Classification of abstract sounds for the application in sound design

MASTER THESIS

to obtain the academic degree Master of Arts (M.A.)

Sound/Vision

Claret Canelon

2338730

HAW HAMBURG | Faculty of Design Media and Information
Department of Media Technology

First supervisor: Prof. Thomas Görne

Second supervisor: MA. Stefan Troschka

Hamburg, 20.08.2019

Abstract

One crucial aspect in the process of sound design is the description of sounds for different applications like, for example, use in a sound collection. Traditional description systems use a series of labels or descriptors included in the metadata of each audio file. Typically, descriptors refer to the source of the sound (for example “door” or “bird”). However, such labels can constraint the description of abstract sounds or any sound that cannot be identified through its source or meaning.

The present study examines different attempts involving relevant taxonomic sound classifications, practical environments and scientific contributions that allow for the description of sounds independent from their sources. The results of this analysis are used to introduce a prototype of a description system specially designed for abstract sounds, illustrated by a practical example.

Keywords: Abstract Sounds; Sound Description; Sound Library; Metadata, Audio Postproduction

Zusammenfassung

Ein entscheidender Aspekt im Prozess des Sounddesigns ist die Beschreibung von Sounds für verschiedene Anwendungen, wie sie beispielsweise in einer Sound Library verwendet werden. Traditionelle Beschreibungssysteme verwenden eine Reihe von Begriffen oder Deskriptoren, die in den Metadaten jeder Audiodatei enthalten sind. In der Regel beziehen sich diese Deskriptoren auf die Schallquelle (z. B. „Tür“ oder „Vogel“). Solche Bezeichnungen sind nicht geeignet für die Beschreibung abstrakter Klänge, die nicht anhand ihrer Quelle oder Bedeutung identifiziert werden können.

Die vorliegende Thesis untersucht verschiedene Ansätze, unter anderem relevante taxonomische Klangklassifikationen, praktische Anwendungen und wissenschaftliche Beiträge, die es ermöglichen, Klänge unabhängig von ihren Quellen zu beschreiben. Die Ergebnisse dieser Analyse werden verwendet, um einen Prototyp eines Beschreibungssystems speziell für abstrakte Klänge vorzustellen, das durch ein praktisches Beispiel dargestellt wird.

Stichwörter: Abstrakte Klänge; Sound Beschreibung; Sound Library; Metadaten, Audio-Postproduktion

Content

List of Figures

Introduction

The recording and editing of sounds make it conceivable to store groups of them in the form of collections or libraries for creative and/or technical uses. Currently, sound collections are a standard tool used in sound design for films, videos, and games. A very important aspect of the organization of sound libraries is the description of each audio file. The usual description in digital sound libraries contains a series of *labels* or separate text *descriptors*, which are integrated into the metadata of every file. These descriptors function to retrieve the sound in every search engine that is able to read the metadata.

Sound descriptors are primarily words that describe a specific source, like *bird* or *car*. Although these kinds of labels are suitable for everyday sounds, a problem appears when one tries to describe or retrieve an abstract sound or "any sound that cannot be associated with an identifiable source"[1] (Merer, Ystad, Kronland-Martinet, & Aramaki, 2011, p. 2). The ambiguous nature of an abstract sound makes the selection of suitable descriptors (and therefore a successful retrieval) difficult.

In the past, several attempts, theories, and studies have been developed to classify sounds regardless of source. Currently, the use of onomatopoeias or words that seem to evoke an interaction over material is common. However, there is not a consensus about how these descriptors are used or written. Some approaches, such as the description through acoustic terms like: "high pitch with a slow decay", makes use of technical vocabulary that is limited to just a certain number of people. Other attempts, such as the description of sounds according to their timbral attributes, seem to combine a mixture of ambiguous terms for the description of a sound.

An alternative to describing an abstract sound is borrowing vocabulary from other sense modalities (Görne, 2017, p. 51). For instance, a sound can be described as big, small, bright or dark. The broad research in the field of crossmodal correspondences confirms that there are some common terms

[1] An extended definition of abstract sounds includes "environmental sounds that cannot be easily identified by listeners or that give rise to many different interpretations depending on listeners and contexts. It also includes synthesized sounds, and laboratory generated sounds if they are not associated with a clear origin (Merer, Ystad, Kronland-Martinet, & Aramaki, 2011). Some other labels include "ambiguous" sounds (Correya, 2017) or "confused" sounds (Bonebright, Miner, Goldsmith, & Caudell, 2005). In this work, we will adopt the term of abstract sounds.

(such as those related to size) that people use to describe an *object or event* in more than one sensory modality. For example, we associate “small” objects with high pitch sounds (Evans & Treisman, 2011). Consequently, the use of a term like “big” or “small” could be used to describe a perceived characteristic of a sound without identifying its source.

The diverse approaches to sound description reveal the complexity behind the concept of auditory domains. On the one hand, sounds refer to the source that originates them, while on the other hand they are perceived sensations that also refer to acoustic energy and motion. For these reasons, there is not a single category for describing all types of sounds in all their conceptual and application environments. Undertaking such an extended framework is not the purpose of this thesis.

However, still the question remains regarding how to describe a sound with an unidentifiable source? To be more precise: How can we describe abstract sounds in a form that could be applied to a sound design environment (for example, in a sound library)? How have the approaches to the sound description in this field been undertaken so far? Further, is it possible to develop a descriptive structure for abstract sounds? How should such a structure look in practice? This framework helps to limit the scope of this study and enable a more tangible subject of exploration.

Within this framework, it is important to clarify the conceptual limits that may appear during the exploration. In this sense, it is necessary to highlight how this process does not and will not consider inventing new descriptors or new words to describe a sound. Making use of the vocabulary that already exists in the various approaches to the auditory domain seems to be a conceptually lucrative direction for this investigation.

To provide answers to the questions outlined above, this study goes through a selection of approaches for sound description and analyses them in reference to the description of abstract sounds. This process aims to borrow the concepts and vocabulary that will allow us to develop a description system and a practical example. Such a description system is developed in the framework for the labeling of abstract sounds in sound design environments and includes an example of a sound library described through this structure.

Objectives and Structure

As previously mentioned, the principal objective of this thesis is to develop a description system for abstract sounds that can serve to organize sound design (for example, that can be used to identify abstract sounds in sound libraries).

To achieve this goal, it is first necessary to present the theoretical and practical background that precedes this proposal. The first chapter: *attempts to classify sounds*, review a range of approaches in the field of sound classification and sound description. The chapter also includes a brief discussion of the contributions and problems that these attempts present regarding the description of sounds with indistinct sources.

The second chapter: *crossmodal correspondences in sound*, brings clarity to the description of sounds using vocabulary that is related to other senses. In its second part, this chapter aims to create a hierarchical structure for the description of abstract sounds.

The third chapter: an *abstract sound library* contains the practical component of the thesis and outlines an example of the theoretical part of work: the description of a sound library through the classification structure proposed in the second chapter. It is expected to show the process and results of the application of the descriptors in the form of a metadata list. This list should serve as the basis for further observation and discussion.

The fourth and final chapter of this paper: *discussion and valuation*, discusses the practical example and the conclusion of this thesis.

I. Attempts to classify sound

It was not until the invention of sound recording that sound could be separated in time from the source that gave rise to it. Along with its technical development, recorded sound was stored on wax, plastic, printed on film and magnetic tape all before someone registered the concerns about how the auditory domain was transforming and consolidated these ideas in a fundamental thesis.

The first considerations came in a moment that recording was not only a medium to register sounds with fidelity. Rather, at the moment it was possible to manipulate them, to re-design them and re-contextualize them. With this realization, the sound was converted into an object of study—a sound object; a phenomenon independent from the time it was originated. It still could be related to events over materials. Sound was also a stimulus related to the perception of other senses. Finally, different descriptors were developed to describe sound from a variety of conceptual perspectives.

The following chapter goes through these relevant attempts to look for clues and contributions that could allow us to describe abstract sounds. First, each respective attempt will be presented, followed by a brief analysis regarding the description abstract sounds. The order of these approaches is conceptual and, in part, chronological. Conceptually, they start from the fundamental theory and continue to a more practical approach.

The first two attempts (*the sound as object* and the *sound as event*) correspond to theoretical backgrounds from taxonomic or schematic organization, the third attempt (*the practical approach*) involves approaches basically for the description of sound in two forms: the labeling in a sound database from the Internet and a lexicon with a hierarchical categories of descriptors.

The musical approach: the sound as object

The first attempt to classify a sound independent from its source arose in the field of the *musique concrete.* In the early 16^{th} century, Pierre Schaeffer published his work: *Treatment of musical objects*. Although he did not use the term abstract sound, Schaeffer was the first to treat sound as a *sound object*, or as a perceptual entity independent of its source and its significance. In order to hear the sound object,[2], it was necessary to develop a certain form of listening. Schaeffer differentiates three types of listening according to the attention of the listener, and referred to them as causal, semantic, and reduced listening.

