Escola das Artes — Universidade Católica Portuguesa

Mestrado em Som e Imagem

A Sound is Worth a Thousand Words:
Exploring the Taxonomic and Causal Link Between Emotions and Sound Objects

Master's Degree in Sound Design | 2015/2016

Diogo da Costa Alves Pinto

Dissertation Advisor: Pedro Pestana PhD

Secondary Supervisor: Vitor Joaquim PhD

April 2016

ACKNOWLEDGEMENTS

The researcher wants to give thanks to everyone who has become a crucial part of this dissertation.

To my family, especially my mother, father and sister for their moral and financial support.

To my better-half Katharina Leppert for all the strength, patience, and loving support.

To my teachers Pedro Pestana PhD and Vitor Joaquim PhD for their guidance.

To my close friends Vanessa Costa, José Góis, Rossana Fonseca and Gabriel Salgado for their helpfulness and relentless concern.

And lastly to all the respondents for their full cooperation, which made them an invaluable part of this dissertation.

ABSTRACT

This investigation's object is the relation between emotion and sound and how the latter can be understood through an emotion-oriented study. Psychological investigations strive to understand how the world affects people and how people, in turn, understand the world on the grounds of their own reflections and interpretations. Thus, an emotional understanding of sound is inevitably linked to the concept of perceived emotion. This dissertation's purpose is to understand whether there is a taxonomic relation between sounds and perceived emotions. To this aim, emotional semantics and proposals for emotional categorization are approached, as well as studies on sound categorization and its relation with experiments between emotion and sound or music. Two studies investigated the aforementioned themes. In Experiment 1, participants rated sound-image pairs in a causal-oriented environment, followed by a similar recall task, with the aim of understanding the connection between the listener and a sound's semantic content. In Experiment 2, participants rated a group of sounds, half of which were masked to hide their semantic content, with the goal of understanding the importance of semantic content in auditory stimuli. Taken together, the data suggest that some emotions cannot be transmitted by sound alone and that it takes a combination of the listener, the context, and the sound's physical features in order to get a complete understanding of perceived emotions.

Keywords: Emotional taxonomy of sound, emotional semantics, sound categorization, affect in sound.

Index

1. INTRODUCTION

1. Purpose

This investigation's object is the relation between emotion and sound and how the latter can be understood through an emotion-oriented study. There is also an aim of understanding if a sound's perceived emotional valence (attractiveness or repulsiveness) is independent from its semantic content (e.g., its source, context, cause, surroundings…) and whether there is a taxonomic[1] relation between sounds and perceived emotions. To this aim, this dissertation will feature phenomenological concepts such as "acousmatic" or "Époché", studies on emotional semantics and proposals for emotional categorization, as well as studies on sound categorization and its relation with experiments between emotion and sound or music.

2. Methodology

The investigation carried out in this dissertation involved several different steps. Initially, there was an aim to analyze past studies or experiments, in order to understand their procedure; their results; the emotional state models that were utilized; the adjectives that were more or less used and their effect on the end result; how many subjects participated in each study; how many (and which) sounds were used and, ultimately, which strengths and weaknesses were encountered in these studies' methodologies and procedures. These analyses were gathered from October 2015 to January 2016 and were used throughout the dissertation, whenever they could help give more information on a specific subject. In a second stage, a multi-modal test (sound-image pairs) was executed. This test was carried out between January and February 2016. In this test, two different sample groups were presented the same sounds accompanied by pictures with different emotional valence, to try and create some misperception with the ultimate objective of understanding the effect of a sound's semantic content in its perceived emotional valence. These auditory stimuli were rated via selection of

[1] Taxonomy is the science of classification

a semantic group from a set of nine different adjective pairs. This experiment had a convenience sample, constituted mostly by individuals with ages between 18 and 30 years of age. The last experiment used the knowledge obtained from the aforementioned multi-modal test and previous studies/experiments' analysis and was carried out in April 2016. In this test, the participants had similar characteristics from the previous one. The stimuli, however, was merely acoustic at this stage. Half of the stimuli was masked in order to hide their semantic content and participants were asked to rate the masked and original stimuli together, with a selection from the same adjectives group from before, but also from a five-point Likert scale for valence.

3. Structure

In TYPES OF LISTENING, many phenomenological and philosophical terms, as well as terms with a firm relation to the study of sound (and its features) that will be referenced throughout this dissertation are explained and clarified.

In EMOTIONAL SEMANTICS, some studies of emotion in music are presented, which allows for an invaluable comparison with the lesser amount of studies of emotion in sound. In this chapter, some emotional psychology terms and theories are introduced, such as The Discrete Emotions Theory and the Self-Assessment Manikin. This chapter ends with a reference to some proposals of emotional taxonomy, which create a connection to the following chapter.

In SOUND TAXONOMY, the focus is set on sound itself and the sound source identification theme is presented. With the relation between sound and its semantic content established, this chapter continues with the combination of both sound categorization and emotional classification, which raises some questions towards the sound's semantic content. This is further explored in EXPERIMENT 1, as was explained in the preceding sub-chapters of this Introduction.

A tentative conclusion from these previous chapters is exposed in the name of this following chapter: "EVENT SIMILARITY AND SOUND SIMILARITY ARE ALTERNATIVE WAYS OF DESCRIBING THE SAME WORLD" (GUYOT, 1996). And it is due to this similarity that this chapter begins to unveil other aspects of sound, other than its semantic content. This assessment of sound's other physical features and their link to perceived emotions lead the investigation to EXPERIMENT 2 which, like EXPERIMENT 1, was already introduced in the previous sub-chapters. In CONCLUSION, the data of the two experiments that were carried out, as well as the data from the studies that were analysed are explored together to get a small look at the bigger picture.

2. TYPES OF LISTENING

1. Everyday Listening, Musical Listening, Descriptive Listening, And Holistic Listening

Gaver (1993) introduced the notion of everyday listening and explained its relation with musical listening. According to him, musical listening focuses on the perceptual characteristics of sound, whereas everyday listening focuses on gathering relevant information about its source and the actions producing it: "(…) as you stand there on the road, it is likely that you do not listen to the sound itself at all. Instead, you are likely to notice that a sound is made by an automobile with a large and powerful engine. Your attention is likely to be drawn to the fact that it is approaching quickly from behind. And you might even attend to the environment, hearing that the road you are on is actually a narrow alley, with echoing walls on each side". Much like Chion's definition of causal listening (see 2.2), everyday listening assumes a focus on events rather than the sounds for themselves. Thus, everyday listening is the way we usually perceive the world: getting from sound the information we need about size, speed, materials interaction, surroundings, and much more, through acoustical properties of sound such as pitch, timbre, amplitude, damping, and reverberation, to name a few. "For example, we do not hear a noisy harmonic complex in combination with a burst of noise, instead we hear a passing car. Likewise we do not hear a double pulse with prominent energy around 2.4 and 6 kHz, but we hear a closing door." (Andringa & Niessen, 2006).

More recently, on the subject of categorisation of everyday sounds and soundscapes, Manon Raimbault (2006) proposed a distinction between the terms descriptive (or everyday) listening, which discriminates the acoustic sources or events in a soundscape, and holistic hearing, which understands soundscapes as a whole, without isolating their sources.

2. Semantic Listening, Causal Listening, And Reduced Listening

Michel Chion (1994) describes three different types (or modes) of listening, which influence the listener's perception of an addressed object. He refers to them as semantic listening, causal listening, and reduced listening, with greater emphasis on the latter.

To Chion, semantic listening exists as more of a mean to interpret a message. Therefore, this mode of listening refers to codes such as spoken language or Morse code. He acknowledges a concrete distinction between semantic listening and causal listening, stating that the latter creates a direct correlation between the sound and its cause. Hence, the purpose of listening becomes the gathering of information about the cause, which can be visible or invisible, recognizable or unknown, and in which a unique item or merely a category (e.g., human or animal source) can be identified.

Acousmatic listening (or reduced listening) constitutes the opposite of direct listening, where sound sources are always present and visible. Acousmatic is a rare word, derived from the Greek, which indicates a noise that is heard without the causes from which it originates being seen. The Larousse dictionary defines acousmatic as the "name given to the disciples of Pythagoras who, for five years, listened to his teachings while he was hidden behind a curtain, without seeing him, while observing a strict silence." The purpose of this was to avoid any visual distractions, disabling their view of his physical appearance. It was more recently used by Pierre Schaeffer (2004) to describe a common experience with consequences that are not always recognized, consisting of hearing sounds with no visible cause, a phenomenon that is unavoidable within such means as the radio, records, telephone, tape recorders, etc. It changes the way we hear. And by isolating sounds from their visual counterparts, conditions are made for reduced listening to take place, allowing for concentration on the sound for its own sake, as a sound object, regardless of its meaning or its causes (although reduced listening can also exist in a direct listening situation, only with greater difficulty).

Some of the characteristic ways in which the acousmatic situation alters the conditions of listening are: the absence of sight and all that comes with it, which allows for the understanding that much of what the listener thought was being heard, was actually only seen and explained by the context; hearing and sight are separated, encouraging listeners to distance themselves from the situational or causal aspects and focus on the sound for itself, the sound object. In fact, by repeatedly listening to the same recording of a sound fragment, the emphasis lays on variations of listening and on new aspects of sound that may arise and towards which attention is drawn (consciously or unconsciously).

To Schaeffer (2004), the acousmatic experience enables the listener to become aware of his own perceptual activity as well as of the sound object. For Schaeffer, reduced (or acousmatic) listening was made possible by radio and recordings, which led to new experiences of sound and the sonorous objects. These technologies now allowed sounds to have a new existence and perception apart from their sources. The sound object is not the instrument that is being played or the magnetic tape: after denying instruments and cultural conditionings, the acoustic properties take the front seat.

In fact, he created musique concrète in 1948 because of the acousmatic nature of radio sounds and because of this, he coined the term "acousmatic experience", an extension of "acousmatic", to describe a new way of hearing: one in which a person would give oneself to listening both exclusively and entirely. Hence, the radio and the tape recorder became Schaeffer's own "curtain of Pythagoras", creating new phenomena to be studied as well as new conditions for observation.

Two terms were present at the origins of musique concrète: "closed groove" and "cut bell". The closed groove experiment consisted in creating a periodic phenomenon with any recorded sound fragment, thus having the ability of repeating it indefinitely. Later, with the tape recorder, it was replaced by the tape loop, which created a similar effect. The closed groove led to the awareness of the sound object and reduced listening, challenging the causal perception and revealing new characteristics to the repeated sound over time. The cut bell experiment also involved interruption of the progress in a recorded sound: by removing a fragment of the resonance of a bell after its attack, evening out its dynamic behaviour and then repeating it using the closed groove technique, "a sound like a flute" can be heard.

Reduced listening consists in listening to the sound as its own event, as a sound object, regardless of its cause, source or meaning. As a very descriptive type of listening, repeated hearings are required to gradually stop attending to the cause and to achieve a complete understanding of the sound object. This occurs in opposition to a more ordinary kind of listening, in which sound is treated as a vehicle. By removing all our habitual references in the listening process, many phenomena that are implicit in our perception are clarified. Reduced listening is therefore connected to the notion of phenomenological reduction or Époché. Époché constitutes a disengagement of regular listening patterns, enabling the perception of the sound object as a medium for its several sonic perceivable dimensions. Reduced listening is also therefore connected to the idea of a sound object, defining each other as the perceptual activity and the object of perception through yet another phenomenological term: Intention. Michel Chion (1983) defines this concept: "If the object transcends every partial experience that I have of it, it is in my experience that this transcendence is formed." Thus, a correlation is formed between a hearing intention and a heard sound object. This means that the sound object only exists by means of reduced listening.

