001. Available at `http://www.soundobject.org`.

[7] F. Avanzini, D. Rocchesso, and S. Serafin. Modeling interactions between rubbed dry surfaces using an elasto-plastic friction model. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-02)*, Hamburg, Germany, 2002.

[8] J. A. Ballas. Common Factors in the Identification of an Assortment of Brief Everyday Sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2):250–267, 1993.

[9] J. A. Ballas and J. H. Howard-Jr. Interpreting the language of environmental sounds. *Environment and Behavior*, 19(1):91–114, 1987.

[10] J. A. Ballas and T. Mullins. Effect of context on the identification of everyday sounds. *Human Performance*, 4(3):199–219, 1991.

[11] L. Barsalou. Ad hoc categories. *Memory & Cognition*, 1:211–227, 1983.

[12] B. Berlin. Ethnobiological classification. In E. Rosch and B. Lloyd, editors, *Cognition and Categorization*, pages 9–26. Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.

[13] E. A. Björk. The perceived quality of natural sounds. *Acustica*, 57:185–188, 1985. Research Notes.

[14] J. Blauert. Cognitive and aesthetic aspects of noise engineering. In *Internoise*, volume 1, pages 5–14, Cambridge, MA, 1986.

[15] J. Blauert and U. Jekosch. Sound-quality evaluation – A multi-layered problem. *Acustica – Acta Acustica*, 83(5):747–753, 1997.

[16] M. Bodden. Perceptual sound quality evaluation. In *Internoise*, Nice, France, 2006.

[17] R. Bresin, A. Friberg, and S. Dahl. Toward a new model for sound control. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 45–49, Limerick, Ireland, Dec. 2001.

[18] P. A. Cabe and J. B. Pittenger. Human sensitivity to acoustic information from vessel filling. *Journal of experimental psychology: human perception and performance*, 26(1):313–324, 2000.

[19] P. Cano, M. Koppenberger, P. Herrera, and O. Celma. Sound effect taxonomy management in production environments. In *Proceedings of 25th International AES Conference London, UK*, 2004.

[20] C. Carello, K. L. Anderson, and A. J. Kunkler-Peck. Perception of object length by sound. *Psychological science*, 9(3):211–214, May 1998.

[21] M. Castellengo and D. Dubois. Timbre ou timbres ? Propriétés du signal, de l'instrument ou construction cognitive ? In *Actes du Colloque Interdisciplinaire de Musicologie/Proceedings of the Conference on Interdisciplinary Musicology (CIM'05)*, Montreal, Canada, march 2005.

[22] C.-Y. P. Chiu and D. L. Schacter. Auditory priming for nonverbal information: implicit and explicit memory for environmental sounds. *Consciousness and Cognition*, 4:440–458, 1995.

[23] C. Classen. Foundations for an anthropology of the senses. *International Social Science Journal*, 153:401–412, 1997.

[24] A. Collins and E. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975.

[25] A. Collins and B. Quillian. Retrieval time from semantic processing. *Psychology Review*, 82, 1969.

[26] A. Corbin. The Auditory Markers of the Village. In M. Bull and L. Back, editors, *The Auditory Culture Reader*, pages 117–125. Berg, Oxford, New York, 2004.

[27] D. Rocchesso and F. Fontana, Eds. *The Sounding Object*. Mondo Estremo, Firenze, Italy, 2003.

[28] R. L. Diehl, M. A. Walsh, and K. R. Kluender. On the interpretability of speech/nonspeech comparisons: a rely to Fowler. *Journal of the Acoustical Society of America*, 89(6):2905–2909, June 1991.

[29] G. Diesendruck, R. Hammer, and O. Catz. Mapping the similarity space of children and adults' artifact categories. *Cognitive Development*, 18(2), 2003.

[30] G. Diesendruck and M. Shatz. Two-year-olds' recognition of hierarchies-evidence from their interpretation of the semantic relation between object labels. *Cognitive Development*, 16(1), 2001.

