Acta Acustica united with Acustica
Vol. 95 (2009) 356 – 372

DOI 10.3813/AAA.918158

The Sound Quality of Car Horns: Designing New Representative Sounds

Guillaume Lemaitre, Patrick Susini, Suzanne Winsberg, Stephen McAdams*
STMS-IRCAM-CNRS, 1 place Igor Stravinsky, 75004 Paris, France. Guillaume.Lemaitre@ircam.fr

Boris Letinturier
SCE-Klaxon, 106 ZI du Clos de la Reine, 78410 Aubergenville, France

We would like to dedicate this article to the memory of our colleague Suzanne Winsberg

Summary
This paper is the second of a two-part study of the quality of car horn sounds. It aims to provide insights into the design of new sounds. It is based on the assumption that hearing a car horn sound warns road users because they recognize the sound of a car horn, i.e. they know what this sound means, and what they have to do as a consequence. The three experiments reported in this paper are grounded in a psychoacoustical framework. They seek to provide car horn builders with recommendations allowing them to create new sounds.
In the first part [1], we studied the perception of the timbre of existing car horn sounds. We found that, from their perception of the sounds, listeners were able to make inferences concerning the different mechanisms causing the sound, and that the perceived differences between the sounds were based on the integration of three elementary sensations, correlated with three acoustical descriptors. In this second part, we focus on the agreement among listeners in categorizing sounds as being members of the car horn category. Membership agreement is operationally defined as the result of a two-alternative forced-choice task. We first study recordings of existing sounds. The results allow us to define relationships that predict membership agreement from a set of acoustical descriptors. To extend these results, we create a new set of sounds in a second step, which we submit to a timbre study similar to the one reported in [1]. We finally study membership agreement for these synthesized sounds. The results allow us to define a methodology to create new car horn sounds.

PACS no. 43.50.-x, 43.66.-x

1. Introduction: Designing sounds that warn road users

Car horns are wildly used and widely down-cried. However, they have the capital (and legal) function of warning road users against potential danger[1]. At the same time, car horn builders wish to tune their sounds to match them to car categories and brand identities. Designing the sound of car horns thus involves a compromise between the need to customize the sounds and the necessity of providing efficient warning signals. To fulfill these constraints, car horn builders wish to create new sounds by means of a new device, made of an electronic synthesizer and a loudspeaker (see [3] for a description of the device). In this context, the goal of this study is to identify the acoustical properties of car horn sounds that allow them to be recognized as such, i.e. that allow the sounds to convey information concerning danger to the listener. This will allow car horn builders to design new sounds still perceived as car warning signals.

1.1. Timbre of current car horn sounds

The first step of this study consisted in investigating existing car horn sounds [1]. A car horn is a self-oscillating electro-acoustical device. Two main categories exist. The first kind (*horn-like* devices) is based on an electrodynamical driver and horn. The second kind (*plate-like* devices) is also made of an electro-dynamical driver, but there is a metal plate attached to the membrane, and no horn. The devices are usually mounted alone (*monophonic* sounds), or in twos or threes, resulting in chords (*polyphonic* sounds).

We studied their timbre by using the psychoacoustical definition of timbre (ANSI definition [4], as summarized by Krumhansl [5] p. 44): "the way in which musical

Received 3 January 2008,
accepted 3 November 2008.

* now at CIRMMT, Schulich School of Music, McGill University, 555 Sherbrooke St. W., Montréal, Québec, Canada H3A 1E3

[1] For instance, according to the French "Code de la route" (articles R313-33, R416-1, http://www.legifrance.fr), i) every motorized vehicle must be provided with an acoustic warning device; ii) any other acoustic signal is forbidden; iii) using the acoustic warning signal is only allowed to warn road users against an immediate danger. Even the United Nations requires a mandatory minimum acoustic level [2]

sounds differ once they have been equated for pitch, loudness and duration". The results of the experiments showed that the perceived dissimilarities between the sounds of car horns are subtended by a small number of independent *auditory attributes* (or perceptual dimensions) defining the timbre space of the car horn sounds. Moreover, there are *acoustical descriptors* correlated with these attributes:

- the first attribute is correlated with modulations of the temporal envelope (roughness),
- the second attribute is correlated with the spectral distribution of energy (spectral centroid). This descriptor has been found in other studies to correspond to the attribute of *brightness*, and
- the third attribute is correlated with fine variations of the spectral envelope (spectral deviation).

Further, a sorting task showed that listeners group together the sounds in categories corresponding closely to the different sound-production mechanisms (type of excitation, type of resonator, number of devices). These categories are defined by the auditory attributes shared by the sounds and thus by the acoustical descriptors.

1.2. Warning signals

There have been different approaches to the perception and design of warning signals (see [6] for an overview). The first idea was that a warning signal has to be **audible** to the listeners for whom it is intended [7]. The issue can be addressed by ensuring a given acoustic level in a specific area [8], or by designing signals with a low detection threshold (considering the design issue as a signal detection paradigm) [9, 10, 11]. However, as noted in [12], the detection threshold is not only modulated by the signal to noise ratio, but also by the perceived relevance of the signal with respect to the perceived dangerousness of the situation.

Indeed, in many contexts (military aircraft, surgery rooms), many different warning signals occur incessantly and concurrently. And it may happen that users become unable to decide whether a warning is really urgent or not, and therefore, if they have to respond or not to this signal. Thus, Edworthy *et al.* [13] [14] have proposed to design warning signals with different levels of **perceived urgency**, relevant to the actual urgency of the situation. Urgency is in this case conceived as an auditory attribute of the sound. Signals made of bursts of harmonic pulses (after [7]) and speech signals [15, 16] were studied. The results of such experiments are mathematical relationships between perceived urgency and acoustical properties (the more urgent signals are high in pitch, with short transients, partials with random frequencies, and an irregular rhythm). However, some results also show that if the perceived urgency can be manipulated by changing the acoustical properties of the sounds, the meaning of the speech signals also has a great influence on the perceived urgency.

So the idea has moved from designing alarms with different urgency levels to warning signals that also **inform** the listener of the reason for this warning [17]. Therefore, three requirements for a warning signal were identified: the signal must be recognized as a warning, the listener must know what it represents and what should be done in consequence. The same idea is found in the analysis of the warning process made by Rogers *et al.* [18] for visual signals: the user must (a) notice the warning, (b) encode the warning, (c) comprehend the warning, and (d) comply with the warning. In order to design warning signals that listeners may comprehend, Edworthy *et al.* [19] designed sounds informing helicopter pilots about the critical evolution of flight parameters (torque, high and low rotor speed, etc.), precursors to warnings: *trendsons* (trend monitoring sounds). To evaluate the comprehension of the trendsons, they asked listeners to rate them along semantic scales.

Rather than, or in addition to, having listeners rate sounds on scales (urgency scale, or other semantic scales), several authors have directly studied how listeners **comply with the warning signals**. Some [20, 21, 22, 23] have measured reaction times of plane pilots performing a tracking task and at the same time having to respond to warning signals (based on Edworthy et al [13]). Similarly, Suied *et al.* [24] measured reactions times of listeners performing a tracking task, for signals used to warn car drivers against too short of a distance between vehicles. Using a driving simulator, Belz *et al.* [25] measured times to brake when drivers were presented with different auditory and visual warning signals. Interestingly, the reaction times were different, according to the relationship between the signal and its meaning: symbolic (arbitrary relationship) or iconic (representational relationship, e.g., the sound of breaking glass used to signal danger of a potential accident). For acoustic signals displayed alone, shorter times to brake were obtained for iconic signals. Graham [26] used a similar paradigm, but participants had to decide how to react to the warning. The results showed that auditory icons[2] lead to more false alarms than symbolic sounds, because the sounds could have several (natural) meanings, whereas symbolic sounds have only one (arbitrary) meaning. Among these symbolic sounds, car horns lead to faster reaction times and fewer false alarms because of their lack of ambiguity.

This indicates that the meaning of a warning comprehended by the listener can play an important part in the warning process. This idea appears also in the results of Guillaume *et al.* [27]. They replicated the experiments of Edworthy *et al.* [13], both with sounds synthesized according to Edworthy's specifications and with recordings of alarms occurring in a plane cockpit. Overall their results fit Edworthy's prediction of perceived urgency on the basis of the acoustical properties of the sounds. However, they also report interesting exceptions: for instance, one sound, identified as a bicycle bell was rated as less urgent than it should have been according to Edworthy's prediction. The authors suggest that this sound was judged as non-urgent,

[2] Auditory *icons* are sounds used to convey a meaning, when there is a relation of similarity between the sound and the meaning (e.g. the sound of an analogical camera meaning that a digital photo has been shot).

because it "is often associated with a low level of threat and may even be associated with relaxation and pleasure" (p. 207).

An original approach is to be found in Vogel's PhD thesis [28]. He studied several warning signals (priority vehicle sirens, car horns, bicycle bells, whistles) in a psycholinguistic framework. This approach is based on the study of the participants' free descriptions of the signals mixed into two different background sequences (traffic noise and public garden). Two experiments are reported. In a first experiment (also reported in [29]), the participants had to describe the sounds at identification threshold. In a second experiment, the signals were played well above threshold. Participants had to describe triads of similar signals, and to compare triads of different signals. The analyses of both these experiments show that the context improves the univocal nature of some signals. They also show that the participants use different types of items to describe or compare the sounds: the name of the sources, the acoustical properties of the sounds and value judgements. While the participants use mostly the names of the sources when they had to compare sounds described as different sources, they describe mostly acoustical properties and value judgements when they describe signals similarly identified. Finally by comparing the descriptions of the significant categories of warning signals and acoustical representations of prototypes of these categories, he infers hypotheses on relationships between acoustical properties of the sounds and the sense given to them.

