Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Acoustics 65 (2004) 763–790
applied acoustics
www.elsevier.com/locate/apacoust

Characterizing the sound quality of air-conditioning noise ☆

Patrick Susini *, Stephen McAdams, Suzanne Winsberg, Ivan Perry, Sandrine Vieillard, Xavier Rodet

Institut de Recherche et Coordination Acoustique/Musique (STMS-IRCAM-CNRS), 1 place Igor Stravinsky, F-75004 Paris, France

Received 27 May 2003; received in revised form 17 February 2004; accepted 19 February 2004
Available online 10 May 2004

Abstract

The aim of the psychoacoustic study presented here was to characterize listeners' preferences for a set of sounds produced by different brands and models of indoor air-conditioning units. In addition, some synthetic sounds, created by interpolation between recorded sound samples, were integrated into the set. The multidimensional perceptual space and the corresponding physical space representative of the sound set were determined with multidimensional scaling (MDS). Then the preferences for different classes of listeners were related to the physical space. The best spatial model yielded by the MDS had three common dimensions and specificities. The three dimensions are correlated with the ratio of the noisy part of the spectrum to the harmonic part (NHR), with the spectral center of gravity (SC) and with loudness (*N*). Two classes of listeners can be distinguished in terms of preference. For one, preference varied primarily with loudness, whereas for the other it varied more with SC and NHR. However, for one class the preference grew with the parameter NHR, while it decreased for the other class. The results replicate under different laboratory conditions and indicate the usefulness of this sound quality assessment approach for characterizing appliance noises.

Keywords: Psychoacoustics; Sound quality; Sound design; Multidimensional analysis; Appliance noise

☆ A preliminary report of portions of these data was presented at the 17th International Congress of Acoustics in Rome [1–3].

* Corresponding author. Tel: +33-1-4478-1609; fax: +33-1-4478-1540.
E-mail address: susini@ircam.fr (P. Susini).

0003-682X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2004.02.003

1. Introduction

The complexity of sounds gives them a multidimensional character from both acoustic and human perceptual points of view. It is necessary in sound quality research to address the components that are related both to sound and to human cognition. As such, the quality of the sounds of our environment is represented, on the one hand, by a set of auditory qualities or attributes that are related to sensory dimensions (loud/soft, bright/dull, etc.), and, on the other hand, by aspects related to pleasantness or even to emotional response (agreeable/disagreeable, stressing/reassuring, etc.) that are strongly dependent on cultural and cognitive differences. Thus, the system of scales often adopted for dealing with sound quality [4–6] is a set of nominal bipolar scales (e.g. high/low) called semantic differentials [7]. However, with the exception of a few studies [8,9], nominal scales often impose judgments with a priori knowledge neither of the perceptual relevance nor of the redundancy of the labels used. The same is true when acoustic quality is characterized with measurement of psychoacoustic descriptors such as loudness, roughness, and sharpness [10]: the relevance and respective contributions to quality can vary from one family of sounds to another. By "family", we mean a corpus of homogeneous sounds produced by the same type of object (vehicle, harmonic or percussive musical instrument, domestic appliance, ventilator, telephone, etc.).

Recently, we have developed a psychophysical approach using a multidimensional scaling analysis of both proximity data to characterize the perceptual component and dominance data to characterize the preference component [11]. Proximity data and dominance data are analysed using the CLASCAL [11,12] and C5SPLN [11,13] programmes, respectively. The two analyses are linked through the acoustic parameters that are hypothesized to underlie each. The most difficult part consists of finding these parameters for a specific family of sounds.

Briefly, proximity data in our case are based on direct judgments of dissimilarity between pairs of sounds and are analysed using the multidimensional scaling analysis CLASCAL [12]. Dissimilarity judgments are modeled as distances in a Euclidean space. Each dimension of the space is hypothesized to represent a perceptual continuum that is common to all the sounds studied, on the one hand, and that is well explained by an acoustic parameter, on the other. Parameters that best explain the dimensions are called objective predictors. In addition, the analysis programme CLASCAL also includes two other properties: the existence of additional dimensions that are specific to individual sounds (called "specificities"), and differences in the perceptual importance of each dimension or the set of specificities between subpopulations of listeners (called "latent classes"), which may be due to differences in perception or in judgment strategy. For example, this analysis has been used to study the perception of synthesized musical timbres by professional and amateur musicians, and non-musicians [14]. The results revealed a three-dimensional space with specificities and five latent classes. Dimensions of the space were quantified in terms of log (rise time), spectral centroid and degree of spectral variation. High specificity values were explained by specific mechanical properties such as the return of the

hopper in the case of the harspichord. However, the five latent classes were not explained by the degree of musical training.