By causal listening, Schaeffer refers to the common listening that concerns the relation of the sound with a source or its cause. In semantic listening, the listener focuses his or her attention on the message, or, in other words, "*die Analyse des kommunikativen Gehalt des Klangs, die identification seiner Bedeutung*" [the analysis of the communicative form of the sound, their identification with their significance] (Görne, 2017, p. 36). Both types of listening could be described as ordinary listening levels, as the listener's most common and immediate interest is the identification of the sound with its cause or the recognition of a message. However, if the focus of the listener is neither the source nor its significance, then it might be a case of *reduced listening* in which the listener concentrates on the attributes of the sound itself.

From this perspective, this theory suggests an abstraction hierarchy according to the interest of the listener that involves a concrete and an abstract level (Figure 1). On the concrete level, the sound is perceptively bound to the source or meaning, while on the abstract level a rupture from both aspects occurs. Schaeffer explains that the sound object does not contradict the source or the meaning, but the observation is implanted elsewhere, beyond these aspects.

The sound object is not an abstract sound but according to Chion *(Chion & Steintrager, 2016, p. 170)* the "*willful and artificial abstraction from the cause and meaning*" of a perceived sound. From this point of view, any sounds from any source including musical objects, phonetic objects,

[2] The term sound object is often debated since the word "sound" conveys an ambiguous meaning as mentioned in p.2. On the one hand, it refers to an effect or sensation, on the other hand to the source that originates it. Throughout this work, the use of the term "sound object" will be limited to refer to Schaeffer's theory.

industrial sounds, and/or birdsongs are sound objects that could be perceived in the frame of reduced listening.

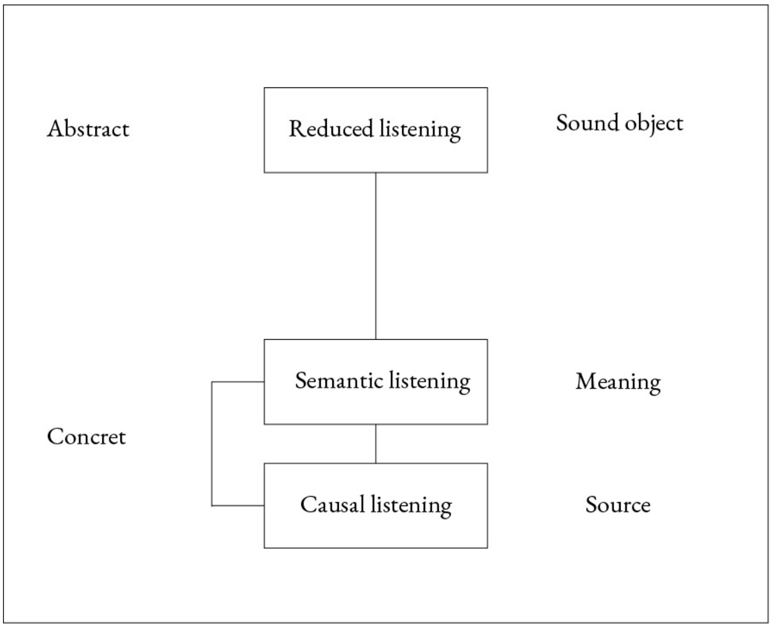

Figure 1 Abstraction hierarchy suggested by reduce listening. (Source: Author)

Schaeffer's treatment does not deny that sounds are also *events* that may be perceived in the form of reduced listening (called *natural listening*). This listening should concern the "repertoire of noises"[3]. For example, at the sound of an opening door, the listener's attention could be focused on one reduced element like the *creaking* of the door, rather than on the sound of its handle or the *squeak* of its *hinge*. But, in this example, our perception will tend to identify the noise with its cause because this noise is familiar to us. An exception occurs when hearing an unfamiliar noise, with no prior reference (an abstract sound). In this case, according to Schaeffer we will be able to perceive the noise as a sound object.

Therefore, Schaeffer finds a limitation to the progressive decoding of sound objects in natural listening, because it deals mainly with the *language of things*. In other words, natural listening leads to the identification of the sound's physical causes coming before the perceived effect. Hence, he outlined a more limited approach to find common criteria for all sound objects. According to Schaeffer, this approach should allow the discovery of the musicality of the sound object without the natural reference to its source. Thus, he designated this variant as "musicianly listening", which

[3] Schaeffer describes noises as "the result of an individual learning process in a group environment" (2017, p. 266).

refers to reduced listening abstractly enough to the level of the sound object and from here, he furthers his treatment through a classification of the sound object inspired in musical structures.

A summary of the modes of listening according to Schaeffer description (2017) is illustrated in Figure 2 in the form of an abstraction hierarchy:

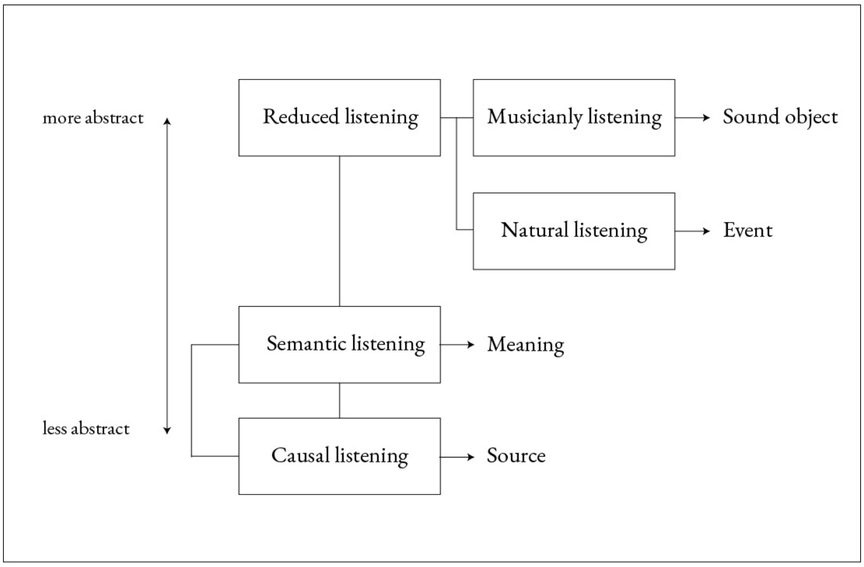

Figure 2 Listening modus und abstraction hierarchies. (Source: Author)

Schaffer presents a typological and morphological classification of the sound object. This classification results in a very complex and completely displayed form. A reduced overview of its form should be adequate this study

The typological aspect should function as general classification and concerns, in its basic description, the aspects of mass and fracture. The mass is described as "the manner in which sound occupies the field of pitch," while the fracture is "the way energy is communicated and appears in duration, closely linked to sustainment"(Schaeffer, North, & Dack, 2017). In this way, a sound could be classified according to how its mass changes over a merely perceptual duration.

The morphological classification in his summarized form concerns the criteria of *matter* and *form*. The matter criteria should classify the sound according to the mass, the *timbre* or the harmonic spectrum, and the grain or the qualitative perception of the microstructure of the object's "surface" equivalent to the haptic perceptual attributes. The form criteria should deal with the *dynamic*, or the intensity of the sound, in particular during the first seconds (attack) and the *allure* should correspond to the idea of vibrato. The variation criteria are added to matter and form to describe the *melodic profile* or the variation of the whole mass in terms of pitch and the *mass profile* or the internal variation of the mass given by the value of thickness. A summary of Schaeffer morphological classification according to Cano, Koppenberger et. al (2004) is shown in Figure 3:

<table>
<tr><td colspan="6">MATTER CRITERIA</td></tr>
<tr><td colspan="2">MASS
Perception of
“noiseness”</td><td colspan="2">HARMONIC
TIMBRE
Bright/Dull</td><td colspan="2">GRAIN
Microstructure
of the sound</td></tr>
<tr><td colspan="6">SHAPE CRITERIA</td></tr>
<tr><td colspan="3">DYNAMICS
Intensity evolution</td><td colspan="3">ALLURE
Amplitude or Frequency Modulation</td></tr>
<tr><td colspan="6">VARIATION CRITERIA</td></tr>
<tr><td colspan="3">MELODIC PROFILE: pitch variation type</td><td colspan="3">MASS PROFILE
Mass variation type</td></tr>
</table>

Figure 3 Schaeffer' Solfege of sound objects. (Source: Cano, Koppenberger, Celma, Herrera, & Tarasov, 2004)

At this stage, it is important to draw some observations from Schaeffer's work in terms of the contributions and limitations that this attempt has to pertain to the description of abstract sounds. Schaeffer's *Treatment of musical objects* provides a foundation for the development of a classification that treats sound independent of references to its physical source. This aim opens the path for the description of abstract sounds because they do not have an identifiable source.

However, the typo/morphological classification lacks a hierarchical structure, leading to ambiguity across different categories. In fact, this taxonomy is often debated because, among other critiques, it is not able to provide a clear outline for which category a particular sound should be categorized in when a greater number of reductive criteria are progressively considered. Additionally, this approach does not aim from the beginning to clarify the ways a sound object could be described (2017, p. 275)—that is, it is unable to provide semantic descriptors. Indeed, labeling a sound “as a fixed mass with unpredictable fracture” is not precisely an example of a clear description, at least not in the context of the sound design application. Therefore, this attempt exposes how important it is to use terms that are easily recognizable for the common listener.

Another difficulty appears when the temporal structure is implicit in the sound object, making it very problematic to delimitate the signal to study. The sound object could be as elemental or broad as the listener is able to sustain the *exercise in power* that supposes the reduced listening.

