CHALMERS

Perceptive approach for sound synthesis by physical modelling

Master's Thesis in the Master's programme in Sound and Vibration

VICKY LUDLOW

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers Room Acoustics Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2008

Master's Thesis 2008:52

MASTER'S THESIS 2008:52

Perceptive Approach for Sound Synthesis by Physical Modelling

Vicky LUDLOW

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Room Acoustics Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2008

Perceptive approach for sound synthesis by physical modelling

Master's Thesis 2008:52

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Room Acoustics Group
Chalmers University of Technology
SE-41296 Göteborg
Sweden

Tel. +46-(0)31 772 1000

Cover:
Logo of the european CLOSED project (Closing the Loop of Sound Evaluation and Design) coordinated by IRCAM

Reproservice / Department of Civil and Environmental Engineering
Göteborg, Sweden 2008

Perceptive approach for sound synthesis by physical modelling

Vicky LUDLOW
Department of Civil and Environmental Engineering
Division of Applied Acoustics
Room Acoustics Group
Chalmers University of Technology

Abstract

This study is part of the European project CLOSED (Closing the Loop Of Sound Evaluation and Design, [http://closed.ircam.fr]). The CLOSED project aims at providing new tools to develop a methodology able to create and evaluate sound design products.
Among these tools, a set of sound synthesis models based on physical parameters have been developed so as to encourage sound creation. These models comprise for instance solid contact models (impact, friction...) and liquid models (bubbles...).
The goal of this master thesis is to contribute to the design of a perceptive interface for these sound synthesis models. It focuses on impact models and on material perception (more specifically on the four following classes: wood, metal, plastic and glass). The aim is then to perform a perceptive classification of synthesised impact sounds according to the material category in order to achieve a mapping between the physical space (the model parameters) and the perceptive space (the four material classes).
The many physical dimensions of the synthesis model and thus the infinity of possible sounds represent a challenge for this approach. Actually, classical perceptive experiment methods are not adapted to deal with such a large sound corpus. A possible field of investigation is active learning techniques that may solve this problem by reducing the required amount of sounds to define the boundaries between the material classes.
This report presents a study on a reduced parameter space, on which both classical perceptive method and active learning techniques can be carried out, so as to evaluate the latter method. Subsequently to this validation phase, a new experimental protocol driven by active learning procedures would allow achieving perceptive experiment on larger sound corpus.

Key words: material perception, psychoacoustic experiment, active learning, physical sound synthesis model, sound classification

Acknowledgements

I first would like to thank my two supervisors at IRCAM, Nicolas Misdariis and Olivier Houix for their friendliness and their availability. Their wise advices guided me throughout my stay at IRCAM.
My thanks also go to Kamil Adiloglu and Robert Annies from NIPG and to Stefano Papetti and Stefano Delle Monache from Univerona, whose patience and motivation allowed a rich collaboration in this multidisciplinary work.
I also would like to thank the Sound Design and Perception Team for their friendship and their helpfulness: Antoine, Guillaume, Hans and Patrick.
I feel very grateful towards Mendel Kleiner, Daniel Västfjäll and Wolfgang Kropp; they have been very encouraging throughout my master program at Chalmers University.
On a more personal side, I thank my parents for their encouragement and their help, for giving me the chance to go and study in Göteborg.
Finally, my thanks go to my master fellow students and especially to Arthur, Clément, Florent, Grégoire and Mathieu, with who I spent a wonderful year in Göteborg.

Contents

1 Introduction

This study is part of a European project named CLOSED (Closing the Loop of Sound Evaluation and Design) and aims at investigating a perceptive approach for sound synthesis by physical modelling, which would permit to control the physical parameters more intuitively. It contributes to the design of a perceptive interface for sound synthesis models.

Previous researches done at IRCAM by Derio [24] and Dos Santos [25] aimed at building such a perceptive by a dimensional approach: the goal of these studies was to define perceptive dimensions related to acoustic descriptors such as brightness or resonance for instance, that would control the sound synthesis model. However, these studies are limited by the fact that they cannot take many parameters into account, and a sound synthesis model can therefore only be characterized by 2 or 3 dimensions. Another possible approach is to consider perceptive classes.

The goal of this study is to perform a mapping between the physical space (the sound synthesis model parameters) and the perceptive space (perceptive classes to be defined). It will focus on material perceptive classes and on impact model sounds.

This approach meets obstacles. Actually, sound synthesis models incorporate many physical parameters and generate an infinite number of possible sounds. Classical perceptive experiments are not adapted to such a huge sound corpus because it demands to many sounds to scan the parameter space.

Machine learning techniques will be investigated so as to define whether they can solve this scale problem. Machine learning techniques would be profitable in the sense that a psychoacoustic experiment should not last more than 1 hour, and the amount of sounds listened by the participant is thus limited. By cleverly choosing the points to be evaluated by participants, it could permit to perform perceptive classification on large sounds corpus within a reasonable amount of time. Figure 1.1 makes this point explicit.

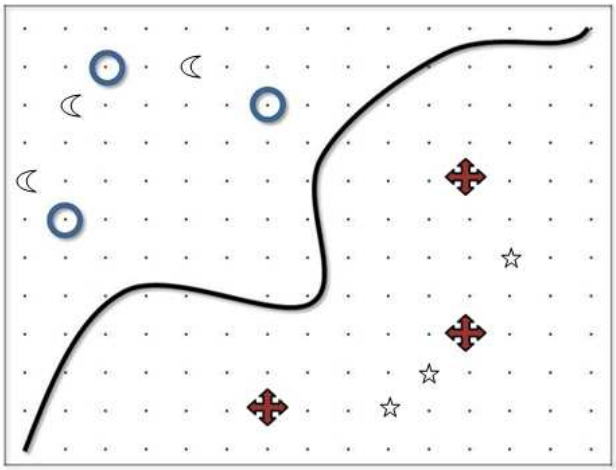

Figure 1.1: *Machine learning techniques driving perceptive classification experiment. The points are sounds in the parameter physical space. The automatic algorithm will not test every sound, especially sounds between two encircled points (stars and moons data points). Actually, the algorithm assumes these points to be "encircled points" as well and will concentrate on unknown regions to define the boundary between circled points and crossed points.*

To evaluate machine learning technique performances, both classical and automatic experiments will be carried out on a restricted corpus.
After having introduced the CLOSED project and detailed the main steps of the project, this report will present a literature review on material identification from impact sounds and then describe the sound synthesis physical models and the machine learning techniques. Results of both classical and automatic experiments will be reported.

1.1 The CLOSED project

The European CLOSED project is a three-year program that began on July 2006 and will be finished on June 2009. This project aims at providing a functional-aesthetic sound measurement tool that can be profitably used by designers. At one end, this tool will be linked with physical attributes of sound-enhanced everyday objects; at the other end it will relate to user emotional response. The measurement tool will be made of a set of easy-to-interpret indicators, which will be related to use in natural context, and it will be integrated in the product design process to facilitate the control of sonic aspects of objects, functionalities, and services encountered in everyday settings. The aim of the CLOSED project is to provide such concepts and tools, toward closing the loop of sound evaluation and design.
The design process implies an iterative loop, which compares the quality of a sound with pre-defined specifications and which evaluates and refines the sound creation until those specifications may be adequately met.

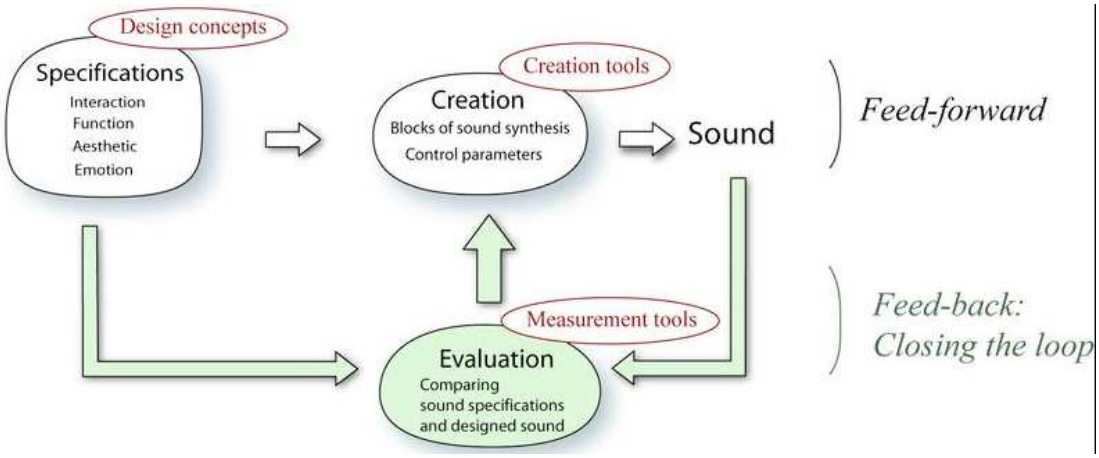

Figure 1.2: *Closing the Loop of Sound Evaluation and Design. The process implies an iterative loop which leads to the adequacy between the sound creation and pre-defined specifications.*

The CLOSED consortium incorporates four different expertises, ranging from physics and signal processing, design, acoustics and psychology of perception, to computer science. The four actors are:

1. IRCAM, Sound Perception and Design team, Institut de Recherche et Coordination Acoustique/Musique, Paris, France. The IRCAM is the coordinator of the project. It works on the sound design and psychoacoutic aspects of the project.

2. UNIVERONA, Vision, Image Processing and Sound laboratory, Dip. di Informatica, University of Verona, Verona, Italy. Univerona develops the physical sound synthesis models used in the CLOSED program.

3. ZHdK, Zürcher Hochschule der Künste, Department of Design and Institute for Cultural Studies in Art, Media and Design, Zurich, Switzerland. ZHdK works on the sound product design research.

4. NIPG, Neural Information Processing Group, Berlin University of Technology, Berlin, Germany. NIPG is the computer science section of the project; it develops active learning algorithms.

The coming section explains with more details the interaction work between IRCAM and the different partners involved in the present study.

1.2 Main steps of the study

A synopsis of this study is presented on figure 1.3. After having chosen the perceptive classes of interest and the sound synthesis physical model, the physical parameters are restricted so as to obtain a reasonable sounds corpus.

The classical perceptive experiment done on 20 participants results in labelled sounds: sounds (and thus the physical parameters of the sound synthesis model) are associated to perceptive classes (the materials: wood, glass, plastic or metal). This

result corpus is divided in groups: group 1 contains 70% of the results data for instance and group 2 contains 30% of the results data.
A machine learning algorithm is composed of two phases (explained in chapter 4): the training phase and the generalisation phase. Group 1 data are used to train the algorithm: it knows the answers given by the 20 participants and draws boundaries between the material classes. The generalisation phase will test whether these boundaries are correct or not. For this purpose, the algorithm will fake a virtual participant: it will be provided group 2 data (sounds corresponding to physical parameters) and will have to decide to which class these sounds belong to. It does not know the answers of the 20 participants.
The answers given by the algorithm for group 2 sounds ("group 2 relabelled") will be compared with the answers of the 20 participants ("group 2 labelled"). This comparison will permit to evaluate the ability of machine learning techniques. If machine learning techniques turn out to be powerful, then new experimental methods based on machine learning techniques would permit to perform large-scale classification perceptive experiments.

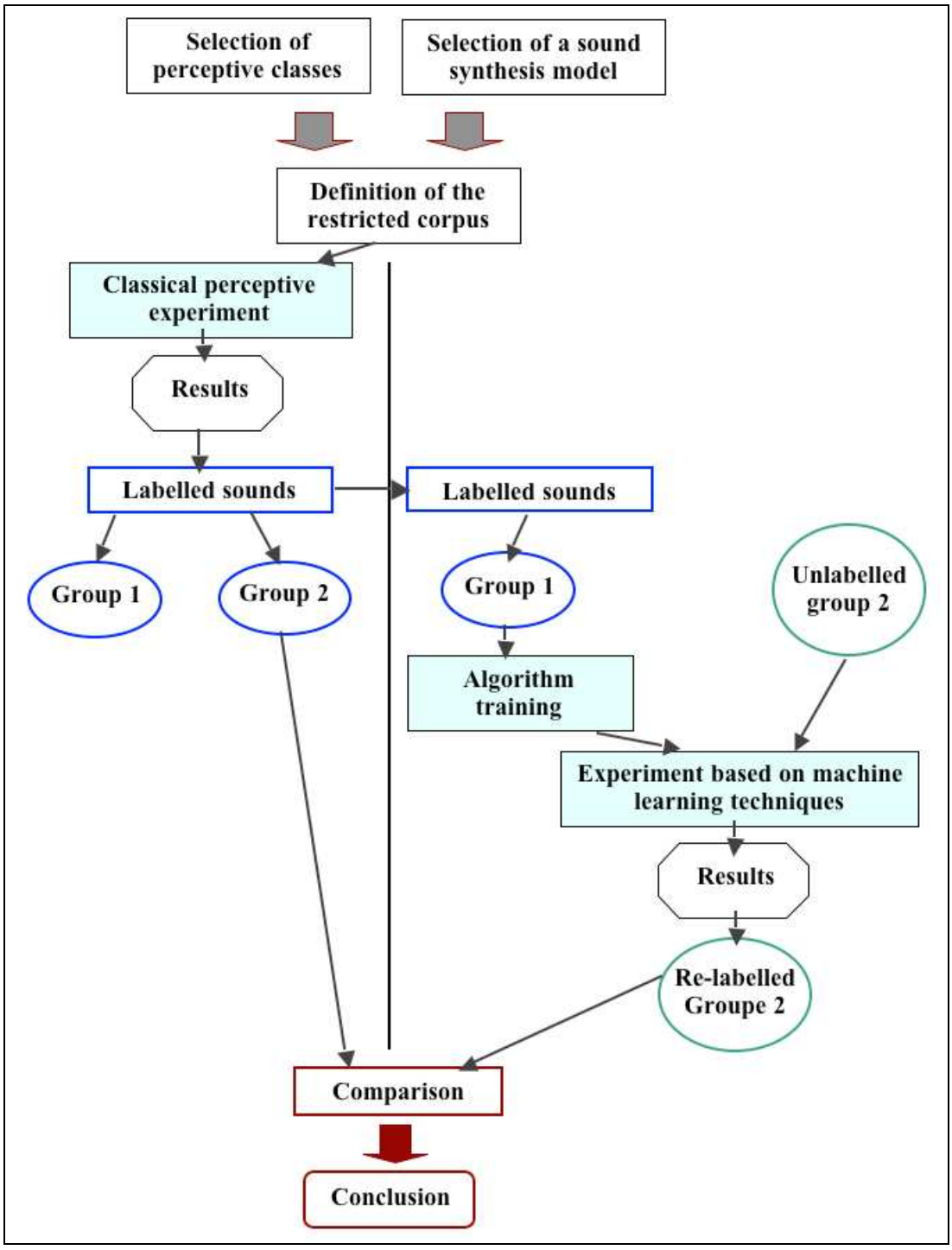

Figure 1.3: *Scheme of the master's thesis project. After having chosen perceptive classes and a sound synthesis model, a classical perceptive experiment will be carried out on a restricted corpus. Some of resulting labelled sounds (group1) will be provided to the machine learning algorithm for its training. A comparison between the results of group2 sounds labelled by participants during the classical experiment and by the algorithm will permit to conclude on the validity of this method.*

1.3 Choosing the perceptive classes

Within the CLOSED project framework, G. Lemaître and O. Houix from the Sound Perception and Design team at IRCAM studied the perceptive organisation of everyday sounds. They investigated in [1] the results of four different studies concerning the classification of everyday sounds (Frédérique Guyot's Ph. D. dissertation [2], Yannick Gérard's Ph. D. dissertation [3], a paper by Michael Marcell and al. [4] and Nancy Vanderveer's Ph. D. dissertation [5]). Bringing together the different subjects' strategies, it appeared that people group sounds together because:

- They share some acoustical similarities (same timbre, same duration, same rhythmic patterns)
- They are made by the same kind of action / interaction / movement
- They are made by the same type of excitation (electrical, electronical, mechanical)
- They are produced by the same object (the same source)
- They are produced by objects fulfilling the same (abstracted) function
- They occur in the same place or at the same occasion

This research led to the conclusions that environmental sounds can be grouped according to three types of similarity:

- The similarity of acoustical properties: ***acoustical similarity***.
- The similarity of the physical event causing the sound: ***causal similarity***.
- The similarity of some kind of knowledge, or meaning, associated by the listeners to the identified object or event causing the sound: ***semantic similarity***.

For instance, a closing car door sound will be grouped together with a sound having the same brightness or the same sharpness if the classification occurs at an acoustical level. At a causal level, this same sound would be grouped together with an object falling on the floor, as both sounds arise from an impact. Finally, at a semantic level, this car door sound would be associated with a car motor sound, since both sounds relate to the car sounds.

This study focuses on the sound classification at an event level. Within this type of classification, two subclasses defined by Carello can be distinguished [16]. The classification can first be performed regarding ***structural invariants***. Structural invariants are invariants that specify the kind of object and its properties under change, for instance the material of the object. The classification can also be performed regarding ***transformational invariants***, which are invariants that specify the change itself (as for instance crumpling, rolling etc).

As for this study, structural invariants will be considered. It investigates the classification of impact sounds with respect to the material of the object that causes the sound. As seen in chapter 2, the materials selected are wood, metal, glass and plastic.

Next chapter reviews some researches about material identification from impact sounds.