3. EMOTIONAL SEMANTICS

1. Emotional Semantics In Music

The distinction between emotion and mood is of great relevance in this matter. Emotion refers to a state that is intense and involves a noticeable stimulus, while a mood is "less intense and its cause is not immediately apparent" (Juslin & Sloboda, 2010). Evolutionist psychologists tend to adopt a perspective in which happiness, sadness, anger, fear, and disgust constitute the basic emotions, although some investigators consider fourteen (Lazarus, 1991), or even sixteen basic emotions (Roseman, Spindel & Jose, 1990). Deryck Cooke (1959) identified sixteen basic terms of vocabulary using examples dating from the middle ages all the way to modern times, connected to which he suggested emotional expressions. For example, an ascending major triad expresses "an outgoing, active, assertion of joy" while a minor triad suggests an "assertion of sorrow, a complaint, a protest against misfortune".

Melvin Rigg (1937) characterized joy, lamentation, longing, and love through short phrases. In this experiment, he played twenty musical excerpts to a hundred students and collected their answers with two different methods: firstly through free description and lastly by asking the students to choose from a set of descriptions (joy, lamentation, hopeful longing, sorrowful longing, and love), among which joy was the most recognized emotional term. Rigg found that the stated emotions were frequently less accurate in the free description and chosen more often when presented among the set of descriptions. Still in the musical emotion recognition realm, Campbell's (1942) emotion designations also found high agreement for joy, gaiety, and assertion, while terms such as calm, sorrow, tenderness, and yearning generated less agreement and also some confusion among each other. In the same year, Watson also found best agreement with the expressions mischievous, happy, exciting, kingly, dignified, and sad, while encountering worst results with terms such as pleading, tragic, and mysterious. These studies show highest agreement of emotional expressions in music for positive emotions, high arousal (see 3.3), and low arousal, as well as very good recognition for negative emotions. In Gundlach (1935), Hampton (1945), Capurso (1952), and Sopchak (1955), best agreement was again found for positive and negative emotions (happy, joyful, triumphant, sadness, despair, melancholy…), high arousal (angry, rage) and low arousal (relaxing, soothing). The terms with less agreement were cruelty, devotion, cruelty, disgust, eroticism, flippancy, hate, horror, irritation, jealousy, pity, whimsy, and worship.

Lists of emotions put together by researchers for a specific study may lack validity and reliability and create difficulties when compared to other studies. "Emotions" and "feelings" are often assumed to be synonyms. Quoting Scherer (2004): "It is suggested that “feelings” can be profitably conceptualized as a central component of emotion, which integrates all other components and serves as the basis for the conscious representation of emotional processes and for affect regulation". Asking listeners to describe their emotional response to a sound event[2] using basic emotions such as joy, sadness, or fear, assumes that the listeners can effectively transpose everyday emotions to an auditory context.

2. The Discrete Emotions Theory

Distinguishing one emotion from another (i.e. emotion classification) is a highly discussed issue in emotion research. Emotion classification has two fundamental points of view: the first believes that emotions are discrete and different constructs, using specific words or word groups to name particular emotions (e.g., happiness, sadness, fear, etc.) (Drossos, Floros, and Kanellopoulos, 2012), while the latter characterizes emotions on a dimensional basis in grouping, looking at emotions as the sum of two or more emotional states, which are illustrated as continuous values (Laurier et al., 2009). The Discrete Emotions Theory claims that there are a small number of universal emotions, which are recognized and shared by all humans (Colombetti, 2009). Thus, this representation model relies on a list of adjectives each describing an emotion and has been used in emotion recognition for long (Schuller et al., 2011; Kim et al., 2010). Ekman (1973) conducted a cross-cultural study with the idea that emotions can be clearly recognized between humans, and concluded that there are six basic and universal emotions: anger, disgust, fear, happiness, sadness, and surprise. By contrast, the second point of view describes emotions as points in a multi-dimensional space and tends to offer a more accurate way to represent emotions.

The discrete emotion model theorists suggest that some adaptive emotional strategies have been developed over time. Darwin (1872) used some of the most common emotion terms in the English language as chapter headings and then showed their functionality, their history and their universality regarding different species as well as different cultures. Tomkins (1962) suggested that the aforementioned theorized basic emotions could trigger muscular

[2] McAdams (1993) referred to sound events as sequences of temporally related sounds such as the sounds accompanying the feeding of a cat or the fixing of a leaky faucet. The sounds that surround the action introduce information about it.

responses (mostly in the face) and ways to measure them.

3. Emotional State Models

The PAD (Pleasure, Arousal, and Dominance) emotional state model (Mehrabian and Russell, 1974) uses three numerical dimensions in order to describe and measure emotional states: the Pleasure-Displeasure Scale measures an emotion's pleasantness; the Arousal-Nonarousal Scale measures an emotion's intensity; the Dominance-Submissiveness Scale measures an emotion's dominance nature, as can be verified in figure 1 (Kim et al., 2010). They also proposed the Semantic Differential Scale as a tool for assessing the three-dimensional structure of objects, events, and situations. It consists of eighteen bipolar adjective pairs, which are each rated along a nine-point scale to generate information on the dimensions of pleasure, arousal, and dominance. Although this method is very informative and widely used, it has some shortcomings. Two of its greatest disadvantages are the heavy effort it requires to measure eighteen different ratings for each stimulus and that the use of a verbal system makes it harder to re-apply the same methodology with non-English speaking subjects.

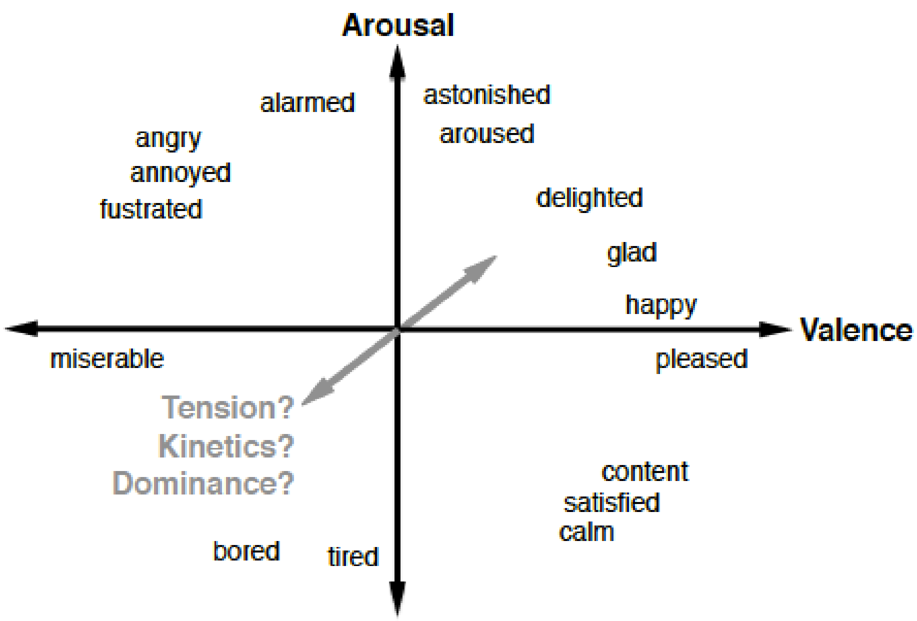

Figure 1: "The Valence-Arousal space, labeled by Russel's direct circular projection of adjectives. Includes semantic of projected third affect dimensions: "tension", "kinetics", "dominance"."

Wilhelm Wundt (1897/1980) suggested a distinction between three different bipolar dimensions of feelings: pleasurable-unpleasurable, arousing-subduing, and strain-relaxation. Harold Schlosberg (1954) renamed these terms to "pleasantness-unpleasantness", "attention-rejection", and "level of activation", respectively. This three-dimensional model had a strong impact on the psychology of emotion. Since there were many difficulties in establishing the attention-rejection dimension in an empirical fashion, feeling was eventually defined by a two-dimensional model, which was formed by valence (pleasantness-unpleasantness) and activation. Thus, the Valence-Arousal model had great representation in the affective sciences and in emotional research in both sound and music, which is due to some of its most practical advantages: reliability and simplicity.

The main variance in emotional meaning is often illustrated in just two dimensions: arousal/activation and pleasure/valence (Smith & Ellsworth, 1985; Yang et al., 2008; Drossos, Floros & Kanellopoulos (2012); Mehrabian & Russell, 1974; Schuller et al., 2012; Osgood, Suci, & Tannenbaum, 1957; Asutay et al., 2012). Research carried out by Björk (1985) showed that a two dimensional approach is, in fact, suited to describe the link between emotion and sound and Yang et al. (2008) conducted an experiment with the goal of understanding the reliability of the Arousal-Valence emotion plane in music. In this experiment, each of the two hundred and fifty-three volunteers was asked to listen to ten random music excerpts and to label the Arousal-Valence values. The stimuli were gathered in a database that included a hundred and ninety-five popular songs from a number of Western, Chinese, and Japanese albums. The experiment's results show that there was a high (95%) agreement between the subjects, even in a test-retest reliability study, conducted two months after the first experiment.

4. Emotional Taxonomy Proposals

Damasio (2006) suggested three different kinds of feelings: those of basic universal emotions (such as happiness, sadness, anger, fear, etc.), those of subtle universal emotions (which are deduced from experience during the course of life), and background feelings, which reflect the "momentary overall condition of the body" (Haverkamp, 2012).

Schuller et al. (2011) proposed some subdivisions on an emotional taxonomy: type of speech would be divided into prompted emotions, non-prompted emotions, or obtained in specific scenarios. It is also proposed that an emotional taxonomy's main categories would be the following: positive emotions, neutral emotions, negative emotions, and the "big n" emotions (e.g., anger, fear, joy, sadness, etc.), with sub-categories adapted to each of their differences. All the main categories can be either pure (e.g., joy) or mixed (e.g., if a mixture of joy and fear is observed) and can be affected by dimensions such as arousal and/or valence.