[31] P. Dourish. Implications for design. In *Proc. ACM Conf. Human Factors in Computing Systems (CHI)*, pages 541–550, Montreal, 2006.

[32] D. Dubois. Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quaterly*, 1:35–68, 2000.

[33] D. Dubois. Perception, representation and knowledge: acoustic phenomena between noise and sounds. In *Proceedings of the EAA Symposium and architectural and urban acoustics TecniAcustica*, Bilbao, Spain, october 2003.

[34] D. Dubois, C. Guastavino, and V. Maffiolo. A cognitive approach to soundscape. Acoustic phenomena between "noise(s)" and "sound(s)". In *Proceedings of the Joint Congress CFA/DAGA'04*, pages 347–348, Strasbourg, France, march 2004.

[35] D. Dubois, C. Guastavino, and M. Raimbault. A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories. *Acta Inf.*, 92(6):382–395, 2006.

[36] E. Edwards, C. Gosden, and R. B. Phillips, editors. *Sensible Objects: Colonialism, Museums and Material Culture*. Berg Publishers Ltd, Oxford, New York, 2006.

[37] V. Erlmann, editor. *Hearing Cultures: Essays on Sound, Listening, and Modernity*. Berg Publishers Ltd, Oxford, New York, 2004.

[38] W. Estes. *Classification and cognition*. Oxford University Press, New York, 1994.

[39] F. Eustache, B. Lechevalier, F. Viader, and J. Lambert. Identification and discrimination disorders in auditory perception: a report on two cases. *Neuropsychologia*, 28(3):257–270, 1990.

[40] M. Eysenck and M. Keane. *Cognitive Psychology. A Student's Handbook*. Psychology Press, 4^{e}edition, 2000.

[41] A. Faure. *Des sons aux mots : Comment parle-t-on du timbre musical ?* Thèse de doctorat, Paris, Ecoles des Hautes Etudes en Sciences Sociales, Paris, 2000.

[42] C. Fellbaum. *Wordnet: an electronic lexical database*. MIT Press, 1998.

[43] A. Findeli. Rethinking design education for the 21st century: Theoretical, methodological, and ethical discussion. *Design Issues*, 17(1):5–17, Winter 2003.

[44] F. Fontana and R. Bresin. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In *Proc. XIV Colloquium on Musical Informatics*, Firenze, Italy, May 2003.

[45] C. A. Fowler. Sound-producing sources as object of speech perception: rate normalization and non-speech perception. *Journal of the Acoustical Society of America*, 88(3):1236–1249, September 1990.

[46] C. A. Fowler. Auditory perception is not special: we see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America*, 89(6):2910–2915, June 1991.

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

[48] W. W. Gaver. *Everyday listening and auditory icons*. PhD thesis, University of California, San Diego, 1988.

[49] W. W. Gaver. How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, 5(4):285–313, 1993.

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

[51] J. J. Gibson. *The senses considered as perceptual systems*. Houghton-Mifflin, Boston, MA, USA, 1966.

[52] B. L. Giordano and S. McAdams. Material identification of real impact sounds: effect of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America*, 119(2):1171–1881, February 2006.

[53] R. Goldstone. The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2):125–157, 1994.

[54] R. Goldstone and A. Kersten. Concepts and Categories. In A. Healy and R. Proctor, editors, *Comprehensive handbook of psychology, Volume Four: Experimental psychology*, pages 591–621. Wiley, New York, 2003.

[55] M. Grassi. Do we hear size or sound ? Balls dropped on plates. *Perception and Psychophysics*, 67(2):274–284, 2005.

[56] C. Guastavino, D. Dubois, J.-D. Polack, and C. Arras. Low frequency perception in urban soundscapes. a cognitive approach. In *Proceedings of the 17th International Conference on Acoustics ICA 2001*, Roma, Italy, 2001.

[57] C. Guastavino, B. F. Katz, J.-D. Polack, D. J. Levitin, and D. Dubois. Ecological validity of soundscape reproduction. *Acta Acustica united with Acustica*, 91:333–341, 2005.