2 Literature review on material's auditory perception

This section presents some researches about the possibility of recovering object material properties from synthesized and natural impact sounds. The goal of this literature review is to point out the material classes studied in such researches, and to know which general classes are the best identified. Moreover, it will give information about some influencing parameters for the material recognition and provide some numerical values of these parameters. This will be useful to define the dynamic ranges of the model physical parameters.

Wildes and Richards defined a mechanical parameter that was supposed to characterise the material of an object from an impact sound. This parameter, related to damping measures, depends on frequency and decay time. Further studies investigated how these two parameters influence material recognition from synthesized impact sounds.

2.1 A mechanical parameter intrinsically related to the auditory perception of materials

In 1988, Wildes and Richards carried out a research [6] that aimed at discovering a physical parameter of the sound following impact that is intrinsically related to material type. They considered a mechanical model of a standard anelastic linear solid composed of two Hookean springs and a Newtonian dashpot. They studied the steady-state and damped behaviour of the solid (i.e. they did not consider the attack), and deduced that from a physical point of view, materials can be characterized using the coefficient of internal friction $\tan(\phi)$, which measures their degree of anelasticity. This parameter, which is supposed to be independent from the material, is defined by equation (2.1).

$$\tan\phi = \frac{1}{\pi f t_e} = Q^{-1} \tag{2.1}$$

Where f is the frequency of the signal, t_e is the time required for the vibration amplitude to decrease to $1/e$ of its initial value, and Q^{-1} the inverse of the quality factor.

The higher *tan(ϕ)*,the greater the damping of the material and the faster the decay time decreases with increasing frequency. In ascending order of *tan(ϕ)*, there is rubber, plastic, glass and metal.
However, the shape independence of the *tan(ϕ)* coefficient is only an approximation. Moreover, the Wilde and Richards model assumes a simple relation of inverse proportionality between frequency and decay time. Further researches on struck bars and plates sounds found that the relationship between these two parameters is quadratic or even more complex than quadratic ([7] and [8] quoted in [9]).
The next section reviews the studies made on synthesized impact sounds to study this *tan(ϕ)* parameter.

2.2 Studies on synthesized impact sounds' material identification

2.2.1 Klatzky, Pai and Krotkov's study [10]

Klatzky and al studied in [10] the effects of damping measures on material identification on synthesized sounds. In order to distinguish the importance of frequency from the importance of and decay time in the material recovering process, they studied two parameters: the frequency and the τ_D parameter, which is defined by equation (2.2).

$$\tau_D = t_e \times f = \frac{1}{\pi \tan \phi} \tag{2.2}$$

This new parameter τ_D is the exponential decreasing decay time scaled by a frequency factor and it is still assumed to be a shape-independent material property (as t_e).
Twenty-five sounds were synthesised from five fundamental frequencies equally spaced in a logarithmic scale ranging from 100 Hz to 1000 Hz (and some partials), and from five τ_D values, equally spaced on a logarithmic scale varying from 3 to 300. They were generated so as to correspond to an ideal bar, clamped at both ends, struck at a point 0.61 of its total length. The physical model was based on additive synthesis principles.
The first two experiments consisted in judging the material similarity between two sounds, in terms of the strength of their feeling that the sounds could come from the same material. All possible pairs of sounds were considered, which means that the thirteen (experiment 1) and fourteen (experiment 2) participants judged 300 sounds combinations. The participants could listen to the stimulus as many times as they

wished. In the first experiment, the initial amplitude was equalized across sounds, which was not the case in the second experiment. The subjects for the two experiments were not the same.
The results were quite identical for both experiments. This led to the conclusion that the signal amplitude does not determine the material recognition.
The study showed that the decay parameter was more influent than the frequency for the material identification, by a factor of approximately two to one. This conclusion however might not be applied generally, as it might partly be due to the fact that the range of interstimulus differences (log scale) was greater for decay than for frequency. Moreover, the number of re-listening was only affected by the decay parameter and not frequency.
In another experiment, Klatzky and al asked fifty participants to directly write down the material they identified. The four response classes were rubber, wood, glass and metal. The sounds were the same as in the previous experiments.
It turned out that τ_D is higher for glass and steel than for rubber and wood. They defined critical values of the decay parameter (logarithmic values) that would lead to half the subjects assigning an object in to a given category. Those values are reported in table (2.1).

	Rubber	Wood	Metal	Glass
Critical log(τ_D) value	0.46	0.5	2.10	2.65

Table 2.1: *Critical logarithmic values of the decay parameter for materials from Klatzky and al study [14]*

These results are consistent with Wildes and Richards study: t_e (τ_D scaled by a frequency factor) was in an increasing order for rubber, wood, glass and metal, with an inversion for metal and glass.
The frequency was more convenient to discriminate the materials within the gross categories: glass was chosen for higher frequencies than metal, and wood for higher frequencies than rubber.
The shape-invariant parameter defined by Wildes and Richards turned out to be a powerful determinant of the perceived material of an object, the time component being more influent than the frequency component.
The frequency was more convenient to discriminate the materials within the gross categories: glass was chosen for higher frequencies than metal, and wood for higher frequencies than rubber.
Other studies investigated the efficiency of the t_e parameter. The next subsection considers the study of Avanzini and Rocchesso, who got different conclusions.

2.2.2 Avanzini and Rocchesso's study [11]

Avanzini's and Rocchesso's experiment was similar to the third Klatzky's experiment, though their physical model provides more realistic attack transient. This should affect the results, as the impact sound can give information of two objects simultaneously: the hammer and the resonator. This phenomenon is called phenomenical scission.

The stimuli were synthesized with five equally log-spaced frequencies from 1000 Hz to 2000 Hz and twenty equally log-spaced quality factors varying from 5 to 5000 (extreme values found for rubber and aluminium). The quality factor is defined by $q_0 = \pi f_0 t_e$. Twenty-two subjects could listen to the 100 stimuli once only, and had to write down the identified material within these four categories: rubber, wood, glass and steel.

Consistently to Wildes and Richards and Klatzky, the quality factor turned out to be the most determining factor, the frequency playing a minor role. q_0 was in increasing order for rubber, wood, glass and steel. Here again, steel and glass are not in the same order as Wildes and Richards conclusion. As for frequency, glass remains more characterised by higher frequencies than metal's frequencies, however wood was chosen for lower frequencies, in contrary to Klatzky's conclusion.

In this experiment, regions for rubber and wood appeared clearly whereas the results were more confused for metal and glass. This was partly explained by participants' verbalisations reporting that "glass" was not clear to them because they could not guess the sound produced when striking a glass bar. Another explanation lies in the physical model: exponential decay envelopes might be a too poor approximation, explaining that materials characterized by longer decays were not correctly identified.

2.2.3 Lutfi and Ho's study [12]

The goal of Lutfi and Ho' study was to precisely determine what acoustic information was used by practiced participants to distinguish the material of synthesized struck-clamped bars. The listeners had to choose the sound object's material regarding to three parameters: decay time, frequency and intensity. On the one hand they had to choose between iron and steel, and on the other hand between glass and crystal. The experiment was based on a correlation procedure: 2 stimuli varied according to one parameter, and the participants had to indicate which of these 2 sounds corresponded to a given material. It led to the results that the eight participants (experienced musicians) largely used the frequency parameters to discriminate the materials, decay time and amplitude having only a second role in the material discrimination.

However, the results showed a quite low performance in the material composition identification, mainly because they tended to give greater weight than warranted to the frequency changes. It thus turned out that frequency was not a significant predictor for the material within the gross categories.
The next section presents Hermes' research about decay time and frequency as sound material predictors. This study provided some rough material regions depending on these two parameters.

2.2.4 Hermes' study [13]

The goal of Hermes' study was to investigate the material perception of simple synthesized impact sound, with regard to two parameters: the centre frequency of the principal mode and the decay time. His study took place in an ecological context, and aimed at attesting whether listeners have a clear mental concept of the material that may have generated the sounds. The sounds consisted of exponentially decaying partials restricted to a limited frequency band. He tested the constancy of the listener by carrying out two experiments with a different set-up regarding the task of the listener and listening conditions.
Hermes' first experiment consisted in a free-identification material in which listeners were asked to write down the material of the object producing the sound. The fifteen listeners could listen to the sound over headphones as many times as they wanted to but once they had asked for the next sound they could not listen to the previous sound again.. The sounds had centre frequencies ranging from 100 Hz to 6.4 kHz (seven equidistant values on a logarithmic scale) and partials time constants varying from 6.25 to 50 ms (seven values as well). The physical model was based on additive synthesis principles. The corpus was composed of these 49 sounds and of 9 practice stimuli covering the range of experimental stimuli.
The most often named materials were wood, metal, glass, plastic and rubber/skin. The results showed that glass, metal and wood are well-defined perceptive categories; glass and wood sounds were more easily identified (the required less re-listening). Glass and metal sounds turned out to have high frequencies, metal's partials having a longer decay time than those of glass. Glass is recognised for frequencies higher than 1600 Hz and characterized by frequencies higher than 3kHz, metal by frequencies between 0.8kHz and 3 kHz. Wooden sounds have lower frequencies around 200 Hz and short decay times around 20 ms. Plastic and rubber sounds are distributed in lower frequencies, with a higher decay times that wood's sounds. Rubber sounds have frequencies lower than 300 Hz. This experiment underlined an uncertainty region (the sounds were re-listened more often than average) around 30 ms for the decay time and around 0.2 kHz to 0.5 kHz for the centre frequency.

As for the second experiment, the thirteen new participants could only listen once to a sound and were under time pressure: the interval time between two stimuli was very short (3 seconds). The sounds were presented over two large loudspeakers, so that room acoustics was part of the stimuli. The listeners were asked to classify the sounds within the five most quoted categories in the first experiment, i.e. wood, glass, rubber, plastic, metal.
For wood and glass, the results were about identical to those in the first experiment. The metal region got smaller in this second experiment: the mid frequencies were containing less metal answers. Metal is then classified somewhat less consistently than glass and wood. On the contrary, the rubber category was larger. The difference was considerable for plastic: it was classified in the middle frequencies, with a longer decay time: it actually fitted with the uncertainty region described in the first experiment.
The results of this experiment are summarized in table2.2. It presents the frequency and decay time values characterising the five materials.

	Frequency (Hz)	Decay time (ms)
Rubber	< 300	15 → 40
Plastic		$t_{plastic} > t_{rubber}$
Wood	200 → 1600	7 → 30
Metal	800 → 3000	$t_{metal} > t_{glass}$
Glass	> 3000	> 17

Table 2.2: *Hermes' investigation on material perception of synthesized impact sound.*

As for Hermes' study, frequency and decay time seemed to be relatively good predictors to discriminate the materials, even within the gross classes. However, these results are only tendencies, and constitute rough regions, which can largely overlap.
The next section presents works in the material recognition from real impact sounds.

2.3 Studies on real impact sounds' material identification

Some studies of real impact sounds material recognition showed that people could discriminate materials very successfully.
Gaver, in [14], investigated on the material recognition of impacted length-varying steel and iron bars. They obtained very high performance between 96% and 99%. The bar length had no effect on the material identification.
Kunkler-peck and Turvey [15] studied shape and material recognition in struck plates sounds. Listeners had to identify if the plate was made of steel, wood or plexiglass. Here again, very high performance recognition was found, with only a

secondary tendency to associate materials with shape.
The differences between the almost perfect results of real impact sounds experiments and the results of synthesized impact sounds experiment were explained by Carello, Wagman and Turvey [16] by the lack of acoustical richness in the synthesized sounds signals, which were thus missing some convenient information for the material discrimination.
Giordano and McAdams in [17] investigated the identification of the material of struck objects of variable size. Twenty-five participants had to judge 2-mm-thick square, struck plates of four different materials (plexiglas, glass, steel and wood) and five different surfaces.
Gross categories, metal/glass on the one hand and plexiglas/wood on the other hand, were identified almost perfectly, independently of the geometry of the plates.
Within each gross category, identifications were based on signal frequencies, glass and wood being associated to higher frequencies than, respectively, metal and plastic. This study concluded that only partials support for the perceptive relevance of $tan(\phi)$. Participants highly failed to identify the material within these gross categories. They tended to associate small plates with glass or wood, and large plates with metal or plexiglas. This observation confirms Lufti's result but is not consistent with the perfect wood/Plexiglas discrimination of Kunkler-peck and al.
Giordano and McAdams found a predictive model based on loudness and frequency, and proposed also an ecological explanation to these results: there may be an ambiguity between the sound of a glass or metal bar. But as listeners are not used to manipulate big glass objects, they associate big objects with metal, and small objects with glass.

2.4 Conclusion of the literature review

In all studies, the coefficient of internal friction $tan(\phi)$ turned out to be a powerful criterion with regards to the material recognition from synthesized impact sounds.
Investigations on the relative importance of frequency and decay time led to different conclusions: Avanzini and Rocchesso, Klatzky and al showed that the decay time was the main influencing parameter whereas Lutfi and Oh proved the contrary. This disagreement has many probable causes. First, the physical models used to synthesize the sounds were different. Then, the experiment procedures were not the same, Lutfi and Oh using a correlation procedure and the others using a similarity procedure (this means that the participants had to identify the material). Moreover, the sampling rates were different for the parameters in each experiment: more space between the frequencies of the stimuli can enhance the role of this parameter for instance.
Consistently to Giordano and McAdams conclusion, it seems that the participants show high performance to distinguish metal and glass on the one hand from plastic and wood on the other hand. Actually, the frequency ordering within these gross

categories is not the same in Klatzky's and Avanzini's investigations for example. Moreover, one can notice that the τ_D values proposed by Klatzky and al were quite similar for rubber and wood on the one hand, and for metal and glass on the other hand. This parameter would then be efficient for a "gross" classification involving two gross classes: glass and metal on the one hand, wood and rubber on the other hand.

As for this research, the selected materials are wood, plastic, metal and glass. Rubber is here considered as a plastic material. As frequency and decay time turned out to be very influencing parameters, they will be given a particular importance. Moreover, gross categories will be investigated.

3 The Sound Design Tool synthesis models

The Sound Design Tool (SDT) package is developed by Univerona and includes the impact sound synthesis model implemented on the software MAX-MSP 4.6. The SDT package is the main software product of a project activity which begun in 2001 with the EU project SOb - the Sounding Object [9].
The physically based sound design tools aims at providing perception-oriented and physically coherent tools. To achieve a very realistic simulation is not the goal of the SDT synthesis models. Actually, cartoonifications are of a high interest: simplifications of sounds that preserve and possibly exaggerate certain acoustic aspects are cheaper computationally speaking and may convey information more effectively. This section presents the low level models, whose the studied impact model belongs to, and then gives an overview of the more complex synthesis models.

3.1 Low-level models

This part will first present the way solid contacts are modelled and then explain more precisely the modal impact models implementation. The models are not simply based on additive synthesis. They take the attack transient into account. More details can be found in [9] and [18].

3.1.1 Generalities about the contact models

The models considered here apply to basic contact events between two solid objects. As the most relevant contact sound events in everyday life come down to impacts and frictions, the provided externals model these two kinds of interactions. The algorithms implemented share a common structure: two solid object models interact through (what is called here) an interactor (see Fig. 3.1).
An interactor represents a contact model or, so to say, the “thing” between the two interacting objects. As for the impact model, it can be seen as the “felt” between the striking object and the struck object, while in the friction model it simulates friction as if the surfaces of the two rubbing objects would be covered with “micro-bristles”.
As for impact sounds, the interactor can be implemented with a non-linear or a linear force. It receives the total compression (the difference of displacements of the two interacting objects at interaction point) and returns the computed impact force. The latter is made of the sum of an elastic component and a dissipative one. The elastic

component is parameterized by the force stiffness (or elasticity) and by a non-linear exponent that depends on the local geometry around the contact area. The dissipative component is parameterized by the force dissipation (or damping weight).

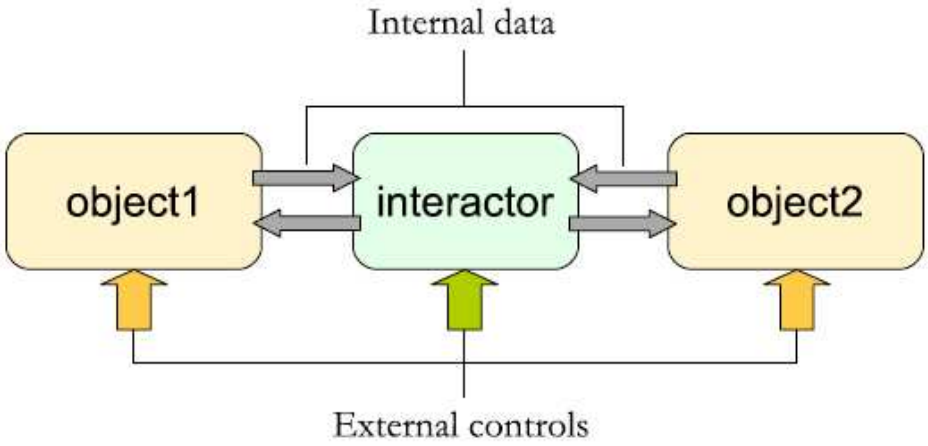

Figure 3.1: *The common structure underlying solid contact algorithms. The interactor represents the contact model between the striking object and the struck object. At each discrete time instant (sample) both objects send their internal states (displacement and velocity at the interaction point) to the interactor, which in turn sends the newly computed (opposite) forces to the objects.*

Three distinct object models are provided:

- ***Modal object***

In the modal description, a resonating object is described as a system of a finite number of parallel mass-spring-damper structures. Each mass-spring-damper structure models a mechanical oscillator that represents a normal mode of resonance of the object. The oscillation period, the mass and the damping coefficient of each oscillator correspond respectively to the resonance frequency, the gain and the decay time of each mode.