The reduced listening practice suggests that auditory objects could be classified according to their abstraction hierarchy, which depends, in this case, on the listener's interest and ability to abstract the cause and meaning of the auditory object. In this hierarchy, the sound object is placed in a prime position that could lead to causal notions (or not). This appreciation pushes one to consider if events or the "repertoire of noises" are in a position to provide a semantic description for abstract sounds or, in other words, to provide a description independent of the source. More details about these concerns are found in the next section: "The ecological approach: the sound as event".

The ecological approach: the sound as event

In the same year as *Treatment of musical objects*, James J. Gibson, a psychologist and researcher, published his work: *The senses considered as perceptual systems* (1966). Gibson introduces an approach that is categorized as *ecological* because it aims to observe sound not in terms of pitch, loudness or duration, but as a medium to provide information about our environment.

In his observations, Gibson points out the fact that the human ear can identify sources' material like water, wind, and solids and, moreover, the mechanical interactions associated with these materials. For example, we are able to identify the sounds of rubbing, scraping and rupturing and associate them with a solid material. Years later, these observations inspired William W. Gaver to create a framework for classifying and describing sounds in terms of their audible source attributes.

From here, Gaver outlined the proposal of a form of listening that he referred to as *everyday listening* or *the experience to hear events rather than sounds (Gaver, 1993)* and distances it from *musical listening* in which the attributes of the sound are related to the sound itself.

Gibson aimed to develop a descriptive and hierarchical structure for characterizing the auditory perception of events. For this purpose, he designed a system of classification for events according to the audible source material that involves vibrating solids, aerodynamic related to gases or air sounds and liquids. From here, he developed a classification and terminology considering the distinctive interactions over those representative materials that indicate an event. The following graphic (Figure 4) illustrates a summarized version of the classification.

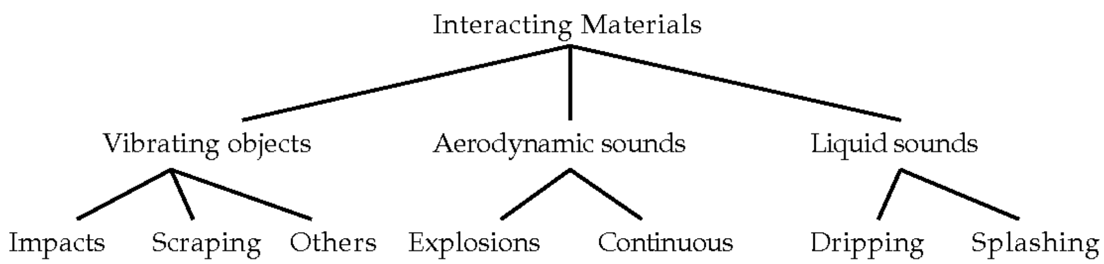

Figure 4 A hierarchical description of simple sonic events. (Source: Gaver, Willian W. 1993a p.14)

This classification is hierarchical; its whole structure is presented in a *map of everyday sounds* (Figure 5). The map outlines four levels of an event: a basic level, a temporal patterned one, a compound one and a hybrid one. The basic level displays relevant groups of actions that involve the basic source material, for example, by solids: deformation, impacts, scraping and rolling. The temporal patterning level is a more descriptive array derived from the first level. For example, the descriptors bouncing, breaking and walking involve a pattern of solid impacts. With compounds, an event that involves more than one sort from the basic level is evident, for example, the bowling effect implicates a rolling as well as an impact event. At the hybrid level are those events in which more than one basic sort of material is involved, for example rain on a surface constitutes a liquid (rain) with a solid (surface).

Gibson was aware of the effect on the soundwave that supposes the interaction of a material. Nevertheless, the framework of everyday listened is oriented to qualitative attributes over other parameters such as pitch or loudness, phase or duration, which Gibson refers to as *primitive components*. From his perspective, there is enough rich information in the world that our descriptions should not be limited to primitive physical dimensions.

The typologies presented by Gaver play a role in the practical application of sound descriptors, his contributions and limitation in this field are following exposed:

Hence, the main contribution of Gaver's typology is how it reveals the sense of materiality and interaction that can be perceived via a sound. However, it does not question the event outside the real source or environment. Especially at the *hybrid level*, in which more specific information about the source is revealed, this classification could result in causal notions

This classification system in the form of a typology can provide suitable and understandable vocabulary (at least in the English language) for the semantic description of events. This happens primarily via the terms placed in the temporal patterning level such as breaking or bouncing, which integrated the descriptors for *source-ambiguous sounds* of the Audio Set Ontology (Gemmeke et al., 2017), a project based on human-labeled audio for the description of audio events.

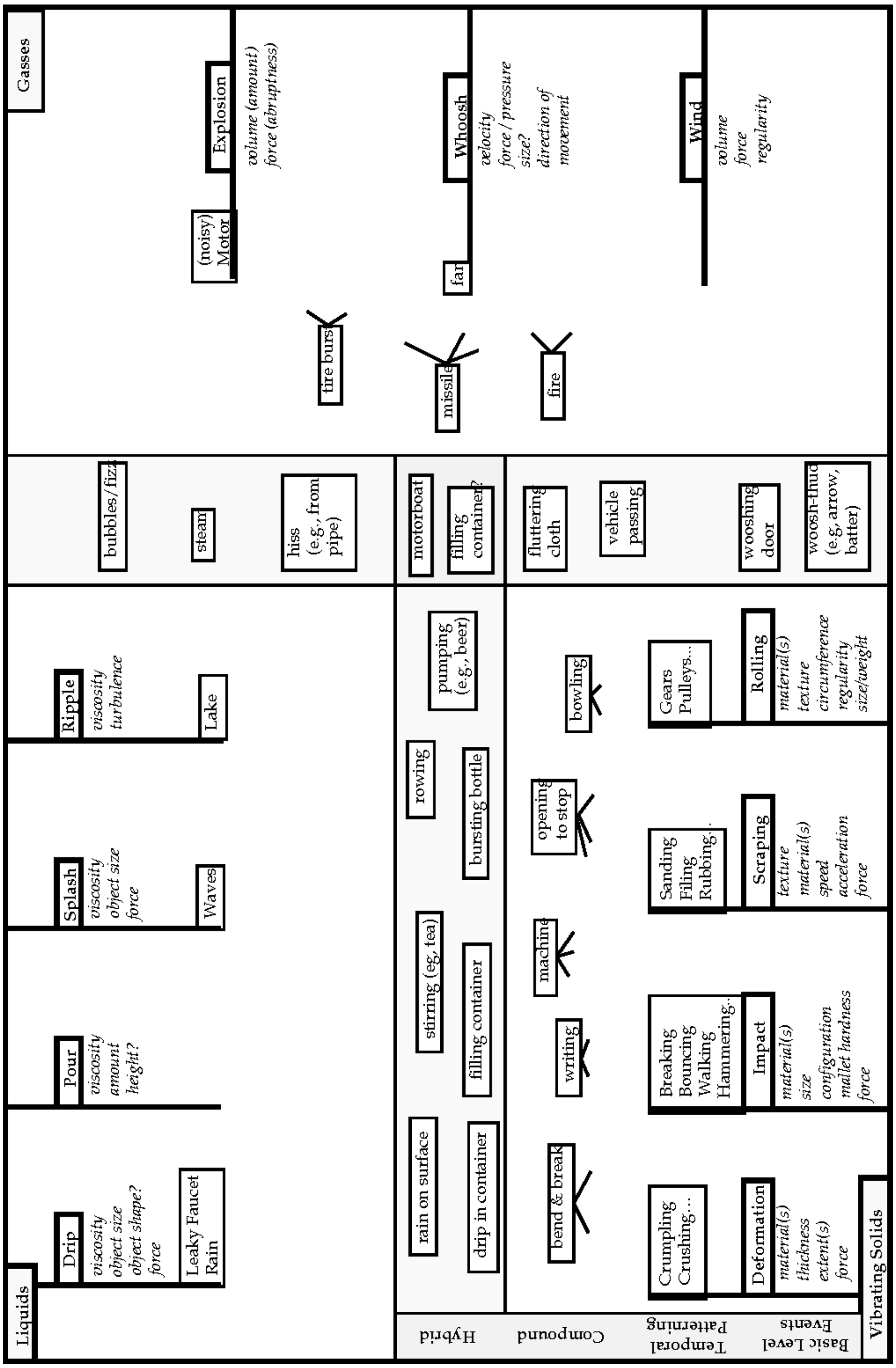

Figure 5 A map of everyday sounds. (Source: Gaver, Willian W. 1993a p.15)

Description from the Author: Three fundamental sources (vibrating solids, liquids and aerodynamics) are shown in the three overlapping sections of the figure. Within each section, basic sound-producing events are shown in bold, and their relevant attributes next to them in italics. Complexity grows towards the center of the figure, with examples showing temporally patterned, compound, and hybrid sounds.