[58] R. Guski, I. Felscher-Suhr, and R. Schuemer. The concept of noise annoyance: How international experts see it. *Journal of Sound and Vibration*, 223(4):513–527, 1999.

[59] F. Guyot. *Etude de la perception sonore en terme de reconnaissance et d'appréciation qualitative: une approche par la catégorisation*. PhD thesis, Université du Maine, 1996.

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

[61] U. Hahn and N. Chater. Similarity and rules: Distinct? exhaustive? empirically distinguishable. *Cognition*, 65(2-3):197–230, 1998.

[62] J. M. Hajda, R. A. Kendall, E. C. Carterette, and M. L. Harshberger. Methodological issues in timbre research. In I. Deliège and J. Sloboda, editors, *Perception and cognition of music*, chapter 12, pages 253–306. Psychology Press, 1997.

[63] J. Hampton. An investigation of the nature of abstract concepts. *Memory and Cognition*, 9(2):149–56, 1981.

[64] J. Hampton. Similarity-based categorization: The development of prototype theory. *Psychologica Belgica*, 35:103–125, 1995.

[65] J. Hampton. Testing the prototype theory of concepts. *Journal of Memory and Langage*, 4:686–708, 1995.

[66] P. A. Houle and J. P. Sethna. Acoustic emission from crumpling paper. *Physical Review E*, 54(1):278–283, 1996.

[67] J. H. Howard and J. A. Ballas. Syntactic and semantic factors in the classification of nonspeech transient patterns. *Perception and Psychophysics*, 28(5):431–439, 1980.

[68] D. Howes. Introduction: Empires of the Senses. In D. Howes, editor, *Empire of the Senses – The Sensual Culture Reader*, pages 1–20C. Berg, Oxford, New York, 2005.

[69] J.-C. Risset. *An Introductory Catalog of Computer-synthesized Sounds*. Bell laboratories, Murray Hill, New Jersey, 1969.

[70] J. O. Smith III. *Physical Audio Signal Processing: For Virtual Musical Instruments and Digital Audio Effects*. `http://ccrma.stanford.edu/∼jos/pasp/`, 2006.

[71] W. Kandinsky. *Point and Line to Plane*. Dover, New York, 1979.

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

[73] P. Klee. The thinking eye. In J. Spiller, editor, *Notebooks, Vol. 1*. Lund Humphries, London, 1961.

[74] L. Komatsu. Recent views of conceptual structure. *Psychological Bulletin*, 12:500–526, 1992.

[75] C. Krumhansl. Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Journal of American Psychological Association*, pages 445–463, 1978.

[76] A. J. Kunkler-Peck and M. T. Turvey. Hearing shape. *Journal of Experimental psychology: human perception and performance*, 26(1):279–294, 2000.

[77] S. Lakatos, S. McAdams, and R. Caussé. The representation of auditory source characteristics: simple geometric sources. *Perception and psychophysics*, 59(8):1180–1190, 1997.

[78] G. Lemaitre, P. Susini, S. Winsberg, B. Letinturier, and S. McAdams. The sound quality of car horns: a psychoacoustical study of timbre. *Submitted to Acta Acustica (manuscript code: AAA2005 _ 415).*

[79] X. Li, R. J. Logan, and R. E. Pastore. Perception of acoustic source characteristics: walking sounds. *Journal of the Acoustical Society of America*, 90(6):3036–3049, December 1991.

[80] A. L. Liberman and I. G. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1–36, 1985.

[81] E. Lupton. Visual Dictionary. In E. Lupton and J. Miller, editors, *The abc's of the Bauhaus and Design Theory*, pages 22–34. Thames and Hudson, London, 2001.

[82] R. A. Lutfi. Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks. *Journal of the Acoustical Society of America*, 97(2):1333–1334, February 1995. Letters to the Editor.

[83] R. A. Lutfi. Auditory detection of hollowness. *Journal of the Acoustical Society of America*, 110(2):1010–1019, August 2001.