- ***Inertial object***

An inertial object simulates a simple inertial point mass. Obviously this kind of objects is useful solely as an exciter for other resonators. The only settable object property is its mass.

- ***Waveguide object***

The digital waveguide technique models the propagation of waves along elastic media. In the one-dimensional case implemented here, the waveguide object models an ideal elastic string.

Having a look at Fig. 3.1, the way two objects interact through an interactor appears evident: at each discrete time instant (sample) both objects send their internal states (displacement and velocity at the interaction point) to the interactor, which in turn sends the newly computed (opposite) forces to the objects. Knowing the new applied

forces, the objects are able to compute their new states for the next time instant. In other words, there's a feedback communication between the three models.
The SDT framework differs remarkably from the approach to physically based sound synthesis found in most existing implementations and literature. The SDT package takes advantage of a cartoonified approach in sound design and implements a feedback network within the interaction

$$\textit{Object 1} \Leftrightarrow \textit{Interactor} \Leftrightarrow \textit{Object 2}$$

with nonlinear characteristics of the interactor. This allows the accurate modelling of complex interactions (e.g. friction) and to output the sound of both the interacting objects. Besides, the continuous feedback approach adopted into the SDT is memory consistent: the system takes record of each previous state during the interaction and manipulation.

3.1.2 Modal Impact models

The sound synthesis model used in this study is an inertial-modal model: the striking object is inertial object and the struck object is modal object. This case is a particular case of a two modal resonator model. This section describes the continuous-time impact model between two modal resonators. The cartoonification approach of this model is schemed on Fig.3.2. The striking object (notified by *h* for hammer) and the struck object (notified by *r* for resonator) are characterised by a mass component, a damping component and a spring component.

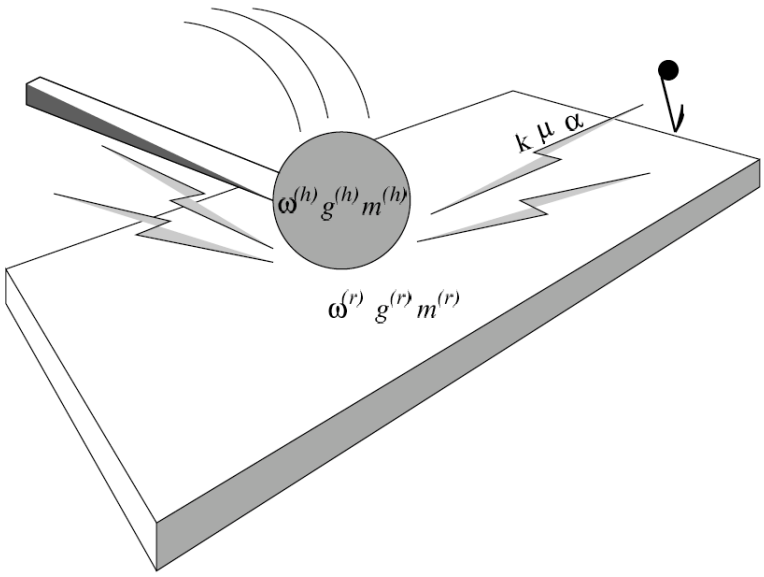

Figure 3.2: *Cartoon impact between two modal resonators. The sound model is controlled through a small number of parameters, which are related either to the resonating objects or to the interaction force.*

Modal objects are characterized by a frequency ω, a mass m and a damping factor g. The interactor parameters are the coefficient shape α that characterizes the surface contact, the elasticity coefficient k and the dissipative factor μ.
The simplest possible representation of a mechanical oscillating system is a second-order linear oscillator of the form:

$$\ddot{x}^{(r)}(t) + g^{(r)}\dot{x}^{(r)}(t) + \left[\omega^{(r)}\right]^2 x^{(r)}(t) = \frac{1}{m^{(r)}} f_{ext}(t) \tag{3.1}$$

Where $x^{(r)}$ is the oscillator displacement, f_{ext} the external driving force, $w^{(r)}$ the oscillator center frequency, $g^{(r)}$ its damping coefficient and $1/m^{(r)}$ controls the inertial properties of the system.
As for the inertial – modal model, the hammer will be characterized by an inertial mass described with null frequency, zero spring constant and zero internal damping (infinite decay time).
Putting N oscillators in parallel, one can get spectrally richer sounds including a set of N partials $\{\omega_l^{(r)}\}$ (l=1…N). The system thus obtained is:

$$\begin{bmatrix} \ddot{x}_1^{(r)}(t) \\ \vdots \\ \ddot{x}_N^{(r)}(t) \end{bmatrix} + G^{(r)} \begin{bmatrix} \dot{x}_1^{(r)}(t) \\ \vdots \\ \dot{x}_N^{(r)}(t) \end{bmatrix} + \left[\Omega^{(r)}\right]^2 \begin{bmatrix} x_1^{(r)}(t) \\ \vdots \\ x_N^{(r)}(t) \end{bmatrix} = m^{(r)} f_{ext}(t) \tag{3.2}$$

where the matrices are given by

$$\Omega^{(r)} = \begin{bmatrix} \omega_1^{(r)} & & 0 \\ & \ddots & \\ 0 & & \omega_N^{(r)} \end{bmatrix}, G^{(r)} = \begin{bmatrix} g_1^{(r)} & & 0 \\ & \ddots & \\ 0 & & g_N^{(r)} \end{bmatrix}, m^{(r)} = \begin{bmatrix} 1/m_1^{(r)} \\ \vdots \\ 1/m_N^{(r)} \end{bmatrix} \tag{3.3}$$

This N-coupled equations system can often be diagonalized using the transformation matrix (3.4) to obtain N decoupled equations.

$$T = \left\{ t_{jl} \right\}_{j,l=1}^{N} \tag{3.4}$$

The new variables are generally referred to as modal displacement. At a given point j, the displacement x_j and velocity v_j of the resonator are given by:

$$x_j = \sum_{l=1}^{N} t_{jl} x_l^{(r)}, \dot{x}_j = \sum_{l=1}^{N} t_{jl} \dot{x}_l^{(r)} \tag{3.5}$$

Assuming that the contact area between the two colliding objects is small (it is ideally a point), Hunt and Crossley ([19] in [9]) proposed the collision model described in equation (3.6) to depict the interaction force. This interaction force depends on the felt compression x and on the compression velocity v.

$$f\left(x(t),v(t)\right)=\begin{cases} k\left[x(t)\right]^{\alpha}+\lambda\left[x(t)\right]^{\alpha}\cdot v(t) & x>0 \\ 0 & x\le 0 \end{cases} \tag{3.6}$$

The compression x is the difference between the displacement of the hammer and the resonator. This means that there is only compression for x>0 and that the two objects are not in contact for x≤0. The k parameter is the force stiffness and the α coefficient characterises the local geometry of the contact area. For instance, its value is equal to 1.5 when both objects are perfect spheres. As for a piano hammer impact, his value varies from 1.5 to 3.5. The parameter λ is the force damping weight, which is related to the viscoelastic characteristic μ by the formula (3.7). It has an influence on bouncing striking object. As for this study, this last parameter is not considered (only simple impacts are taken into account).

$$\mu=\frac{\lambda}{k} \tag{3.7}$$

The continuous-time equations of the whole system are then:

$$\begin{cases} \ddot{x}_j^{(r)}(t)+g_j^{(r)}\dot{x}_j^{(r)}(t)+\left[\omega_j^{(r)}\right]^2 x_j^{(r)}(t)=\dfrac{1}{m_{jm}^{(r)}}(f_e^{(r)}-f),(j=1...N^{(r)}) \\ \ddot{x}_i^{(h)}(t)+g_i^{(h)}\dot{x}_i^{(h)}(t)+\left[\omega_i^{(h)}\right]^2 x_i^{(h)}(t)=\dfrac{1}{m_{il}^{(h)}}(f_e^{(h)}-f),(i=1...N^{(h)}) \\ x=\sum_{j=1}^{N^{(r)}} t_{mj}^{(r)} x_j^{(r)}-\sum_{i=1}^{N^{(h)}} t_{li}^{(h)} x_i^{(h)} \\ v=\sum_{j=1}^{N^{(r)}} t_{mj}^{(r)} \dot{x}_j^{(r)}-\sum_{i=1}^{N^{(h)}} t_{li}^{(h)} \dot{x}_i^{(h)} \\ f\left(x(t),v(t)\right)=\begin{cases} k\left[x(t)\right]^{\alpha}+\lambda\left[x(t)\right]^{\alpha}\cdot v(t) & x>0 \\ 0 & x\le 0 \end{cases} \end{cases} \tag{33.8}$$

Where the terms $f_e^{(h)}$ and $f_e^{(r)}$ represent external forces, $N^{(r)}$ and $N^{(h)}$ the number of modes for the resonator and the hammer.

Considering that the striking object is an inertial object, the equations driving the impact model are then:

$$
\begin{cases}
\ddot{x}_j^{(r)}(t) + g_j^{(r)} \dot{x}_j^{(r)}(t) + \left[\omega_j^{(r)}\right]^2 x_j^{(r)}(t) = -\dfrac{1}{m_{jm}^{(r)}} f, (j = 1,2) \\
\ddot{x}^{(h)}(t) = \dfrac{1}{m_l^{(h)}} f \\
x = \sum_{j=1}^{2} t_{mj}^{(r)} x_j^{(r)} - t_l^{(h)} x^{(h)} \\
v = \sum_{j=1}^{N^{(r)}} t_{mj}^{(r)} \dot{x}_j^{(r)} - t_l^{(h)} \dot{x}^{(h)} \\
f(x(t)) = \begin{cases} k[x(t)]^{\alpha} & x > 0 \\ 0 & x \le 0 \end{cases}
\end{cases}
\tag{3.9}
$$

The time-continuous system is discretized using an impulse invariance transform, which means that the impulse response is the same (invariant) at the sampling instants.

Figure 3.3 shows the MAX-MSP interface for the two modal-resonators impact model.

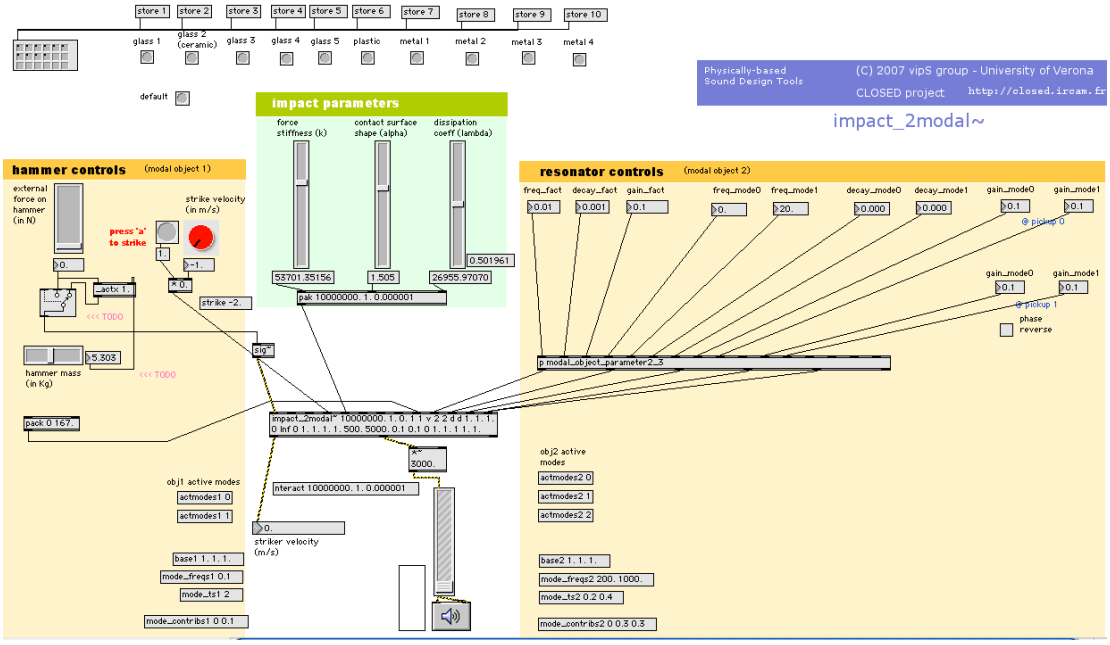

Figure 3.2: *Cartoon impact between two modal resonators. The sound model is controlled through a small number of parameters, which are related either to the resonating objects or to the interaction force.*

The remaining controlling parameters are, for the hammer: the hammer mass *m* and the external force, which is equal to zero (this means that there is no bouncing). As for the resonator, the parameters are the two modes frequencies, decay times and gains. The interactor parameters are the stiffness force, the α coefficient and the dissipation coefficient, which is not influent given that there is no bouncing.

For computational reasons, only one pick-up point is studied, the interaction point. This is not problematic given that there is no sound artefact.
In the same view of limiting the number of parameters to be controlled (limitations of the machine learning algorithms), the hammer mass and the interactor stiffness and shape coefficient parameters will be settled. Actually, a listening working session reveals that these parameters are obviously not as influent as the resonator parameters for the output sound.

3.2 Higher-level models

The expression "higher level" indicates more complex and structured algorithms, corresponding to somewhat large-scale events, processes or textures. In a way, that matches the meaning of the expression "high-level" in Computer Science, where it often denotes languages similar to those of human beings. Of course, in order to achieve that, high-level languages are indeed more complex and structured than low-level ones.
The higher-level algorithms here discussed implement temporal patterns or other physically consistent controls (e.g. external forces) superimposed to low-level models. The low-level used for higher level models is the inertial-modal model, that was chosen for this reason.

3.3 Definition of presets

A listening session of real wood, glass, metal and plastic impact sounds was organized with Stefano Delle Monache and Stefano Papetti from Univerona. Stefano Papetti is working on the physical sound synthesis algorithms, whereas Stefano Delle Monache is more involved in sound design, controlling the validity of the models.
The sounds we found in an existing library (Cd Audio Soundscan V2 Vol.61 SFX Toolbox) were not simple impact sounds; there were more complex sounds recorded in the everyday life. A spectral analysis on Audiosculpt (web link on http://forumnet.ircam.fr/691.html?L=1) permitted to isolate the main frequencies, decay times and gain patterns of the sounds. These parameter values were then implemented on the physical impact model.
The goal of these presets were first to validate the impact model, then to study the dynamic ranges of the parameters in order to reduce them and finally to form the sound corpus. Actually, the presets served as a basis for the sound corpus formation.
The values of the parameters of the presets are presented in table 3.1.

	f1 [Hz]	f2 [Hz]	t1[s]	t2[s]	g1[ ]	g2 [ ]
Metal 1	340	1215	0.09	0.32	80	100
Metal 2	570	1100	0.18	0.24	100	100
Metal 3	390	1025	0.12	0.23	100	100
Metal 4	570	1700	0.27	0.4	95	100
Glass 1	1900	3500	0.085	0.066	100	90
Glass 2	1300	3500	0.048	0.041	100	95
Glass 3	1400	3200	0.047	0.088	100	90
Glass 4	1450	3200	0.04	0.08	100	90
Glass 5	2195	5100	0.085	0.1	100	85
Wood 1	419	874	0.002	0.006	100	100
Wood 2	430	368	0.013	0.006	93	102
Wood 3	430	251	0.013	0.016	92	94
Wood 4	419	967	0.013	0.008	100	100
Plastic 1	1500	750	0.003	0.002	100	80

Table 3.1 *Values of the parameters' presets for every material samples and for each parameter.*

Roughly, glass sounds have higher frequency components than metallic sounds, and lower decay time. Wood impact sounds have low decay times and frequency relatively low frequencies. As for plastic impact, only one sample was available. However, it seems to be characterised by very low decay times.
The dynamic ranges of the parameters were defined as follow:

Parameter	Dynamic range
Frequency (Hz)	[150 ; 5100]
Decay Time (s)	[0.001 ; 0.405]
Gain ( )	[80 ; 110]

Table 3.2 *Dynamic ranges of the parameters of the sound synthesis models chosen to define the restricted corpus.*

4 Basis of machine learning techniques

Machine learning techniques build algorithms that allow machines to "learn", i.e. algorithms able to improve their performance based on previous results.
The final aim of this study is to evaluate whether a machine learning algorithm can drive a perceptive classification experiment. It would allow achieving perceptive experiment on larger sound corpus. This implies that the algorithm is capable of cleverly choosing the sounds (i.e. sets of physical parameters) to be presented to the listener with regards to the material classes boundaries. This ability to cleverly choose a point is called **active learning**.
This chapter first explains the basic principles of machine learning techniques and then presents three different machine learning techniques implemented by Kamil Adiloglu and Robert Annies from NIPG: the active perceptron algorithm, support vector machines (SVM) and probabilistic generative models (PGM). SVM and probabilistic generative models are passive learning techniques. Their results were used as a comparison basis in order to estimate the active perceptron algorithm performance.
More details about machine learning techniques can be found in Bishop's book [20].

4.1 Basic structure of an active learning algorithm

A typical artificial intelligence (AI) example is the recognition of hand-written digits. The goal is to build a machine that will take a hand-written digit as input (corresponding to a 28*28 pixel image for instance, i.e. a 784 real numbers vector) and that will produce the identity of the digit 0… 9 as output.
A machine learning approach to this problem consists in two main steps.

- First, the training phase, during which a large set of digits $\{x_1,\ldots,x_n\}$ called training set is used to tune the model parameters. The target vector *t* represents the identity of the digit. The ***training phase***, or *learning phase*, consists in determining a function y(x) which takes a digit image x as input and that generates an output vector y, encoded in the same way as the target vectors. In this case, there is a target vector *t* for each digit image x.
- The second step, called ***generalization***, is the ability of the model to classify new digit images that differ from those used during the training phase.