1.3. Perception of environmental sounds

As indicated by some of the results reported above, what a listener identifies as the cause of a sound might influence a lot what he or she associates with this sound. Several experimental studies have investigated the ability of listeners to spontaneously identify the cause of environmental sounds.

Sound events and environmental sounds
A first notable result is that there are sounds that are not even distinguished from their source. For instance, a series of psycholinguistic experiments on soundscape[3] perception [31, 32, 33, 34, 35, 36, 37, 38] showed that listeners perceive differently *amorphous* sound sequences ("background noises") and sound sequences in which listeners are able to identify emerging *sound events* (see [39, 40] for a discussion).

Conversely to the former sort of sequences, listeners describe sequences with sound events by referring to the identified sound sources, and with reference to how the source affects them in their everyday lives. Sound events are not distinguished from their sources and listeners appraise the source and the values they associate with the source. These findings coincide with Gaver's ecological distinction between *musical listening* (when listeners focus on qualities of the acoustic stimulus) and *everyday listening* (when listeners identify the properties of the events causing the sound: interaction, material, shape) [41, 42].

This idea of sound event is close to the definition of *environmental sound* proposed by Vanderveer [43, pp.16–17]:

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

In the same study, Vanderveer investigated how listeners identify and describe environmental sounds. The results showed that they mostly described: 1) the action, 2) the object of the action or 3) the place where the action took place.

Perception of the cause of the sounds
Thereafter, many publications have studied the perception of environmental sounds, and have reported the listener's ability to recover auditorily the properties of the events causing the sounds. Some of these properties were related to the **objects** causing the sound: the *length* of wooden rods dropped on the floor [44], the *thickness* of struck bars made of wood or metal [45], the *shape* (square, rectangular or circular) and the *materials* of struck hung plates [46], the *shape* of a ball dropped on a plate [47], the categories of *materials* (metal and glass vs. wood and Plexiglas) of recorded struck plates [48]. Others were related to the **action**: discrimination between *bouncing* or *breaking* events (glass objects falling) [49] or the ability of blindfolded participants to fill a vessel to a normal drinking level or to the brim [50].

One important question raised by these results (especially when it comes to design) is to identify the acoustic information used (or needed) by the listeners to recover these properties. Synthesized sounds (physical modeling simulating the physics of the events) of impacted bars of different materials and multidimensional techniques allowed McAdams et al. [51] to identify perceptual dimensions correlated with physical parameters.

However, another series of experiments using synthesized sounds (struck bars) [52, 53, 54, 55] showed that listeners do not optimally use the available acoustic information to decide upon the material or the hollowness of struck bars. Using recorded sounds, it is sometimes difficult to identify a clear correlation between acoustic properties and the perceived event properties [56], or to reveal stereotypical relationships between acoustical properties and listeners' responses (e.g. slow, loud and low frequency sounds systematically associated male hand-clappers [57] or walkers [58]).

[3] The term "soundscape" was introduced in the late 70's by the Canadian composer R. Murray Schafer [30], who defined soundscape as the auditory equivalent to landscape. Beside Schafer's project, the term soundscape perception is used in a scientific context to characterize how inhabitants perceive, experience and appraise their sonic environment.

Factors in environmental sound identification
Therefore, it can be assumed that both the acoustic properties of the sound (i.e. the information present within the sound) and the context and the knowledge of the listener are responsible for the recognition of a sound. This question has been explored thoroughly in a series of papers published by Ballas and Howard. The main idea of these authors is that the perception of environmental sounds shares similarities with the perception of language (though the parallels having to be considered carefully [59]).

Identification of sounds results from both a bottom-up process (recovering of the information available in the sound and in the context) and a top-down process (using previous knowledge and expectations): "It is not only what we hear that tells us what we know; what we know tells us what we hear [60][4]. In [60], they showed that the syntax and the semantics of sound sequences influence their memorization (organized and meaningful sound sequences are better memorized).

In [61], they reported homonym-type sounds: sounds being discriminated, but confused when listeners have to identify their cause. In this case, the context helped listeners to choose among the alternative causes of the sounds [59].

An imposing series of experiments reported in [62] showed that the identification performance is influenced by several factors, including acoustic variables, ecological frequency (the frequency with which a listener encounters a specific sound in everyday life), causal uncertainty (measured as the amount of reported alternative causes for a sound) and sound typicality. Actually, acoustic variables accounted for only about half of the variance in identification time and accuracy. Therefore, their results suggested that sound identifiability is related to many other factors than acoustical ones. Some of these factors (context independence - when the sound can be identified easily without context -, the ease of using words to describe the sound) have been studied thoroughly in [28].

1.4. Approach used in this paper

The function of a car horn sound is to warn people against a danger. Therefore, this review leads us to analyze the case of car horn sounds in light of the different approaches to warning signal design. Car horn sounds are warning signals used by car drivers to warn other drivers about a danger. There is no automatic reaction to be undertaken: when hearing a car horn sound, road users have to localize the potential danger, and to decide how to react (the same analysis may be found as well in [27]). The first requirement is therefore that these sounds must be audible in a road traffic background noise. This requirement is actually already addressed by the law: car horn sounds must be very loud broadband sounds [2]. Since, at the time of this study, the manufacturers are not allowed to supply cars with several different sounds, it is not possible to imagine for the time being a system that would allow the driver to choose between different sounds, with different urgency levels.

Rather, to warn road users against a danger requires that, when hearing the sound of a car horn, they must understand immediately the meaning: "danger". This is the second requirement.

Even when heard out of any context, car horn sounds are the sounds identified the most rapidly [62, 63] and almost perfectly [62, 64]: they have a low causal uncertainty [62], a low ambiguity of meaning [28], and lead to the shortest reaction times [26] (among the sounds studied in the paper). These results indicate that the association of meaning with car horn sounds is very strong. But to be associated with this meaning a sound must first be recogn ized as a car horn.

In practical terms, we base our study on the assumption that hearing a car horn sound warns road users because they recognize the sound of a car horn, because they know what this sound means and what they have to do as a consequence. This leads us to reformulate our problem: designing new sounds that still warn road users is equivalent to designing sounds that are still recognized as car horn's. This is also suggested by Vogel [28]: when introducing new warning signals, care must be taken that th ese sounds are not too different from the already-existing ones: the more the new signals are different from the already-existing ones, the more the road users will need time to learn their meaning.

The third requirement is that these sounds must not be confounded with other sounds [65].

This study situates itself within a psychoacoustical framework. Because our aim is to provide car horn builders with acoustical specifications, we will base our study on the perceptual dimensions of timbre and pitch and the related acoustical descriptors revealed in the previous article. Indeed these descriptors provide us with a tool allowing us to study the acoustical properties of the sounds, and are based on what listeners perceive. Specifically these descriptors have shown their ability to account for the perception of the mechanical causes of car horn sounds.

We will first seek to identify among existing car horn sounds, which are the best items of this category, by measuring the agreement of the participants on the membership of each sound in the category of car horn sounds (Section 2). By observing the gradient of this *membership agreement* within the acoustical descriptor space, we will be able to relate it to the acoustical descriptors, and thus to provide specifications for the design of new sounds. We will then test the generality of these relationships and their relevance for the creation of new sounds by studying synthesized sounds (sections 3 and 4).

2. Experiment 1: Agreement on the membership of current car horn sounds

The timbre of existing car horn sounds was studied in a previous paper [1]. In the first experiment of the current

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

study, we will ask listeners whether they perceive these sounds as coming from car horns (even though all the sounds tested are real car horn sounds heard on a daily basis by road users). We will therefore measure the agreement among the listeners on the membership of each of these sounds in the category "car horn".

2.1. Experimental setup

Method

Participants: Twenty-nine participants (14 men and 15 women) volunteered as listeners and were paid for their participation. They were aged from 18 to 34 years old. Most of them were students from the various universities of Paris. Thirteen were musicians (from amateur to nearly professional level), and the other sixteen had no musical education. Three were audio specialists. All reported having normal hearing.

Stimuli: Twenty-two car horn sounds were chosen so as to sample the nine classes obtained from the sorting task in the previous study (see the lower panel of Figure 2). The car horns were recorded in an anechoic chamber (see [1] for the details of the recordings). All sounds lasted approximately 550 ms. They had been previously equalized in loudness in a preliminary experiment. Listeners were asked to adjust the level of each sound so that they perceived it at the same loudness as a reference sound (1 kHz pure tone at 83 dB SPL). Their loudness is therefore 83 phons.

Apparatus: The test took place in the IAC sound-attenuated rooms at IRCAM. The experiment was run on a Personal Computer under Linux, and the graphical interface was implemented under Matlab. The sounds were amplified through a Yamaha P2075 amplifier and sent to Sennheiser HD 520 II headphones.

Procedure: For each sound, we studied the agreement of the participants on the membership in the car horn category by means of a two-alternative forced-choice (2AFC) procedure: participants listened to each sound, and had to answer the question, "Do you recognize a car horn sound? Yes or no".

The sounds were played in random order. The listeners had to answer by clicking one of two icons labeled "yes" and "no" (see Appendix A for the verbatim of the instructions). The 2AFC task actually amounts to a binary categorization in which the participants categorize each sound in one of two categories: "car horn" or "not car horn".