4. SOUND TAXONOMY

1. The Role Of Source Identification And Categorization Of Sound Events, Environmental Sounds[3], And Soundscapes

Schubert (1975) stated that the primary task of the auditory system is the identification of the sound sources and their behaviour. In fact, studies on human-made sounds support Gaver's everyday listening theory (Ballas, 1993; Guyot, Castellengo, & Fabre 1997; Susini, Misdariis, Winsberg, & McAdams, 1998; Vanderveer, 1980), suggesting a causal taxonomy. In order to understand these causal taxonomies, attention was drawn to urban soundscapes, in which noise is emitted simultaneously by a wide variety of sources. Guastavino (2007) investigated everyday listening of urban soundscapes in a mail survey with seventy-seven participants. The main categories identified were human sounds, traffic noise, natural sounds, and music. Of these, the ones that originated positive judgements were human sounds (except when reflecting anger), natural sounds, and human-made music (musician), whereas mechanical sounds and indirect music (loud-speakers, car radio) gave rise to negative judgements. Within the mechanical sounds category, positive judgements were attributed to electric cars and public transportation noise, in relation to environmental concerns. This clearly shows the importance of the sound source when creating a taxonomy or attributing emotional value to an acoustic phenomenon. On a similar note, Gygi (2007) showed that the subjects take physical aspects into account, but similar sources (vocalizations, water-based sounds, rhythmic impacts, and mechanical sounds) were also categorized together. Dubois (2000) stated that an acoustic phenomenon can be classified according "either to the source that produces it or to the action generating the noise (this is the case, for instance, with the squeaking of a door, which can be categorized either with “noises of doors” – fermeture, claquement, ouverture d’une porte ‘shutting, slamming, opening of a door’ – or with other instances of grincement ‘squeaking’ – of doors, of windows, or of steps)." It was also shown that participants categorized sounds based on the sound source more often than based on its

3 Van DerVeer (1980), Jenkins (1985), Ballas and Howard (1987), and Marcell et al. (2007), define environmental sounds as being: non-musical and non-linguistic sounds; produced by real events; usually more complex than machine-generated sounds (e.g., pure tones); potentially inanimate (e.g., machines), animate (e.g., human made), natural (e.g., rain), artificial (e.g., car horn), dynamic, and carrying information about their source and their surroundings.

physical features. Bergman et al. (2008) made sound sources unidentifiable, by removing the meaning of environmental sounds by means of frequency smearing using different window lengths and frequency bands, which kept the sound's psychoacoustic properties showing that content has a bigger impact on emotional reactions than form.

Results from Maffiolo (1999) and Guastavino & Cheminée (2003) aid the distinction between sound events with clearly identifiable sources, and ambient noise, with a blurry background noise. The first are described with clear reference to their source, either by referring the object or the part of the object that generates noise. "These metonymies — substituting the name of the source producing sound for the name of the sound itself — indicate confusions between sounds and sources producing the sound, and further suggest that the acoustic phenomenon is not abstracted from the object generating the sound." (Guastavino, 2007). In the descriptions of ambient noise, however, the object source is mentioned less while a majority of the utilized terms refer to the physical properties of the sound, showing a broader kind of listening than the one verified from the sound events with recognizable sources.

Bonebright (2001) conducted an experiment that rated seventy-four everyday sounds in a seven-point bipolar adjective scale. The stimuli consisted of seventy-four sounds made by objects that individuals in the USA would have exposure to on a regular basis. Dull-sharp, round-angular, and relaxed-tense were positively correlated with one another. Rough-smooth was positively correlated with unpleasant-pleasant, and both of them were negatively correlated with soft-loud.

VanDerveer (1980) asked participants to write a short sentence identifying thirty everyday sounds. The participants always tried to identify the sounds according to their sources, except when they could not clearly identify it, which became noticeable even in their mistakes, since they would confuse similar sounding events as clapping with dropping a book, but rarely with tearing paper. In a similar experiment, Gaver (1988) asked participants to describe seventeen sounds. They were tested individually and asked to go to into as much detail as possible. The conclusion was similar to VanDerveer's but with added accuracy: some participants were able to distinguish the sound of someone running upstairs from someone running downstairs, most participants correctly described chalk writing on a chalkboard, some participants were right about the size of objects dropped into water and most said that a cup was being filled by just hearing a pouring liquid. In a curious note, the sound of someone walking on a floor covered with newspapers was correctly interpreted by one of the participants, who eventually rejected this answer because it was "too implausible". This shows just how far the rationalization involved in everyday (or causal) listening can influence test results, which are looking for descriptions of sound's physical properties and not their relation to their surroundings or the everyday world.

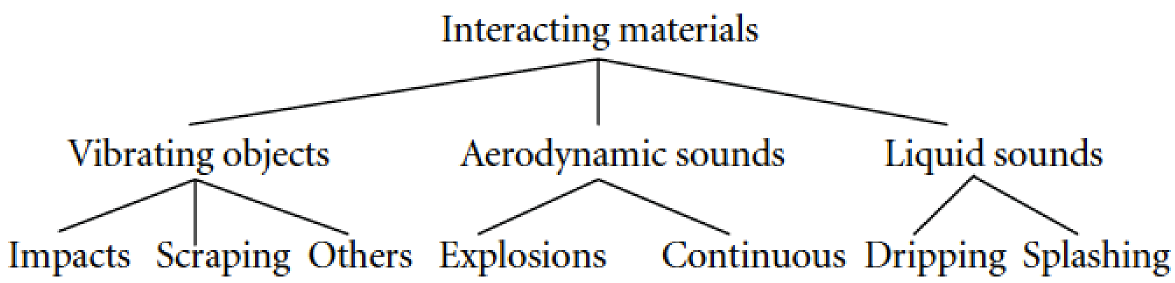

Figure 2: Brief taxonomy of basic sound events adapted from Gaver (1993). Used with permission.

Gaver (1993) proposed a taxonomic description of basic sound events, which included categories such as vibrating objects, aerodynamic sounds, and liquid sounds, which can be verified in figure 2 (Gygi and Shafiro, 2010). From these basic sound events, complex sounds are formed, such as patterned sources (repetition of a basic event), complex sources (more than one sort of basic level event), and hybrid sources (involving more than one basic sort of material). Gygi and Shafiro (2010) went one step further with the decision tree on figure 3.

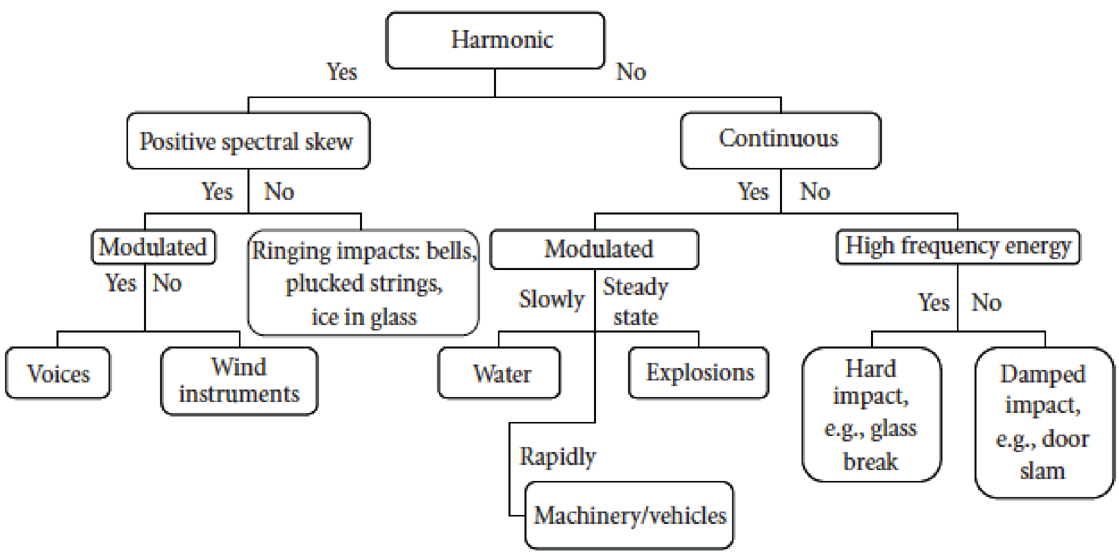

Figure 3: Decision tree proposal for sound recognition. Used with permission.

2. An Emotional Taxonomy Of Source Identifiable Sounds

Solomon (1958) made an attempt to describe complex sounds within a rating scale instrument: fifty subjects rated twenty different sounds on fifty seven-point bipolar scales (composed by fifty adjectives alongside their antonyms). Seven clusters were created for dividing these scales and named: magnitude, aesthetic-evaluative, clarity, security, relaxation, familiarity, and mood.

According to Table 1 (Solomon, 1958), the "positive" ends of psychological dimensions should consist of terms such as heavy, beautiful, clear, mild, loose, familiar, and colourful. In the opposite ("negative") side of the scale, reside terms as light, ugly, intense, tight, strange, and colourless.

Factor I—Magnitude	
heavy-light	84[a]
large-small	79
rumbling-whining	76
wide-narrow	73
low-high	71
Factor II—Aesthetic-evaluative	
beautiful-ugly	67
pleasant-unpleasant	60
good-bad	59
pleasing-annoying	58
smooth-rough	54
Factor III—Clarity	
clear-hazy	58
definite-uncertain	52
even-uneven	51
concentrated-diffuse	43
obvious-subtle	43
Factor IV—Security	
mild-intense	53
gentle-violent	51
calming-exciting	48
safe-dangerous	48
simple-complex	48
Factor V—Relaxation	
relaxed-tense	64
loose-tight	57
soft-hard	36
gentle-violent	34
mild-intense	33
Factor VI—Familiarity	
definite-uncertain	39
familiar-strange	35
wet-dry	34
active-passive	30
steady-fluttering	30
Factor VII—Mood	
colorful-colorless	41
rich-thin	34
happy-sad	31
deliberate-careless	30
full-empty	22

Table 1: Bipolar scales used to define psychological dimensions (factors).

Ozcan Vieira, Van Egmond, and Jacobs (2014) carried out two experiments with the goal of determining the domain of domestic product sounds alongside their constituting categories and to better understand the aforementioned domain and categories. The participants were asked to freely group the sounds they considered similar. The stimuli consisted in twenty-two domestic product sounds under five seconds longs taken from various sound effect CDs. Four categories were formed: short duration sounds caused by an impact between product parts, digitally produced alarm-like sounds, sounds which are consequences of engines with high RPM (Revolutions Per Minute), as well as small rotating and rubbing mechanical product parts, and sounds that are caused by the heating of liquids. In the second part of this experiment, participants were asked to rate experienced similarities between pairs of sounds within the following categories: air, cyclic, liquid, and mechanical sounds. As expected, the similar sound pairs had the highest similarity rating. Among the dissimilar sound pairs, the air-cyclic sound pair had the highest similarity rating, while the air-liquid pair had the lowest similarity rating.

Schuller et al. (2012) performed an experiment with the goal of creating an emotional sound database. To this aim, four students rated three hundred and ninety sounds in Valence and Arousal and then wrote down the perceived emotion. The sound stimuli were divided into eight categories from the free online engine FindSounds.com: Animals, Musical Instruments, Nature, Noisemaker, People, Sports, Tools, and Vehicles. The results show that agreement was much higher for valence than for arousal and that both arousal and valence are highly correlated with loudness, but the correlation with valence is a negative one. This illustrates the idea that loud sounds are not pleasant.