Dominance data are based on binary choice of preference between the same pairs of sounds and are computed on a matrix of proportions indicating the proportion by which one sound is preferred over another. This matrix is analysed using the analysis programme C5SPLN [13]. The proportions of preference are associated with "utility" functions that account for the contribution to preference of the objective predictors over the set of sounds studied. Two types of functions are tested, either additive or multivariate. The "utility" function obtained is called the preference map.

The hypothesis underlying this approach is that the dissimilarity and the preference judgments are based on the same acoustic or psychoacoustic parameters. The approach can be applied to a family of sounds such as musical instruments, cars, and domestic appliances among others, and takes into account differences within the population of tested individuals. The advantage of this approach is that it allows the study of acoustic quality without requiring a priori selection of the perceptual and preference dimensions to be studied for a given family of sounds: instead these dimensions are derived from the perceptual data.

Using this approach, the present article studies experimentally the perceptual structures that underlie comparison and preference judgments on sounds of interior air-conditioning units. The final goal is to propose objective guidelines for characterizing the acoustic quality of air-conditioning systems, and to produce a valuable tool for the process of sound design in order to create new sounds according to perceptual characteristics.

This article will be presented in four parts. First, a description of the constitution of the corpus of sound samples is presented. The interest of creating hybrid sounds with digital analysis/synthesis techniques and reintroducing them into the perceptual tests is also discussed. Second, using the selected sound corpus, a dissimilarity experiment is performed with 50 subjects. The structure that underlies the perceived differences between the sounds is revealed and represented by an Euclidean space. Third, acoustic and/or psychoacoustic parameters (called objective descriptors) that best explain the organisation of the sounds along the perceptual dimensions of the Euclidean space are developed and analysed. Fourth, a preference experiment is performed with two groups of 100 subjects in two different laboratories and the functional relations between listeners preferences and the objective predictors are established.

2. Constitution of the sound corpus

The constitution and selection of the sound corpus takes place in four steps.

2.1. Recording of sound samples

Firstly, the study of noises produced by indoor air-conditioning units requires the recording of different brands and models in order to cover a representative range of

existing noises. Seventeen units drawn from 10 brands were recorded under identical conditions in the anechoic chamber of the French Laboratoire National d'Essai (LNE) (Appendix A). Two types of air-conditioning units were recorded: free-standing (VCV) and ceiling-mounted (SPLIT). In both cases, a vertical reflecting plane was installed to simulate the wall against which the unit was placed and a horizontal reflecting plane was installed to simulate the floor for free-standing units and the ceiling for ceiling-mounted units (see Figs. 12 and 13). DAT recordings were made with a B&K 4179 microphone connected to a B&K 2660 preamp for the main recording (right channel) and a B&K 4165 microphone connected to a B&K 2639 preamp for a second, backup recording (left channel). The microphones used for the two channels did not have the same characteristics. The 4179 (right channel) was a low-noise microphone for recording low-level signals without contamination by electrical noise, but its pass band is limited to 12 kHz. The 4165 (left channel) is noisier but its passband is greater than 20 kHz. For both recordings, a B&K 2636 amplifier was used. The recordings were made at a representative point at which a user would be seated in the room in which the unit was installed, and for different operating speeds of the units. In general, the units included three speeds, with the exception of three units that had two, four and five speeds, respectively. In all, 47 noises were recorded.