Most of the time, the original input variables are pre-processed in order to transform them into some new space of variables where the pattern recognition problem will be

easier solved and to speed up the computation as well. This is called ***feature extraction***.

There are two kinds of training:

- ***Online learning***: the adaptation of model parameters (the y function) is done by using one observation at a time, in consecutive steps. The order of presentation influences the training. This effect should be negligible with many observations. Online learning is used when the training data arrive during the training phase. Online learning is necessary for active learning. Actually, the algorithm has to re-evaluate the model at each step to cleverly choose the next point.
- ***Batch learning***: all observations from training set are applied at once to adapt the model parameters. There cannot be active learning with such a learning phase, but only passive learning: the sounds are not cleverly chosen.

Online learning is characteristic of active learning. Actually, active learning chooses observations step by step and needs a model estimation at each step. This estimation influences the next point to be chosen.

One can distinguish three machine-learning families.

- ***Supervised learning*** problems deal with applications in which the training data are composed of input vectors along with their corresponding target. The problem is there to find a model that attribute the correct target value for a given input vector. This includes classification tasks in which the aim is to assign each input vector to one of finite number of discrete categories (as in the digit example) and regression tasks in which the desired output consists of one or more continuous variables.
- As for ***Unsupervised learning*** problems, the training data consists of a set of input vectors x without any corresponding target values. Unsupervised learning problems count clustering tasks (the goal of such problems is to discover groups of similar examples within the data) as well as density estimation tasks (the goal is to determine the distribution of data within the input space) and visualization tasks (the aim is to project the data from high- dimensional space down to two or three dimensions for the purpose of visualization).
- In ***Reinforcement learning*** problems, the machine can also produce actions that affect the state of the world and receive awards or punishment. The aim is to find suitable actions to take in a given situation in order to maximize rewards in the long term. The two components of such problems are exploration (trying new kinds of action) and exploitation (using actions that are known to yield a high reward).

As for this study, the accurate machine learning family for the classification problem is supervised learning.

Three supervised machine learning techniques were compared in this study: the perceptron algorithm, SVM and PGM.

4.2 The perceptron algorithm

4.2.1 General algorithm

The active learning program used in this study is based on the two-class linear discriminant model developed by Rosenblatt in 1962: the perceptron algorithm. The algorithm consists in finding the model ω that assigns an output value $\tilde{y}$ to an input vector x. as for this study, x is a six-dimension vector (the two frequencies, the two decay times and the two gains). The output value is the material class.

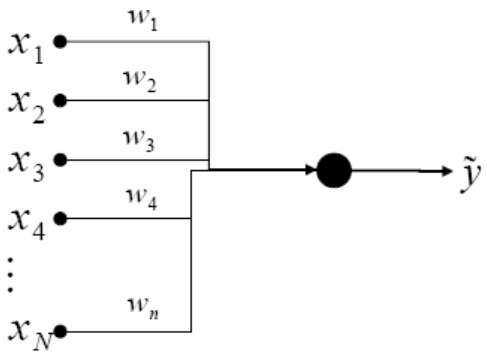

Figure 4.1: *The perceptron algorithm. The model is defined by the weight vector ω.*

The input vector *x* is first transformed using a fixed nonlinear transformation to give a feature vector $\phi(x)$ (feature extraction to facilitate the classification task). The vector $\phi(x)$ is then used to construct a generalized linear model of the form:

$$y(x) = f(w^T \phi(x)) \tag{4.1}$$

Where the nonlinear activation function f is given by a step function *f(a).*

$$f(a) = \begin{cases} +1 & a \geq 0 \\ -1 & a < 0 \end{cases} \tag{4.2}$$

Typically, the vector $\phi(x)$ includes a bias $\phi_0(x) = 1$. The goal is then to find the weight vector *w* such that patterns x_n in class C_1 will have $w^T\phi(x_n) > 0$ whereas patterns x_n in class C_2 will have $w^T\phi(x_n) < 0$. Using the $t \in \{-1, +1\}$ target coding scheme, the classification task is then to satisfy equation (4.3) for all patterns.

$$w^T \phi(x_n) t_n > 0 \tag{4.3}$$

The error function for the classification task is defined by the perceptron criterion that associates zero error with any pattern that is correctly classified, whereas for a misclassified pattern x_n it tries to minimize the quantity $-w^T\phi(x_n)t_n$. M denoting the set of all misclassified patterns, the perceptron criterion is therefore given by:

$$E_p(w) = -\sum_{n \in M} w^T \phi_n t_n \tag{4.4}$$

If the training data are linearly separable, the perceptron algorithm guarantees a solution in a finite number of steps. However, the solution depends on the initialization of the parameters and on the order of presentation of the data points.

The perceptron algorithm can only deal with two classes. To obtain the four-materials classification, the one-versus-the-rest approach was used: one classifier for each material. Consequently, the algorithm answered the following question: *is this sound in the plastic class or not?*

The perceptron algorithm has a simple interpretation: it cycles through the training patterns in turn, and for each pattern x_n it evaluates the perceptron function presented in equation 4.1. If the pattern is correctly classified, then the weight vector remains unchanged, whereas if it is incorrectly classified, then for class C_1 it adds the vector $\phi(x_n)$ onto the current estimate of weight vector w while for class C_2 it subtracts the vector $\phi(x_n)$ from w.

Figure 4.2 (extracted from [20]) illustrates how the perceptron algorithm converges. This algorithm is a linear classifier.

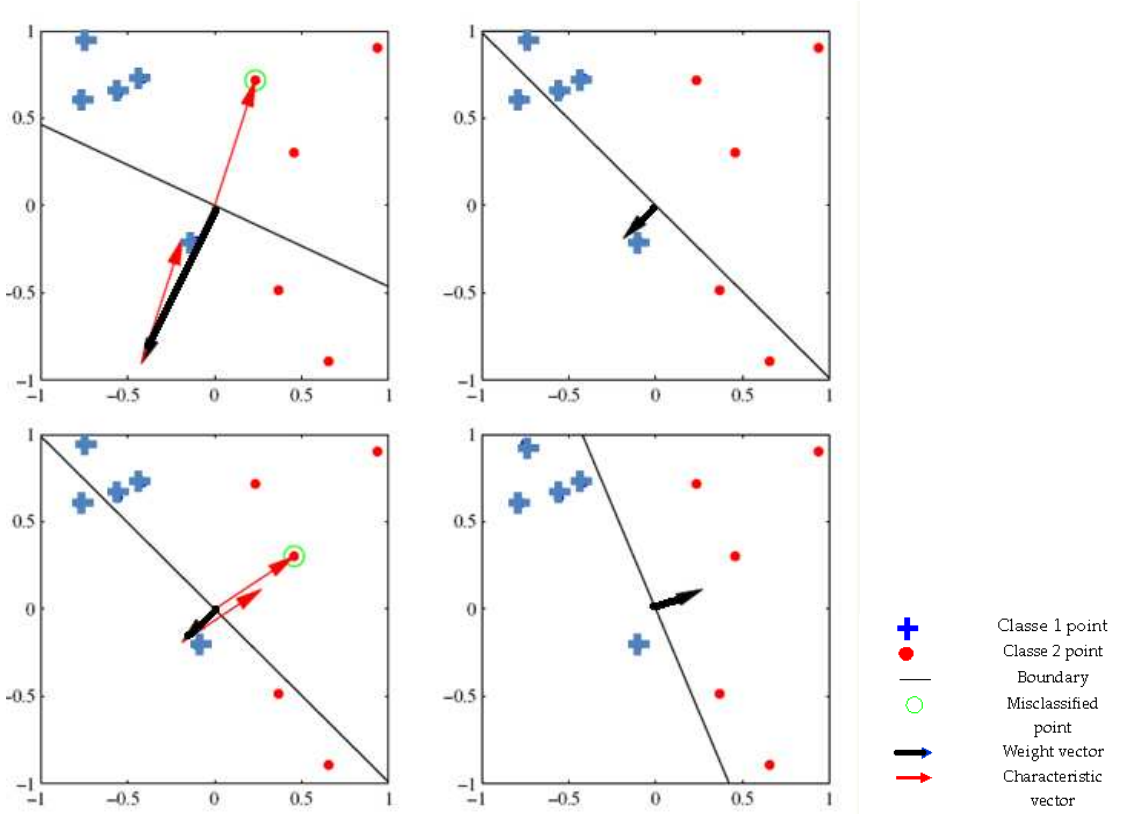

Figure 4.2: *Illustration of the convergence of the perceptron learning algorithm, showing data points from two classes (circles and crosses) in a two-dimensional feature space ($\phi 1$, $\phi 2$). The top left plot shows the initial parameter vector w shown as a black arrow together with the corresponding decision boundary (black line), in which the arrow points towards the decision region which classified as belonging to the red class. The data point circled in green is misclassified and so its feature vector is added to the current weight vector, giving the new decision boundary shown in the top right plot. The bottom left plot shows the next misclassified point to be considered, indicated by the green circle, and its feature vector is again added to the weight vector giving the decision boundary shown in the bottom right plot for which all data points are correctly classified. Figure extracted from Bishop's book [20].*

4.2.2 Active perceptron algorithm

NIPG added an "active" component to the perceptron algorithm. This means that the next sound to be chosen by the algorithm during the training phase will not be chosen randomly (which is the case for passive learning) but cleverly, i.e. so that it conveys new information. Hence, during the experiment, the algorithm will choose the next sound to be heard by the participant so that the classification can be performed with fewer sounds.
Fig. 4.3 explains how the algorithm chooses the points to be studied. Vol (V) is the volume function. In the diagram it is the length of an arc on the circle (the bold read

and black line). In higher dimensions it is a part of the sphere surface and is quite difficult to estimate.

The bold lines are version spaces. A version space is the subset of all hypotheses that are consistent with the observed training examples. This set contains all hypotheses that have not been eliminated as a result of being in conflict with observed data. The goal is then to reduce this version space, so as to obtain the best classifier.
V is the version space after a learning iteration and V+ the version space after the next iteration (learning step), which should shrink the version space. On figure 4.5 for instance, the circle point will be chosen because the weight vector cuts the version space. Depending on the label of the point (plus or minus) the next version space is selected (the V+ bold line).
The cross point does not divide the previous version space: it does not convey information, and thus is not selected.
To classify a data point, the formula (4.5) calculates for each class the probability that the point belongs to this class and just takes the greatest value out of the four classifiers.

$$\arg\max_{k}\left\{\frac{Vol\left(V^{+}{}_{C_k}(x)\right)}{Vol\left(V_{C_k}(x)\right)}\right\} \tag{4.5}$$

Ambiguities can appear when 2 or more classifiers say to 100% (V+ = V) that the point belongs to their class.

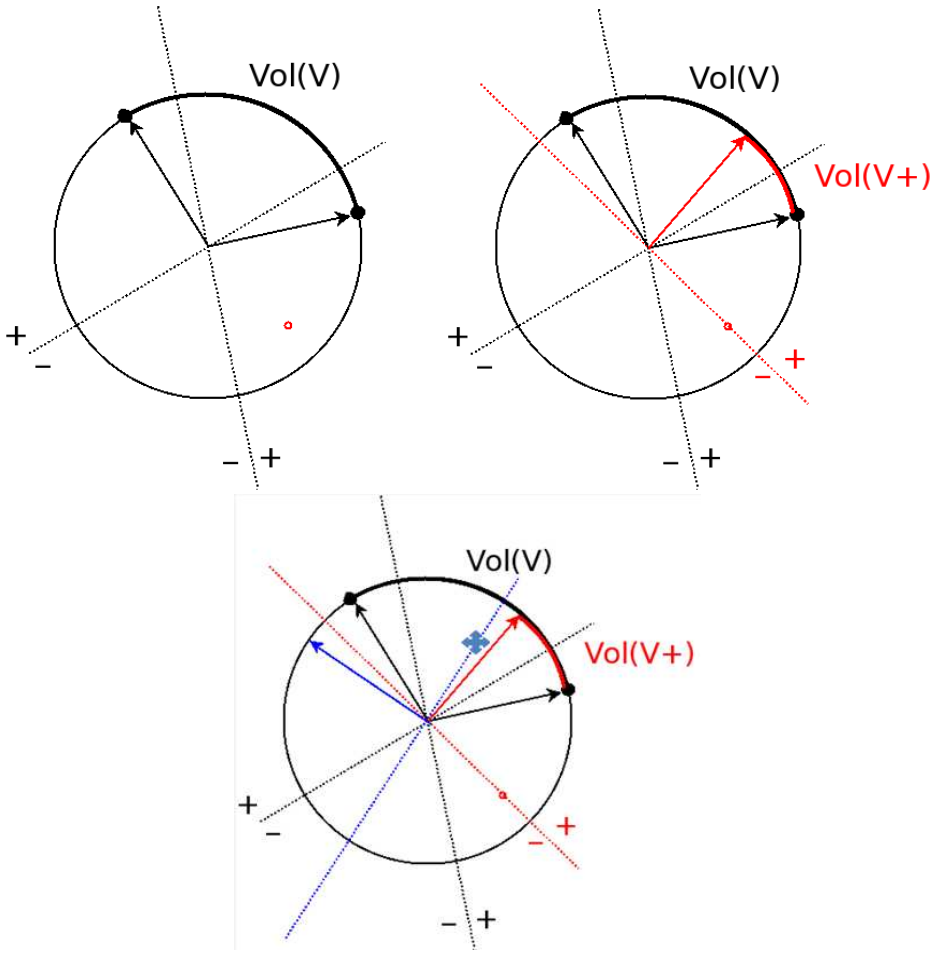

Figure 4.3: *Active learning algorithm. On the left figure, each point on the circle (in our case a 6-dimensional sphere) represents a linear boundary: the dotted line. This boundary divides the circle in two halves. Data points on either side get the class label +1 and - 1 respectively. The algorithm learns from a sample of labelled data points and has some idea where to put those dotted lines best, such that the learned data points get the correct labels. Since the sample is finite there will be always region between the data points where one can put several boundaries that would label the points equally correct, this is called the Version space and is depicted on the figure by the bold line on the circle. Anywhere inside one can find classifiers that are consistent with the training set. The target is to minimize the bold line as much as we can, by choosing points that are located such that the normal vector of the updated classifier (the arrow) falls inside the version space and divides it. Because the label of that training point is known the version space shrinks (see the right figure with the red bold line). The data point was 'interesting' for the algorithm because it conveyed new information: the version space could be shrunk. The blue point (bottom plot) is not interesting: the resulting weight vector does not cut the version space.*

4.3 Support Vector Machines

Support Vector Machines (SVM) are passive learning techniques. They are based on batch training. SVM are a set of supervised learning techniques used for classification and regression.

In a 2-classes SVM problem (for instance a sound has to be classified in the wood category or the glass category), a data point is viewed as a p-dimensional vector (a list of p numbers, in our case: a 6-dimensional vector); the goal is to separate and classify the data with a $p - 1$-dimensional hyperplane. Such a classifier is called a linear classifier. However, SVM can also deal with nonlinear boundaries. A 2-classes classification problem can be resolved by using linear models on a form presented in equation (4.6).

$$y(x) = w^T \phi(x) + b \tag{4.6}$$

Here, $\phi(x)$ denotes a fixed feature-space transformation, b is an explicit bias and w is the weight vector (perpendicular to the hyperplane). The training data set comprises N input vectors $x_1 \ldots x_N$, with corresponding target values $t_1, \ldots, t_N$ where $t_n \in \{-1, 1\}$, and new data points x are classified according to the sign of $y(x)$. As seen on figure 4.3, there are many possibilities to separate two classes.

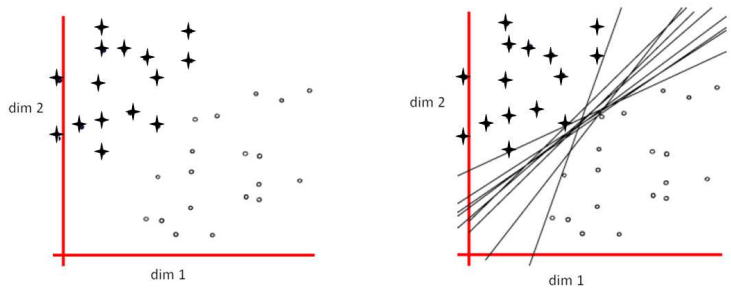

Figure 4.3 *Many classifiers (black lines) to separate Class 1 (black points) from Class 2 (white points) in a 2-dimension space.*

In SVM, the decision boundary (the line that separate the two classes) is chosen to be the one for which the margin is maximized, the margin being defined as the perpendicular distance between the decision boundary and the closest of the data points (see Fig. 4.3 extracted from [20]). This allows getting the minimum probability of error relative to the learned density model.