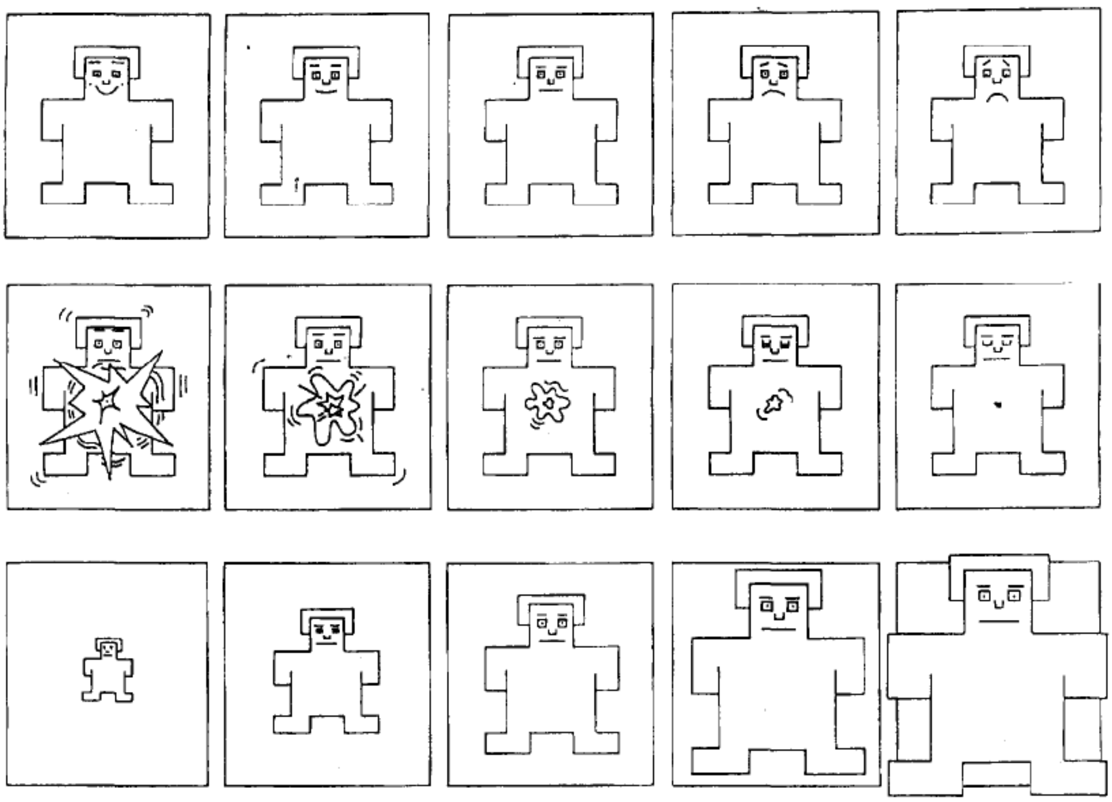

Figure 4: The Self-Assessment Manikin (SAM). The top panel valence assesses valence, the middle panel assesses arousal and the bottom panel assesses dominance.

Bradley and Lang (2000) also performed two experiments that investigated emotional reactions to naturally occurring sounds. While many studies on physiological responses to sounds tended to rely on short sets of stimuli (Gang and Teft, 1975; Pallmeyer, Blanchard, and Kolb, 1986; Meyers and Smith, 1986), the number of sounds in this study is more extensive and including a broad range of semantic categories such as erotica, bombs, or animal and human vocalizations. The goal of Experiment 1 (Bradley and Lang, 2000) was to describe the two-dimensional (Pleasure and Arousal) distribution of a collection of sounds that engage a broad range of emotional responses in comparison to the distributions previously obtained for picture stimuli, and to understand the relationship between these affective dimensions. The sound stimuli's characteristics are comparable to that of a previously studied set of pictures, which was then used for reference and will henceforward be referred to as IAPS (i.e., International Affective Picture System, Lang, Bradley, & Cuthbert, 1999). In Experiment 1, a hundred and sixteen subjects rated sixty sounds on the dimensions of pleasure, arousal, and dominance using the Self-Assessment Manikin (on figure 4, (Bradley and Lang, 1994)), an affective system created by Lang (1980), which was followed by a free recall task.

In most dimensions, the results were very similar to the IAPS' pleasure ratings. Figure 5 (Bradley and Lang, 2000) shows the distribution of affective space for the picture stimuli, illustrating their progression from a calm arousal state toward either a high-arousal pleasant or a high-arousal unpleasant quadrant. Hence, the results of Experiment 1 indicate that most sounds rated as highly pleasant or highly unpleasant were also rated high in the arousal dimension. Experiment 1 also showed that sounds with similar semantic content to pictures from the IAPS were located very close together in the affective space.

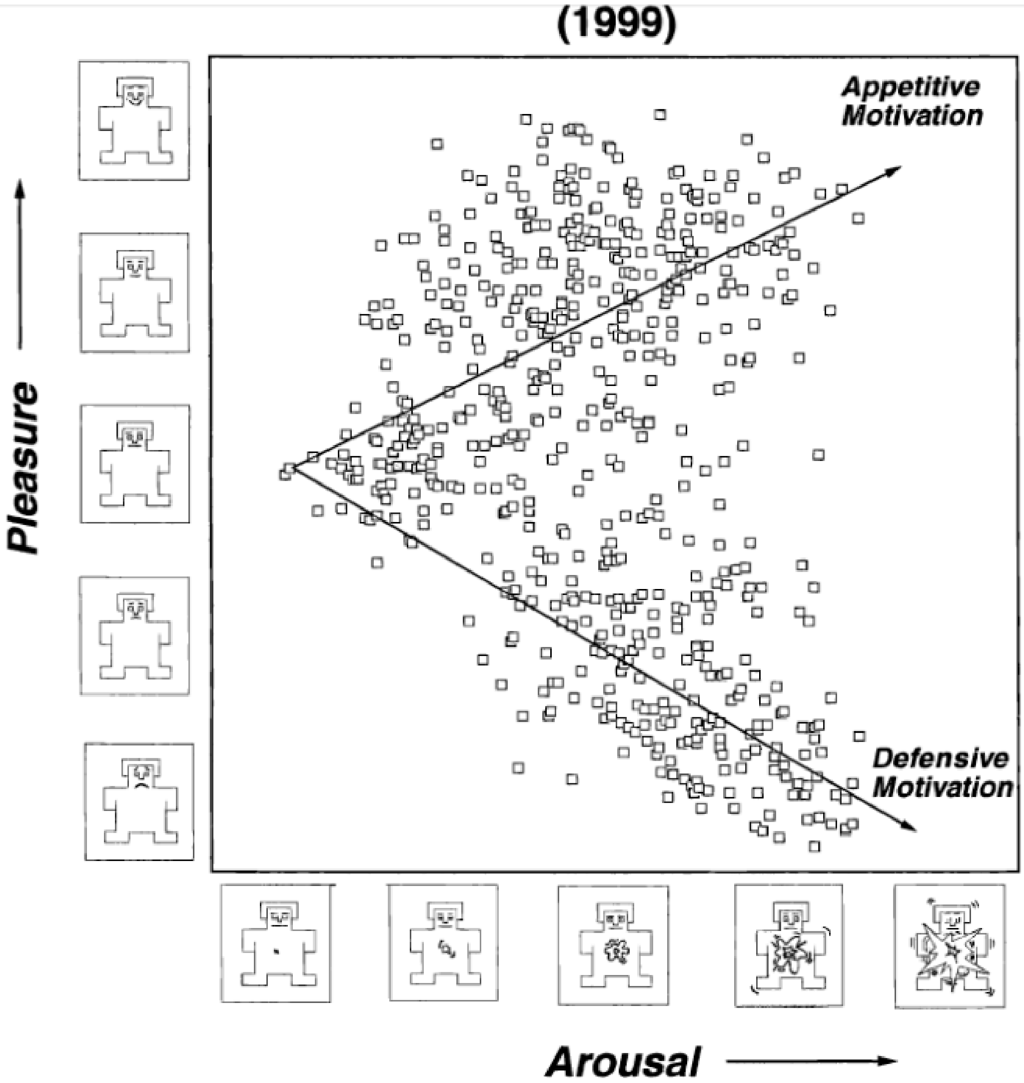

Figure 5: Distribution of affective space for the picture stimuli in Bradley and Lang's Experiment 1 (2000).

In Experiment 2 (Bradley and Lang, 2000), autonomic (heart rate, skin conductance), somantic (facial EMG), and startle reflexes were measured and recorded while subjects were listening and rating sounds, in order to understand effects of a priori valence on physiological reactions to affective sounds, as well as any variation between participants' reports of pleasure and arousal and their physiological response. While there were no significant effects involving sound intensity, it was verified that the a priori valence of the sound stimulus affected both pleasure and arousal ratings. Unpleasant sounds generated larger startle reflexes, and larger heart rate deceleration. Pleasant and unpleasant sounds were rated as more arousing than the neutral sounds. Figure 6 (Bradley and Lang, 2000), compares the results obtained between male and female subjects and then creates a relation concerning the visual and auditory stimuli.

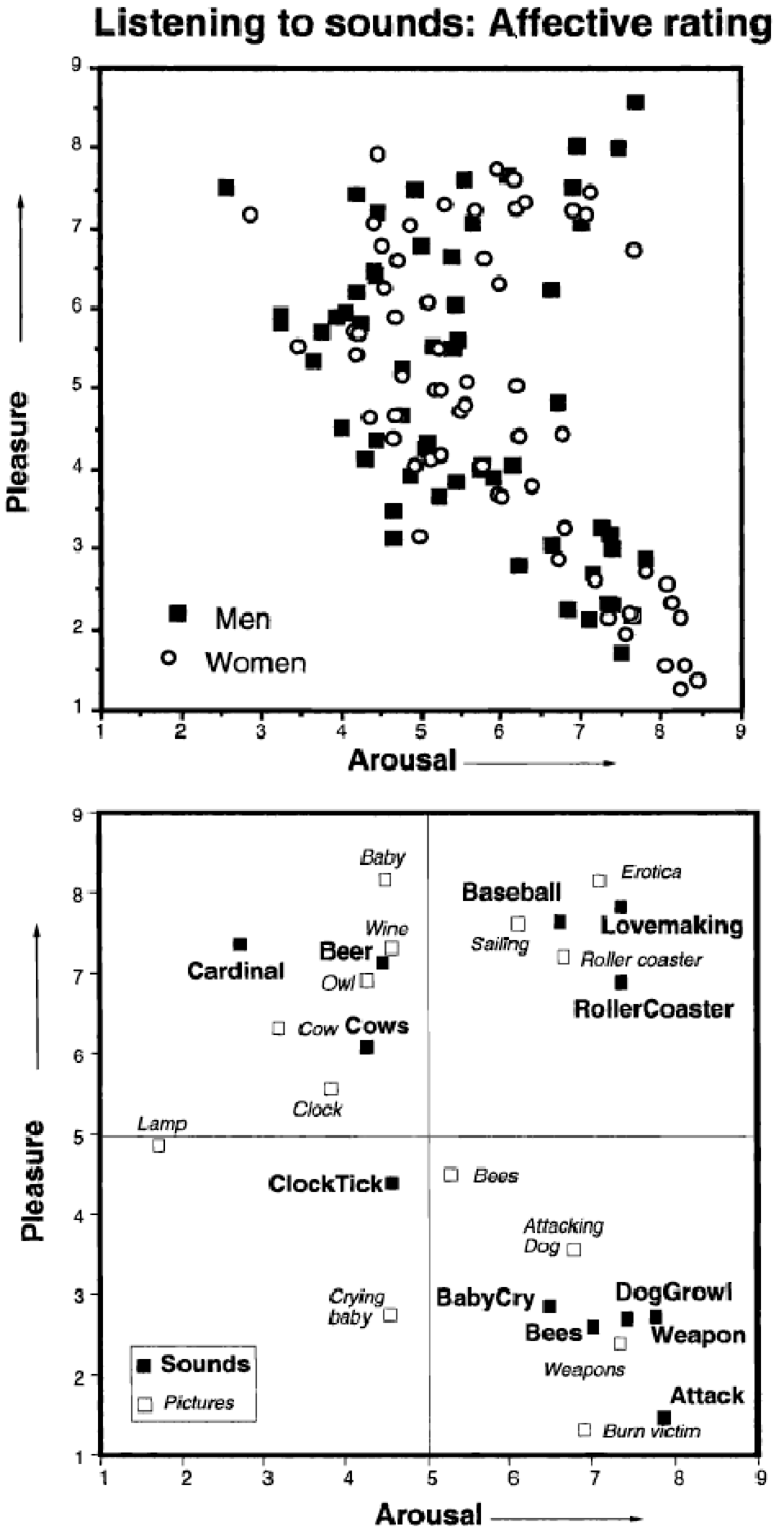

Figure 6: The top panel shows a distribution of the 60 sounds rated in Experiment 1 within the two-dimensional affective space (Arousal and Valence). The bottom panel compares results achieved with stimuli with similar semantic content from the IAPS database and the IADS database.