Coding results: Two variables are derived from this experiment. For each sound, we count how many participants gave a positive answer ("yes"). We call this variable the *membership agreement*. We also count, for each participant, how many sounds were rated as a car horn. We call this variable the *positive answer rate*. This latter variable is only used to compare the participant strategies.

2.2. Results

Participant strategies

One way to study the participant's response strategy is to compare how they divided the 22 sounds into two categories. As all the sounds are genuine car horn sounds, we expect that most of them will be categorized as car horns. The positive answer rates range from 39% to 100%, with a median of 61%. The cumulative distribution (Figure 1) shows that 83% of the participants have categorized more than 50% of the sounds as car horn sounds.

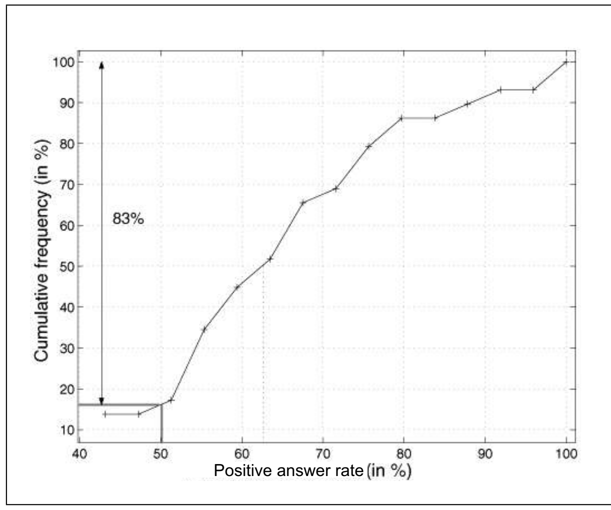

Figure 1. Cumulative distribution (in %) of the positive answer rate across participants.

A single-sample t test was performed [66] to test the null hypothesis that "the average positive answer rate is 50%". The test result is that the null hypothesis can be rejected ($t(28)=4.3$, $p < 0.01$). This leads us to conclude that participants did not make two equal partitions of sounds. The participants thus neither balanced their answers, nor answered randomly. Furthermore, since not all the sounds were categorized as car horns, we can expect to observe differences among the values of membership agreement for the different sounds.

Agreement on the membership of the sounds

Observing the agreement among the participants to categorize a sound as a car horn or not, three cases can be highlighted:

- There is consensus among the participants to categorize the sound as produced by a car horn. The membership agreement τ is close to 100%. The sound is *representative* of the category of car horn sounds.
- There is consensus among the participants to categorize the sound as not produced by a car horn. The membership agreement τ is close to 0%. The sound is *not representative* of the category of car horn sounds.
- There is no consensus among the participants. The membership agreement τ is around 50%.

Hence thresholds of membership agreement have to be set, to decide whether each sound falls into one of these cases. This is done by means of an exact binomial test [66], which tests the null hypothesis for each sound: "the membership agreement is 50%". If the null hypothesis can be rejected, the sound is either representative or non-representative. Otherwise, there is no consensus among the participants to categorize the sound as a car horn or not. A

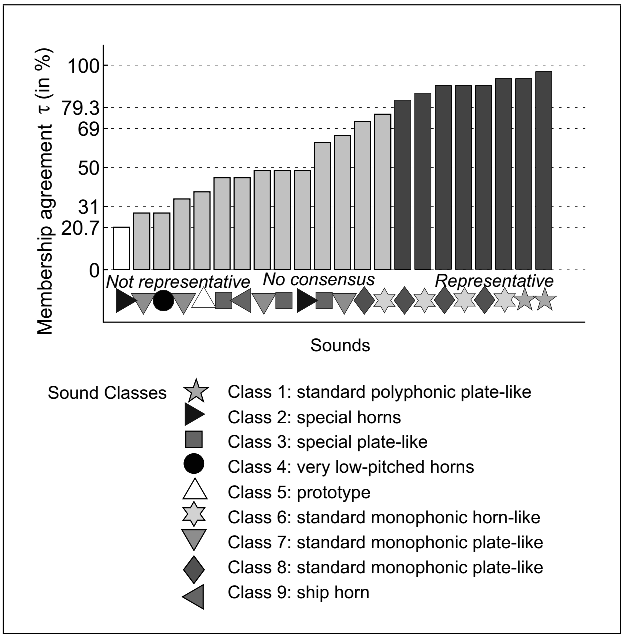

Figure 2. Membership agreement of each of the 22 recorded sounds in the category "car horns". Symbols refer to the nine classes of sounds highlighted in [1], reported in the lower panel of the figure. Gray scale codes the results of the exact binomial test. White and dark grey: $p < 0.05/22$, light grey: $p > 0.05/22$.

simple exact binomial test with an α value of 0.05 leads to two thresholds: τ_1=34% (11 among 29) and τ_2=66% (19 among 29: $\tau \leq \tau_1 \rightarrow$ non-representative, $\tau \geq \tau_2 \rightarrow$ typical). However, as several tests are compared, a more conservative significance criterion has to be chosen. This is done by performing a Bonferroni procedure [66]. The significance criterion becomes $\alpha_{adjusted} = 0.05/22$, and the membership agreement thresholds $\tau_{1adjusted} = 20.7\%$ and $\tau_{2adjusted} = 79.3\%$.

Membership agreement is represented for each sound in Figure 2.

The representative families

Figure 2 shows that every sound in category 1 (standard polyphonic plate-like), and almost all of the sounds in category 6 (standard polyphonic horn-like) and category 8 (standard monophonic plate-like) are representative. This indicates that the polyphony and the spectral characteristics due to the plate act as a kind of signature of car horn sounds. Listening to them reveals that they indeed sound like a caricature of car horns.

For these three categories, the membership agreement is rather homogeneous (category 1: 93-96%; category 6: 76-90%; category 8: 72-90%), whereas the values are much more spread over the categories for which there is no consensus (e.g. category 2: 21-52%), or the categories of non-representative sounds.

For instance, within category 7 (standard monophonic horn-like), one sound was categorized as a car horn by only 30% of the participants, whereas another one was categorized by more than 65% of the same participants, although they were judged to be perceptually close to one another in the dissimilarity rating task. The lack of consensus thus reveals that the listeners were actually unable to decide whether the sounds did or did not belong in the category of car horn sounds, and could even give different responses for sounds rated as being similar. They are ambiguous, as the post-experimental interviews revealed: participants declared that they did not know what to answer for some sounds. They could possibly have been car horn sounds, but they could also have been emitted by other sound sources, such as trumpets, car alarms, ambulance sirens, etc.

2.3. Relation to the acoustical descriptors

We now seek to link up membership agreement and acoustical properties of the sounds. The original LCREG (Latent Class REGression with spline transformations) algorithm developed by Winsberg [67, 68] is used. This technique aims to build a predictive additive model $\hat{y}$ of a dependent variable y based on spline functions of the independent variables x_i: $\hat{y} = \sum_{i=1}^{n} S_{o_i}^{m_i}(x_i)$, where $S_{o_i}^{m_i}$ is a spline function of order o_i defined for m_i knots. Each spline function is a linear combination of the B-spline basis for the given order and set of knots. In our case, for each sound the dependent variable y is the membership agreement, and x_i are acoustical descriptors. For a given model, LCREG maximizes the likelihood to get the parameters. The best model is then chosen among competing models using the Bayesian Information Criterion (BIC) [69], a log-likelihood measure of model fit that takes into account the number of degrees of freedom in the model.

We test several acoustical descriptors in addition to those revealed in the previous study [1] (we assume indeed that identification may be based on other perceptual attributes than those used to rate dissimilarities between the sounds). The best model found by the algorithm includes as independent variables *roughness*, *spectral deviation*, and *fundamental frequency*[5]. Fundamental frequency is related to pitch perception, and is therefore not a dimension of timbre (according to the definition used in this study). The model predictions are significantly correlated with the measured membership agreement ($r(20) = 0.9$). Figure 3 represents the three additive functions of the model.

It indicates that the sounds leading to higher membership agreement are those with high roughness values, low spectral deviation values, and a fundamental frequency of around 480 Hz. The first condition corresponds to polyphonic sounds. The second condition corresponds to the sounds of the plate-like horns. The third condition can be related to the fact that most of the horns sold in Europe are tuned to a fundamental frequency around 440 Hz (this can be related to the concept of ecological frequency developed by James Ballas [62]). Listeners tend to favor sounds they are used to listening to. This can be further visualized in Figure 4 which represents the positions of the category

[5] For polyphonic sounds, we take the fundamental frequency of the lowest note.

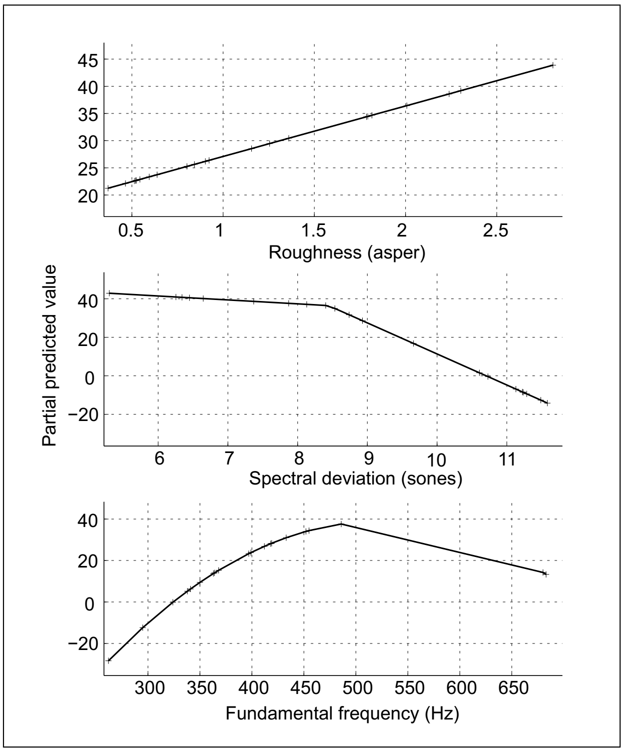

Figure 3. Additive predictive model of membership agreement.