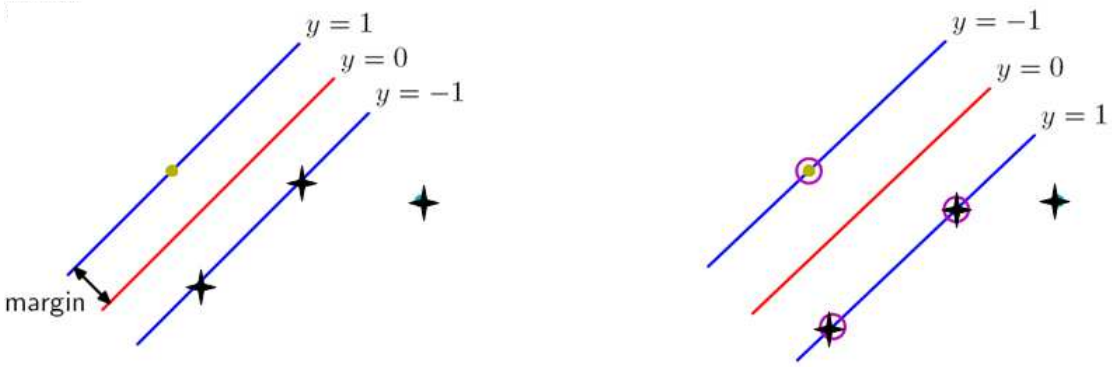

Figure 4.4 *Maximum margin boundary. The margin is defined as the perpendicular distance between the decision boundary and the closest of the data points, as shown on the left figure. Maximizing the margin leads to a particular choice of decision boundary (y=0 line), as shown on the right. The location of this boundary is determined by a subset of the data points, known as support vectors, which are indicated by the circles.*

The maximum margin solution is found by solving the equation (4.7).

$$\arg\max_{w,b}\left\{\frac{1}{\|w\|}\min_{n}\left[t_n\left(w^T\phi(x_n)+b\right)\right]\right\} \tag{4.7}$$

In the limit $\sigma^2 \rightarrow 0$ (σ^2 being the probability of error relative to the learned model) the optimal hyperplane is shown to be the one having the maximum margin. As σ^2 is reduced, the hyperplane is increasingly dominated by nearby data points relative to more distant ones. In the limit, the hyperplane becomes independent of data points that are not support vectors.

This method assumes there is no overlapping i.e. that the classes are completely separated. In practice, however, the class-conditional distributions may overlap and in this case exact separation of the training data can lead to poor generalization. The SVM is modified so as to allow some training points to be misclassified. Data points are allowed to be on the "wrong side" of the margin boundary but with a penalty that increases with the distance from that boundary. The penalty is a linear function of this distance that introduces slack variables ξ. $\xi_n=0$ for data points that are on or inside the correct margin boundary and $\xi_n = |t_n - y(x_n)|$ for the other points. The problem to achieve the classification by soft maximum margin problem is then given by equation (4.8).

$$\min_{n}\left\{C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|w^2\|\right\} \tag{4.8}$$

Where the parameter $C > 0$ controls the trade-off between the slack variable penalty and the margin.

SVM is fundamentally a two-class classifier. However, several methods were proposed for combining multiple two-class SVMs in order to build a multiclass classifier.

One commonly used approach was developed by Vapnik, 1998 and is called the ***One-versus-the-rest approach.*** One constructs K separate SVMs, in which the k^{th} model $y_k(x)$ is trained using the data from class C_k as the positive examples and the data from the remaining K-1 classes as the negative examples. The problem of this method is that it can lead to inconsistent results, in which an input can be assigned to multiple classes simultaneously. Moreover, the training sets are unbalanced: if there are ten classes comprising an equal amount of training data, then the individual classifiers are trained on data sets comprising 90% negative examples and only 10% positive ones, the symmetry of the original problem is lost.

The ***One-versus-one approach*** is another method to build a multiclass classifier. It consists in training K(K-1)/2 different 2-class SVMs on all possible pairs of classes, and then to classify test points according to which class has the highest number of "votes". This method avoids the symmetry problem but more computation is required, and it can lead to ambiguities in the results.

As for this study, the selected method is the one-versus-the-rest approach. Hence, four classifiers run parallel during the experiment: a first classifier wood/not wood, a second one metal/not metal etc...

4.4 Probabilistic generative models (PGM)

Probabilistic generative models are based on the assumption that the four materials likelihood functions respond to a Gaussian distribution (or normal distribution).

A random variable follows the Gaussian distribution if its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \tag{4.9}$$

Where m is the expected value and σ is the standard deviation.

From answers of the participants in the classical perceptive experiment, this model will calculate the mean and the standard deviation of the four materials in the 6-dimensions physical space. Probabilistic generative models are thus batch learning techniques: the algorithm needs all the observations at the same time to define the mean and the standard deviation. Figure 4.5 illustrates this technique for a 2-dimensions case.

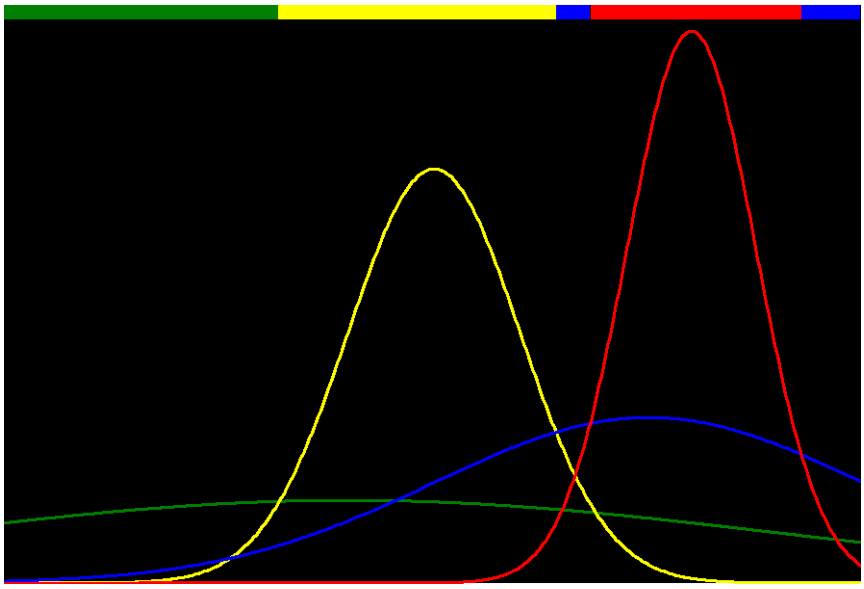

Figure 4.5 *Generative probabilistic models. Each curve corresponds to a material gaussian's distribution.*

After the training phase, for a given sound (a 6-dimensions point in the physical space), all material probabilities will be compared. The sound is assigned the material class that has the highest probability:

$$\underset{k}{\operatorname{argmax}}\{p(x|C_k)\} \tag{4.10}$$

Where $p(x|C_k)$ is the probability that sound x belongs to the class C_k.

A Gaussian classifier can form quadratic discrimination borders, which is not the case of the SVM program used for the comparison, which can only form linear planes in the case considered in this study.

4.5 Summary

This study will thus compare the results of four machine learning techniques, as seen on Figure 4.6. Active and passive perceptron algorithms will be studied as well as SVM and PGM. The only method able to drive an experiment in real time is the active perceptron algorithm, as it is based on online learning. The other methods are used to better evaluate the active perceptron algorithm performances.

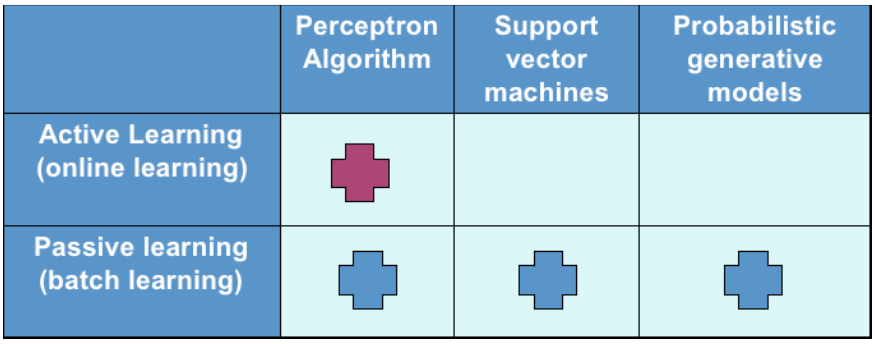

Figure 4.6 *Machine learning techniques studied in this project.*

5 The classical psychoacoustic perceptive experiment

As for this classification experiment, 20 participants were asked to identify the material of the object creating the sound they heard. This was a 4 AFC experiment (4 alternatives forced choice): the listener had to choose one of the four proposed materials.
The objective of this experiment was to get significant labelled sounds, i.e. to associate each set of physical parameters $\{f_1\ f_2\ t_1\ t_2\ g_1\ g_2\}$ to the probability of this sound to belong to a material class among {wood, metal, plastic, glass}.
Analysis of the results permitted to study the distribution of the material classes in the physical parameter space.
The labelled sounds were then used to train and evaluate the perceptive experiment based on machine learning techniques.

5.1 Composition of the sounds corpus

The sounds corpus is composed of 372 sounds that were generated from the presets (c.f. section 3.3). Sounds were chosen so as to cover the dynamic ranges of all parameters and so that the four material classes are a priori well represented. Figure 5.1 illustrates the distribution of the 372 sounds in the frequency space, the decay time space and the gain space.
As one can notice, the sounds are not equally distributed over the dynamic ranges. This is explained by the fact that a listening working session revealed that many materials could be identified for instance in the lower frequencies domain. Thus, studying more sounds in this area would permit to define finer boundaries between materials.
The sounds were not equalised in loudness. Actually, this research focuses on environmental sounds, and keeping the sounds at their natural level is important: when exciting objects of the same size and of different materials by the same exciter, there will be different sound levels. Thus, an assumption here is that every material has natural specificities and a normalisation of the sounds could spoil these specificities. Moreover, Giordano and McAdams showed in [17] that the loudness was an influencing factor for the perception of the materials: glass was not associated with loud sounds for instance.

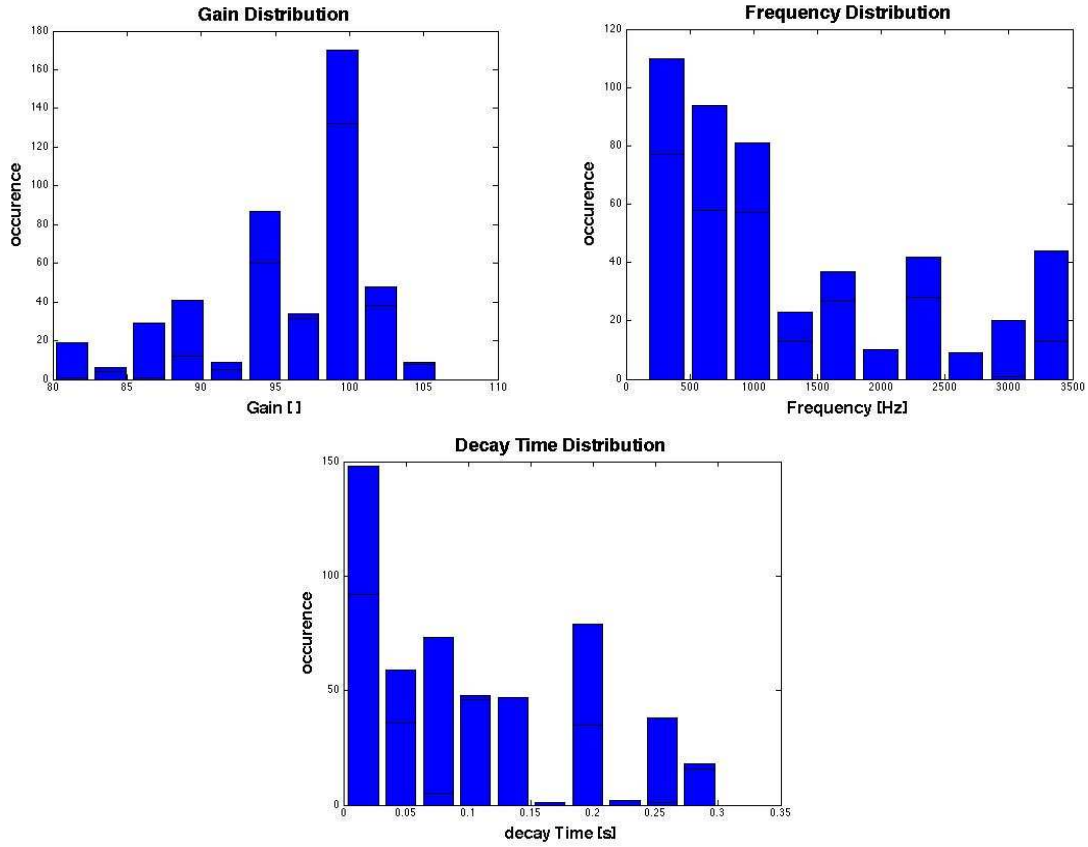

Figure 5.1: *Distribution of the sounds in the physical parameter spaces: the top left plot shows the distribution of the sounds in the frequency space, the top right in the gain space and the bottom plot, in the decay time space.*

5.2 Proceeding the experiment

The experiment took place in a soundproof booth and the sounds were emitted by loudspeakers Yamaha MSP5. The experiment was implemented on psiexp, software developed at IRCAM, and running on a MAC Pro. The sounds were played through Max-MSP 4.6 software.

12 women and 8 men individually participated on voluntary basis to the perceptive experiment. They all reported normal hearing.

The experiment was composed of two steps. To get familiar with the computer interface, it begun with a training phase in which 5 randomly chosen sounds were presented to the listener. Then, the 372 sounds were presented in a random order.

All the sounds were not presented to the listener before the experiment because of time reasons. Moreover, no example of natural sounds was given: there is no information about the geometry of the object causing the sound, and giving a natural sound would have biased the listeners.

Only one re-listening was possible to assure that the participant answered spontaneously. The experiment lasted approximately 45 minutes.

5.3 Experiment interface

Figure 5.2 presents the experiment interface. For each sound, the participant chooses among the four materials, and has to valid his answer to get to the following sound. To avoid a context effect due to the interface, the material labels had various positions depending on the participant.

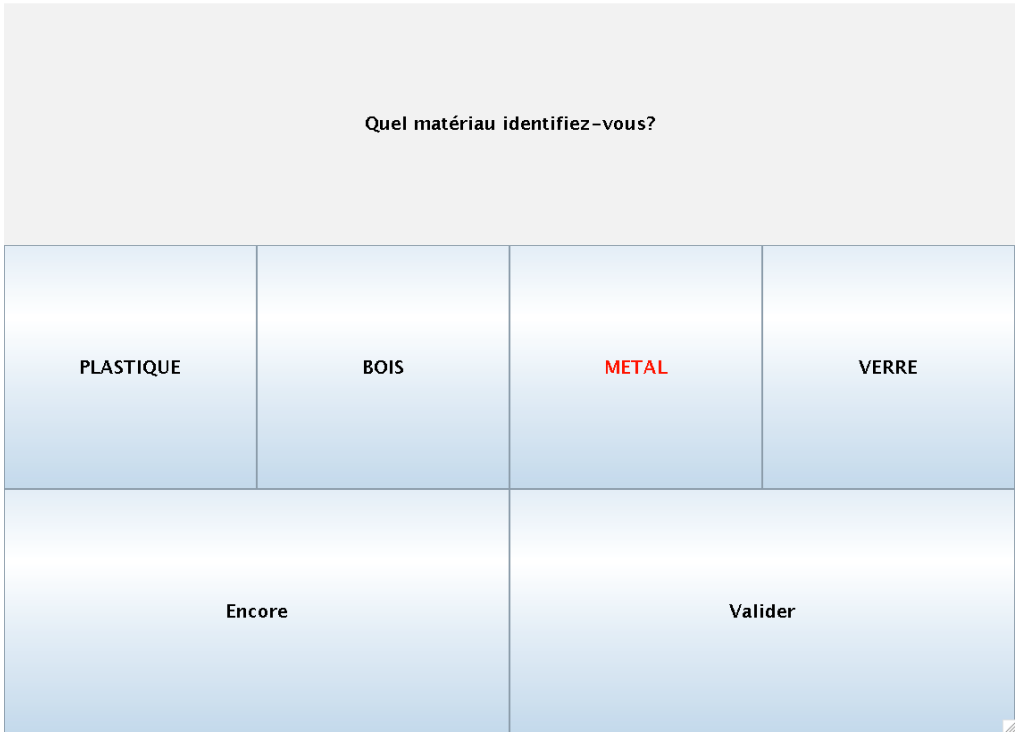

Figure 5.2: *Classical perceptive experiment interface on psiexp.*

5.4 Results of the classical perceptive experiment

5.4.1 Viewing data in 2-D spaces

Figures 5.3 till 5.5 show the 372 labelled sounds in the frequency, decay time and gain spaces. Three participant's results are provided for each space: participants 1,13 and 14. There is good agreement between participants 1 and 13, and poor agreement between participants 13 and 14.
In the frequency space (f_1 vs. f_2), one can observe that plastic and wood tend to have lower frequency than glass, but metal spreads over the frequency space.