Some researchers consider Gaver's classification to have a limitation for the description of abstract sounds (Correya, 2017). One argument, for instance, is that two physically different events can produce perceptually similar sounds, as the Foley recording practice exposes. For example, it is possible to record the sound of crumpling a cigarette paper to simulate the creaking sound of a fire. On the other hand, the fact that *abstract sounds do not convey univocal meaning* could be seen as an advantage because it makes it possible to use them for diverse purposes *according to the aim of the experience (Merer et al., 2011).*

The practical approach

Meanwhile, in the field of practical application, several alternatives have been developed to classify and provide descriptors for sounds independent of their source. Most of them are established according to the needs in each respective field of study or sound category; other attempts aim to cover common perceptual characteristics to provide a shared vocabulary for users, engineers, etc.

One of these approaches is classification according to the timbral attributes of a sound. Research in this field has been gaining space in the field of quality description and sound retrieval (Correya, 2017). A recent study from (Pearce, Brookes, & Mason, 2017) identifies the timbral descriptors implemented in numerous studies and compares them with the frequency of use in the collaborative database of sounds *freesound.org*. The following graphic should illustrate the 40 most frequently terms associated with timbral attributes adopted by users to retrieve a sound:

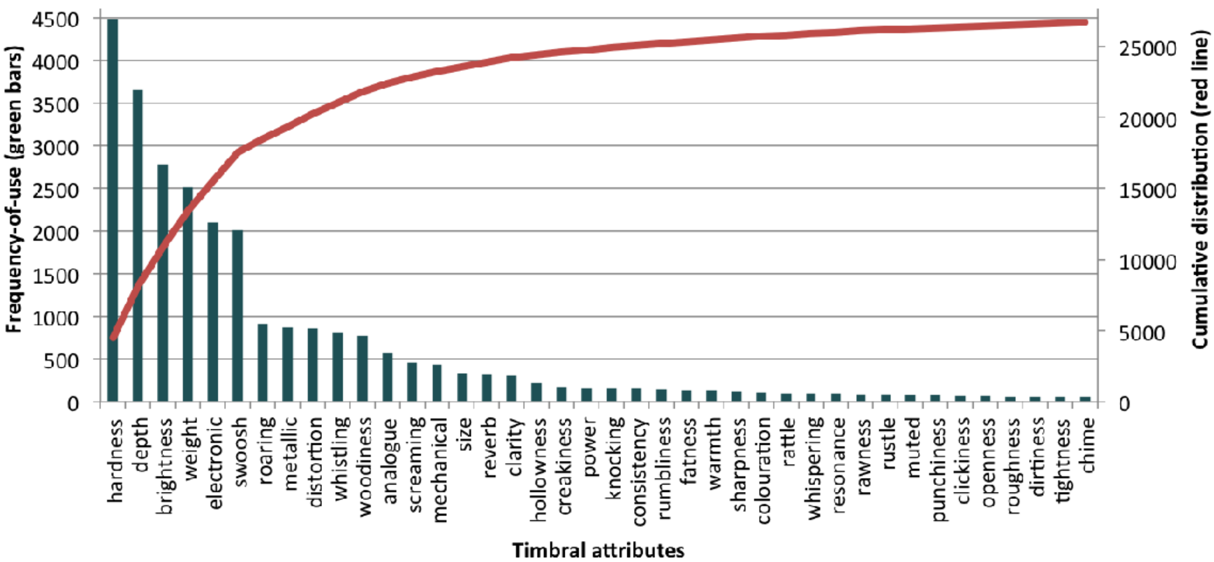

Figure 6 Frequency-of-use and cumulative distribution for the 40 most frequently searched timbral attributes used in freesound.org. (Source: AES Conference on Semantic Audio, Erlangen, Germany, 2017 June 22 – 24 p.7)

Looking at the graph above, it is noteworthy that the presented attributes refer to multiple attributes' categories; some terms entail words used to describe attributes from other senses such as size or brightness. Other words express an interaction of material like swoosh[4] or rattle, whereas others still seem to carry overly connotative or ambiguous associations such as power or dirtiness.

[4] Notice that swoosh may also carry an onomatopoeia connotation without consensus in the form is written.

Certainly, timbre is one of the most, if not the most, debated sound attributes as it is not a homogeneous concept (Chion & Steintrager, 2016, p. 174). Timbre seems to involve an imprecise series of sound attributes without a hierarchical distinction and is open to almost all descriptors with a few exceptions like loudness or pitch. Although classification according to timbral attributes seems to cover a large number of descriptors, it lacks categories and delimitations, making it worthwhile to explore other attempts that could provide a schematic organization.

Such an approach is found in a previous report from (Holm Pedersen, 2008) called "The Semantic Space of Sounds". In this case, there is a "Lexicon" of sound is presented as a "tool", which can convey a semantic distinction. The report starts with the idea of dividing sound descriptors into three broad categories of words: those that refer to the perceptual attributes of a sound, those that refer to the source, and those that are associated with a sound. In particular, the focus of the Lexicon lies in the words related to the perception which are further divided into perceptual (auditory) characteristics, affective characteristics and connotative characteristics.

The lexicon in its complete semantic classification presents the following classes:

1) Direct sound descriptors (e.g. loud, bassy, shrill)

2) Words relating to perceptions from senses other than hearing (e.g. bright, dark, colorless)

3) References to events and sound sources (e.g. howling, roaring, rattling)

4) Changes or differences in perceptions (e.g. colored, compressed, muffled)

5) Affective responses to sounds (e.g. pleasant, annoying, boring)

6) Connotative associations (e.g. sporty, luxurious, powerful)

7) Onomatopoeia (e.g. woof-woof, yap-yap)

8) Attributes

The classes 1-4 adhere to the words related to the first category or *perceptual* characteristic. Classes 5-6 include *affective* and *connotative* characteristics. Class 7 contains words that imitate sounds. Class 8 should contain a range of words in the form of substantive characteristics (e.g distance). Descriptors are primarily adjectives (Dark, bright). According to Pedersen, adjectives may be used as word anchors for scaling the attributes.

The semantic organization of the lexicon suggests an abstraction scale that is more reduced in the ways it goes up. Pedersen states that class one and two are marked by fundamental attributes and mention that words in class three and four may be described by words from the first two classes. Likewise, words in groups 5-7 may be described by words from the first four classes.

The idea of the lexicon, according to its author, was to provide a guide for finding, defining and scaling auditory attributes. Indeed, there is much to be developed related to the definitions and categorization because some words seem to convey different classes. For example, the descriptors “sharp” and “cold” are not except for indistinctness. Sharp is considered by the lexicon to be an attribute outside the terms related to other senses. However, from a semantical point of view, it could be related to a haptic sensation. Other approaches are needed, especially in the category of sounds related to other senses.

It is precisely in this category that a series of scientific studies have come into being. Some studies explored the relationship between auditory, visual and haptic sensory attributes to describe the same object or event. The studies are framed in multisensory integration, specifically in the field of crossmodal correspondences. The aim of going through these crossmodal correspondences is to contribute to a less unambiguous description in this category. The next chapter has the purpose of deep in this theme.

II. Crossmodal correspondences in sound

Preliminary Concepts

It is possible to borrow words from other senses to describe a sound. In particular, these words are useful at the moment to describe an unfamiliar noise or an abstract sound when one lacks a point of reference to a source. In such situations, a sound can be described as bright, small, etc. with a perceived attribute of the sound. These associations are a standard topic of research in the field of crossmodal correspondences.

Studies of crossmodal correspondences reveal that the brain selects and correlates information from different senses to identify an object or event among many others in our field. This correlation should contribute to a clearer representation of our environment and should also improve the speed of detection of objects and events to modulate our response to them.

Crossmodal correspondences do not evaluate the sound according to its source nor based upon the attributes themselves, but as an auditory perception that concerns with information from other sensory inputs. To say that a sound is small or big is just the semantic expression of an exchange of perceptual information that has already occurred, for example between the attribute of pitch and the visual size of an object. This association of features between senses provides a vocabulary that also describes sound attributes. But what does it mean to describe a sound as bright or dark, round or sharp? And what are the concerns of using these correspondences as sound descriptors?

On this point it is important to consider the current situation concerning the theme: on the one hand, recent and widely addressed studies on crossmodal correspondences reveal a growing interest in the subject from a scientific approach. On the other hand, as was observed in the last chapter, some crossmodal associations have already been adopted in the practice for sound descriptors in different categories. Some of them integrate a blend of attributes that conform to the field of timbre, while others have been included in an abstraction scale between the words used for direct sound description and words that refer to an event.

Throughout this chapter, the focus will lie on crossmodal correspondences that deal with auditory perceptions, chiefly those presented between auditory and vision[5]. The subchapter: *Crossmodal correspondences as sound descriptors,* aims to explore how words from the other senses match to the characteristics of the sound. This will be done first from a scientific point of view and then complemented with observations from practical application. This is expected to contribute to a less arbitrary selection of descriptors for the second part of this chapter: *a classification for abstract sounds* that concerns the creation of a structure of classification for abstract sounds that integrates the vocabulary of crossmodal correspondences.

Crossmodal correspondences as sound descriptors

Size: Big and Small

The attribute of visual size is correlated with the perceived pitch of a sound. According to research, people associate smaller objects with high pitch tones and larger objects with low pitch tones (Gallace & Spence, 2006). Pitch and size are a natural association in terms of the real relation of attributes in our environment: the larger an object is, the bigger its resonance cavity and the lower its resonant frequency. Thus, the descriptor "small" is a consistent label for a high pitch sound, while the descriptor "big" is appropriate for a low pitch one.[6]

Brightness: Bright and Dark

Brightness is metaphorically correlated with the perceptual pitch of a sound. Studies reveal that people match high pitch tones with brighter surfaces. Additionally, it has been observed that the perceived loudness can also interact with the pitch in the perception of brightness (Marks, 1987). However, the aspect of pitch seems more decisive than loudness in this association.