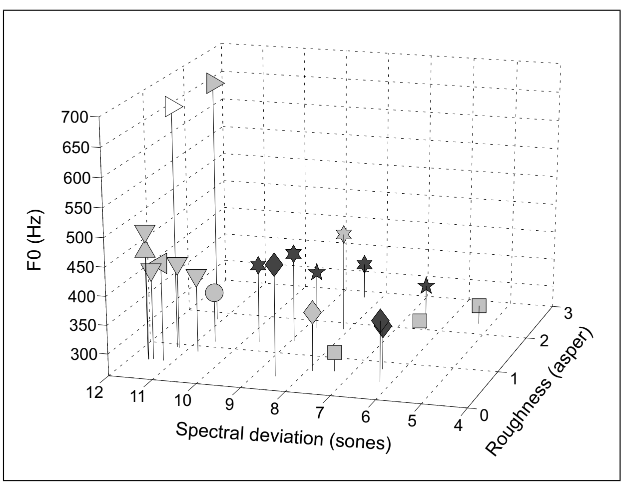

Figure 4. Representation of the sounds in the descriptor space: *roughness*, *spectral deviation* and *fundamental frequency*. The color of the symbols codes the results of the exact binomial test: black = representative sounds; gray = ambiguous sounds; white = non-representative sounds.

of sounds in the space formed by the three descriptors and the results of the exact binomial test.

In this figure the two categories of polyphonic sounds are represented by star symbols. They correspond to the upper values of roughness. The monophonic plate-like sounds, represented by diamond symbols, are located at lower values of spectral deviation and roughness. The horn-like sounds, represented by triangles or circles are located at higher values of spectral deviation. Representative sounds are roughly spread over a hyperplane around 480 Hz. Interestingly, they are located at the centre of the space.

2.4. Discussion

We can draw several conclusions from this experiment. Qualitatively speaking, it is clear that the standard polyphonic sounds are representative of the category of car horns, whether they are horn-like or plate-like. The standard plate-like sounds are representative, whether they are monophonic or polyphonic. Two main criteria hence emerge to characterize what makes a sound representative of car horns: polyphony and plate resonance. The standard monophonic horn-like sounds are ambiguous. They can be confused with other sound sources such as trumpets or alarms (according to informal interviews with the subjects). This is quite an astonishing result, inasmuch as most of the high-end cars are provided with horn-like hooters. This indicates that for most people, car horns are still associated with the old rough plate-like sounds, which actually define a caricature of car horn sounds, mainly because they are unlikely to be confused with another sound source.

This conclusion leads us to qualify our first assumption: some car horn sounds are indeed almost always identified. But some others, although they are regular car horn sounds, are likely to be confounded with other sound sources, when heard in a context-free situation. One sound was even judged as non-representative of the category. Although this sound is currently mounted on cars, it has a very high pitch, quite different from the sounds usually heard on the street.

Quantitatively speaking, the membership agreement can be specified and predicted by means of three descriptors: roughness, spectral deviation and fundamental frequency. It must be noted that fundamental frequency is not related to timbre. It is rather related to the sensation of pitch. This experiment therefore shows the importance of the pitch of the car horn sounds. This perceptual dimension did not appear in the timbre study of the car horns, because we explicitly asked listeners to rate the dissimilarities between the sounds without taking pitch into account (which has been demonstrated to be feasible [70]). The predictive model does not rely on spectral centroid. This may indicate that although the car horn sounds are perceived with different brightnesses, these differences do not change their belongingness to the car horn category.

To test the validity of this model, we repeated the same experiment with synthesized sounds in the next section.

3. Experiment 2: Synthesizing new sounds

The previous results were obtained using only recorded car horn sounds. This means that the relationships between the membership agreement and the acoustical descriptors are only tested for the range of the acoustical descriptors covered by these sounds. To extend these results, and to generalize the relationships to descriptor values outside the range of the current recorded sounds, we synthesized new sounds. These new sounds had to respect two constraints. First, they had to share the same perceptual dimensions as the recorded ones, and to have a more extended range of values than the previously tested sounds, in order to investigate the relationships between membership agreement and acoustical descriptors. Second, they had to be perceptually close to the categories of existing car horn sounds, in order to not be set apart from the existing ones. Because the results of the previous study [1] showed that listeners perceive the mechanism causing the sound, we took care to preserve these aspects.

Synthesizing sounds fulfilling these constraints also helped us to test a possible methodology for the design of new car horn sounds.

3.1. Creation of a new set of sounds

The three descriptors correlated with the perceptual dimensions shared by the car horns are related to the spectral properties of the sounds (spectral centroid and spectral deviation) and to short-term temporal properties (roughness). Yet even this latter property can be seen as spectral, since it may result from the mistuning of the harmonic partials of the spectrum. Thus, we can assume that the perceived dissimilarities between the current car horn sounds are based only on spectral and harmonic differences. And we can further assume that these sounds share identical temporal properties, and particularly have an identical temporal envelope. Therefore, to create new sounds close to the current ones, we have to create sounds with the same temporal envelope. Extending the range of the timbre dimensions can thereby be achieved by modulating the spectral and harmonic properties.

Synthesis method

Car horns, like most of musical instruments, can be thought of as an excitation mechanism (the membrane, set into vibration by an electromechanical or pneumatic system), and a resonator (the plate or the horn). Three phenomena are worth considering. First, it may be assumed that the resonator, acting like a filter, has a strong influence on the spectral envelope, and thereby, on the spectral centroid and spectral deviation of the sounds. Second, the particular excitation mechanism of the car horns forces the membrane to vibrate with a nearly square wave movement. Thus, the sounds produced have a very rich and harmonic spectrum in steady state. However, when car horns are not firmly fixed to the body of the car (which occurs after the car has been driven for a while), the device is not free to vibrate in the proper way, which causes the fundamental frequency to move slightly (therefore detuning the chords when horns are mounted in twos or threes). This slightly shifts the frequencies of the partials from a perfect harmonic series. This phenomenon is assumed to be responsible for the roughness of the sounds. Third, listening carefully to the car horn sounds reveals that the harmonic steady state takes time to become established and to release. Furthermore, we might suspect that these transient parts of the sounds, and particularly the non-harmonic noise, are very important for the recognition of car horns. Because the perceived dissimilarities between the current car horn sounds do not rely on any temporal property, we can therefore assume that these properties are identical for all the sounds of the category. They must be kept identical for the synthesized sounds, if we want them to be perceptually close to the recorded ones.

This analysis of the sound production mechanism of the car horns led us to propose a synthesis model in four parts:

- A. A nearly harmonic excitation source, made of a sum of N normalized sinusoids, the frequency of which are integer multiples i of a fundamental frequency ω_0 added to an inharmonicity term ϵ_i. The excitation is then: $\sum_{i=1}^{N} sin(i\omega_0 + \epsilon_i)t$.
- B. A temporal envelope $T_i(t)$ defining the temporal evolution of each of the sinusoids.
- C. A non-harmonic noise $n(t)$.
- D. A filter defining the amplitude of each of the sinusoids A_i.

The synthesis model is therefore defined by the following model:

$$S(t) = \sum_{i=1}^{N} A_i \sin\big((\mathrm{i}\omega_0 + \epsilon_i)t\big)T_i(t) + \alpha n(t). \qquad (1)$$

The different parameters of the model are modulated to create sounds with different descriptor values:

- The fundamental frequency is adjusted by varying ω_0.
- The roughness is adjusted by varying each inharmonicity term ϵ_i.
- The spectral deviation and the spectral centroid are adjusted by varying the number of sinusoids N and their amplitudes A_i.

The temporal envelope $T_i(t)$ of each partial and the non-harmonic noise $n(t)$ are kept constant for all the sounds. They are actually computed from two recordings of car horn sounds (one horn-like and one plate-like, both being rated as representative in the first experiment).

The temporal envelopes of each partial and non-harmonic noise are extracted using the ADDITIVE algorithm [71]. The signal to noise ratio α is kept constant and adjusted to 18 dB, as measured from the recorded sounds. The principle of the synthesis method is summarized in Figure 5.

Synthesized sounds

Forty-six new sounds were synthesized. We chose 19 sounds from among them. They were all 550 ms in duration and were equalized in loudness in a preliminary experiment. We report in Table I the range of the descriptors for these 19 sounds as well as for the 22 sounds previously tested. There are synthesized sounds with higher values of roughness, lower and higher values of spectral centroid, lower values of spectral deviation and lower fundamental frequencies than the recorded sounds. We assume that the timbre of these sounds is defined by the same perceptual dimensions as the recorded sounds. To test the assumption we subjected these new sounds to the same experimental procedure as in [1] (see section 3.2).