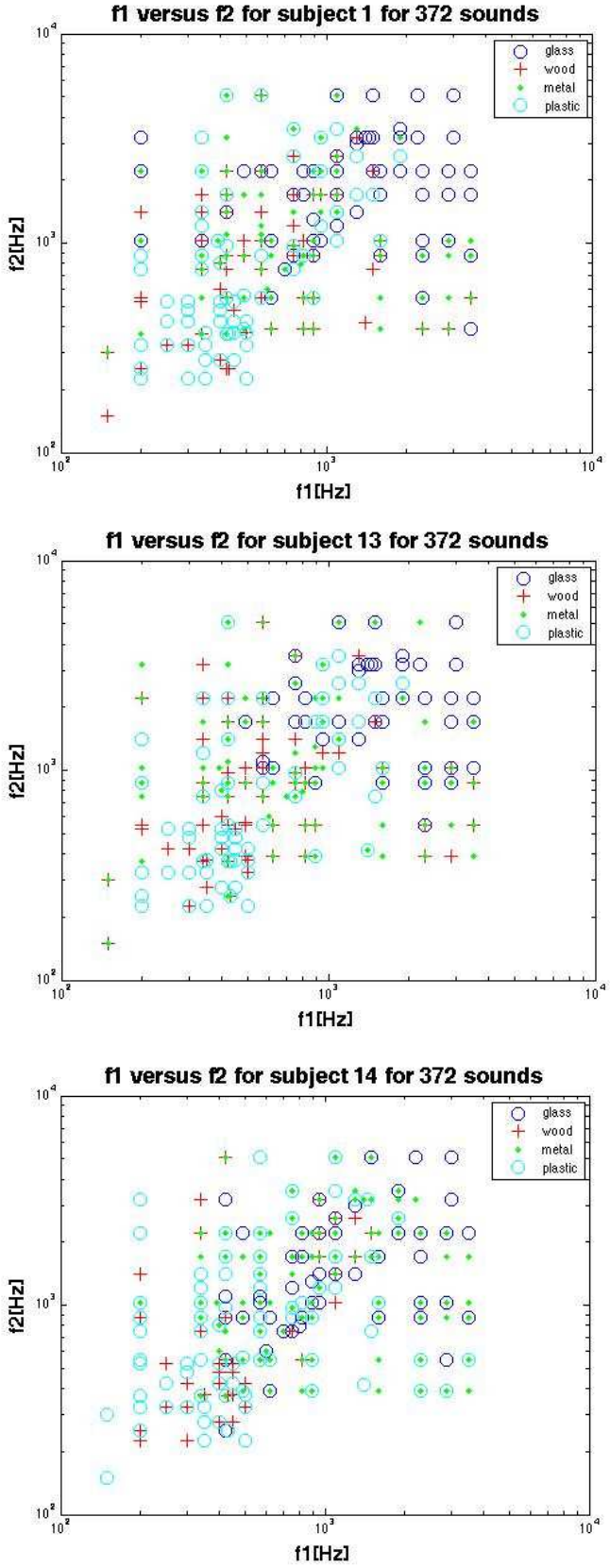

Figure 5.3: *Projected data results of the classical experiment in the frequency space $\{f_1, f_2\}$ for 372 sounds.*

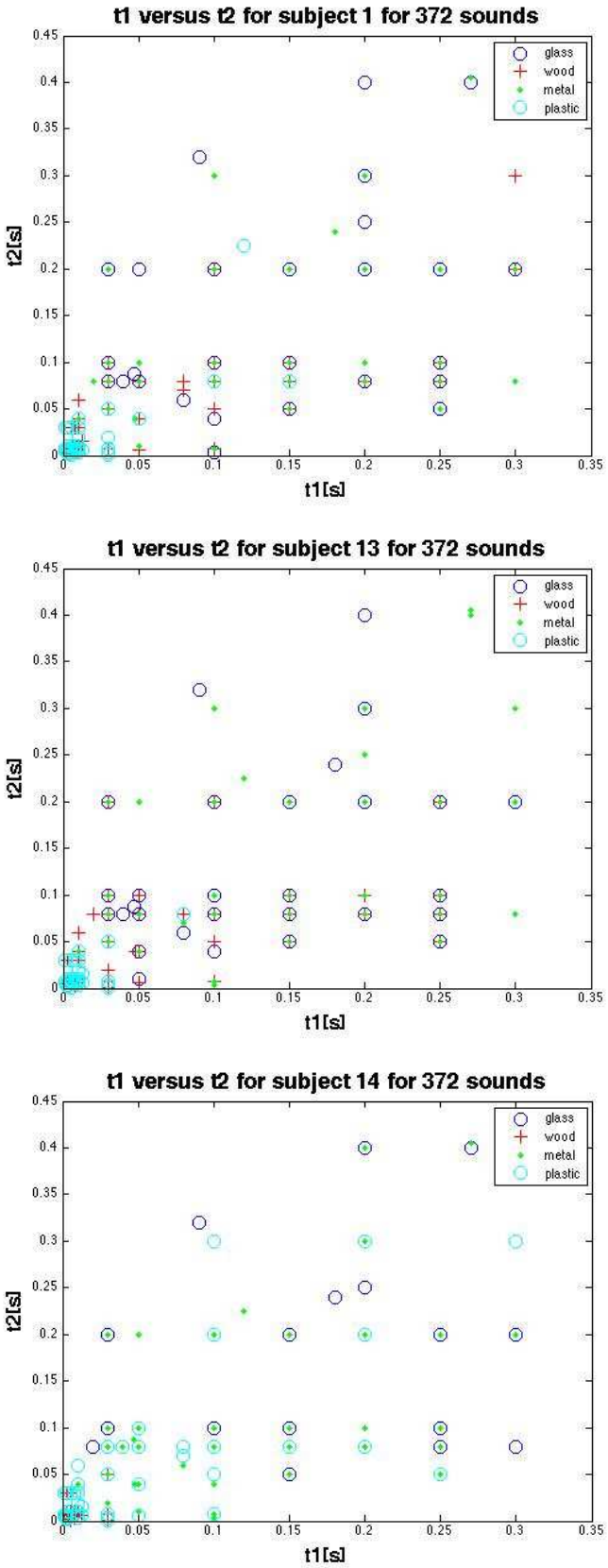

Figure 5.4: *Projected data results of the classical experiment in the decay time space $\{t_1, t_2\}$ for 372 sounds.*

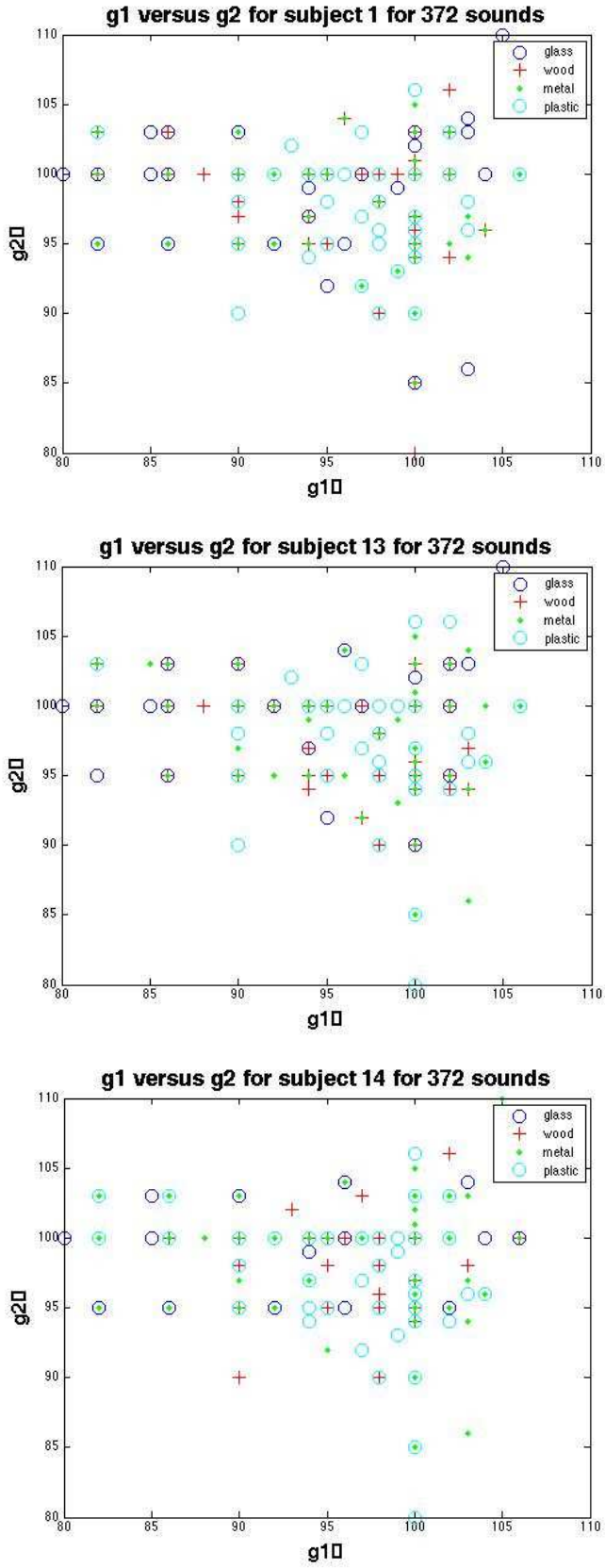

Figure 5.5: *Projected data results of the classical experiment in the gain space $\{g_1, g_2\}$ for 372 sounds.*

In the decay time space, plastic and wood are characterised by low values. Metal and glass are less well-defined, which confirms Avanzini and Rocchesso study that concluded that rubber (plastic) and wood are more defined regions. Depending on the participant, wood and plastic are not positioned in the same order. As for the gain space, one can notice that no particular pattern can be derived. The data are kind of randomly spread over the space.

Figure 5.6 shows the data projected in the best angle with regards to the classes' separation. Linear discriminant analysis is used to obtain the most pertinent axis, which are linear combinations of the model's physical parameters.

In this space, the gross-class wood/metal is obvious: for all participants, the red and the yellow data points are mixed together. As for the gross class plastic/wood, data are more or less separated depending on the participant.

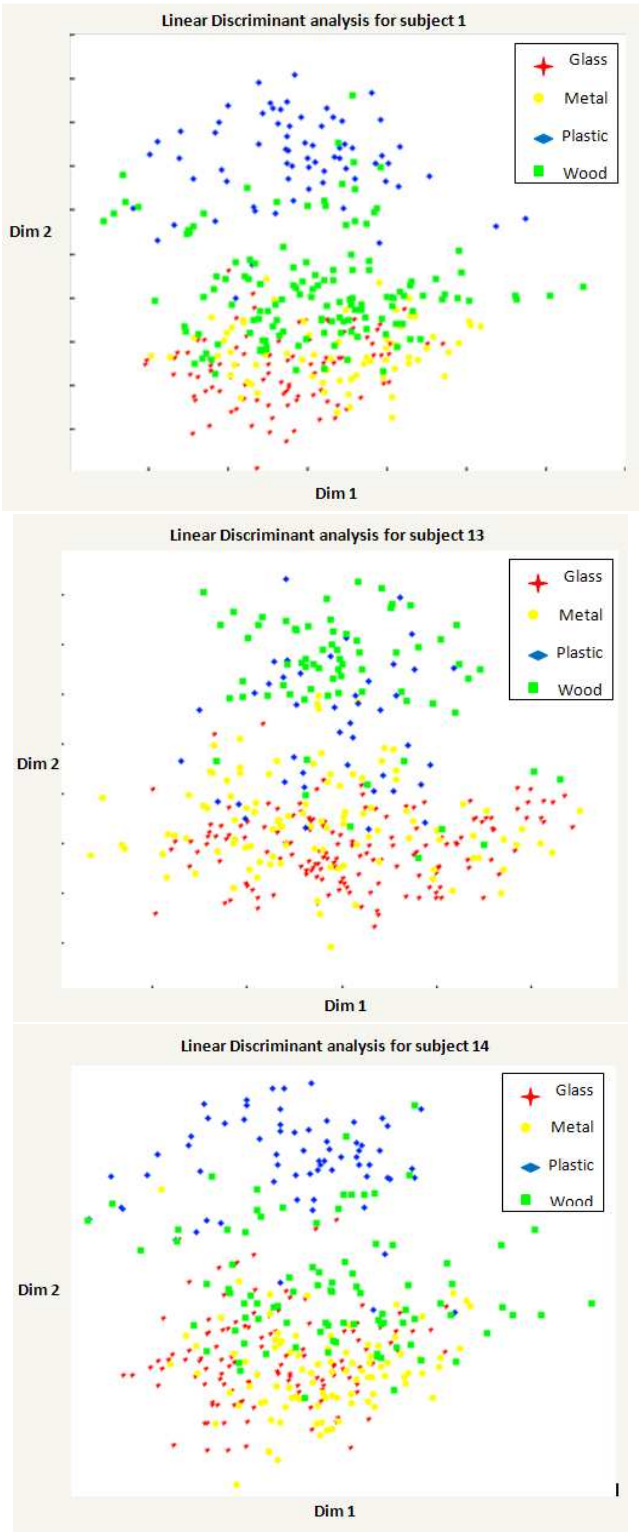

Figure 5.6: *Data projected in the best separating-class space, calculated by Linear Discriminant Analysis, for three different subjects.*

5.4.2 Degree of identification of sounds and agreement between participants

Figure 5.7 represents the degree of identification of the 372 sounds. For each sound, the participants' answers distribution over the four materials was calculated, and the maximum material percentage was defined as the identification degree of the sound.

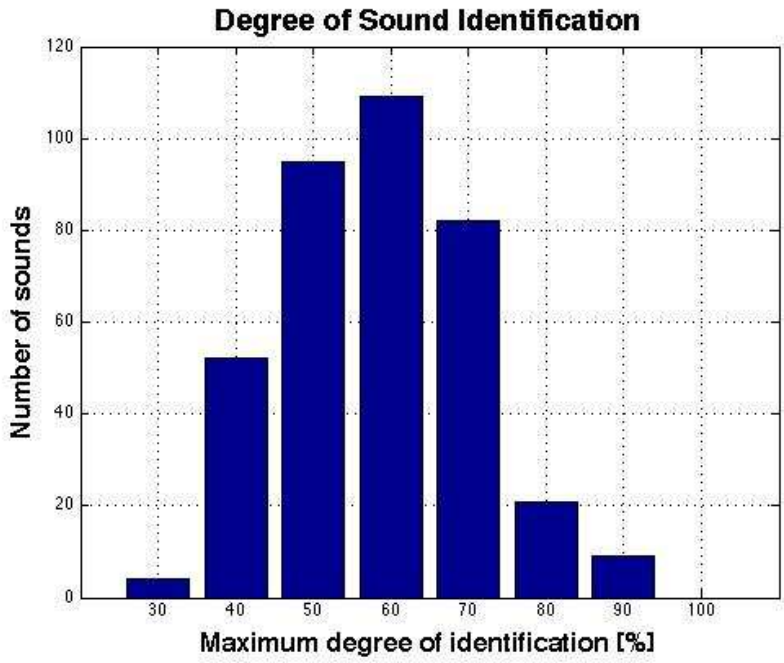

Figure 5.7: *Sounds degree of identification for the classical perceptive experiment.*

This graphic indicates that no sound was identified with a 100% score. 109 sounds were identified at 60%, and most sounds were recognized between 40% and 70%.
The agreement between participants is given by the kappa coefficient. As this study considers 4 classes, the kappa were calculated with Cohen's Kappa adapted to many raters by Fleiss [21]. Kappa is defined by relation 5.1

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \tag{5.1}$$

Where P_0 is the observed agreement and P_e is the proportion of expected agreement. Kappa is comprises between -1 (complete disagreement) and +1 (total agreement). Table 5.1 gives the kappa values obtained for the 20 participants. Kappa represents the chance-corrected probability that 2 judges agree, i.e. that two participants choose the same material over.
The results show fair global agreement, and slight agreement for wood. Best agreement is found for plastic, but its kappa remains low.

Global Kappa	0.23 (error: 6.8e-6)
Kappa_wood	0.16
Kappa_metal	0.22
Kappa_plastic	0.32
Kappa_glass	0.23

Table 5.1 *Kappa values for 20 participants, four material classes and 372 sounds.*

The weak value of this kappa coefficient could be partly due to material confusion, i.e. to a good identification of gross categories. To calculate the material confusion of participants, the two maxima of the participants' answers distribution over material were counted and summed up. Results are presented on graph 5.8.

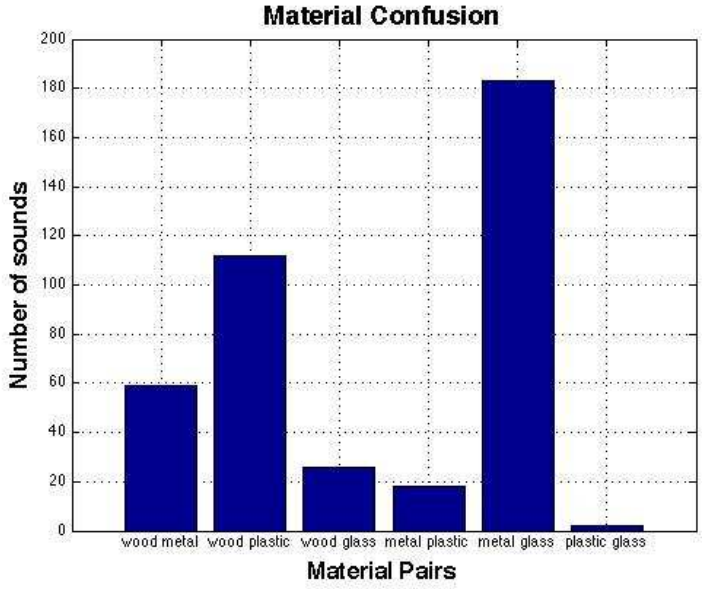

Figure 5.8: *Material confusion: the x-axis shows all possible pairs of materials, and the y-axis indicates the number of sounds for which the two participants' answer maxima correspond to the material pairs.*

This graph underlines confusions between metal and glass, as in Girdano and McAdams' paper ([17]). Actually, 182 sounds out of the 372 were identified by the participants as being either glass or metal. The other gross class is composed of plastic and wood and counts 112 sounds. Confusion is visible as well for wood and metal (59 sounds). These results are consistent with the literature, which highlighted the perception of the following gross classes: wood/plastic and metal/glass. The kappa coefficient for the two gross classes is

Kappa_gross_classes = 0.55 ***(error: 2.9e-5)***

The identification of gross-classes material is thus characterised by moderate agreement. This value is better than the global kappa but remains somewhat weak. The physical model can explain these results. Actually, it relies on cartoonification principles, which involve very simple physical modelling algorithms. Thus, some sounds can be hardly identified, and the participants' answers can be characterised by chance. The Chi-square test permits to identify these sounds.