3. Exploring A Causal Deniability Within An Emotional Taxonomy Of Sound

The most usual means of reporting perceived emotional expressions are free phenomenological reports[4], choice among provided descriptive terms by the investigator, or ratings of how well descriptive terms apply to the excerpt in question. These different options have a very high effect on the possible described results. When listeners are asked to provide free descriptions, the reported emotions are many and with a lot variabilities between each other. Thus, instead of free reports, investigators tend to use lists of descriptive terms and ask the subjects to choose the appropriate ones. Both Rigg (1937) and Juslin (1997) found that the stated emotions were frequently less accurate when the free description method was used and that there was less variability when applying the forced choice method. Also, listener agreement appears to be greater for some emotions than other, suggesting that music (and sound) can convey some emotions, but not others, as can be seen on Table 2 (Juslin and Laukka, 2004).

Emotion	Freq.	Emotion	Freq.	Emotion	Freq.
Joy	99% (98%)	Pride	71% (69%)	Curiosity	46% (63%)
Sadness	91% (91%)	Pain	70% (86%)	Boredom	45% (47%)
Love	90% (89%)	Desire	69% (74%)	Disappointment	43% (49%)
Calm	87% (89%)	Hope	67% (70%)	Guilt	42% (43%)
Anger	82% (83%)	Nostalgia	67% (76%)	Satisfaction	42% (57%)
Tenderness	82% (86%)	Fear	63% (79%)	Admiration	37% (37%)
Longing	77% (71%)	Contempt	55% (53%)	Jealousy	35% (42%)
Solemnity	76% (73%)	Tiredness	55% (52%)	Sympathy	34% (39%)
Anxiety	75% (90%)	Regret	53% (56%)	Shame	31% (39%)
Hate	74% (69%)	Expectancy	51% (66%)	Trust	30% (33%)
Humour	74% (87%)	Confusion	49% (65%)	Interest	29% (44%)
Loneliness	73% (79%)	Disgust	47% (51%)	Humiliation	28% (31%)
Tension	72% (89%)	Surprise	47% (68%)	Other	10% (16%)

Note. Numbers within parantheses indicate the corresponding results from a questionnaire study of 135 musicians who received the same question (Lindström et al., 2003).

Table 2: Answers to the question "What emotions can music express?" Used with permission.

[4] Phenomenological reports describe personal experiences.

Thus, the first aim of the following experiment is to understand whether the listener has a primal need to connect a sound to its cause. The second goal is to understand if the abovementioned link is ever questioned or if it constitutes an automatic reaction. To this aim, Experiment 1 will show the same sound accompanied by different pictures to try to generate some confusion in the participants. The beginning of Chapter 6 will conclude Experiment 1, displaying studies with results that link sounds' semantic content to the elicited emotions of the listeners. After this section, Chapter 6 will also introduce Experiment 2, which will be presented later in this dissertation.

5. EXPERIMENT 1

1. Participants

Forty-two Portuguese volunteers (twenty-three males) took part in the test. Their ages varied between eighteen and thirty-two years (24.8 on average).

2. Procedure

There were two separate groups in this experiment. Both groups rated the same ten sounds alongside ten different pictures. Ten out of the total twenty pictures correspond to the sound source, while the other ten pictures correspond to events or objects that the investigators consider to produce a similar enough sound to create ambiguity.

The listening experiment took place in an acoustically isolated studio, where each participant was tested alone. Sounds were reproduced through closed-back headphones and participants rated the sounds by checking boxes using paper and pen. The stimulus order of presentation was varied and no significant effect was found. The participants were given instructions before the beginning of the experiment and were able to hear each sound as many times as they wanted to.

This experiment had two different stages. In the first stage, a group was assigned to each participant. Two months later, all fourty-two volunteers were called back and assigned to the other group, for further confirmation on the initial responses.

3. Stimuli

The auditory stimuli used in this experiment were either selected from the Auditory Lab database or recorded by the investigator. Http://www.auditorylab.org/ is a large freely accessible database of environmental sounds, which include several variations on basic auditory events. They are complemented with information such as the sound's location, source, spectral centroid, duration, harmonicity, time, and distance from the microphone, recording level, and semantic attributes. The criteria for selection were to make sure that the chosen stimuli would be able to trigger more than one emotional response. Every recording was edited so it would last two seconds. A five millisecond fade-in and fade-out was applied to prevent any abrupt start or ending to the sounds. All sounds were stereo files in WAV-format (44.1khz, 16bit).

The twenty pictures were extracted from Google Images and were not altered in any way.

Participants were asked to associate one of the provided pairs of adjectives with each of the sound-image pairs. The nine pairs of adjectives were: humiliating/embarrassing, comforting/serene, triumphant/exciting, happy/cheerful, majestic/emphatic, sad/melancholic, humorous/whimsical, sacred/serious, and aggressive/intense. An additional reason for this selection of adjectives is that their meaning does not change when literally translated to Portuguese. This is significant since all participants were Portuguese native speakers. The following table shows the sound-image pairings in both groups, which were randomly selected:

Sound	Image in group A	Image in group B
Shaking water bottle	Shaking water bottle	Plunger in a toilet
Bubbles in a milkshake	Water boiling in a pan	Bubbles in a milkshake
Rain	Rain	Human body covered in wasps
Watermelon being cut	Watermelon being cut	Goat with a knife aiming at its neck
Teeth being brushed	Teeth being brushed	Toilet brush
Breaking raw spaghetti	Fireplace	Breaking raw spaghetti
Closing the metal lock on a briefcase	Opening the slide on a semi-automatic pistol	Closing the metal lock on a briefcase
Cracking a lobster	Athlete injuring his leg	Cracking a lobster
Spray	Spray	Smoke grenade
Pouring water into a glass	Toilet	Pouring water into a glass

Table 3: Sound-image pairings for groups A and B.

4. Results

Table 1 shows the emotions with most agreement for each of the groups and in both test rounds. The emotional perception of the acoustic stimuli changed along with the visual stimuli in most scenarios. Most results are very linear: "Shaking water bottle" was described as "Humorous, Whimsical" when the picture showed a man shaking a water bottle but described as "Humiliating, Embarrassing" when the picture showed a plunger in a toilet. "Bubbles in a milkshake" and "Rain" both had high agreement for "Aggressive, Intense" and

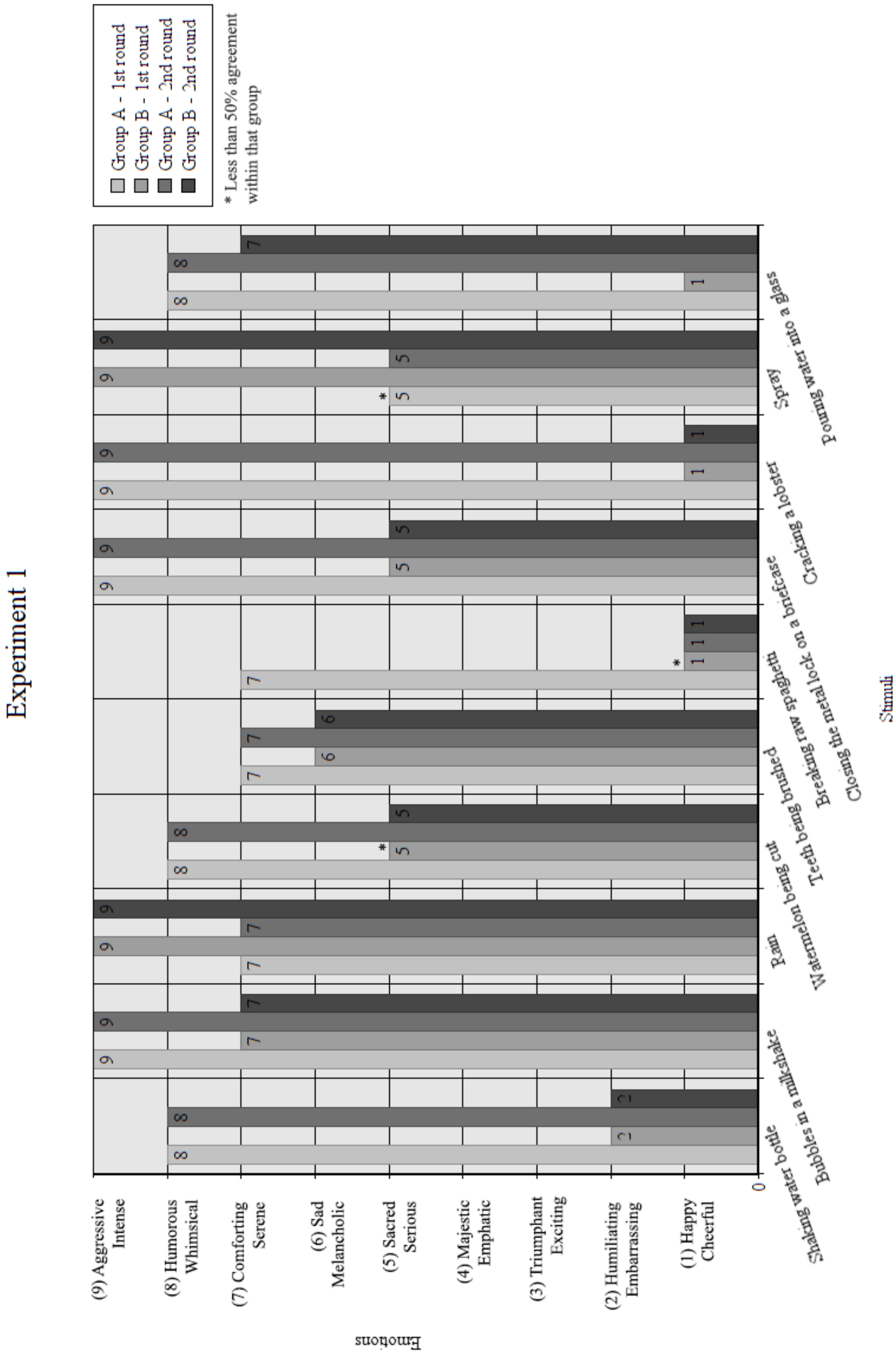

Figure 7: Results with the highest agreement for each group

"Comforting, Serene" in the different groups. "Closing the metal lock on a briefcase" and "Spray" both had high agreement for "Aggressive, Intense" and "Sacred, Serious" in the different groups, with a small exception for Group A on the first round of "Spray". In this group, "Sacred, Serious" could not gather more than 50% agreement. The second highest voted emotion was "Happy, Cheerful", which further shows the gap in emotional perception between the former and "Aggressive, Intense", the emotion terms with highest agreement in the other group. "Watermelon being cut" was described as "Humorous, Whimsical" and "Sacred, Serious" for Group A and B, respectively. The second highest voted emotion for "Watermelon being cut" in Group B on the first round (where "Sacred, Serious" could not gather more than 50% agreement) was "Aggressive, Intense", which further shows the gap in emotional perception between the former and "Humorous, Whimsical", the emotion terms with highest agreement in the other group. "Teeth being brushed" had highest agreement with "Comforting, Serene" in Group A and with "Sad, Melancholic" in Group B. "Cracking a lobster" had highest agreement with "Aggressive, Intense" for Group A and with "Happy, Cheerful" in Group B. "Pouring water into a glass" had highest agreement with "Humorous, Whimsical" in Group A and with "Happy, Cheerful" for Group B in the first round and "Comforting, Serene" for the second group of the same group.