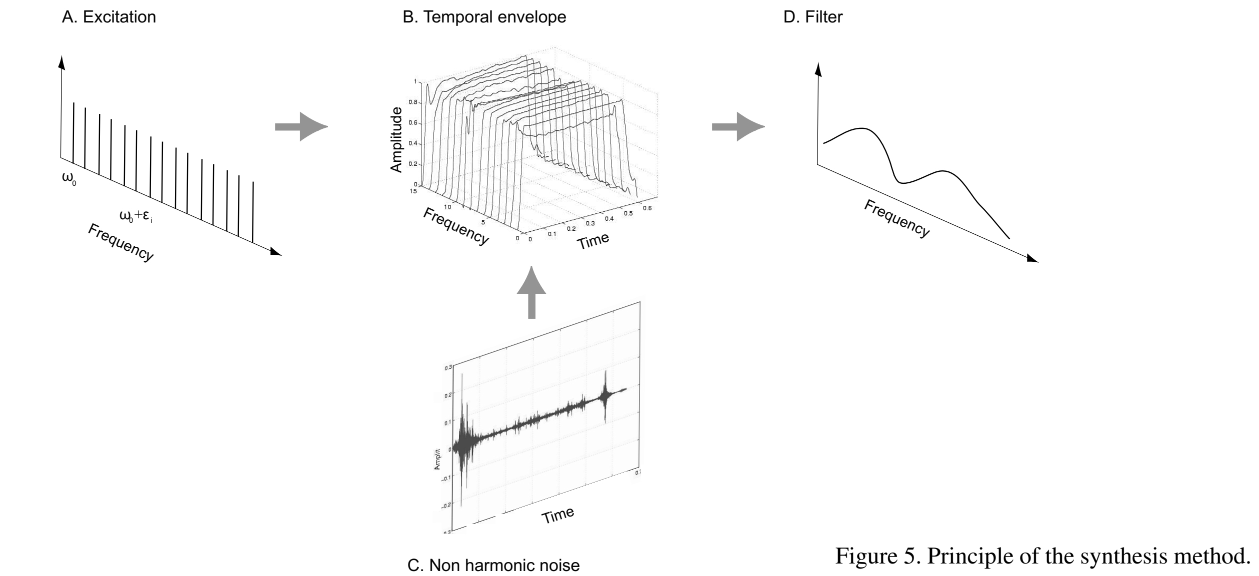

Figure 5. Principle of the synthesis method.

Table I. Range of the acoustical descriptor values for the set of 22 sounds tested in [1], and for the set of 19 synthesized sounds tested in Experiments 2 and 3. *r*: Roughness (asper), *c*: Spectral centroid (Hz), *d*: Spectral deviation (sones), *f*: Fundamental frequency (Hz).

	Recorded sounds		Synthesized sounds	
	Min.	Max.	Min.	Max.
r	0.37	2.81	0.53	2.91
c	1380	3790	1180	4560
d	5.30	11.6	2.85	11.6
f	262	683	198	500

3.2. The perception of the timbre of the new sounds

We perform a dissimilarity rating experiment to investigate the timbre of the synthesized sounds. According to the psychoacoustical definition of timbre used in this study, timbre is what allows a listener to differentiate two sounds that have been equalized in duration, loudness, and pitch. Following the multidimensional scaling approach (see [1] for a rationale of the method), we first collect dissimilarity judgments.

Dissimilarity judgments
Participants: Thirty participants (15 men and 15 women) volunteered as listeners and were paid for their participation. They were aged from 22 to 43 years old. All reported having normal hearing. The majority of the participants were students from the various universities in Paris. Thirteen were musicians (from amateur to nearly professional level), and the other 17 had no musical education. None of them was considered to be an audio specialist.
Stimuli: Nineteen sounds were chosen from among the 46 synthesized sounds. Four recorded sounds that had already been tested in the previous experiments were also included to make sure that recorded sounds and synthesized sounds would not be set apart. They were played at the same level as in the previous experiment (83 phons).
Apparatus: Same as in previous experiment.
Procedure: Participants all received written instructions explaining the task (see Appendix B). They were told that they were to make judgments on the timbre. The meaning of the word timbre (neither pitch, nor perceived duration, nor loudness) was explained to them. Particular emphasis was placed on ignoring pitch [70].

All 253 different pairs (AB or BA pairs are considered as equivalent) among the 23 sounds were presented. At the beginning of the session, the participant listened to all of the samples in random order to get a sense of the range of variation possible. Next, six training trials were presented to familiarize the participant with the rating task. On each trial, a pair of sounds was presented, separated by a 500-ms silence. The participant saw a horizontal slider on the computer screen with a cursor that could be moved with the computer mouse. The scale was labeled "Very Similar" at the left end and "Very Dissimilar" at the right end. A rating was made by moving the cursor to the desired position along the scale and clicking on a button to record it in the computer.
Coherence of the responses: The correlations between the responses of the participants ranged from 0.11 to 0.72. One participant was removed from subsequent analyses, because of the poor correlation of his judgments with the other participants (the correlation between his judgements and the other participants ranged from $r(21) = 0.11$, $p<0.01$ to $r(21) = 0.36$, $p<0.01$).

CLASCAL analysis
CLASCAL, a multidimensional scaling (MDS) technique, is described in detail in [72]. Here we only give a short description. In the CLASCAL model, dissimilarities are modeled as distances in an extended Euclidean space of R dimensions. In the spatial representation of the N stimuli, a large dissimilarity is represented by a large distance. The

CLASCAL model for the distance between stimuli i and j postulates common dimensions shared by all stimuli, specific attributes, or *specificities*, particular to each stimulus, and latent classes of subjects. These classes have different saliences or weights for each of the common dimensions and for the whole set of specificities. For the t^{th}latent class, the distance between two sounds i and j within the perceptual space is thus computed according to

$$d_{ijt} = \sqrt{\sum_{r=1}^{R} w_{tr}(x_{ir} - x_{jr})^2 + v_t(S_i + S_j)}. \qquad (2)$$

In this equation d_{ijt} is the distance between sound i and sound j, t is the index of the T latent classes, x_{ir} is the coordinate of sound i along the r^{th} dimension, w_{tr} is the weighting of dimension r for class t, R is the total number of dimensions, v_t is the weighting of the specificities for class t, and S_i is the specificity of sound i.

The class structure is latent, i.e. there is no a priori assumption concerning the latent class to which a given subject belongs. The CLASCAL analysis yields a spatial representation of the N stimuli on the R dimensions, the specificity of each stimulus, the probability that each subject belongs to each latent class, and the weights or saliences of each perceptual dimension for each class. We found a spatial model with two dimensions, specificities, and two latent classes (see Figure 6).

We chose the model configuration by comparing BIC [69] across models, as well as by performing Hope's (Monte Carlo) test [73].

The two classes of participants
Table II displays the weights of the two latent classes of participants over the two dimensions of the spatial model. The most noticeable difference between the two classes is that participants in class 2 weight the two dimensions more overall than do participants in class 1. This indicates that they used a larger range of the slider to rate the dissimilarities. The second difference between the two classes is that participants in class 2 weight the specificities more than the dimensions, conversely to participants in class 1, i.e. they placed more emphasis on the particularity of each sound than on the shared properties of all the sounds. We did not find any relation between the biographical data of the participants (gender, age, musical skills) and the belongingness to the latent classes.

Perceptual dimensions and acoustical descriptors
The first dimension of the spatial model is correlated with roughness ($r(21) = -0.8$, p<0.01), and the second dimension is correlated with spectral centroid ($r(21) = -0.8$, p<0.01). No dimension is correlated with spectral deviation even in non-optimal (according to BIC values) models with higher dimensionality. The dissimilarities between the recorded sounds are consistent with the previous data. The four recorded sounds are not set apart from the synthesized ones, which indicates that synthesized sounds are perceptually close to the recorded ones (see Figure 6). The two sounds with the highest specificity values have excessively audible noise transients.

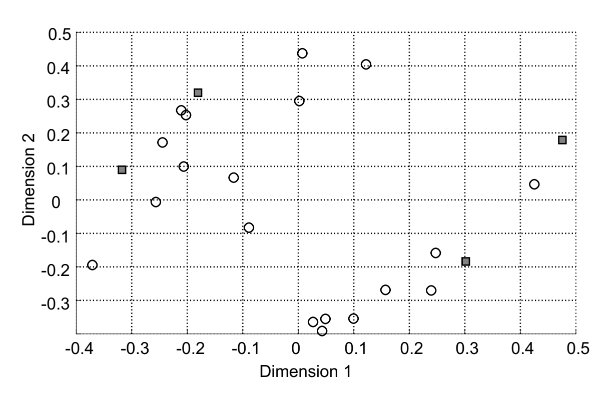

Figure 6. Representation of the perceptual space obtained by the Clascal analysis in Experiment 2. Circles represent synthetic sounds. Gray squares represent recorded sounds.

Table II. Weights of the two latent classes of participants over the two perceptual dimensions. Class1: 23 participants, class 2: 6 participants.

	Dimension 1	Dimension 2	Specificity
Class 1	0.83	0.92	0.71
Class 2	1.17	1.08	1.29
Total	2	2	2

Discussion
The assumption that the timbre of synthesized sounds is defined by the same perceptual dimensions as the recorded sounds is only partially supported, because none of the dimensions of the synthesized sounds is correlated with spectral deviation. One hypothesis could be that these sounds really do not differ according to a perceptual dimension related to spectral deviation. But they were created such that the range of this descriptor is wider than for the real sounds. Listeners should have been able to hear the differences.

In [1], we assumed that combinations of the sensations correlated with spectral centroid and spectral deviation were used by the listeners as cues that help them to distinguish between horn-like and plate-like sounds. As our synthesis method did vary both descriptors arbitrarily, combinations of these descriptors were no longer related to resonating phenomenon, and it may have become difficult for the participants to hear dissimilarities due to variations of spectral deviation alone (see Caclin et al. [74] for a discussion on a similar phenomenon). They may have focused on more obvious differences due to roughness or spectral centroid variations. Moreover, listening to the stimuli reveals that the sounds with the highest specificity values have more audible transient noises. These particularities may also have pushed listeners to concentrate only on strong dissimilarities.