5.4.3 The Chi-square test

The Chi-square test is a statistic test that permits to test the null hypothesis, i.e. in our case, to know whether a sound's material was randomly chosen by the participants. The Chi-squared statistic is given by

$$\chi^2 = \sum_m \frac{(Observed\ \ frequencies - Expected\ \ frequencies)^2}{Expected\ \ frequencies} \tag{5.2}$$

Where m is an index for the material class. The expected frequency to test the null hypothesis is the random probability: 25% for four materials. If the Chi-square test is above a certain value defined by tables, depending in the number of classes and the desired precision, then the null hypothesis can be rejected.
A Chi-square test is done for every sound. This implies 372 chi-square test. The probability that a sound is wrongly selected by the statistical test increases with such a huge number of sounds. Therefore, a correction has to be applied to the threshold. Benjamini [22] and Bonferroni [23] both proposed corrections. Table 5.2 gathers the results of the test. The significant sounds are sounds that reject the null hypothesis.

Correction	Number of significant sounds	Kappa values for 4 classes	Kappa values for 2 classes
No correction	**372**	0.23	0.55
Benjamini	**196**	0.30	0.63
Bonferroni	**53**	0.45	

Table 5.2 *Results of Kappa and Chi-square tests for different corrections and different classes.*

Bonferroni's correction is very conservative. Only 53 sounds are selected by this method. As for Benjamini's correction, it selects 196 sounds having a fair agreement: the kappa value is equal to 0.3.

Considering gross classes with the 196 selected sounds, the agreement among participants is very high. Actually, the kappa is then equal to

Kappa_gross_classes = 0.63 ***(error: 1.8e-5)***

The gross classes are thus well identified.

Agreement between participants reaches 0.45 for 4 classes and 53 sounds (selected by Bonferroni's correction), which is a moderate agreement.
The next section presents the projection in 2-D spaces of the sounds selected by Bonferroni's correction. Classification patterns should be more visible on these values, as there is a better agreement between the participants.

5.4.4 Selected data viewed in 2-D projections

A first observation is that there are no many wood sounds: this means that wood is not well represented in the most significant sounds, and thus that it is not very well recognized (the answer percentage distribution is not high for wood sounds).
Material regions clearly appear in the frequency space: with increasing frequencies, there are plastic, metal and glass. When wood is represented (participant 14), its frequency values are located between plastic's and metal's values.
As for the decay time space, plastic and glass regions are visible. Plastic are identified for very low decay times, and glass for higher values. As for wood, participant 14 locates it below plastic. Metal is not well defined in this domain.
In the gain space, there is no particular pattern: the sounds are distributed all over the space.

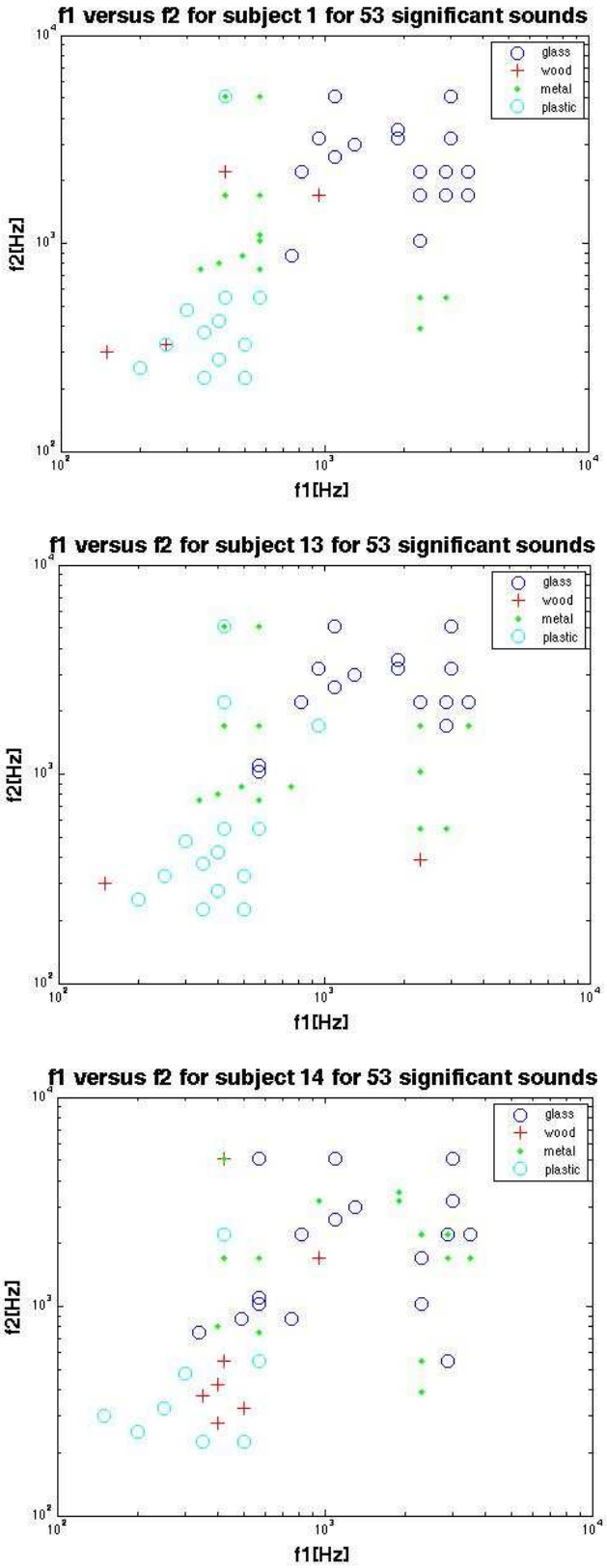

Figure 5.9: *Projected data results of the classical experiment in the frequency space $\{f_1, f_2\}$ for 53 sounds.*

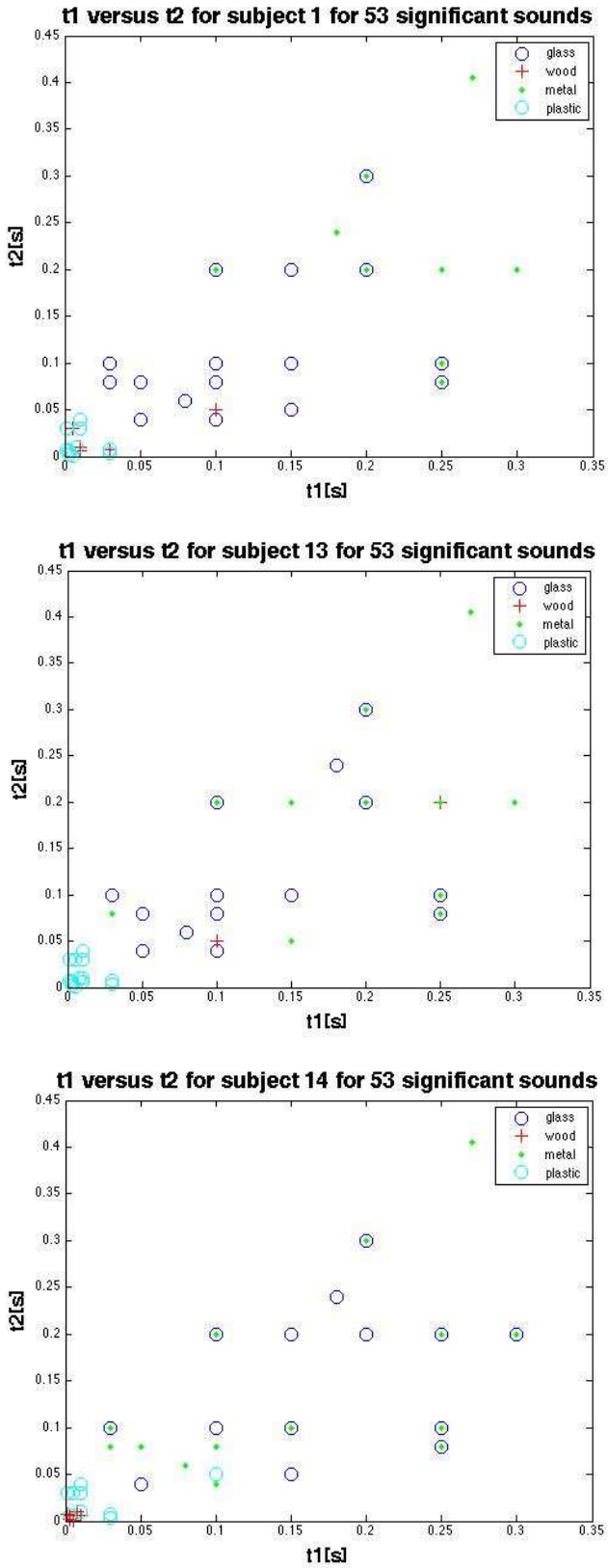

Figure 5.10: *Projected data results of the classical experiment in the decay time space $\{t_1, t_2\}$ for 53 sounds.*

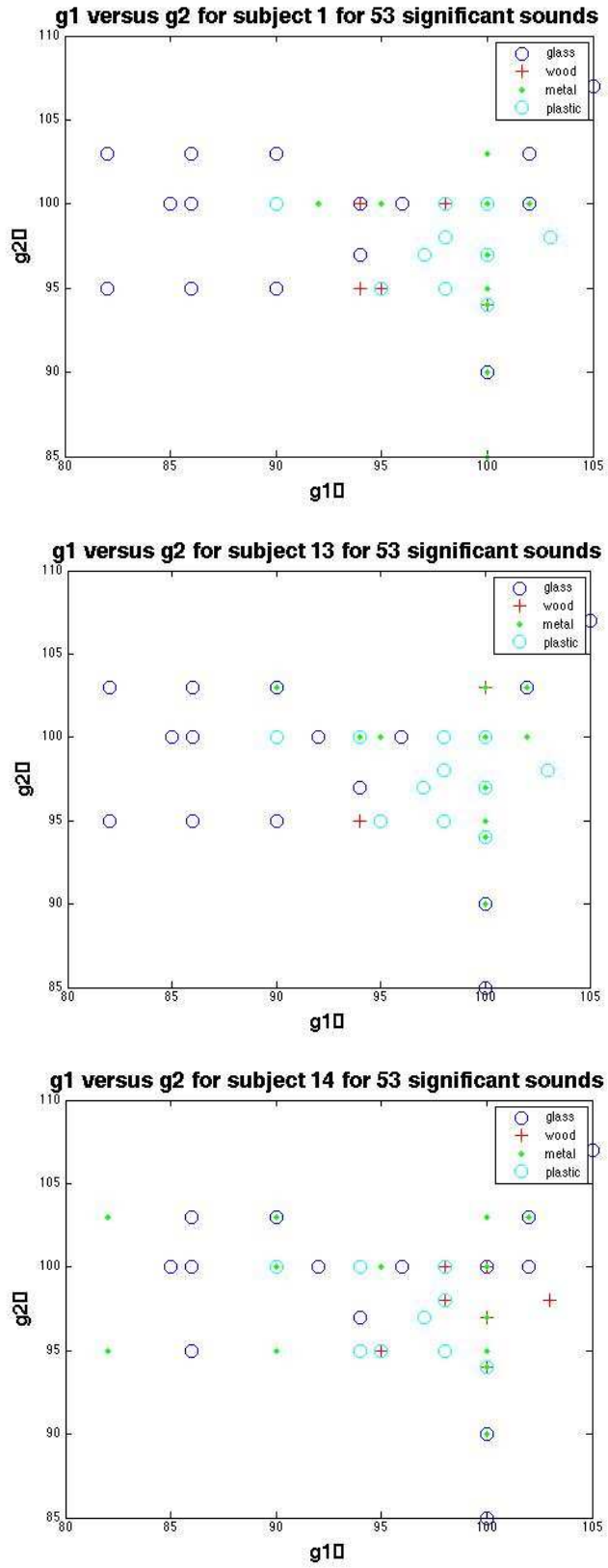

Figure 5.11: *Projected data results of the classical experiment in the gain space $\{g_1, g_2\}$ for 53 sounds.*

5.4.5 Conclusions

Some rough regions appear in the frequency and decay time spaces, especially for plastic and glass. Metal is better defined in the frequency space than in the decay time space. Wood got the worst agreement between participants, and was slightly characterised in both spaces. In increasing frequencies, the material ordering is plastic, wood, metal and glass and in increasing decay times, the order is wood, plastic and glass. These results are mainly consistent with the researches described in the literature part, with some inversions between wood and plastic and between metal and glass.

However, the agreement between participants is not very strong. Actually, this study highlights confusions between wood and plastic on the one hand, and metal and glass on the other hand. These gross classes had been observed by Giordano and McAdams in [17].

For these two gross classes, the agreement among participants is much higher, especially with the 196 sounds rejecting the null hypothesis.

Two sets of data will be tested in the next experiment based on machine learning techniques. On the one hand, noisy data composed of the 372 sounds (4 classes and 2 classes), and on the other hand, more significant data composed of the 196 sounds selected by Benjamini's correction 4 and 2 classes). These two sets will be used to train and evaluate the machine learning algorithms.

6 Experiment based on machine learning techniques

The objective of this experiment is to evaluate the ability of machine learning techniques for a perceptive classification experiment. In the long term, this would allow machine learning algorithms to drive perceptive experiments so as to realise them on large-scale sounds corpus. Driving an experiment involves an active component that will cleverly choose the next sound to be presented to the participant.
This chapter evaluates active learning techniques as well as passive learning techniques in order to have some comparison basis.

6.1 Proceeding of the experiment

As seen in the introduction (see Figure 1.3 on page 4), the results of the classical perceptive experiment were given to the machine learning algorithm. Beforehand, data were divided in two groups, group 1 and group 2. The results consist, for each of the 20 participants, in an input vector (the 6 physical parameters) and an output vector (the associated material). Group 1 sounds of each participant was used to train the machine learning algorithms: the algorithm knows the "good" answer, i.e. the material given by the participant. This group was composed of 272 sounds for the initial corpus and 96 sounds for the more significant sounds corpus. As for SVM and PGM methods, all sounds of group 1 will be used for the training. As for the perceptron algorithm, the point is to evaluate whether active learning techniques converge faster by cleverly choosing the points. Therefore, the algorithm will train with 35 data of group 1: these 35 points are randomly chosen by the passive perceptron algorithm and are cleverly chosen by the active perceptron algorithm.
Then, a virtual participant was simulated to test the generalisation ability of the algorithm. The algorithm is given an unlabelled data from group 2 (100 data) and has to decide in which class this points belong to.
Finally, the group 2 results of both classical and automatic experiments (the real participant and the simulated one) are compared to evaluate the machine learning algorithm.
The experiment was done separately for 20 virtual participants, so as to avoid inter-individual differences. Then, the mean error on all participants was calculated.

6.2 Experiment results

This section first presents the machine learning results for 372 sounds, i.e. for noisy data and then for 196 classes. Active learning (active perceptron algorithm) will be compared to passive perceptron algorithm, support vector machines and probabilistic generative models.

6.2.1 Results with the initial corpus (372 sounds)

Figure 6.1 presents the error for each virtual participant of active learning and passive learning algorithms (perceptron algorithm) for a four materials classification and for the initial corpus of sounds. This error is the error on the 100 testing sound after a training of 35 data. Actually, passive perceptron and active perceptron algorithms have the same results for a large amount of sounds. The difference is important at the beginning of the training: active learning is advantageous if better results are obtained for fewer sounds. As for active learning, the 35 sounds were cleverly chosen within the group 1. As for passive learning, 35 sounds were randomly chosen within the 272 sounds of group 1.
Concerning Support vector machines and Probabilistic generative models, the training was done on the 272 sounds.

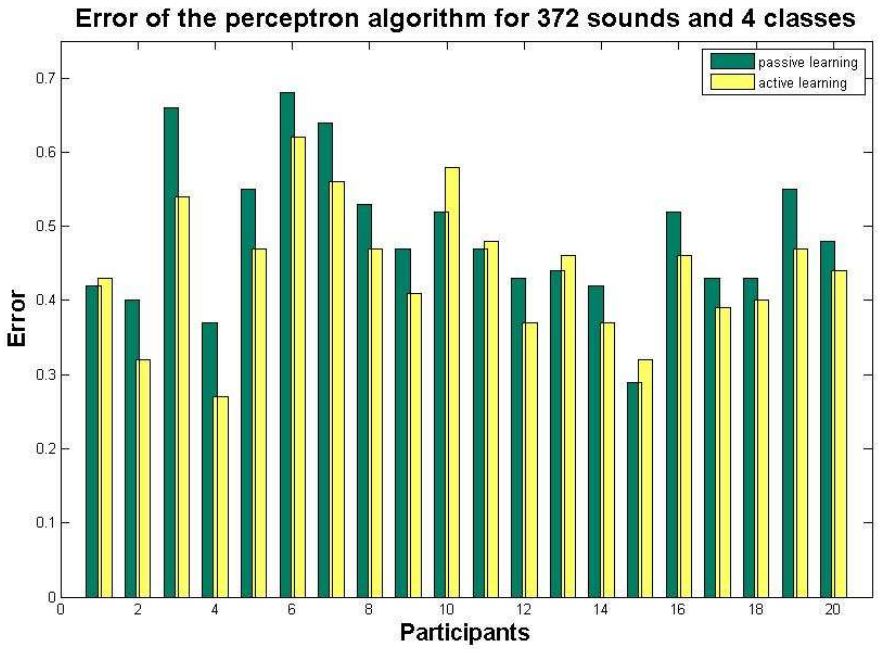

Figure 6.1: *Error of active and passive perceptron algorithm for 4 classes and 372 sounds for each virtual participant.*

Some important differences in the results are visible among the 20 participants.