Conversely, darkness is a relevant attribute, according to the lexicon presented by Pedersen. Darkness is defined in the lexicon as " *the sensation produced by low-frequency components in*

[5] Mainly crossmodal associations can be found in this field. (Spence, 2011)

[6] The fact size is a natural association could explain why Chion considers the notion of pitch inappropriate for the description of sound since they should reveal behind their false testimony, causalist notions (Chion & Steintrager, 2016, p. 173). Previously, Shaeffer enounced the association between mass and the attribute of pitch.

sounds". (Holm Pedersen, 2008). As adjectives, "bright" and "dark" represent two poles of the same scale related to the perceived pitch of a sound.

Shape: Round and Sharp

The pitch of a perceived sound could awake metaphoric associations with the shape of an object. High-pitched sounds are associated with sharp visual images and low-pitched tones with smooth rounded forms (Karwoski & Odbert, 1938). Sharpness is included as an attribute in the lexicon presented by Pedersen (2008) and is defined as the sensation produced by high-frequency components in sounds. The sharpness sensation is based on the relative balance of the sound spectrum, independent of the fine structure. (Holm Pedersen, 2008, p. 15)

However, sharpness could be a metaphoric perception and semantic related to the perception of a metallic material.

Dry and Wet

The association between a perceived sound and the perception of texture has been studied in several studies related to the perception of food (Chauvin, Younce, Ross, & Swanson, 2008). Although the words dry and wet are not used as descriptors in these studies, the dryness of food is associated with attributes like crunchiness or crispness. In every case, the textural property is seen to be associated with the perceived fracture information that the sound conveys.

The words "dry" and "wet" could lead to a semantic ambiguousness in the sound description, because wet and dry are poles used to designate the level of reverberation from a sound in technical language.

Elevation: Up and Down

The direction of movement has also been correlated with the notion of pitch. According to some studies, a high pitch is associated with a higher position in space and lower pitch with lower positions in space. (Clark & Brownell, 1976). Additionally, changes in the pitch can affect the perception of visual motion (Maeda, Kanai, & Shimojo, 2004). In this sense, we associate upward movement with an ascending pitch and vice versa.

Therefore, descriptors in this category include: upwards, downwards, up and down.

Creating a hierarchical description for abstract sounds

The concern of this subchapter is the creation of a structure that allows us to describe abstract sounds in sound design applications, particularly in sound collections. The structure presented is a hierarchical classification of descriptors. The main category is the descriptors from crossmodal correspondences or descriptors from other senses, the other categories are based on the approaches reviewed in the first chapter including Gaver (1993) and Pedersen's Lexicon (2008).

As previously exposed in the first part of this chapter, descriptors from crossmodal correspondences could be related to two categories: a) primary components like pitch or loudness as well as b) events. These associations suggest that crossmodal correspondences descriptors could be a midpoint between these categories in an abstraction hierarchy. Such a hierarchy implies that the descriptors from a subordinate category can be seen as broadly cross-referenced with descriptors in the upper category. For example, the word "crispy" that designates an event (fracture) could be described as a "bright" sound in the upper category. The following figure illustrate this structure in its basic form:

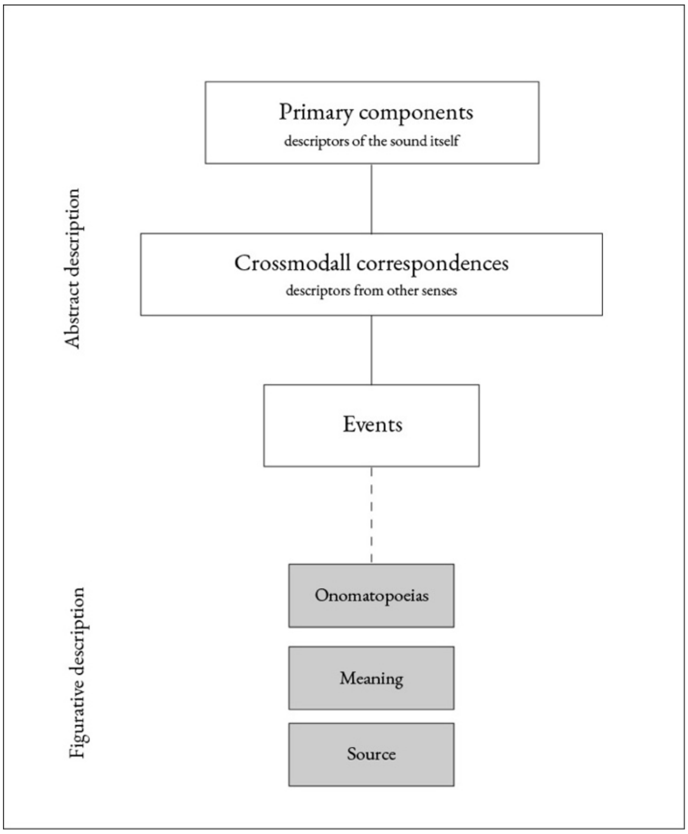

Figure 7 A hierarchical categorization of sound descriptors. (Source: Author)

Note: Below to the event s and in grey color are found the categories that allow a more figurative description related to the source. These categories might not be suitable for the description of abstract sounds but are placed in this illustration as a reference

In some cases, this "midpoint" of crossmodal descriptors can also represent the upper category in this structure. This is because words that refer to sound themselves like consonant, discordant, disharmonious, glide, or tremolo found in the Pedersen Lexicon (2008) are words that have been developed for musical description. However, "*musical sounds are not representative of the range of sounds we normally hear" (Gaver, 1993, p. 2)*. Hence, descriptors of sound itself could be too *reduced* for the description of non-musical sounds. Some other descriptors in this category (like boomy, high pitch, or balanced) make use of a technical vocabulary from a specialized field beyond sound design like, for example, sound quality description.

On this point, we might discriminate the common language descriptors from the technical vocabulary. The main difference is the use of *Jargons* in the technical language. Jargon is defined as words and phrases used by particular groups of people that are not generally understood. Mainly sound descriptors from the sound itself seem to belong to technical language. There is a natural resistance to use descriptors in this category. Certainly, the description of a sound in itself is a relatively new phenomenon because *sound (until the recording was created) has always been linked in time to the energetic source that gave rise to it (Schaeffer et al., 2017, p. 51).*

In the case of an event's descriptors, the border between technical and common language is not precisely outlined. Events descriptors are generally understood[7], but they specify precise phenomena. A creaking sound is clearly differentiated from a rubbing sound. According to Chion (2016), the use of this verbal precision is an essential means of refining and cultivating perception and could be used to empower listening because they allow us to hear not the object, but the objects of our sensations.

It is precisely the "*object of our sensations"* that concerns the description of abstract sounds[8], as the unambiguous recognition of a source is not possible. Thus, this structure does not originate from an ecological approach in which the event leads us to the identification with a real source in our

[7] at least in the sound design field.
[8] of course, it could happen that the perceived effect cannot be described more than through the association with a source. Every sound including an abstract one is not completely free of figurativeness. However, this association would bring us to the same problem that motivates this work.

environment. In this sense, an event could evoke some metallic characteristics without necessarily revealing a familiar metallic source: it could comes from a metallic gate but also a synthesizer.

Thus, the perceived phenomena in events could refer to a sense of materiality as well as a sense of interaction. As previously observed in the first chapter (see p.9), events can convey information about both attributes. However, the relation material/interaction is not equal in every auditory event. While some events seem to convey more information about an interaction, some others convey more information about a material. In this regard, the classification could then be expanded by dividing events descriptors into two categories: those that refer to the material and those that refer to interaction over materials.

The following figure should serve as an illustration for this expansion:

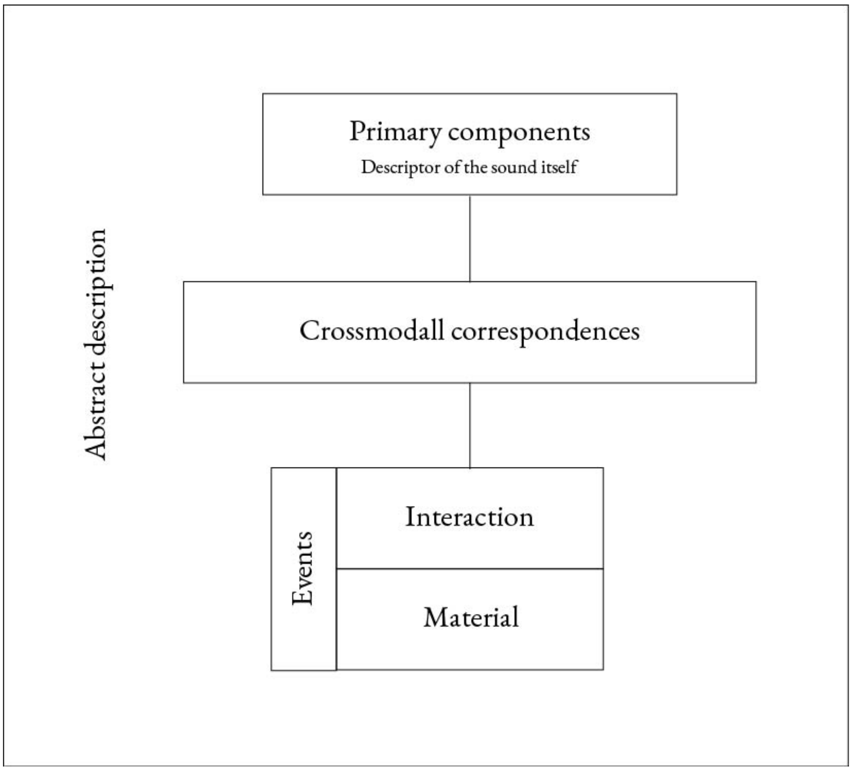

Figure 8 A hierarchical categorization of descriptors for abstract sounds. (Source: Author)

Note: Events descriptors are here divided into those that refer to interaction and those that refer to the material. The relation to a source is at this point unnecessary.