However, it should be stressed that the absence of a perceptual dimension related to spectral deviation does not negate the fact that the first two perceptual dimensions are still present.

4. Experiment 3: Agreement on the membership of synthesized sounds

These new sounds are presented in a 2AFC experiment similar to the one reported in Section 2. In order to test the consistency of these measures, we include in the test the 22 recorded sounds tested in the first experiment.

4.1. Experimental setup

The experiment took place during the same session of experiment 2. The participants[6] began either with experiment 2, or with experiment 3. The order (Exp. 2 Exp. 3 vs. Exp. 3 Exp. 2) was counterbalanced across the participants. None of them had taken part in the previous experiment 1. The apparatus and the procedure were exactly the same as in section 2. The 19 synthesized sounds were tested as well as the 22 recorded sounds.

4.2. Results

Participant strategies and consistency of the measure
The set included roughly as many recorded as synthetic sounds. The latter were designed to explore the limits of the perceptual space, so we expected that a lot of them would not be categorized as car horn sounds. The positive answer rates range from 32% to 80% with a median of 51%.

A single-sample t test (testing the null hypothesis: "the average positive answer rate is 50%") confirmed that participants partition the set equally ($t(30)=0.97$, $p>0.05$).

To determine the consistency of the results, the measures of membership agreement for the recorded sounds in this experiment were compared to those found in the experiment described in section 2. Figure 7 represents the regression of the membership agreement for the previous experiment (homogeneous set of recorded sounds) onto the membership agreement measured in this experiment (mixed set of both recorded and synthetic sounds).

The correlation coefficient is 0.9 (df=20, $p<0.01$), and the measures of membership agreement are smoothly spread over the regression line (slope: 0.9, origin ordinate: 5.2%).

The perfect consistency regression line (slope: 1, origin ordinate: 0%) falls within the 95% confidence interval of the regression. This allows us to conclude that the measure of membership agreement is not influenced by the kind of set tested. The indicator can be compared for the two experiments.

The representative sounds
The measures of membership agreement are represented in Figure 8. They range from 3.2% to 96.8%. For each sound, the result of an exact binomial test is represented by the gray scale.

There is no consensus among the participants for categorizing most of the synthetic sounds in one of the two categories. Only one of the synthetic sounds was representative of car horns. For recorded sounds, among the eight sounds that were representative when they were tested in the homogeneous set (not counting the reference sound), seven are again representative when tested in this mixed set. The qualitative results described in section 2 remain unchanged.

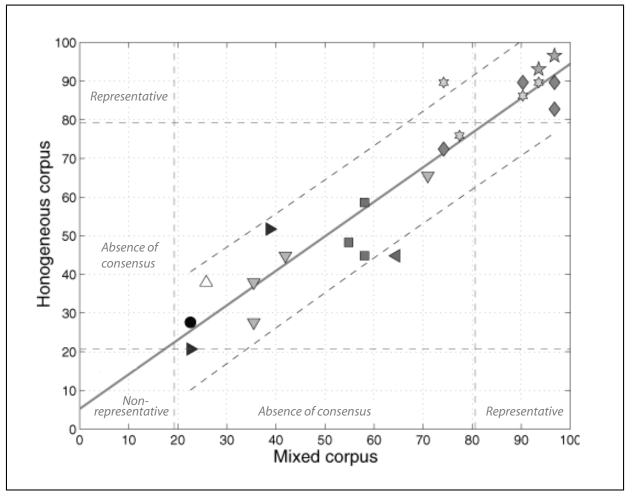

Figure 7. Linear regression between the membership agreement measured for the recorded sounds, when they were part of an homogeneous set of recorded sounds, and when they are part of a mixed set of both recorded and synthetic sounds. The dashed lines define the 95% confidence interval around the prediction. The horizontal and vertical dashed lines correspond to the thresholds fixed by the exact binomial test τ_{1ajust} and τ_{2ajust}.

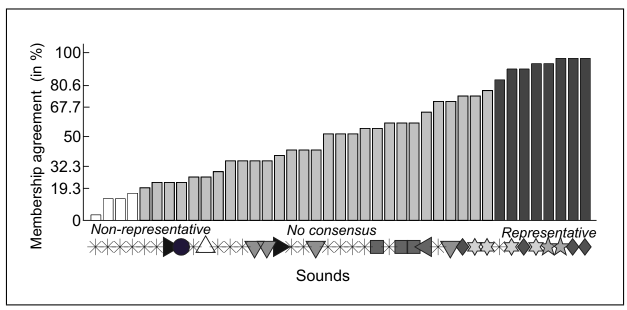

Figure 8. Membership agreement measured for the mixed set of both recorded and synthetic sounds. Symbols refer to Figure 2. Stars represent the synthetic sounds. Results of the binomial test: $p < 0.05/41$ for white and dark grey bars, $p > 0.05/41$ for light grey bars.

4.3. Relation to acoustical descriptors

Following the method described in section 2, we relate the measures of membership agreement to the acoustical descriptors by means of the LCREG algorithm. The best predictive additive model is again based on roughness, spectral deviation and fundamental frequency. Correlation between predicted and measured membership agreement is ($r(39) = 0.7$, $p< 0.01$). Figure 9 represents the spline functions defining this model.

[6] There was one more participant in Experiment 3: there are therefore 31 participants in this experiment

The first function, indicating the contribution of roughness to the membership agreement, is qualitatively identical to the first function of Figure 3. Sounds the most often associated with the category "car horns" are those with the highest roughness values. The second function predicts that sounds with a spectral deviation value around 7 sones lead to the highest membership agreement. The model depicted in Figure 3 predicted that the highest membership agreement would have been obtained for sounds with a spectral deviation lower than 9 sones. But none of the sounds tested had a spectral deviation lower than 5 sones. Hence this experiment allows us to extend the model to lower values of spectral deviation. The third function predicts that sounds with a fundamental frequency of around 350–400 Hz are those that are most often categorized as car horns.

To better visualize the localization of the representative sounds in the descriptor space, this space is represented in 3D in Figure 10.

As in Figure 4, the representative sounds are located at the center of the space. Two areas can be distinguished: one corresponding to the polyphonic sounds (plate- or horn-like), and one corresponding to the monophonic plate-like sounds. The unique representative synthetic sound is located close to the area corresponding to the monophonic plate-like sounds. These two areas are defined for a fundamental frequency around 350-400 Hz and for a spectral deviation value between 6 and 9 sones. Roughness in itself does not allow segregation between representative and non-representative sounds. Rather, it has to be combined with spectral deviation. This is mainly due to the monophonic plate-like sounds, which are representative, whereas monophonic horn-like sounds are ambiguous. Hence, to be representative, a sound may possibly have a low roughness (i.e. monophonic), but only if it has a high spectral deviation value (i.e. plate-like). In other cases, the roughness value must be high.

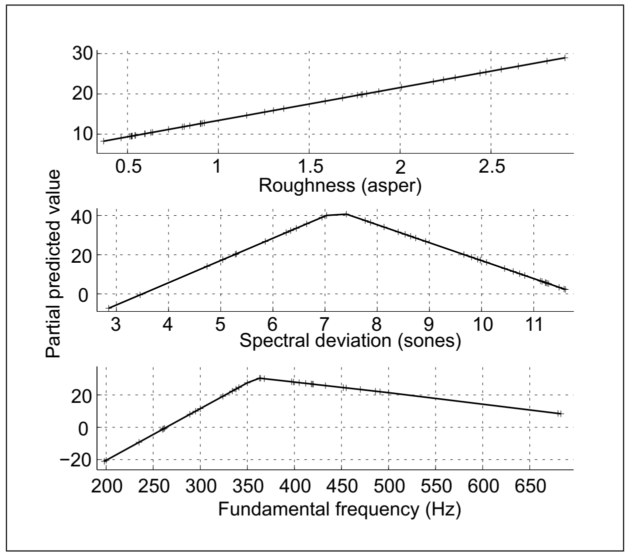

Figure 9. Predictive model of the membership agreement for the set of mixed recorded and synthetic sounds. The crosses represent the sound samples used.

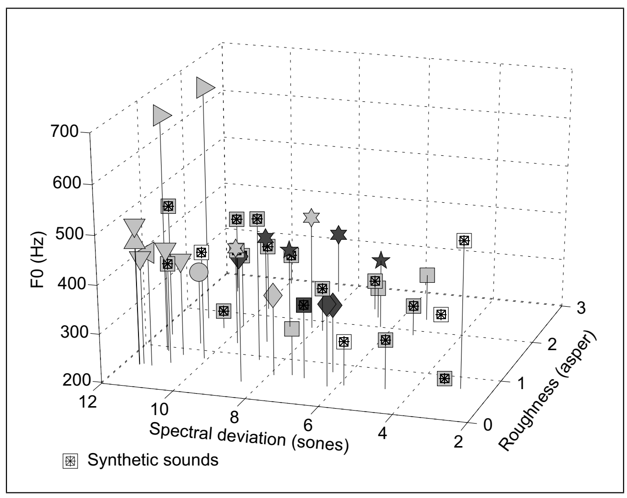

Figure 10. Localization of the sounds in the descriptor space (Roughness, spectral deviation, fundamental frequency). The results of the exact binomial test are coded by the color of the symbols: black = representative, white = non-representative, gray = no consensus. Symbols refer to Figure2.