"Breaking raw spaghetti" is the only sound stimulus, which does not change noticeably along with the visual stimulus. The second highest voted emotion in Group B on the first round (where "Happy, Cheerful" could not gather more than 50% agreement) was "Comforting, Serene", which was the same as Group A in the same round despite having a different visual stimulus. This can all probably be explained by the selected visual stimuli. It is likely that the participants have a positive emotional perspective on both visual actions: breaking raw spaghetti and a lit fireplace.

"Triumphant, Exciting" and "Majestic, Emphatic" did not achieve high agreement with any of the stimuli.

This experiment's results show a clear link between a sound's semantic content and the listener's elicited emotion. The following chapter will explore some similar conclusions.

6. "EVENT SIMILARITY AND SOUND SIMILARITY ARE ALTERNATIVE WAYS OF DESCRIBING THE SAME WORLD" (GUYOT, 1996)

1. Affective Acoustic Ecology And Concluding Experiment 1

Drossos, Floros, and Kanellopoulos (2012) define affective acoustic ecology as the relation between the sound events that surround the listener and the emotions that these can cause on him. In their study, the International Affective Digital Sounds (IADS) database was used. This database's affective annotation follows an emotions' model frequently used in MRI (Magnetic Resonance Imaging) and MER (Microelectrode Recording) researches, which can be divided in two categories: discrete and continuous. Examples of a continuous approach to emotions' models are dimensional models, such as the SAM (Self-Assessment Manikin). Discrete models describe emotions through specific or groups of words i.e., basic emotions (e.g., happiness, sadness, anger) or lists of adjectives. Discrete models are granted integrity since a wide range of studies can use the same words to describe an emotion. The following list of adjectives represent an alternative discrete modeling approach to the latter. In this case, instead of single words describing basic emotions, word groups are used, with several synonyms in the same group. Hevner (1936) created eight adjectives groups (Figure 8), while Li & Ogihara (2003) added some more adjectives and created thirteen different groups (Table 4).

A	cheerful,gay,happy	H	dramatic, emphatic
B	fanciful, light	I	agitated, exciting
C	delicate,graceful	J	frustrated
D	dreamy,leisurely	K	mysterious, spooky
E	longing, pathetic	L	passionate
F	dark, depressing	M	bluesy
G	sacred, spiritual		

Table 4: Thirteen adjective groups.

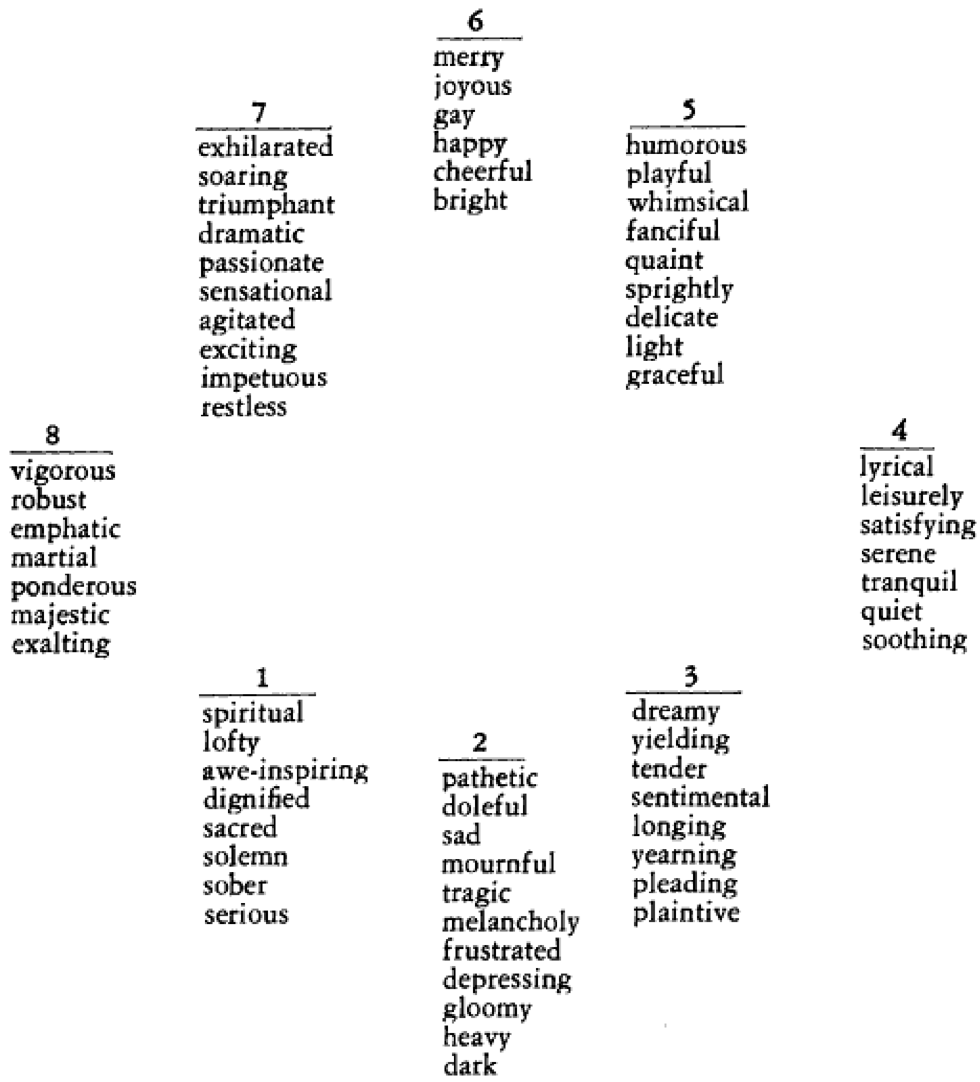

Figure 8: Eight adjective groups.

The goal of the experiment was to analyze the sound events in the IADS database. The sounds' technical features were extracted and classification algorithms were employed in accordance to the Arousal-Valence emotional plane. The accuracy obtained from the employed classification algorithms was below 50%. Typical results for accuracy classification derived by existing published works had the scores: 82%, 83,5%, 73%, or 67%. This shows that the semantic content in these sounds is the sole responsible for the elicited emotions of the listeners. In addition, it can be observed that there is no high relation between the energy of a signal and the arousal dimension, unlike what is verified in many experiments.

2. After Sound Events: Associating Perceived Emotions With Sound's Physical Features

A vast section of the research connecting affective reactions to everyday sounds has been focusing on sound properties and their link to basic emotions (Västfjäll and Kleiner, 2002). However, Tajadura-Jimenez and Västfjäll (2008) suggest that a larger number of categories (e.g., physical, psychological, spatial, and cross-modal determinants) would describe this relation more accurately. Drossos et al. (2013), for example, conducted an experiment with the goal of investigating the connection between rhythm-related characteristics of common sound events and the listener's arousal. The sound stimuli consisted of one hundred and sixty seven sound events with semantic content from the IADS database. To this aim, three training algorithms were conducted in the classification task: ANN implementations, Logistic Regression, and the K-Nearest-Neighbour technique. High accuracy results were obtained (the lowest accuracy score was 71.26%), which demonstrate that the rhythm of a sound stimulus can affect the listener's arousal, illustrating the idea that the connection between the two is applicable to sound events other than music.

Kirandziska and Ackovska (2012) carried out an experiment in order to achieve a model of positive and negative emotions according to some specific sound properties such as tempo, amplitude, and pitch. The stimuli consisted of twenty sound excerpts. In all of them, an actor was recorded saying, "I feel good". Each sound was segmented so that each segment had information for about twenty-five milliseconds. Then, several algorithms were used for extracting pitch, amplitude, and tempo. From the information gathered from the previous algorithms, a classification model was built and used for testing. The results indicate a close relation between negative emotions and low mean amplitude, high max amplitude, and low mean pitch, which have some correlation with the results found in some neuron and psychological studies that associate negative emotions with high max amplitude and low tempo.

Scherer et al. (1991) split emotions into ten different types: happiness, pleasant, fear, activity, anger, potency, boredom, disgust, surprise, and sadness. A set of sound features was attributed to each of these emotions. Table 5 (Scherer et al., 1991) shows the main sound characteristics for both positive and negative emotions.

Emotion type	Tempo	Pitch	Amplitude
Positive	Fast tempo	Pitch contour down Large pitch variation	Law mean amplitude
Negative	Slow tempo	Pitch contour up High pitch Small pitch variation	High mean amplitude

Table 5: The main sound characteristics for both positive and negative emotions.

Västfjäll (2012) performed an experiment with the aim of understanding whether emotional reactions to tone and noise complexes vary in the two-dimensional emotional experiential space and to find self-reported and physical correlates to valence and activation. Tone and noise complexes varied in both valence and activation. It was also shown that valence reactions are mainly affected by loudness and that the activation reactions are mainly affected by perceived sharpness of the sound. The sounds used in the aforementioned experiment were all stationary and devoid of meaning, showing that some physical characteristics can help the understanding of how sounds with little affective meaning can induce an emotional response in the listener, as was shown in previous research (Bradley and Lang, 2000).

In the following experiment, the aim is to understand how important the semantic content of auditory stimuli really is, in relation to the listeners' perceived emotions. At this stage, the stimuli will be merely acoustic, as it will be explained in the following chapter. The chapter that follows Experiment 2 is the final chapter of this dissertation: Conclusion. This final chapter will present investigations and studies that combine auditory semantic content with sound's physical properties and emotional studies. Furthermore, both Experiment 1 as well as Experiment 2 will be discussed in combination with the abovementioned investigations.

7. EXPERIMENT 2

1. Participants

Fifty-three Portuguese volunteers (twenty-six males) took part in the test. Their ages varied between eighteen and fourty-six years (25.8 on average).

2. Stimuli

The sixty sounds used in this experiment were randomly selected from the International Affective Digital Sounds (IADS) database. This database includes six-second long sound events, which correspond to a large variety of everyday situations. In order to mask the sounds' more obvious causal features, the investigator altered half of the sounds through frequency manipulation using SPEAR[5].

3. Procedure

There were two separate surveys in this experiment. Each survey presented thirty sounds in a random order; fifteen of which were altered. For each sound excerpt, participants were asked to rate the stimulus' valence in a five-point Likert scale[6] with the following options: "Very Negative", "Negative", "Neutral", "Positive", and "Very Positive"; and they were also asked to associate the sound excerpt with one of the provided pairs of adjectives, which were already used in Experiment 1: humiliating/embarrassing, comforting/serene, triumphant/exciting, happy/cheerful, majestic/emphatic, sad/melancholic, humorous/whimsical, sacred/serious, and aggressive/intense. Initially, there was also a 5-point Likert scale rating the stimuli's arousal but it was removed because the Portuguese expressions for arousal and valence are often mixed up.