[84] R. A. Lutfi, E. Oh, E. Storm, and J. M. Alexander. Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians and non musicians. *Journal of the Acoustical Society of America*, 118(1):393–404, July 2005.

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

[86] R. Lyon. Product sound quality – from perception to design. *Sound and Vibration*, pages 18–22, March 2003.

[87] V. Maffiolo, D. Dubois, S. David, M. Castellengo, and J.-D. Polack. Loudness and pleasantness in structuration of urban soundscapes. In *Proceedings of Inter-noise 98*, New Zealand, 1998.

[88] B. Malt, S. Sloman, S. Gennari, M. Shi, and Y. Wang. Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, 40(2):230–262, 1999.

[89] S. McAdams. Recognition of auditory sound sources and events. In S. McAdams and E. Bigand, editors, *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford University Press, 1993.

[90] S. McAdams, A. Chaigne, and V. Roussarie. The psychomechanics of simulated sound sources: material properties of impacted bars. *Journal of the Acoustical Society of America*, 115(3):1306–1320, March 2004.

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

[92] D. Medin, R. Goldstone, and D. Gentner. Respects for similarity. *Psychological Review*, 100(2):254–278, 1993.

[93] D. Medin and M. Schaffer. Context theory of classification learning. *Psychological Review*, 85(3):207–238, 1978.

[94] C. Mervis and E. Rosch. Categorization of natural objects. *Annual Review of Psychology*, 32(1):89–115, 1981.

[95] G. Miller. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 6, 1969.

[96] L. Moholy-Nagy. *Vision in Motion*. Paul Theobold, Chicago, 1969.

[97] G. L. Murphy and D. L. . The role of theories in conceptual coherence. *Psychological review*, 92(3):289–316, 1985.

[98] R. Nosofsky. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1):39–57, 1986.

[99] K. Otto and K. Wood. *Product Design: Techniques in Reverse Engineering and New Product Development*. Prentice Hall, Englewood Cliffs, NJ, 2000.

[100] S. Poitrenaud, J. Richard, and C. Tijus. Properties, categories, and categorisation. *Thinking & Reasoning*, 11(2):151–208, 2005.

[101] J.-D. Polack, M. Castellengo, V. Maffiolo, C. Guastavino, and B. F. G. Katz. Soundfield reproduction: the limits of physical approach. In *Proceedings of the Joint Congress CFA/DAGA'04*, Strasbourg, France, march 2004.

[102] E. Pothos. The rules versus similarity distinction. *Behavioral and Brain Sciences*, 28(01):1–14, 2005.

[103] M. Raimbault and D. Dubois. Perceptual judgements about urban soundscapes. In *Proceedings of Euronoise 2003*, Naples, Italy, 2003.

[104] M. Raimbault and C. Lavandier. Sound ambient environment of urban places: comparison of sound appraisal factors with acoustical parameters. In *Proceedings of the 3^{rd} European Congress on Acoustics - Forum Acusticum*, Sevilla, Spain, September 2002. Reproduced in Acta Acustica united with Acustica, vol. 88, suppl. 1 S70, 2002.

[105] M. Rath and D. Rocchesso. Continuous sonic feedback from a rollling ball. *IEEE Multimedia*, 12(2):60–69, 2005.

[106] B. Rehder and R. Hastie. Category coherence and category-based property induction. *Cognition*, 91(2):113–153, 2004.

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

[108] J.-F. Richard. *Les activités mentales*. Armand Colin, 4e edition, 2005.

[109] L. Rips. Similarity, typicality, and categorization. In S. Vosniadou and A. Ortony, editors, *Similarity and analogical reasoning*, pages 21–59. Cambridge University Press, Cambridge: England, 1989.

[110] L. Rips and A. Collins. Categories and resemblance. *Journal of Experimental Psychology: General*, 122(4):468–486, 1993.

[111] E. Rosch. Principles of categorization. In E. Rosch and B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.

[112] E. Rosch and C. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.

[113] E. Rosch, C. Mervis, W. Gray, D. Johnson, and P. Boyes-Bream. Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439, 1976.