Hence, how is this structure expressed in the practical application in sound design? For example, in the description of a sound library? Such a description is expected in the Metadata of each audio file to be formed in the following principle:

Primary component (s) (if relevant), Crossmodal descriptor (s), Interaction descriptor(s), Material descriptors (s)

Figure 9 Overview of a metadata structure for the description of abstracts sounds. (Source: Author)

In the last figure, every category descriptor includes an (s). This means that different words from the same category could be added to provide more information. Additionally, it is appropriate to cover the description across the different hierarchies. The reason for this is the advantages that could lie in the moment of recalling a sound through the Metadata. Depending on which category the descriptor is associated, the search might be more general or specific. It is to be expected that the higher the category to which the descriptor belongs, the more general and broader the search will be, and vice versa. This aspect will be discussed in greater detail after the experimental application of this structure in the next chapter.

Another concern that can be discussed, for example, is why connotative or affective category descriptors are not involved in this structure. In this sense, it is important to consider that, for a search engine that can read audio file metadata, descriptors are visually presented before or at the same time that the sound is reproduced. In this condition, affective and connotative descriptors could strongly influence perception, thereby leading to an expectation. This can limit the creative use of the sound.

The next chapter, *Description of abstract sounds in a sound library*, aims to apply this structure to a practical example.

III. Description of abstract sounds in a sound library

Preliminary concepts

Having established in the last chapter a hierarchical structure for the description of abstract sounds, the goal of this chapter will be to illustrate this structure by way of a practical example. The description will be applied in a sound library of abstract sounds. First, however, we will present some considerations about the sound files that are integrated into this sound library. Later, we will present a simple model as a guide for a description of every audio file and apply this description. Finally, we will present a metadata list of the sound library containing all the descriptors for every file. This metadata will be used as material for the discussion in the last chapter.

The first consideration of the sounds to be used in this practical example relates to the type of library itself and how the sounds are classified.

Different studies, including Murray Schafer (Schafer, 1993), borrow the notion of "*figure and ground"* from Gestalt psychology, and apply this concept from an environmental perspective by studying the "soundscape." Schafer describes the figure as the focus of interest and the ground as the context. Concerning the relation between them, he describes the figure as a signal or *soundmark*, and the ground as the ambience around the figure.

In the practice of sound design, primarily sound collections are differentiated into two general categories: isolated sounds and ambient sounds. The libraries containing isolated sound or "figures" correspond in practice to the "effects libraries", being differentiated from the "ambience libraries". Each type of library's effect or ambience could have a different approach to technical parameters like loudness, duration, and channels. The temporal structure and/or duration of the file is especially distinct. For example, these isolates sound or effects are shorter in duration than those that are conceived as background or ambience. Therefore, it was important from the conception to deliberate about this aspect to choose the material to be described in the practical example.

In this work, the samples of this experiment can be considered as sound effects—that is, as foreground sounds and not as ambience sounds. They are short clips with a duration between 1 to

4 seconds. The main reason for this was that a more definite shape with less duration would be less susceptible to the changes to ambience sounds across its duration.

A guide for a hierarchical description

In the previous chapter, we presented a hierarchical classification of categories for the sound description (see Figure 8 and Figure 9). In this section, the goal is to develop a guide that allows us to classify abstract sounds following the structure previously revealed in a simple manner. This guide should serve as a reference for the application in the exemplar of a sound library of abstract sounds. A selection of descriptors was added to the guide. These descriptors were compiled from the typologies of Gaver (1993) and the Pedersen Lexicon (2008).

In the center, were added those descriptors from other senses that were already defined in the second chapter (see p.17). In all of the categories, the field "others" was established as an open field for an alternative descriptor in the corresponding category.

File name: ______________

Primary components
descriptors of the sound itself

Loud	Fast	Intense	Sweeping
Soft	Wide	Pulsating	Other: ______

Crossmodal Correspondences
descriptors from other senses

Small	Bright	Sharp	High	Upwards	
Big	Dark	Round	Low	Downwards	Other: ______

Interaction

Deformation	Impact	Scraping	Rolling	Splashing	Wooshing	Hissing
Chrushing	Breaking	Rubbing	Bowling	Driping	Fizzing	Rustling
Crumpling	Bouncing	Sanding	Purring	Bubbling	Chirping	Crashing
Creacking	Hammering	Gurgling	Clattering	Squeaking	Buzzing	Other: ______

Material

Liquid	Gas	Electric
Solid	Metalic	Other: ______

Figure 10 A guide for the application of metadata of abstract sounds. (Source: Author)

As is clear from the graphic, not all of the descriptors found in the attempt are displayed in this basic guide. Some of those that are not shown include the descriptor f sound itself that seems to belong to other categories like rough or airy, as well as those descriptors that seem to be primarily for musical descriptions like tremolo or vibrato.

However, the omitted descriptors in the corresponding category are not excluded in the practice if they are relevantly characteristic of the sound. For this basic form, the decision to omit is not arbitrary. A smaller number of descriptors with an open field correspond to a more manageable quantity during the practical workflow.

Application of the description

The following chapter should serve as an example for the first experimental application of the hierarchical description system for abstract sounds, in line with the system already outlined (Figure 8 and Figure 9). For this example, the former guide for the application of metadata was used as a reference for possible labels for the description.

For this attempt, the sound collection contains 50 audio files with short-term abstract sounds. It was essential to note that the files did not have any metadata description before the application of the descriptive system. The reason for this was to avoid the former labels' perception of the audible object affects the sound description. Given this condition, it was not possible to work with sound files from established sound collections or sound databases on the Internet, as that would require retrieval through a description (and hence, the files would already have a metadata description).

Accordingly, the samples selected for this application came from two sources:

a) Already produced abstract sounds. *(20 files)*
b) Freely designed sound effects produced in a studio for this study *(30 files)*

The files were all short-term sounds with different levels of complexity, some of which presented a more complex structure or a more significant change over time compared to others. In the selection, some sounds presented more synthetic characteristics (similar to sinus waves—which they were not) compared to others. However, all of them fit within the definition of abstract sounds (see p.1)because they could not provide a clear origin and meaning.

The description was applied in an attempt to cover all categories and the maximal number of words in the same category for each file. However, it was not mandated that a category be used if they were not able to provide relevant descriptors for the sound.

In order to get greater reliability in the description, each sound was listened to more than one time: from 3 to 5 times, and with and silent pause between 2 and 4 minutes before the next file was reproduced.

The figures below show the results in the form of a metadata list:

RecID	Filename	Description	Duration
1	CC_ABS_01.wav	Rough, Dry, Cracking, Crushing, Dropping	0:02
5	CC_ABS_02.wav	Rough, Dry, Cracking, Crushing, Rustle	0:02
7	CC_ABS_03.wav	Bright, Downward, Electric	0:02
2	CC_ABS_04.wav	Loud, Sharp, Round, Splashing, Liquid	0:01
51	CC_ABS_05.wav	Sharp, Bright, Clattering, Metallic	0:02
3	CC_ABS_06.wav	Round, Downward, Purring, Bubbling, Gas	0:02
4	CC_ABS_07.wav	Fast, Bright, Sharp, Crunchy, Creaking, Electric	0:01
8	CC_ABS_08.wav	Sweeping, Bright, Sharp, Fizzing, Hissing, Gas	0:02
9	CC_ABS_09.wav	Soft, Sweeping, Round	0:04
10	CC_ABS_10.wav	Intense, Loud, Bright, Sharp, Fizzing, Hissing	0:07
11	CC_ABS_11.wav	Bright, Sharp, Crispy, Crumpling, Rustling	0:02
12	CC_ABS_12.wav	Sharp, Crushing, Crashing, Rustling, Splashing	0:04
13	CC_ABS_13.wav	Soft, Small, Sharp, Crushing, Crashing, Rustling	0:01
14	CC_ABS_14.wav	Soft, Fast, Small, Round	0:01
15	CC_ABS_15.wav	Small, Crispy, Rustling, Rubbing	0:01
16	CC_ABS_16.wav	Wide, Bright, Sharp, Hight, Hissing, Fuzzing, Gas	0:04
17	CC_ABS_17.wav	Pulsating, Bright, Hight, Spinning, Electric	0:04
18	CC_ABS_18.wav	Fast, Soft, Small, Clattering, Purring	0:01
19	CC_ABS_19.wav	Fast, Soft, Small, Round, Purring, Bubbling	0:01
20	CC_ABS_20.wav	Rough, Dry, Scraping, Rustling	0:02
21	CC_ABS_21.wav	Soft, Fast, Small, Bright, Spinning	0:01
22	CC_ABS_22.wav	Soft, Fast, Small, Round, Bouncing, Squeaking	0:01
23	CC_ABS_23.wav	Soft, Fast, Round, Bubbling, Liquid	0:01
24	CC_ABS_24.wav	Rough, Sharp, Impact, Breaking, Creaking, Cracking	0:01
25	CC_ABS_25.wav	Fast, Sharp, Bright, Whooshing, Electric	0:01
26	CC_ABS_26.wav	Fast, Small, Sharp, Bright, Electric	0:01
27	CC_ABS_27.wav	Fast, Sharp, Down, Scraping, Rubbing, Fizzing	0:01
28	CC_ABS_28.wav	Pulsating, Sharp, Electric	0:02
29	CC_ABS_29.wav	Pulsating, Small, Bright, Hammering, Chirping, Electric	0:01
30	CC_ABS_30.wav	Rough, Sharp, Electric	0:01
31	CC_ABS_31.wav	Fast, Hight, Small, Bright, Spinning, Chirping, Electric	0:02
32	CC_ABS_32.wav	Fast, Small, Round, Splash, Bubbling, Liquid	0:01
33	CC_ABS_33.wav	Sweeping, Small, Round, Whooshing, Spinning	0:02
34	CC_ABS_34.wav	Fast, Small, Chirping, Electric	0:01
35	CC_ABS_35.wav	Sweeping, Up, Rustling, Spinning, Electric	0:03
36	CC_ABS_36.wav	Soft, Small, Round, Rubbing, Bouncing	0:03
37	CC_ABS_37.wav	Fast, Sweeping, Clattering, Electric	0:02
38	CC_ABS_38.wav	Small, Round, Bubbling, Spinning	0:03
39	CC_ABS_39.wav	Fast, Down, Dripping, Crumpling, Scraping, Rustle	0:01
40	CC_ABS_40.wav	Soft, Rough, Crumpling, Breaking, Cracking	0:02
41	CC_ABS_41.wav	Fast, Downward, Sharp, Bright, Electric	0:02
42	CC_ABS_42.wav	Soft, Fast, Rough, Small, Spinning	0:01
43	CC_ABS_43.wav	Fast, Bright, Creaking, Bouncing, Spinning	0:01
44	CC_ABS_44.wav	Soft, Dark, Dripping, Breaking, Creaking, Cracking	0:04
45	CC_ABS_45.wav	Fast, Bright, Small, Clattering, Spinning	0:02
46	CC_ABS_46.wav	Fast, Round, Small, Rubbing, Spinning	0:02
47	CC_ABS_47.wav	Small, Round, Dripping, Bubbling, Rolling, Liquid	0:02
48	CC_ABS_48.wav	Fast, Sweeping, Rough, Up, Fizzing, Hissing, Whooshing	0:02
49	CC_ABS_49.wav	Rough, Fast, Sharp, Crumpling, Dropping, Fizzing, Hissing	0:02
50	CC_ABS_50.wav	Small, Bright, Chirping	0:03

Figure 11 Metadata description of abstract sounds library.

IV. Discussion

This section aims to discuss the results of the descriptive system for abstract sounds constructed in the last chapter, and present additional considerations for its practical application. The first concerns are given by observing the metadata list presented in the last chapter.

As shown in the metadata list (Figure 11), each file contains more than one category of descriptor in the metadata. However, the number of descriptors differs widely from one file to another: while some files contain just three descriptors, others have more than five. The second observation is that some sounds contain event descriptors that are associated with dissimilar materials. For example, in the following description, while splashing is associated with a liquid material, crashing and rustling are related to solid materials:

CC_ABS_12	Sharp, Crushing, Crashing, Rustling, Splashing

Figure 12 Sound description with dissimilar events

The explanation of both these observations has to do with how complex the information that a sound conveys is. Here, it is important to note that, although abstract sounds cannot provide specific information about their causes or context, they can still convey other kinds of information.

Regarding the number of descriptions[9], those sounds that suggest more changes over time contain a larger description or number of labels with less variation. The example is consistent with the statement that auditory information is concealed with the changes that the sound has over a period of time (Görne, 2017, p. 26). In this case, sounds with more variation over time can carry a significant number of descriptors, and vice versa.

The second observation confirms a position previously mentioned—namely that abstract sounds do not convey unequivocal meaning (see p.12). Hence, they can have different descriptors that, at first glance, may seem to be incongruent[10]. This distinction is not necessarily a defect description or misinterpretation[11], at least in the context of the sound design environment. On the contrary, different possible interpretations allow for a broader spectrum of applications, according to the

[9] Our practical example
[10] That is not the case in the traditional description system.
[11] Gaver consider this "misinterpretation" as an illusion of the sort exploited by Foley artists creating sound effects (Gaver, 1993, p. 14)

judgment of the sound designer and listener. It is also interesting that the sounds with more synthetic characteristics (similar to sinus waves) were not able to provide ambiguous information about the material and were in most cases described in this category as "electric" material.

Other observations from the last description include the number of files that each category involves. In this respect, the category that concerns the major number of files was the descriptors from crossmodal associations. In this category, all 50 files contain borrowed words from other senses. Meanwhile, 44 files contain events descriptors, 32 files have descriptors from the sound itself, and 15 files use descriptors that referred to a material. During the process of labeling, in the open field (others) the label *rough* was added as an additional descriptor related to other senses due to the association between the perceived sounds with the haptics attribute (texture). Some other additional descriptors include s*pinning* and *chirping* by events category.

As mentioned in the last chapter (see p.22), it was expected that by applying this structure it would be found that the higher the abstraction hierarchy the category belongs to, the larger the number of files that could be associated with it. The results reveal that the position in the hierarchy is not necessarily a relevant factor for the association with a larger number of files if the descriptors that describe the sound itself is used as the higher position in the scale. However, crossmodal descriptors are associated with primary components like pitch or loudness as well as event descriptors (see p.19) and it is not surprising that, when both categories were involved, they could describe a broader range of sounds.

Due to their wide-ranging application, descriptors from crossmodal correspondences can be useful in this structure as a starting point for sound design in audiovisual projects. Mainly descriptors of this category are associated with information from the visual domain like brightness, size or sharpness. The fact that they can involve a major number of files pro descriptors can be useful if the user aims to search for unexpected sounds in the library. Otherwise, they could be combined with other descriptors from other categories for a more specific retrieval[12]. For the enlargement of

[12] Normally search engine allows to retrieve the sound with multiple descriptors by adding a comma (,) between every descriptor or the word "and" for example "bright, crushing" or "bright and crushing". This action will limit the retrieval of the files in those that contain both (or more) descriptors. Such an action can be useful for a more reduced retrieval with a smaller number of displayed files.

descriptors in this category, further investigations are crucial to identify the relationship between auditory input and other senses that can provide a vocabulary to describe sounds.

Some variables in the practical workflow are appropriate to consider the proposed hierarchical description structure. In this sense, here are three variables to be discussed. The first two correspond to variables found in the relationship between sound designer and user, the third variable corresponds to the information given by the duration of the sound file itself.

Concerning the first variable, it is important to illustrate that, in a context in which the classification or description is made by humans, the sound description modulates the interaction between: a) the person who classifies the sounds and b) the person who tries to retrieve the sound. In this panorama, there occurs an interchange of information in which the only medium is the label or description (Figure 13). **The descriptor is not the information that is searched, but the cue to find this information**. The situation supposes a semiotic codification first by a) looking for information that the sound involves and later by b) trying to get access to this information. The first codification is descriptive, the second prescriptive.

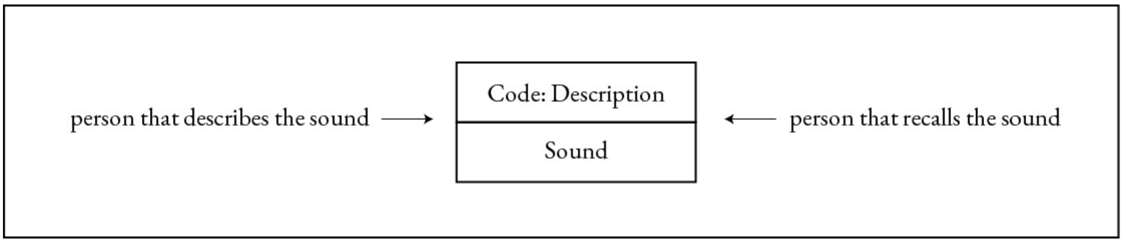

Figure 13 Interaction by the sound description. (Source: Author)

In this scenario, a usual constraint for the retrieval of an abstract sound is given by the practice of **listening for references**. In this sense, it might happen that though b) is trying to retrieve an abstract *creaking* effect, it may not be able to codify it in something more than like a "*door*" using this last word as a label. Conversely, a referential description can be insufficient to involve the different possible interpretations that an abstract sound can carry. For example, the creak could be interpreted by someone else as like an "*animal*" or like a "*machine*".