SurveyGizmo hosted the surveys, which were sent via email from the investigator to the participants. Instructions were provided throughout the survey and participants were able to

[5] SPEAR (Sinusoidal Partial Editing and Resynthesis) is an audio editing and analysis software.

[6] Likert scales are self-report methods often used in questionnaires that consist of bipolar scaling methods gathering emotion concepts.

hear each sound as many times as they wanted to. Only solicited answers were accepted and partially completed surveys were declined.

4. Results

Table 6 shows the valence options with highest agreement and compares the altered stimuli with the non-altered ones. Only four out of the thirty altered stimuli had positive feedback, while thirteen out of the thirty non-altered stimuli had positive feedback. Neutral results were very similar in both (ten on the altered stimuli and nine on the non-altered stimuli).

Valence	Altered Stimuli	Non-altered Stimuli
Very Negative	1	1
Negative	15	7
Neutral	10	9
Positive	4	11
Very Positive	0	2

Table 6: Valence options with the highest agreement in both altered and non-altered stimuli.

Table 7 shows the results with highest agreement with the pairs of adjectives. The non-altered stimuli show very broad results, while the altered stimuli had no answers with high agreement for the following pairs of adjectives: "Humiliating, Embarrassing", "Triumphant, Exciting", "Majestic, Emphatic", and "Humorous, Whimsical", which leads to the belief that there are some emotions that cannot be induced by a sound's physical properties.

Pairs of Adjectives	Altered Stimuli	Non-altered Stimuli
Happy, Cheerful	5	5
Humiliating, Embarrassing	0	3
Triumphant, Exciting	0	3
Majestic, Emphatic	0	1
Sacred, Serious	3	2
Sad, Melancholic	5	3
Comforting, Serene	1	4
Humorous, Whimsical	0	2
Aggressive, Intense	16	7

Table 7: Pairs of adjectives with the highest agreement in both altered and non-altered stimuli.

Table 8 displays the names of the sounds that were extracted from the IADS database, shows which stimuli were altered and which stimuli were not altered, and reveals which were the answers with highest agreement in valence and adjective pairs for each stimulus.

This experiment's results show a noticeable disparity in the range of answers with highest agreement between altered and non-altered stimuli. The investigations featured in the next chapter will explore this disparity, as well as the link between a sound's semantic content and the listener's elicited emotion, which was suggested in Experiment 1.

Stimuli	Altered	Valence	Pairs of Adjectives	Stimuli	Altered	Valence	Pairs of Adjectives
Attack1	No	Very Negative	Aggressive, Intense	**HeartBeat**	No	Neutral	Sad, Melancholic
Attack2	Yes	Negative	Aggressive, Intense	**Helicopter1**	No	Neutral	Aggressive, Intense
Attack3	Yes	Negative	Sad, Melancholic	**Jackhammer**	Yes	Negative	Aggressive, Intense
BabyCry	No	Negative	Sad, Melancholic	**Jet**	No	Neutral	Majestic, Emphatic
Beer	No	Positive	Happy, Cheerful	**Laughing**	No	Very Positive	Happy, Cheerful
Bomb	No	Negative	Aggressive, Intense	**Lawnmower**	No	Neutral	Aggressive, Intense
BoyLaugh	Yes	Neutral	Happy, Cheerful	**ManWheeze**	No	Negative	Aggressive, Intense
Brook	No	Positive	Comforting, Serene	**Musicbox**	No	Positive	Happy, Cheerful
BrushTeeth	Yes	Neutral	Happy, Cheerful	**NoseBlow**	No	Negative	Humiliating, Embarrassing
CarHorns	Yes	Negative	Aggressive, Intense	**Office1**	No	Negative	Aggressive, Intense
Carousel	Yes	Positive	Happy, Cheerful	**Paint**	No	Positive	Humorous, Whimsical
CarWreck	Yes	Negative	Aggressive, Intense	**Panting**	No	Neutral	Aggressive, Intense
Cattle	No	Positive	Humorous, Whimsical	**Paper2**	Yes	Negative	Aggressive, Intense
ClagGame	Yes	Neutral	Sad, Melancholic	**Pig**	Yes	Neutral	Sad, Melancholic
CorkPour	Yes	Neutral	Happy, Cheerful	**Polaroid**	No	Negative	Happy, Cheerful
CourtSport	No	Positive	Comforting, Serene	**Rain1**	No	Neutral	Sad, Melancholic
Crowd2	No	Positive	Triumphant, Exciting	**Robin**	Yes	Positive	Comforting, Serene
DentistDrill	Yes	Positive	Aggressive, Intense	**Seagull**	No	Very Positive	Comforting, Serene
Electricity	Yes	Negative	Aggressive, Intense	**Sink**	Yes	Positive	Sacred, Serious
EngineFailure	No	Negative	Humiliating, Embarrassing	**Thunderstorm**	Yes	Neutral	Aggressive, Intense
EroticCouple	Yes	Negative	Aggressive, Intense	**TireSkids**	Yes	Negative	Aggressive, Intense
EroticFem1	No	Positive	Triumphant, Exciting	**Toilet**	No	Neutral	Humiliating, Embarrassing
EroticMale1	No	Positive	Happy, Cheerful	**Train**	No	Positive	Triumphant, Exciting
Explosion	Yes	Neutral	Aggressive, Intense	**Typewriter**	Yes	Neutral	Aggressive, Intense
Fan	No	Neutral	Comforting, Serene	**Victim**	Yes	Negative	Aggressive, Intense
FemaleCough	Yes	Negative	Humiliating, Embarrassing	**Vomit**	Yes	Negative	Aggressive, Intense
Fight2	Yes	Negative	Happy, Cheerful	**Walking**	No	Neutral	Sacred, Serious
GlassBreak	Yes	Negative	Aggressive, Intense	**War**	Yes	Very Negative	Aggressive, Intense
Growl1	Yes	Neutral	Sacred, Serious	**Wind**	Yes	Negative	Sad, Melancholic
GunShot	Yes	Neutral	Sacred, Serious	**Writing**	No	Positive	Sacred, Serious

Table 8: Stimuli selected from the IADS database, altered and non-altered, followed by the answers with the highest agreement for both valence and adjective pairs.

8. CONCLUSION

1. Somewhere New: Where Sound's Causal And Physical Properties, The Listener, And The Context Come Together To Unveil Specific Emotions

Asutay et al. (2012) carried out an experiment in order to show that auditory-induced emotions depend not only on the physical characteristics of a specific sound, but also on the meaning attributed to its cause, which reinforces Experiment 1's suggestion that there is a link between a sound's semantic content and the listener's elicited emotion. In the first part of the conducted experiment, participants rated all the stimuli that were processed and then rated the original stimuli presented in a different order without knowing that there were two sets of stimuli. In the second part of the experiment, a different group of participants rated only the processed sounds, which suffered both spectral and temporal changes, in order to make them unidentifiable. Since emotional responses are often learned over time, they inevitably generate associations with life events when the right stimulus is showed. By reducing a sound source's identifiability, the investigators tried to interrupt these associations, similarly to the masking performed in Experiment 2 earlier in this dissertation. A third set of participants was asked to rate how they felt during each processed sound. A nine-point SAM scale was used in all three settings of the experiment, in which the participants rated how annoying sounds were from 1 (not at all annoying) to 9 (very much annoying) and also on a perceived loudness scale from 1 (not loud) to 9 (very loud). The stimuli consisted of eighteen sounds selected from the IADS database. They were all six-second long recordings of everyday events, none of which contained music or erotica. The selected sounds were: Screaming Woman, Alarm, Bees, Growling Dog, Pigs, Jackhammer, Cuckoo Clock, Cat, Ticking Clock, Helicopter, Frogs by a Lake, Toilet Flush, Cows, Carousel, Rollercoaster, Beverage, Bottle Opening, Laughter, and Applause. Parallel to what was observed previously in Experiment 2, it was verified that identifiability decreased dramatically due to processing (91% to 11%). Priming the processed stimuli caused identifiability to increase significantly, to around 78% correct identification. Normative mean valence and arousal rating can be verified in Table 9 (Asutay et al., 2012). The reduction in identifiability was expected to cause larger changes in auditory-induced emotion for stimuli labelled pleasant or unpleasant, compared to neutral. Indeed, "one can say that processed stimuli were emotionally neutral" (Asutay et al. 2012).

Category	Sound	Normative mean valence rating	Normative mean arousal rating
Unpleasant	Screaming woman	1.74	8.07
	Alarm	2.38	7.96
	Bees	2.58	7.01
	Growling dog	2.73	7.77
	Pigs	3.56	6.38
	Jackhammer	3.66	5.60
Neutral	Cuckoo clock	4.23	6.19
	Cat	4.31	5.47
	Ticking clock	4.38	4.56
	Helicopter	5.19	5.50
	Frogs by a lake	5.49	4.51
	Toilet flush	5.57	3.37
Pleasant	Cows	6.09	4.27
	Carousel	6.21	5.43
	Rollercoaster	6.90	7.36
	Beverage bottle opening	7.13	4.43
	Laughter	7.32	6.37
	Applause	7.69	5.57

Table 9: Auditory stimuli and respective normative mean valence and arousal ratings.

Indeed, in Experiment 2, twenty-five out of thirty altered auditory stimuli were considered to be either negative or neutral. There was also much less variability in pairs of adjectives, when compared to the stimuli that did not have their semantic content masked. Nevertheless, and although several studies have tried to find a link between some of sound's physical properties and specific emotions (Västfjäll, 2003; Berglund, Berglund, & Lindvall, 1975; Landström et al., 1995; Björk, 1986; Hiramatsu, Takagi, Yamamoto, 1983; Björk, 1999), Neuhoff (2004) suggested that sound events, alongside their specific context, may constitute a more accurate approach to what is actually perceived by the listener. Indeed, it was verified in Experiment 1 that the semantic content of the utilized stimuli created some intense changes in perception (p.e., high agreement for Agressive, Intense for "Water boiling in a pan" and high agreement for Comforting, Serene for "Bubbles in a milkshake"). It is also argued by Juslin & Västfjäll (2008) that the listener and the context are as important as a sound's physical characteristics when it comes to understanding the underlying emotions in sound.

Although neuropsychology and neuroscience constitute invaluable assets in the present and future of this field of study, it is of great importance to understand that as long as there is not full understanding of emotion in human minds, classifying it with high accuracy will remain an elusive goal.

REFERENCES AND BIBLIOGRAPHY:

Andringa, T. C., & Niessen, M. E. (2006). Real-world sound recognition: A recipe. *Proceedings of LSAS*, 106-118.

Asutay, E., Västfjäll, D., Tajadura-Jiménez, A., Genell, A., Bergman, P., & Kleiner, M. (2012). Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design. *Journal of the Audio Engineering Society*, *60*(1/2), 21-28.

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of experimental psychology: human perception and performance*, *19*(2), 250.

Ballas, J. A., & Howard, J. H. (1987). Interpreting the language of environmental sounds. *Environment and behavior*, *19*(1), 91-114.