4.4. Discussion

The addition of synthetic sounds to the set allows us to generalize the conclusions drawn from the recorded sound set. First of all, duplicating the measure of membership agreement demonstrates that this measure is stable. Here again, the representative car horn categories are the standard polyphonic sounds (both plate- and horn-like) and the standard monophonic plate-like sounds.

We demonstrate again that fundamental frequency plays an important role in predicting the membership agreement in the car horn category. Membership agreement is thus related to both the timbre and the pitch of the sounds.

Finally, we generalize the description of the representative sounds. The additive regression model applied to synthetic sounds and recorded sounds allows us to define more precisely the combinations of descriptors that describe the sounds categorized as car horn sounds without ambiguity. Representative sounds are those with the largest values of roughness, a spectral deviation around 7 sones, and a fundamental frequency around 400 Hz. These quantitative results are important for the design of new sounds.

However, we shed a light on the importance of the ambiguity phenomenon. A large number of synthetic sounds, as well as some recorded sounds, are not categorized as car horn sounds, not only because they are perceptually different from the sounds most often categorized as car horns, but also because they possess specific properties that would lead listeners to identify them as other sound sources. This is coherent with the CLASCAL analyses of dissimilarity ratings of both recorded and synthetic sounds, which showed that sounds were compared according to specificities in addition to continuous dimensions. However, we are not able from these experiments to predict possible associations with other categories of sound sources.

5. General discussion and conclusions

This paper concerns the design of new car horn sounds. Warning is the main function of car horns. This function must be preserved when the sounds are tuned according to the customer's wishes. The review of the literature on warning signal design reported in Section 1 leads us to base our approach on the following assumption: hearing a car horn sound warns road users because they recognize the sound of a car horn, they know what this sound means, and they know what they have to do as a consequence. Therefore, the experimental studies reported in this paper seek to identify acoustic properties that are responsible for a sound (among sounds sharing common dimensions with current car horns) to be categorized as coming from a car horn.

Following a paradigm that was originally designed to study the timbre of musical sounds, we have defined in [1] the timbre of car horn sounds as the integration of three continuous perceptual dimensions (shared by all the sounds) and specificities (particular to each sound). The continuous perceptual dimensions were correlated with appropriate acoustical descriptors. Latent class analysis revealed that different classes of participants weighted the dimensions and specificities differently. The latent classes were not related to any recorded biographical data concerning the participants (age, gender, musical skills).

The experiments reported in this paper were focused on measuring the agreement of some groups of listeners on the membership of sounds in the car horn category. The membership agreement was operationally defined as the results of a 2AFC task. These measures were analyzed so as to highlight three types of agreement: there might be an agreement among listeners to categorize a sound as a car horn. In this case, we called the sound representative. Conversely, there might be an agreement of the listeners to categorize a sound as not coming from a car horn. In this case, we called the sound non-representative. When there is no agreement among the participants, analyses of the participants' comments suggested that the sounds were ambiguous.

We first measured in Section 2 the membership agreement of the 22 recorded sounds tested in [1]. Whereas most of the sounds were representative, we observed a gradient of membership agreement. By means of a multispline regression technique, we were able to relate the membership agreement to three acoustical descriptors of the sounds: roughness, spectral deviation and fundamental frequency. Furthermore, the results showed that some categories of car horn sounds were systematically categorized as car horns.

To generalize these conclusions to descriptor values outside the range of the recorded sounds (thereby testing a methodology for the design of new sounds), we created in Section 3 a set of new sounds. The synthesis method was designed so as to preserve the temporal properties of the sounds (particularly the transient parts), suspected to underlie the recognition of the car horns, and so as to create sounds sharing the same perceptual dimensions as the recorded ones. To test this last assumption, we subjected the created sounds to a dissimilarity rating task. The analysis revealed that these sounds shared only two of the three dimensions of the recorded sounds.

Finally in Section 4, we performed a 2AFC task that aimed to measure car horn membership agreement for a mixed set of recorded and synthesized sounds. These results confirm those of Section 2 and lead to a generalization of the predictive model.

Several conclusions are to be drawn from these results. From a general standpoint, a major result of the first part of the study [1] was the importance of the perception of the sound-producing mechanism. The categorization experiment indeed revealed that categories of car horns built by listeners closely correspond to the different kinds of devices. These categories were preserved both in dissimilarity judgments and in 2AFC tasks.

These experiments have suggested that when sounds are not categorized as car horns, it is not only because they are dissimilar to the sounds most often categorized as car horns, but also because they may be identified as another sound source. This has shed light on the problem of identification ambiguity. Ambiguity occurs when a sound may be associated with distinct categories of sound sources. In our case, ambiguity must be avoided: if a sound is confused with another sound source, it may fail to convey the warning message.

However, our results only allow us to predict how "close" a sound is to what we may call a shared representation of what a car horn's sound is. It does not predict if the sound may be identified as another sound source. As our experimental tasks were based on sound comparisons, we have emphasized the properties shared by the sounds. The acoustical descriptors that we used are correlated with the common perceptual dimensions: they describe the shared properties of the sounds. A more efficient description should also include idiosyncratic properties, because these distinctive properties may explain why some sounds can be confused with other sound sources. The CLASCAL analyses performed on both synthetic and recorded sounds have included not only common dimensions, but also specificities, which are individual properties of the sounds (this is similar to Tverky's contrast model [75], which aims to include both common and distinctive features). These specificities may indicate possible misinterpretations of the sound. It is however difficult to determine to which acoustical properties these specificities correspond and then to make any a priori predictions.

The question of sound source identification is, however, only partially addressed by this study. Indeed, all the reported experiments were done in a laboratory, without any acoustic, visual or situational context. The question is still open as to how these sounds would be categorized if they were heard on the street. However, we can assume that the context would play two roles. Firstly, the acoustic background noise would raise the detection threshold of the sounds. The issue of detection of the sounds is, however, already addressed by the very high level imposed by law.

Secondly, it can be assumed that a road traffic situation would lower the ambiguity of some sounds. Indeed, if a sound heard in a laboratory can be confused with a musical instrument sound, and the more so when this laboratory is located in a institution devoted to music, it can be assumed that the same sound heard in a road traffic situation would have less chance to be confused with a trumpet call. As noted by Vogel [28], listening to warning signals in context can improve the univocal nature of these signals. We can therefore assume that our results are more conservative than the real situation.

Another interesting finding of our experiments is that listeners do not use the same perceptual dimensions when they have to judge the dissimilarity between the sounds, and when they have to categorize a sound as a car horn. Indeed, the timbre study reported in [1] revealed three perceptual dimensions correlated to roughness, spectral centroid and spectral deviation. When listeners had to categorize the same sounds, our analyses concluded that they had based their judgments on roughness, spectral deviation and fundamental frequency. It is easy to explain why fundamental frequency did not appear in the timbre study: in this experiment listeners were specifically asked to not base their judgment on pitch, and other studies have shown that they are able to do so [70]. Because the wide majority of car horn sounds that can be heard nowadays have a fundamental frequency in the region of 440 Hz, introducing sounds with lower or higher fundamental frequencies may lead listeners to judge them as unusual, and makes this descriptor a good predictor of the membership agreement. More puzzling is the fact that spectral centroid did not seem to have been used to categorize the sounds as car horns, even when we introduced sounds with more extreme values of this descriptor. The brightness of the sounds (a sensation correlated to spectral centroid) appears very often in the multidimensional study of timbre (see [1] for a review), but seems therefore to be of minor importance to categorize car horn sounds, with respect to the other sensations related to the modulations of fundamental frequency, roughness and spectral deviation. Spectral deviation, on the other hand, did not appear as a perceptual dimension of the timbre of the synthesized sounds, certainly because some synthesis artifacts overwhelm the more subtle differences due to variations of spectral deviation. Yet the analyses of the third experiment predicted that listeners use a sensation correlated to this descriptor to categorize the sounds as car horns or not. This again indicates that listeners can weigh their sensations differently according to what they have to judge.

Going back to the framework of sound design, these results are useful, despite the reservations expressed in the above paragraphs. Car horn builders will continue to design broadband, loud, harmonic sounds. Hence, tuning new sounds may be conceived as choosing values of the descriptors of the car horn sounds. Our synthesis method easily allows a car horn builder to design a new sound and to compute the descriptor values. With the results of the studies of the agreement on the membership of the sounds, the sound designer is thus able to predict whether such a sound will be close to the sounds best recognized as car horns.

Appendix: Experimental instructions provided to the participants

A. 2AFC experiment (Experiments 1 and 3)

Goal of the study

The goal of this experiment is to answer the question: "Do you recognize a car horn sound ?" for each sound of a set.

Procedure

You will sit in front of a computer screen. You will hear a set of sounds played one after the other. For each sound you will have to answer the question: "do you recognize a car horn sound ?" Two buttons are displayed on the interface, labeled with "yes" and "no". To indicate your answer, you will have to click on one of these buttons. The sounds are only played once. When you have entered your answer, the next sound will be played after a pause. Try to answer spontaneously.

Note

The sounds may originate from different sources. We are interested in *your* opinion, so there is no "correct answer". Do not try to balance the amount of "yes" and "no" answers. You can even answer "yes" for every sound, or "no" for every sound, if this is what you hear.

B. Dissimilarity rating experiment (Experiment 2)

Goal of the study

The goal of this experiment is to study the perception of the timbre of a set of sounds. Your task is to judge the dissimilarity that you perceive between two sounds.