In the presented structure, the description is supposed to emplace the perception at the same level of abstraction of the sound. In other words, to use an abstract description that allows the possible interpretations suitable for an abstract sound. This emplacement requires a certain level of **reduced**

listening that is **not at the same level presented by Schaeffer** (2017) but at the level of the categories that have already been presented. With this action, a higher level of abstraction in this structure is sought to reach a **consensus point** of both subjects a) and b) that is less probable in a referential description.

Language is the focal point of another usual debate found in almost all studies and descriptive systems that involve semantic description. In this context, it is important to note that the intention of this study is not to provide descriptors or vocabulary for describing sounds but to create an adequate structure that can be used for the description of abstract sounds. The structure could be adapted for each language. It might happen that there are not the same descriptors in every language, but they might still be bound to the same categories. Further explorations in this area are outside the scope of this study.

Finally, future studies should focus on how the variable for any sound description might be the temporal structure or duration of the sound. The material used for this practical example corresponds to a temporary short-term auditory event. However, in the case that the temporal structure is extended through repetition of an event or by the use of files with a larger duration, the description might likewise be more disposed to musical descriptors from the sound itself as well as to affective connotations. Descriptors found in the Pedersen lexicon (like quiet, relaxing, scary) might be only conceivable in a larger temporal structure. Future studies in this field might contribute to identify the relationship between temporal structure and sound description.

Summary

In this study, a hierarchical structure for the description of abstract sounds was developed. This hierarchical classification can be used to describe a sound independent from its source(s). The classification includes the following descriptive categories: descriptors of the sound itself, descriptors from other senses or crossmodal correspondences, event descriptors that are divided into those that refer to interaction and those that refer to the material.

This hierarchical description system can be used to describe and sort abstract sounds in a sound collection or sound library for the application in a sound design context. The structure of the description can be useful for retrieving specific sounds or a broader number of files for creative uses. The description also allows for the description of sounds with different possible interpretations that the abstract sound can carry in multiple (or all) categories.

Some components still need to be explored, particularly those that concern abstract sounds with longer duration and other variables that come from human factors like the interaction between the person who makes the description and the user who retrieves the sound. For future variables and overviews of this structure, it will be necessary to adapt and value this structure in other contexts, paying attention to the human factors described above, other sound collections and potentially other languages.

The presented study serves as a primary alternative for the description of abstract sounds or sounds that have ambiguous meaning and are thus difficult to categorize and organize. This study comes to the conclusion that while we might have enough words to describe sounds, we need to use them in an organized way to be able to retrieve precisely what we are looking for.

Glossary

Label: a word or phrase indicating that what follows belongs in a particular category or classification. (“Definition of label | Dictionary.com,” n.d.)

Descriptor: A word or phrase that describes, identifies, or labels an attribute or a characteristic. (Holm Pedersen, 2008)

Metadata: Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. (Source:(Hodge, 2004)

Search engine: a program that searches for and items, files or content in a database that corresponds to keywords or labels specified by the user.

Musique concrète: (French: “concrete music”), an experimental technique of musical composition using recorded sounds as raw material. The technique was developed about 1948 by the French composer Pierre Schaeffer and his associates at the Studio d’Essai (“Experimental Studio”) of the French radio system. The fundamental principle of musique concrète lies in the assemblage of various natural sounds recorded on tape (or, originally, on disks) to produce a montage of sound. During the preparation of such a composition, the sounds selected and recorded may be modified in any way desired—played backward, cut short or extended, subjected to echo-chamber effects, varied in pitch and intensity, and so on. The finished composition thus represents the combination of varied auditory experiences into an artistic unity. (“Musique concrète | musical composition technique,” n.d.)

Vibrato: A tremulous or pulsating effect with rapid variations in pitch produced by instrumental or vocal tone.

Pitch: in music, the position of a single sound in the complete range of sound. Sounds are higher or lower in pitch according to the frequency of vibration of the sound waves producing them. A high frequency (e.g., 880 hertz [Hz; cycles per second]) is perceived as a high pitch and a low frequency (e.g., 55 Hz) as a low pitch.(“Pitch | Definition, Frequency, & Music,” n.d.)

Stimuli: Stimuli may be anything that evokes a response from an assessor when presented with the stimuli. Such stimuli may stimulate one or many of the senses e.g. hearing, vision, touch, olfaction or taste.(Holm Pedersen, 2008)

Bibliography

Bonebright, T. L., Miner, N. E., Goldsmith, T. E., & Caudell, T. P. (2005). Data collection and analysis techniques for evaluating the perceptual qualities of auditory stimuli. *ACM Transactions on Applied Perception*, *2*(4), 505–516. https://doi.org/10.1145/1101530.1101550

Cano, P., Koppenberger, M., Celma, O., Herrera, P., & Tarasov, V. (2004). Sound effects taxonomy management in production environments. *25th AES International Conference. Metadata for Audio*, 8. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=12831

Chauvin, M. A., Younce, F., Ross, C., & Swanson, B. (2008). Standard scales for crispness, crackliness and crunchiness in dry and wet foods: Relationship with acoustical determinations. *Journal of Texture Studies*, *39*(4), 345–368. https://doi.org/10.1111/j.1745-4603.2008.00147.x

Chion, M., & Steintrager, J. A. (2016). *Sound: An acoulogical treatise*. Durham: Duke University Press.

Clark, H. H., & Brownell, H. H. (1976). Position, direction, and their perceptual integrality. *Perception & Psychophysics*, *19*(4), 328–334. https://doi.org/10.3758/BF03204238

Correya, A. A. (2017). Retrieving Ambiguous Sounds Using Perceptual Timbral Attributes In Audio Production Environments. *Zenodo*. https://doi.org/10.5281/zenodo.1098587

Definition of label | Dictionary.com. (n.d.). Retrieved July 23, 2019, from www.dictionary.com website: https://www.dictionary.com/browse/label

Evans, K. K., & Treisman, A. (2011). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6–6. https://doi.org/10.1167/10.1.6

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203. https://doi.org/10.3758/BF03193720

Gaver, W. W. (1993). What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, *5*(1), 1–29. https://doi.org/10.1207/s15326969eco0501_1

Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., … Ritter, M. (2017). Audio Set: An ontology and human-labeled dataset for audio events. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 776–780. https://doi.org/10.1109/ICASSP.2017.7952261

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Oxford, England: Houghton Mifflin.

Görne, T. (2017). *Sounddesign: Klang, Wahrnehmung, Emotion*. München: Hanser.

Hodge, G. M. (2004). *Understanding metadata.* Bethesda MD: NISO Press.

Holm Pedersen, T. (2008). *The semantic space of sounds: Lexicon of sound-describing words*. DELTA.

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, *14*(23), R990–R991. https://doi.org/10.1016/j.cub.2004.11.018

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394. https://doi.org/10.1037/0096-1523.13.3.384

Merer, A., Ystad, S., Kronland-Martinet, R., & Aramaki, M. (2011). Abstract Sounds and Their Applications in Audio and Perception Research. In S. Ystad, M. Aramaki, R. Kronland-Martinet, & K. Jensen (Eds.), *Exploring Music Contents* (pp. 176–187). Springer Berlin Heidelberg.

Musique concrète | musical composition technique. (n.d.). Retrieved July 23, 2019, from Encyclopedia Britannica website: https://www.britannica.com/art/musique-concrete

Pearce, A., Brookes, T., & Mason, R. (2017). Timbral Attributes for Sound Effect Library Searching. *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=18754

Pitch | Definition, Frequency, & Music. (n.d.). Retrieved August 5, 2019, from Encyclopedia Britannica website: https://www.britannica.com/art/pitch-music

Schaeffer, P., North, C., & Dack, J. (2017). *Treatise on musical objects: Essays across disciplines*. Oakland, California: University of California Press.

Schafer, R. M. (1993). *The soundscape: Our sonic environment and the tuning of the world*. Rochester, Vt. : [United States]: Destiny Books ; Distributed to the book trade in the United States by American International Distribution Corp.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Acknowledgment

First, I would like to offer my sincere gratitude to my supervisor, Prof. Thomas Görne, for all the support throughout this master's thesis. I am very thankful for his time, patience, knowledge, and feedback gave throughout my study and practical projects in sound design. He was always open and helpful when it came to all the concerns that are part of being an international student.

I am very thankful to my second supervisor, Stefan Troschka, for his insightful comments and observations, as well as for encouragement during an essential moment of this work.

My sincere thanks to the generous and essential support from DAAD (German Academic Exchange Service) for my master's fellowship. I am especially grateful for my contact person, Nadia Ruiz, for her cordial and helpful assistance.

I would like to thank my family—my mother and my father in his memory—for giving me the moral support that allows me to achieve the most significant goals in my life, including this study. Thank you for lovingly showing me the values of dreams, commitment, sharing and being thankful.

Statutory declaration

I hereby declare that the thesis with the title "Description of abstract sounds for the application in sound design" has been composed by myself autonomously and that no means other than those declared were used. In every single case, I have indicated parts that were taken out of published or unpublished work, either verbatim or in a paraphrased manner, as such through a quotation.

This thesis has not been handed in or published before in the same or similar form.

Place, date

Claret Canelon