The mean errors of both methods over participants are:

Active mean error = 0.44

Passive mean error = 0.49

This error is very high: this means that the active learning perceptron algorithm has 44% chance to fail the classification. However, the error turns out to be better than the passive algorithm error.

As shown on figure 6.2, support vector machines have worst results than active learning but Gaussian classifiers are more powerful. This is explained by the quadratic boundary of this last algorithm, whereas SVM and perceptron algorithms can only draw linear boundaries.

Perceptron Algorithm	Support Vector Machines	Probabilistic Generative Models
44%	46%	40,7%

Figure 6.2: *Mean error for the 20 participants of perceptron algorithm, SVM and probabilistic generative models for 372 sounds and 4 classes.*

Figure 6.3 shows the perceptron algorithm error for 372 sounds and the two gross classes (wood/plastic and metal/glass).

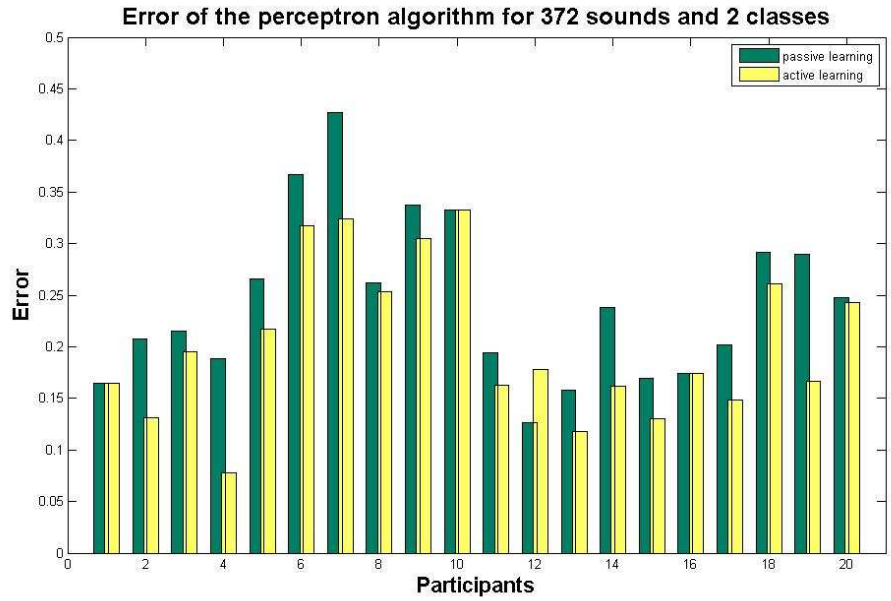

Figure 6.3: *Error of active and passive perceptron algorithms for 2 classes and 372 sounds for each virtual participant.*

The mean errors of both methods over participants are:

Active mean error = 0.20

Passive mean error = 0.24

The error for gross classes is much better than for the four materials. Active learning still presents better results than passive learning. However, 20% of the sounds were misclassified.

6.2.2 Results with significant data (196 sounds)

In this section, machine learning algorithms performance is evaluated for more significant data, i.e. for data selected by the Chi-square test corrected by Benjamini [22]. The training set is composed of 96 sounds (35 chosen by the perceptron algorithm), and the test set of 100 sounds.
Figure 6.4 presents the results for four classes.

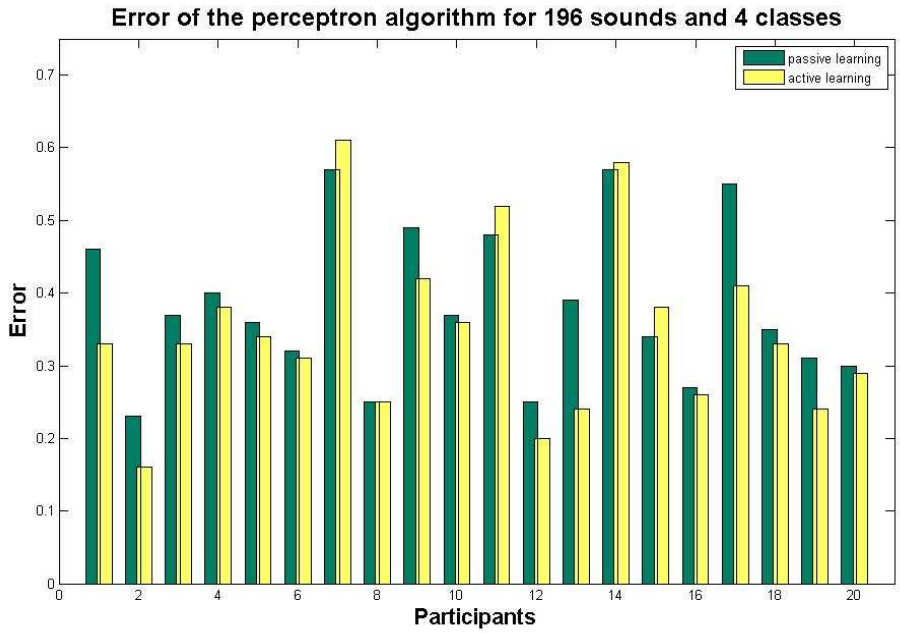

Figure 6.4: *Error of active and passive perceptron algorithms for 4 classes and 196 sounds for each virtual participant.*

The mean errors of both methods over participants are:

Active mean error = 0.35

Passive mean error = 0.38

Here again, active learning shows better results than passive learning. However, the error is very high. More than one third on the data was misclassified.
Such poor performance is obtained with SVM and probabilistic generative models, as shown on picture 6.5.

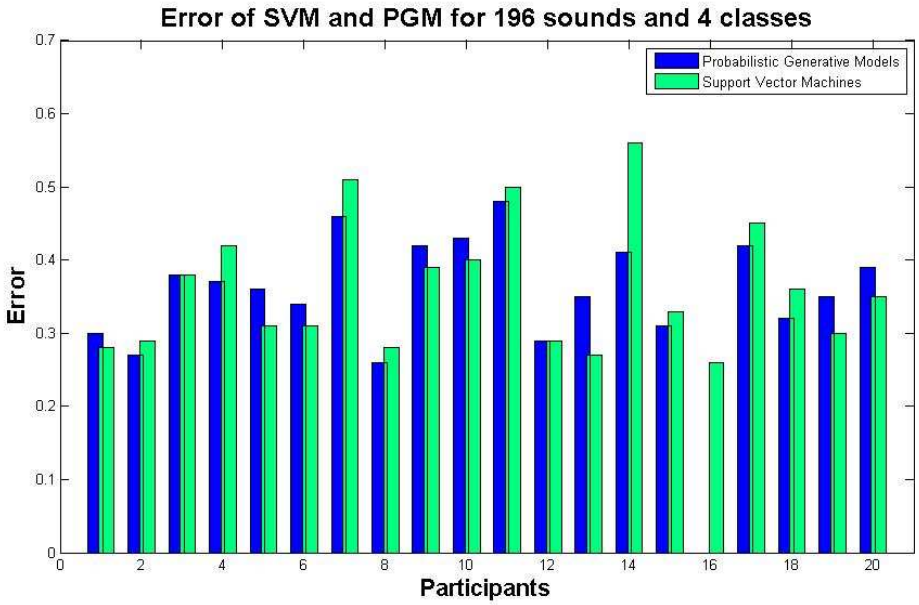

Figure 6.5: *Error of SVM and PGM for 196 sounds and 4 classes for each virtual participant.*

The mean errors of both methods over participants are:

SVM mean error = 0.35

PGM mean error = 0.36

Better results are obtained with the identification of the two gross-classes, as shown on Figure 6.6.

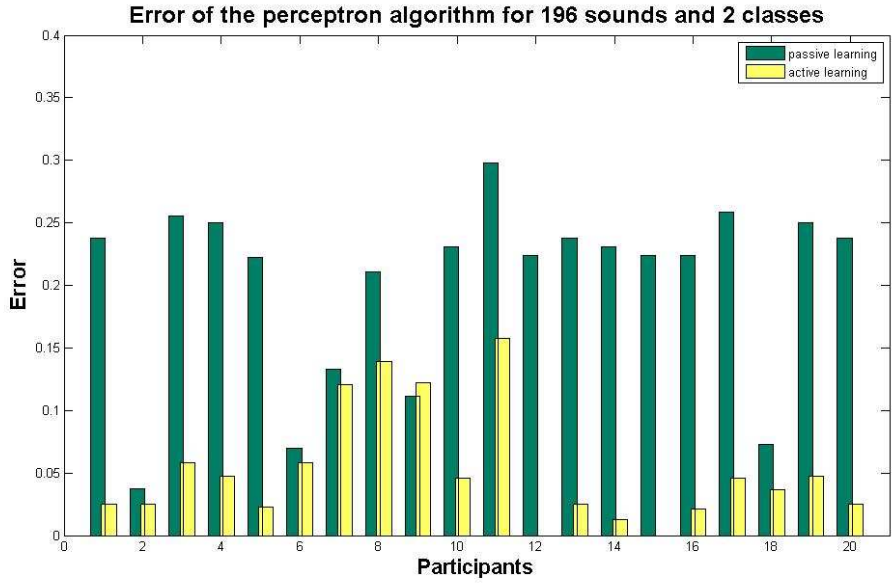

Figure 6.6: *Error of active and passive perceptron algorithms for 4 classes and 196 sounds for each virtual participant.*

The mean errors of both perceptron algorithm methods over participants are:

Active mean error = 0.05

Passive mean error = 0.20

Active learning performs the classification with 5% error, and much better than passive learning.

SVM and PGM also give very good results with cleaned data and 2 gross classes.

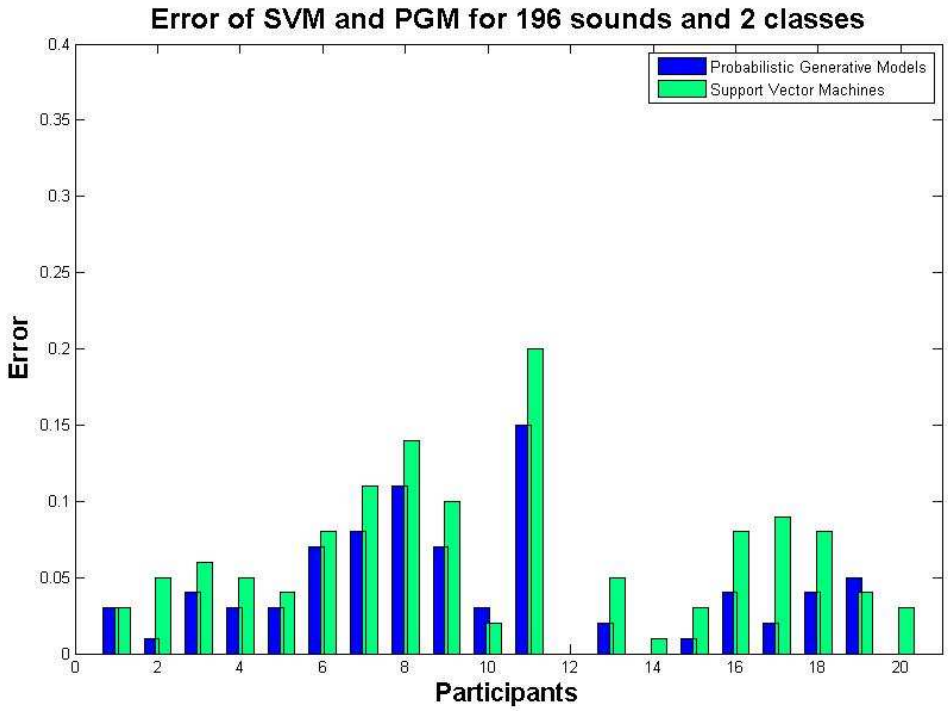

Figure 6.5: *Error of SVM and PGM for 2 classes and 196 sounds for each virtual participant.*

The mean errors of these two methods are:

SVM mean error = 0.04

PGM mean error = 0.06

6.2.3 Conclusion

These experiments show that active learning helps converge more rapidly than passive learning. Active learning algorithm show very good results for cleaned data and gross classes. However, it remains slightly less efficient than probabilistic generative models. This is due to the fact than the perceptron algorithm only draws linear boundaries between material regions, whereas probabilistic generative models provide a quadratic boundary.

7 Conclusion

Finally, the classical perceptive experiment showed little consensus among participants for the four material classes identification but significant results. However, it highlighted much better agreement for gross material classes: wood and plastic on the one hand, metal and glass on the other hand. Statistical analysis based on Chi-square statistic permitted to obtain a restricted sounds corpus composed of the most significant sounds.

Experiments based on machine learning techniques were given these two corpuses to investigate the algorithms performance with both "noisy" and "cleaned" data. The perceptron active learning algorithm showed an important error of 35% for the restricted corpus and four classes, but this is explained by the poor performance of participants to identify the four material classes. Active learning turned out to be much more efficient with the 2 gross classes' identification: only 5 % error was obtained.

A possible improvement of this machine learning technique is the implementation of quadratic boundaries between the material classes instead of linear boundaries. This would permit active learning to be more efficient with noisy data.

The limit of these methods is the necessity to have a training phase, which compels to perform a preliminary experiment to tune the model.

Finally, this multidisciplinary work introduced a promising experimental method for psychoacoustic tests. This would permit to build a perceptive interface for sound synthesis models, usable by sound designers.

The next step would be to implement, in real time, a classification perceptive experiment based on active learning techniques on a large-scale corpus. Another possible work is to use active learning for a work based on acoustic descriptors.

Bibliography

[1] P. Susini, N. Misdariis, O. Houix, G. Lemaître. *Everyday sound classification – Experimental classification of everyday sounds.* Deliverable 4.1 Part 2 of the CLOSED project, IRCAM, June 2007.

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

[3] Y. Gérard. *Mémoire sémantique et sons de l'environnement.* PhD thesis, Université de Bourgogne, 2004.

[4] M. M. Marcell, D. Borella, M. Greene, E. Kerr, S. Rogers. *Confrontation naming of environmental sounds.* J. of clinical and experimental neuropsychology, 22(6):830-864, 2000.

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

[6] R. P. Wildes, W. A. Richards. *Recovering Material Properties from Sound.* In W. Richards, editor, Natural Computation. Cambridge, MA:MIT Press, 1988.

[7] A. Chaigne, C. Lambourg. *Time-domain simulation of damped impacted plates: I. theory and experiments.* J. of the Acoustical Society of America, 109(4):1422-1432, April 2001.

[8] A. Chaigne, C. Lambourg, D. Matignon. *Time-domain simulation of damped impacted plates: II. Numerical models and results.* J. of the Acoustical Society of America, 109(4):1433-1447, April 2001.

[9] D. Rocchesso, F. Fontana, editors. *The Sounding Object*. Mondo Estremo, 2003. Freely available from http://www.soundobject.org/.

[10] R. L. Klatzky, D. K. Pai, E. P. Krotov. *Perception of Material from Contact Sounds.* Presence 9(4), 399-410, 2000.

[11] F. Avanzini, D. Rocchesso. *Controlling material properties in physical models of sounding objects.* Proceedings of the International Computer Music Conference, La habana, Cuba, 91-94, 2001.

[12] R.A. Lufti, E. Oh. *Auditory discrimination of material changes in a struck-clamped bar.* J. of the Acoustical Society of America, 102(6), 3647-3656, 1997.

[13] D. J. Hermes. *Auditory Material Perception.* IPO Annual Progress Report 33, 1998.

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

[15] A. J. Kunkler-Peck, M. T. Turvey. *Hearing Shape.* J. exp. Psych. Hum. Percept. Perform 26(1), 279-294, 2000.

[16] C. Carello, J. B. Wagman, M. T. Turvey. *Acoustical Specification of Object properties.* J. Anderson, B. Anderson editors, Moving Image theory: Ecological consideration, Southern Illinois University Press, Carbondale, 2003.

[17] B. L. Giordano, S. McAdams. *Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates.* J. of Acoustical Society of America 119(2), February 2006.

[18] S. D. Monache, D. Devallez, C. Drioli, F. Fontana, S. Papetti, P. Polotti, D. Rocchesso. *Algorithms for ecologically-founded sound synthesis: library and documentation.* Deliverable 2.1 of the CLOSED project, Univerona, May 2005.

[19] K.H. Hunt, F.R.E. Crossley. *Coefficient of restitution, Interpreted as Damping in Vibroimpact*. ASME J. Applied Mech., pages 440-445, June 1945.

[20] C.M. Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006.

[21] J. L. Fleiss. *Measuring nominal scale agreement among many raters.* Psych Bull, 76, 378-382, 1971.

[22] Y. Benjamini, D. Yekutieli. *The control of false discovery rate in multiple testing under dependency.* Annals of Statistics 29, 1165-1188, 2001.

[23] S. P. Wright. *Adjusted P-Values for simultaneous Inference.* Biometrics, Vol. 48, 1005-1013, 1992.

[24] M. Derio. *Approche perceptive dans un logiciel de synthèse sonore par modélisation physique: Modalys.* Master's thesis, Université du Maine, 2005.

[25] R. Dos Santos. *Interface perceptive de contrôle dans un logiciel de modélisation par synthèse physique.* Master's thesis, Université Pierre et Marie Curie, Jussieu Paris VI, 2006.