Using the same recording equipment, a 1 kHz, 44 dB SPL reference signal was recorded with a B&K 4231 calibrator and a B&K 2636 attenuating amplifier (right and left channel). This signal was subsequently used to calibrate the sound levels for the laboratory experiments. The list of sounds, operating speeds, and sound levels is shown in Table 1. The different units are indicated by letters. The sounds were transferred from the DAT tape to a NeXT workstation with a sampling rate of 44.1 kHz and 16-bit resolution. They were shortened to 6 s in duration and raised cosine ramps of 50 ms duration were applied to the beginning and ending of each sound sample. The signal from the low-noise microphone (right channel) was used for the listening tests. This mono-signal was presented binaurally with identical signals presented to the two ears.

The sound samples are characterized by a noisy part related to the air flow from the device that is more or less colored by the baffles (acoustic filtering), and by an harmonic part related to the motor and presenting more or less prominent spectral lines. The initial analyses of the sound files revealed a presence in all of the recordings of a high-amplitude acoustic component at 100 Hz (see Appendix B).

2.2. Initial selection of sounds with a classification paradigm

A reduction of the initial sound corpus was performed to keep only a selection that was representative of the variety of sounds produced by the different brands and to minimize the duration of the dissimilarity and preference test sessions. A choice was made to maximize the variation of timbral differences and to minimize differences in loudness between samples, since timbre remains relatively constant but loudness varies considerably with the distance of the observer from the appliance. The loudnesses were not equalized, but it was decided to choose samples roughly

Table 1
List of recorded sounds

Sound number	Unit	Speed	Level (dBA)
1	a	3 (fast)	47.8
2	a	2 (medium)	41.9
3	a	1 (slow)	32.9
4	b	3 (fast)	37.6
5	b	2 (medium)	30.8
6	b	1 (slow)	26.2
7	c	3 (fast)	40.6
8	c	2 (medium)	38.5
9	c	1 (slow)	37.1
10	d	3 (fast)	46.6
11	d	2 (medium)	42.0
12	d	1 (slow)	34.9
13	e	3 (fast)	46.2
14	e	2 (medium)	38.5
15	e	1 (slow)	30.3
16	f	3 (fast)	45.6
17	f	2 (medium)	38.9
18	f	1 (slow)	37.7
19	g	3 (fast)	46.5
20	g	2 (medium)	34.7
21	g	1 (slow)	32.2
22	h	3 (fast)	38.6
23	h	2 (medium)	33.4
24	h	1 (slow)	27.9
25	i	3 (fast)	37.3
26	i	2 (medium)	26.8
27	i	1 (slow)	18.7
28	j	2 (fast)	35.0
29	j	1 (slow)	30.4
30	k	2 (fast)	34.5
31	k	1 (slow)	28.6
32	l	3 (fast)	39.4
33	l	2 (medium)	32.7
34	l	1 (slow)	24.5
35	m	only one speed	31.9
36	n	4 (fast)	34.2
37	n	3 (medium)	30.6
38	n	2 (slow)	28.6
39	n	1 (very slow)	25.2
40	o	3 (fast)	35.5
41	o	2 (medium)	30.2

Table 1 (*continued*)

Sound number	Unit	Speed	Level (dBA)
42	o	1 (slow)	23.7
43	p	3 (fast)	37.1
44	p	2 (medium)	34.4
45	p	1 (slow)	31.0
46	q	2 (fast)	37.3
47	q	1 (slow)	28.5

Columns 2–4 indicate the units with different letters, the operation speed, and the acoustic pressure in dBA, respectively. For reasons of confidentiality, the brand names and models cannot be divulged.

similar in loudness. The first selection of the sounds was based on a preliminary experiment considering both timbre and loudness.