Procedure

You will sit in front of a computer screen.
There are 23 sounds in the test. They all last about half a second. At the beginning of the test, you will be provided with 23 buttons, which allow you to listen to the 23 sounds and to familiarize yourself with them. Then you will be provided with each one of the 23 possible pairs of sounds among the 23 sounds. For each pair of sounds, the interface looks similar: there are two buttons labeled "listen again" and "validate", above a cursor with the labels "very different" and "very similar" at each extremity. When you click on the "listen again" button, you can hear the two sounds. You can listen to the pair of sounds as many times as you wish. The cursor allows you to rate the dissimilarity between the sounds. When your are sure of your rating, click on the "validate" button. This moves to the next pair of sounds. Before the real test, you will be provided with six pairs of sounds to familiarize yourself with the interface in the presence of the experimenter.

Remark on the notion of timbre

You have to group together sounds with similar *timbre*. Timbre is what allows you to distinguish between two sounds having the same duration, the same intensity and the same pitch. For instance, two musical instruments playing the same note, with the same intensity and of the same duration do not sound identical. What distinguishes them is referred to here as “timbre”. Timbre may also be called the “color”, “texture”, . . . of the sound. These sounds are supposed to have the same intensity. You may however feel that certain sounds are louder than others. We ask you to not take into account intensity in your judgments.

Similarly, the sounds do not all have the same pitch. They “play different notes”. Here again, we ask you to not include these differences of pitch in your judgments, but rather to focus on timbre.

Acknowledgements

This study has been founded by SCE Klaxon and by the French Ministry of Industry.

References

[1] G. Lemaitre, P. Susini, S. Winsberg, B. Letinturier, S. McAdams: The sound quality of car horns: a psychoacoustical study of timbre. Acta Acustica united with Acustica **93** (2007) 457–468.

[2] United Nation Organization, Economic and Social Council, Uniform provisions concerning the approval of audible warning devices and of motor vehicles with regard to their audible signals: Regulation 28 to the TRANS/WP .29/343/Rev.13 agreement. http://www.unece.org/trans/main/wp29/wp29wgs /wp29gen/wp29fdoc/TRANS-WP29-343r13efr.pdf http://www.unece.org/ trans/, 2000.

[3] G. Lemaitre, B. Letinturier, B. Gazengel: Electroacoustical model of an electrodynamic horn loudspeaker used as a new car horn. Applied Acoustics **69** (2008) 47–59.

[4] USA acoustical terminology S1.1-160. American Standard Association, 1960.

[5] C. Krumhansl: Why is musical timbre so hard to understand? – In: Structure and perception of electroacoustic sound and music. S. Nielzen, O. Olsson (eds.). Elsevier, 1989, 43–53.

[6] N. A. Stanton, e. J. Edworthy: Human factors in auditory warnings. Ashgate Publishing, 1999.

[7] R. D. Patterson: Guidelines for auditory warning systems on civil aircraft. Civil Aviation Authority, 1982. Paper 82017.

[8] C. Laroche, L. Lefebvre: Determination of acoustic features for reverse alarms: field measurements and the design of a sound propagation model. Ergonomics **41** (1998) 1203–1221.

[9] H. Fleischer, J. Blauert: Audibility of some specific publicadress warning signals in typical environmental noise situations. Applied Acoustics **27** (1989) 305–319.

[10] J. D. Binnington, A. W. McCombe, M. Harris: Warning signal detection and the acoustic environment of the motorcyclist. British Journal of Audiology **27** (1993) 415–422.

[11] M. Ebata, H. Baba: Detection of the ambulance siren while driving. Proceedings of Inter-noise 1997, Budapest, Hungary, 1997, 1009–1014.

[12] M. R. Lehto, J. D. Papastavrou: A signal-detection theory based perspective on design of warnings. Perceptual and Motor Skills **86** (1998) 720–722.

[13] J. Edworthy, S. Loxley, I. Dennis: Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. Human Factors **33** (1991) 205–231.

[14] E. J. Hellier, J. Edworthy, I. Dennis: Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. Human Factors **35** (1993) 693–706.

[15] E. Hellier, B. Weedon, J. Edworthy, K. Walters: Using psychophysics to design speech warnings. Proceedings of the IEA 2000/HFES 2000 Congress, 2000, 698–701.

[16] B. Weedon, E. Hellier, J. Edworthy, K. Walters: Perceived urgency in speech warnings. Proceedings of the IEA 2000/HFES 2000 Congress, 2000, 690–693.

[17] J. Edworthy, N. Stanton: A user-centred approach to the design and evaluation of auditory warning signals. 1: Methodology. Ergonomics **38** (1995) 2262–2280.

[18] W. A. Rogers, N. Lamson, G. K. Rousseau: Warning research: An integrative perspective, human factors. 42 **2000** (102-139).

[19] J. Edworthy, E. Hellier, R. Hards: The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals. Ergonomics **38** (1995) 2341–2361.

[20] J. L. Burt, D. S. Bartolome, D. W. Burdette, J. R. Comstock: A psychophysiological evaluation of the perceived urgency of auditory warning signals. Ergonomics **38** (1995) 2327–2340.

[21] E. C. Haas, J. C. Casali: Peceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. Ergonomics **38** (1995) 2313–2326.

[22] E. C. Haas, J. Edworthy: Designing urgency into auditory warnings using pitch, speed and loudness. Computing and Control Engineering Journal (1996) 193–198.

[23] E. C. Haas, J. Edworthy: The perceived urgency and detection time of multitone auditory signals. – In: Human Factors in Auditory Warnings. N. A. Stanton, J. Edworthy (eds.). Ashgate Publishing Ltd., 1999.

[24] C. Suied, P. Susini, S. McAdams: Evaluating warning sound urgency with reaction times. Journal of Experimental Psychology: Applied **14** (2008) 201–212.

[25] S. M. Belz, G. S. Robinson, J. G. Casali: A new class of auditory warning signals for complex systems: Auditory icons. Human Factors **41** (1999) 608–618.

[26] R. Graham: Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. Ergonomics **42** (1999) 1233–1248.

[27] A. Guillaume, L. Pellieux, V. Chastres, C. Drake: Judging the urgency of nonvocal auditory warning signals: perceptual and cognitive processes. Journal of experimental psychology: Applied **9** (2003) 196–212.

[28] C. Vogel: Étude sémiotique et acoustique de l’identification des signaux sonores d’avertissement en contexte urbain. PhD thesis, Université de Paris 6, 1999.

[29] C. Vogel, D. Dubois, J.-D. Polack, M. Castellengo: Perception of warning signals in urban context. Proceedings of Inter-noise 98, Christchurch, New-Zealand, November 1998, 1203–1206.

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

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

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

[33] M. Raimbault, C. Lavandier: Sound ambient environment of urban places: comparison of sound appraisal factors with acoustical parameters. Proceedings of the 3rd European Congress on Acoustics – Forum Acusticum, Sevilla, Spain, September 2002., Reproduced in Acta Acustica united with Acustica **88** (2002) Suppl. 1 S70.

[34] M. Raimbault, D. Dubois: Perceptual judgements about urban soundscapes. Proceedings of Euronoise 2003, Naples, Italy, 2003.

[35] M. Raimbault, C. Lavandier, M. Bérengier: Ambient sound assessment of urban environments: field studies in two french cities. Applied Acoustics **64** (2003) 1241–1256.

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

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

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

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

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

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

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

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

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

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

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

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

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

[49] W. H. Warren, R. R. Verbrugge: Auditory perception of breaking and bouncing events: a case study in ecological acoustics. Journal of Experimental Psychology: Human perception and performance **10** (1984) 704–712.

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

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

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

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

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

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

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

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

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

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

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

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

[62] J. A. Ballas: Common factors in the identification of an assortment of brief everyday sounds. Journal of Experimental Psychology: Human Perception and Performance **19** (1993) 250–267.

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

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

[65] H. Höge, A. Schick, S. Kuwano, S. Namba, M. Bock, H. Lazarus: Are there invariants of sound interpretation? The case of danger signals. Proceedings of the 5th international congress on noise as a public health problem, Stockholm, 1988, 253–258.

[66] D. C. Howell: Statistical methods for psychology. PWS-Kent, 1992.

[67] S. Winsberg: Multiple regression with tensor product splines. Presented at the European Meeting of the Psychometric Society, University of Leuven, Belgium, July 1989.

[68] S. Winsberg: Multiple regression with tensor product splines. Presented at the 1990 Meeting of the Psychometric Society, Princeton, New Jersey, June 1990.

[69] G. Schwarz: Estimating the dimensions of a model. The annals of Statistics **6** (1978) 461–464.

[70] J. Marozeau, A. de Cheveigné, S. McAdams, S. Winsberg: The dependency of timbre on fundamental frequency. Journal of the Acoustical Society of America **114** (2003) 2946–2957.

[71] Depalle, Garcia, Rodet, Woehrmann, Perry: The additive analysis-synthesis package. Ircam www.ircam.fr/ equipes/ analyse-synthese/ DOCUMENTATIONS/ additive.

[72] S. Winsberg, G. D. Soete: A latent class approach to fitting the weighted Euclidian model CLASCAL. Psychometrika **58** (1993) 315–330.

[73] A. C. A. Hope: A simplified Monte-Carlo significance test procedure. Journal of the Royal Statistical Society B **30** (1968) 582–598.

[74] A. Caclin, S. McAdams, B. K. Smith, S. Winsberg: Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. Journal of the Acoustical Society of America **118** (2005) 471–482.

[75] A. Tversky: Features of similarity. Psychological review **84** (1977) 327–352.