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, *12*(5), 1161.

Berglund, B., Berglund, U., & Lindvall, T. (1975). Scaling loudness, noisiness, and annoyance of aircraft noise. *The Journal of the Acoustical Society of America*, *57*(4), 930-934.

Bergman, P., Sköld, A., Västfjäll, D., & Fransson, N. (2009). Perceptual and emotional categorization of sound. *The Journal of the Acoustical Society of America*, *126*(6), 3156-3167.

Bergman, P., Västfjäll, D., Asutay, E., Tajadura-Jiménez, A., Sköld, A., Genell, A., & Fransson, N. (2008). Emoacoustics: sound character versus source meaning in emotional responses to sounds. In *Icben 2008*.

Björk, E. A. (1985). The perceived quality of natural sounds. *Acta Acustica united with Acustica*, *57*(3), 185-190.

Björk, E. A. (1986). Laboratory annoyance and skin conductance responses to some natural sounds. *Journal of sound and vibration*, *109*(2), 339-345.

Björk, E. A. (1999). Startle, annoyance and psychophysiological responses to repeated sound bursts. *Acta Acustica united with Acustica*, *85*(4), 575-578.

Bonebright, T. L. (2001). Perceptual structure of everyday sounds: A multidimensional scaling approach.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli.*Psychophysiology*, *37*(02), 204-215.

Bruner, J. S. (1990). *Acts of meaning*. Harvard University Press.

Campbell, I. G. (1942). Basal emotional patterns expressible in music. *The American Journal of Psychology*, 1-17.

Capurso, A. (1952). *Music and your emotions*. Liveright Pub. Corp.

Chion, M. (1983). Guide to sound objects. *Pierre Schaeffer and Musical Research (English translation by Dack, J. and North, C.). Buchet/Chastel*.

Chion, M. (1994). The three listening modes. *The Sound Studies Reader*, 48-53.

Colombetti, G. (2009). From affect programs to dynamical discrete emotions. *Philosophical Psychology*, *22*(4), 407-425.

Cooke, D. (1959). The language of music. Oxford University Press.

Damasio, A. R. (2006). *Descartes' error*. Random House.

Darwin, C. (1872). The origin of species by means of natural selection.

Drossos, K., Floros, A., & Kanellopoulos, N. G. (2012, September). Affective acoustic ecology: towards emotionally enhanced sound events. In *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound* (pp. 109-116). ACM.

Drossos, K., Kotsakis, R., Kalliris, G., & Floros, A. (2013, July). Sound events and emotions: Investigating the relation of rhythmic characteristics and arousal. In *Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on* (pp. 1-6). IEEE.

Dubois, D. (2000). Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quarterly*, *1*(1), 35-68.

Ekman, P. (1973). Cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review*, 169-222.

Gabrielsson, A., & Juslin, P. N. (2003). *Emotional expression in music*. Oxford University Press.

Gang, M. J., & Teft, L. (1975). Individual differences in heart rate responses to affective sound. *Psychophysiology*, 12, 423–426.

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

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

Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *61*(1), 54.

Guastavino, C., & Cheminée, P. (2003). Une approche psycholinguistique de la perception des basses fréquences: Conceptualisations en langue, représentations cognitives et validité écologique. *Psychologie française*, *48*(4), 91-101.

Gundlach, R. H. (1935). Factors determining the characterization of musical phrases. *The American Journal of Psychology*, 47(4), 624–643.

Guyot, F. (1996). *Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative: une approche par la catégorisation* (Doctoral dissertation).

Guyot, F., Castellengo, M., & Fabre, B. (1997). Étude de la catégorisation d'un corpus de bruits domestiques. *D. Dubois, éditeur, Catégorisation et cognition: De la perception au discours*, 25.

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

Gygi, B., & Shafiro, V. (2010). Development of the database for environmental sound research and application (DESRA): Design, functionality, and retrieval considerations. *EURASIP Journal on Audio, Speech, and Music Processing*.

Hampton, P. J. (1945). The emotional element in music. *The Journal of general psychology*, *33*(2), 237-250.

Haverkamp, M. (2012). Visual representations of sound and emotion. IV International Conference Synaesthesia: Science And Art, Almeria, 16-19th February 2012

Hevner, K.. (1936). Experimental Studies of the Elements of Expression in Music. *The American Journal of Psychology*, *48*(2), 246–268.

Hiramatsu, K., Takagi, K., & Yamamoto, T. (1983). Experimental investigation on the effect of some temporal factors of nonsteady noise on annoyance. *The Journal of the Acoustical Society of America*, *74*(6), 1782-1793.

Jenkins, J. J. (1985). Acoustic information for objects, places, and events. *Persistence and* *change*, 115-138.

Juslin, P. N. (1997). Can results from studies of perceived expression in musical performances be generalized across response formats?. *Psychomusicology: A Journal of Research in Music Cognition*, *16*(1-2), 77.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217-238.

Juslin, P. N., & Sloboda, J. (Eds.). (2010). *Handbook of music and emotion: Theory, research, applications*. OUP Oxford.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, *31*(05), 559-575.

Kane, B. (2007). L'Objet Sonore Maintenant: Pierre Schaeffer, sound objects and the phenomenological reduction. *Organised Sound*, *12*(01), 15-24.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., ... & Turnbull, D. (2010, August). Music emotion recognition: A state of the art review. In *Proc. ISMIR* (pp. 255-266).

Kirandziska, V., & Ackovska, N. (2012). „Sound features used in emotion classification ". In *The 9th International Conference for Informatics and Information Technology* (pp. 91-95).

Landström, U., Åkerlund, E., Kjellberg, A., & Tesarz, M. (1995). Exposure levels, tonal components, and noise annoyance in working environments. *Environment International*, *21*(3), 265-275.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson & T. A. Williams (ed.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ: Ablex.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Instruction manual and affective ratings. *The center for research in psychophysiology, University of Florida*.

Laurier, C., Lartillot, O., Eerola, T., & Toiviainen, P. (2009). Exploring relationships between audio features and emotion in music. ESCOM, Conference of European Society for the Cognitive Sciences of Music.

Lazarus, R. S. (1991). *Emotion and Adaptation.* New York: Oxford University Press.

Li, T., & Ogihara, M. (2003, October). Detecting emotion in music. In *ISMIR*(Vol. 3, pp. 239-240).

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). “Expressivity comes from within your soul”: A questionnaire study of music students' perspectives on expressivity. *Research Studies in Music Education*, *20*(1), 23-47.

Maffiolo, V. (1999). *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain. Structuration des représentations mentales et influence sur l'appréciation qualitative. Application aux ambiances sonores de Paris* (Doctoral dissertation).

Marcell, M., Malatanos, M., Leahy, C., & Comeaux, C. (2007). Identifying, rating, and remembering environmental sound events. *Behavior research methods*, *39*(3), 561-569.

McAdams, S. (1993). Recognition of auditory sound sources and events.*Thinking in sound: the cognitive psychology of human audition (S. McAdams and E. Bigand, eds.), Oxford University Press*.

McDermott, J. M. (2008). *Evolutionary computation applied to the control of sound synthesis* (Doctoral dissertation, University of Limerick).

Mehrabian, A., & Russell, J. A. (1974). An approach to environmental psychology. Cambridge, MA: MIT Press.

Meyers, M., & Smith, B. D. (1986). Hemispheric asymmetry and emotion: Effects of nonverbal affective stimuli. *Biological Psychology*, 22, 11–22.

Neuhoff, J. G. (2004). *Ecological psychoacoustics*. Elsevier Academic Press.

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

Ozcan, E., & Jacobs, J. (2014). Product sounds: Basic concepts and categories. *International Journal of Design*, *8*(3).

Pallmeyer, T. P., Blanchard, E. B., & Kolb, L. C. (1986). The psychophysiology of combat-induced post-traumatic stress disorder in Vietnam veterans. *Behavioral Research and Therapy*, 24, 645–652.

Parsons, C. E., Young, K. S., Craske, M. G., Stein, A. L., & Kringelbach, M. L. (2014). Introducing the Oxford Vocal (OxVoc) Sounds database: a validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Front. Psychol*, *5*(562), 10-3389.

Raimbault, M. (2006). Qualitative judgements of urban soundscapes: Questionning questionnaires and semantic scales. *Acta acustica united with acustica*, *92*(6), 929-937.

Rigg, M. (1937). An experiment to determine how accurately college students can interpret the intended meanings of musical compositions. *Journal of Experimental Psychology*, *21*(2), 223.

Roseman, I. J., Spindel, M. S., & Jose, P. E. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of Personality and Social Psychology*, *59*(5), 899.

Schaeffer, P. (2004). Acousmatics. *Audio culture: Readings in modern music*, 76-81.

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them?. *Journal of new music research*, *33*(3), 239-251.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and emotion*, *15*(2), 123-148.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological review*,*61*(2), 81.

Schubert, E. D. (1975). The role of auditory perception in language processing. In D. D. Duane & M. B. Rawson (Eds.), *Reading, perception, and language.* Baltimore: York Press, 1974.

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, *53*(9), 1062-1087.

Schuller, B., Hantke, S., Weninger, F., Han, W., Zhang, Z., & Narayanan, S. (2012, March). Automatic recognition of emotion evoked by general sound events. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (pp. 341-344). IEEE.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*, 813–838.

Solomon, L. N. (1958). Semantic approach to the perception of complex sounds. *The journal of the Acoustical Society of America*, *30*(5), 421-425.

Solomon, L. N. (1959). Search for physical correlates to psychological dimensions of sounds. *The Journal of the Acoustical Society of America*, *31*(4), 492-497.

Sopchak, A. L. (1955). Individual differences in responses to different types of music, in relation to sex, mood, and other variables. *Psychological Monographs: General and Applied*, *69*(11), 1.

Susini, P., Misdariis, N., Winsberg, S., & McAdams, S. (1998). Caractérisation perceptive de bruits. *Acoustique et Techniques*, *13*(4), 11-15.

Tajadura-Jiménez, A., & Västfjäll, D. (2008). Auditory-induced emotion: A neglected channel for communication in human-computer interaction. In *Affect and Emotion in Human-Computer Interaction* (pp. 63-74). Springer Berlin Heidelberg.

Tomkins, S. S. (1962). Affect, imagery, consciousness: Vol. I. The positive affects.

VanDerveer, N. J. (1980). *Ecological acoustics: Human perception of environmental sounds* (Doctoral dissertation, ProQuest Information & Learning).

Västfjäll, D. (2003). *Affect as a component of perceived sound and vibration quality in aircraft*. Chalmers University of Technology.

Västfjäll, D. (2012). Emotional reactions to sounds without meaning. *Psychology*, *3*(8), 606.

Västfjäll, D., & Kleiner, M. (2002). Emotion in product sound design. *Proceedings of Journées Design Sonore*.

Watson, K. B. (1942). The nature and measurement of musical meanings. *Psychological Monographs*, *54*(2), i.

Wold, E., Blum, T., Keislar, D., & Wheaten, J. (1996). Content-based classification, search, and retrieval of audio. *MultiMedia, IEEE*, *3*(3), 27-36.

Wundt, W. (1980). *Outlines of psychology* (pp. 179-195). Springer US.

Yang, Y. H., Lin, Y. C., Su, Y. F., & Chen, H. H. (2008). A regression approach to music emotion recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, *16*(2), 448-457.