Five expert listeners, all members of the psychoacoustics team at IRCAM, participated in the preliminary experiment. The noises were presented over headphones (Sennheiser HD420) at a level corresponding to normal functioning at a distance of 1.2 and 1.5 respectively for the VCV and SPLIT units (the position of the microphone with respect to the units in the anechoic chamber, Appendix A). In order to reproduce the level accurately, the noises were calibrated with a 1-kHz pure-tone signal recorded at a level of 44 dB SPL at the time at which the units were recorded. Listeners were asked to classify the sounds according to similarities in timbre and loudness. The data representation corresponds to a two-dimensional categorization scheme with loudness on the horizontal axis and timbre on the vertical axis. No constraint was imposed as to the number of categories or the number of sounds per category. The frequency across listeners at which two sounds were placed in the same timbre or loudness category was determined. This gave a matrix of individual proximities. A hierarchical cluster analysis was performed on this matrix. The distance metric between stimuli is Euclidean, and the cluster analysis technique is complete linkage (farthest neighbor). The results are presented as a tree structure in which the relative proximity between samples across listeners is represented for each of the grouping criteria (timbre and loudness) (Fig. 1). The height at which two sounds are joined in the cluster tree indicates the distance that separates them. Three main classes were found for both loudness and timbre. After listening comparisons, an intermediate loudness category was selected for further testing by the experimenter. This choice was designed to limit the possible predominance of loudness, while keeping a wide range of timbre variation and having a representative sample of the range of brands for the units under consideration. This procedure resulted in a final group of 18 sound samples shown in bold-faced type in Fig. 1.

2.3. Preliminary testing with a dissimilarity paradigm

The aim of the third stage in the constitution of the noise corpus was to determine approximately the perceptual structure underlying the 18-noise corpus in

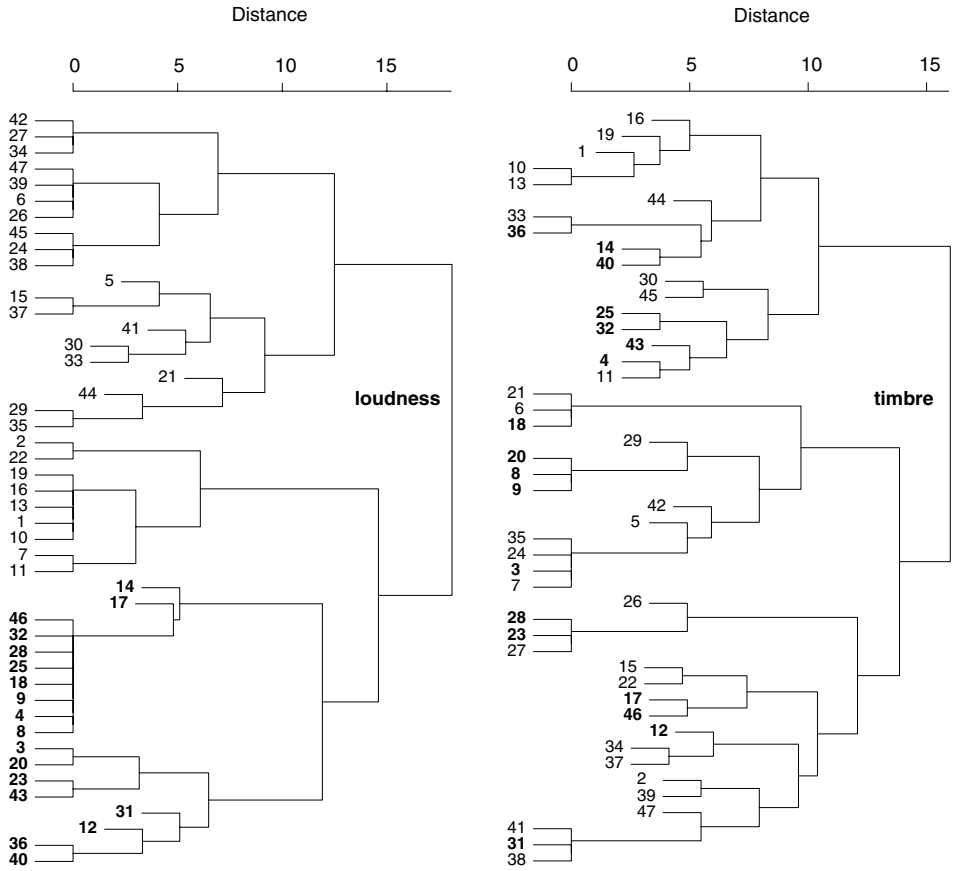

Fig. 1. Hierarchical clustering representation of the sounds with respect to loudness (a, left panel) and timbre (b, right panel).

terms of common perceptual dimensions. It is important that the noise corpus be relatively evenly distributed along each dimension in the multidimensional space in order to characterize appropriately the continuous acoustic nature of the dimension. If such is not the case, the corpus has to be redefined to reduce such inhomogeneities.