[114] B. Ross and G. Murphy. Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, 38(4):495–553, 1999.

[115] A. Sasse and H. Knoche. Quality in context – an ecological approach to assessing QoS for mobile TV. In *Proc. ISCA Workshop on Perceptual Quality of Systems*, Berlin, 2006.

[116] T. Sato, T. Yano, M. Björkman, and R. Rylander. Road traffic annoyance in relation to average noise level, number of events and maximum noise level. *Journal of Sound and Vibration*, 223(5):775–784, 1999.

[117] S. Sattah and A. Tversky. Additive similarity trees. *Psychometrika*, 42(3):319–345, 1977.

[118] R. M. Schafer. *The tuning of the world.* Random House Inc., 1977.

[119] B. Schulte-Fortkamp and D. Dubois. Recent advances in soundscape research. *Acta Acustica united with Acustica*, 92(6):viii–v, 2006. Preface to the special issue.

[120] B. Schulte-Fortkamp and W. Nitsch. On soundscapes and their meaning regarding noise annoyance measurements. In *Proceedings of Inter-noise 99*, Fort Lauderdale, Florida, USA, 1999.

[121] R. Shepard. Analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27:125–140, 1962.

[122] R. Shepard. Analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27:219–246, 1962.

[123] S. Sloman. Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, 35(1):1–33, 1998.

[124] S. Sloman and L. Rips. Similarity as an explanatory construct. *Cognition*, 65(2-3):87–101, 1998.

[125] V. Sloutsky. The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6):246–251, 2003.

[126] E. Smith, E. Shoben, and L. Rips. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81(3):214–241, 1974.

[127] E. E. Smith and S. A. Sloman. Similarity- versus rule-based categorization. *Memory & cognition*, 22(4):377–86, 1994.

[128] G. Storms. Concept representation and the structure of semantic memory. In P. De Deyn, E. Thiery, and R. D'Hooge, editors, *Memory: Basic concepts, disorders, and treatment.* Acco, Leuven, 2003.

[129] P. Susini, S. McAdams, and S. Winsberg. A multidimensional technique for sound quality assessment. *Acustica united with Acta Acustica*, 85:650–656, 1999.

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

[131] E. Tulving. Episodic and semantic memory. In E. Tulving and W. Donaldson, editors, *Organisation of Memory*. Academic Press, London, 1972.

[132] A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.

[133] K. van den Doel. Phisically-based models for liquid sounds. In *Proc. ICAD 04*, Sydney, July 2004.

[134] K. van den Doel, P. G. Kry, and D. K. Pai. Foleyautomatic: Physically-based sound effects for interactive simulation and animation. In *Proc. ACM SIGGRAPH*, August 2001.

[135] N. J. Vanderveer. *Ecological acoustics: human perception of environmental sounds*. PhD thesis, Cornell University, 1979.

[136] D. Västfjäll. Sound quality evaluation of urban traffic soundscapes. In *Proceedings of Euronoise, Naples 2003*, 2003. Paper ID: 430-IP.

[137] W. H. Warren. The dynamics of perception and action. *Psychological review*, 113(2):358–389, 2006.

[138] W. H. Warren-Jr. and R. R. Verbrugge. 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, 1984.

[139] L. Wittgenstein. *Philosophical investigations*. The MacMillan Company, New York, 1953.

[140] T. Yamauchi and A. Markman. Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3):776–795, 2000.

[141] W. A. Yost, L. D. Braida, W. W. Hartmann, G. D. K. Jr., J. B. Kruskal, R. E. Pastore, M. B. Sachs, R. D. Sorkin, and R. M. Warren. Classification of complex nonspeech sounds. Final report of the panel on classification of complex nonspeech sounds, National Research Council, Washington, DC, USA, 1989. National Academy Press.

[142] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer-Verlag, Berlin/Heidelberg, 1999.

[143] F. Zwicky and A. Wilson, editors. *New Methods of Thought and Procedure: Contributions to the Symposium on Methodologies*. Springer, Berlin, 1967.