2.3.1. Apparatus

The level of presentation of the sounds was calibrated with a 1-kHz sine tone adjusted to be 44 dB SPL at the earphones. The level at the earphone was measured with a B&K 2209 sound level meter (fast response time) using a metal flat-plate coupler made in-house. The experimental session was run using the PsiExp v2.5 [15] software environment running on a NeXT computer equipped with IRCAM's ISPW sound processing card and the Max synthesis software. The sounds were sampled at

44.1 kHz with 16-bit resolution and were converted to audio signals with ProPort DACs including anti-aliasing filters. The stimuli were amplified by a Yamaha P2075 stereo amplifier and presented diotically (the same signal in the two ears) over a Sennheiser HD420 headset. The listeners were seated in a double-walled IAC sound booth.

2.3.2. Procedure

The same group of five expert listeners who performed the preliminary classification experiment also participated in a preliminary dissimilarity rating experiment. No information concerning the nature of the sounds was given to the listeners prior to testing. The experiment was performed in one session by each listener individually. At the beginning of the session, listeners heard all of the sounds once each in random order to give them an idea of the range of variation. All pairs of different sounds were then presented in random order. The order of presentation of the sounds of a pair (i–j or j–i) was selected randomly for each listener and each pair. A one-second silence separated the two sound samples. A horizontal cursor on the computer screen in front of the listener could be displaced with the computer mouse along a continuous scale labeled "similar" on the left and "different" on the right. The listener was required to move the cursor at least once before validating his or her rating by clicking on a button on the computer screen. Once the rating was validated, the next trial was presented automatically. On each trial, the sounds presented and their order of presentation were recorded along with the coded value of the cursor (0 at the far left and 100 at the far right).

2.3.3. Results

The data were analyzed with the MDS programme CLASCAL [12] with a single latent class of listeners. The analysis determines the spatial model for which the interstimulus distances best fit the rated dissimilarities and represents each sound as a point in a space. Of course, with so few listeners, the relative positions of the sounds were not reliably estimated, but the aim was to anticipate and adjust the distribution of the sounds before running the full experiment with 50 listeners. The MDS analysis gave a three-dimensional solution without specificities (Fig. 2). The sounds are distributed unevenly in the space, leaving holes where no sounds are found.

2.4. Creation of hybrid sounds

Due to the uneven spatial distribution, the fourth stage involved the integration of new sounds into the corpus that were created by an analysis/synthesis technique [16]. Another reason for this approach was to test whether it was possible to create new sounds from the perceptual distances among noises derived from the same family of appliances. The analysis/synthesis technique interpolates between the spectral envelopes of different sounds in the previously obtained space. The new sounds thus obtained are hybrids of the original sounds. A new space was thus developed by adding the new sounds. The different steps in the analysis/synthesis procedure are

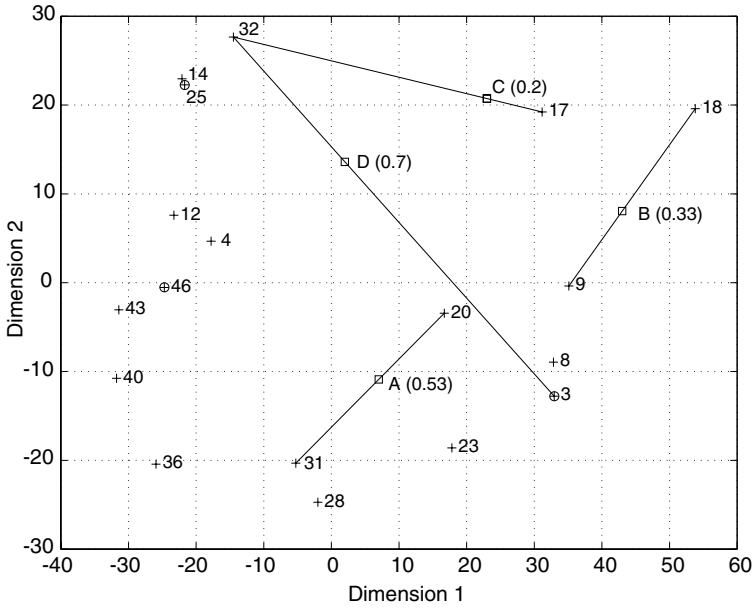

Fig. 2. Projection of the preliminary perceptual space obtained with the data from five listeners onto the first two dimensions. The labels correspond to the sounds. Sounds A, B, C, and D are hybrid sounds created by interpolation between recorded sounds. The number in parentheses for these sounds indicates the interpolation coefficient α (see text). The circles show sounds that were eliminated from the final stimulus set.

described in detail in Appendix B. Briefly, an interpolation plane was defined in order to constrain an equal distribution along each of the dimensions. Each interpolation consisted in placing a new sound between two sounds of the original space (Fig. 2), this new sound being created by a model of additive synthesis. The new sound corresponded to a linearly weighted combination of the spectral characteristics of the two originals as a function of their position in the perceptual space and the desired position of the new sound. Four such positions were defined in the three-dimensional space in order to equilibrate the distribution of the sounds along each dimension.

Fig. 2 presents the interpolation plane projected into the two-dimensional space corresponding to dimensions 1 and 2 of the three-dimensional space. The new sounds are indicated by the labels A, B, C and D. For each one, the value α of the coefficient of interpolation in three dimensions with respect to the two sounds from which it is derived is indicated. The interpolation relation is given by $I = \alpha X + (1 - \alpha)Y$, where I corresponds to the new sound (A, B, C, D), and X and Y are the two parent sounds. The computation of the interpolation coefficients corresponds to the ratio of the distance between X and I ($[X–I]$) over the distance between X and Y in the 3-D space.

The four new sounds replaced three of the original 18 sounds in order to keep the number of stimuli to be compared within reason. Thus, sounds 3, 25 and 46 were eliminated. The labels of the resulting 19 sounds are thus 4, 8, 9, 12, 14, 17, 18, 20, 23, 28, 31, 32, 36, 40, 43 from the initial corpus, and A, B, C, D for the four new sounds. Finally, the same five expert listeners listened to the new corpus of sounds to determine whether they would notice any anomalies. One person did not notice a change, two people remarked that the sounds were more homogeneously distributed, and two others spoke of "hybrid sounds" but did not qualify them as unnatural.

3. Multidimensional scaling of perceptual dissimilarities

The aim of this step was to determine the perceptual structure underlying the comparison of the sound samples using a dissimilarity rating paradigm. The CLASCAL MDS programme was used to model the perceptual structure.

3.1. Method

The apparatus and procedure were identical to those in the preliminary dissimilarity experiment.

3.1.1. Listening panel

Fifty listeners (29 men and 21 women with an average age of 28) were recruited for the experiment. No specific selection criterion other than self-reported normal hearing was used.

3.1.2. Stimuli

The stimuli consisted of the 19 sound samples described previously (15 originals and 4 hybrids). The stimulus duration was reduced to 3 s in order to minimize the duration of the experimental session for each subject.

3.2. Results

Dissimilarity judgments obtained from all subjects were analysed using the analysis programme CLASCAL [12]. This analysis was performed to reveal:

- the common perceptual dimensions shared by the whole set of sounds, i.e. the auditory attributes used by the subjects to make their judgments;
- the existence of perceptual dimensions that are specific to individual sounds (specificities), i.e. not shared by other sounds;
- the number of subpopulations (latent classes) of listeners, which are distinguished according to the perceptual weights they assign to the various dimensions and specificities.

The steps of the analysis are described briefly here, but more detailed accounts are available [11,12,14]. The initial Monte Carlo analyses were performed on the null model (mean dissimilarities) to determine the number of latent classes of listeners.

Given the number of listeners tested (50), tests on from one to seven classes were performed. The analysis yielded five latent classes. Spatial models from one to six dimensions without specificities and from one to four dimensions with specificities were then tested with five latent classes. A Bayesian information criterion (BIC) is used to select the most parsimonious model that best fits the data. The lowest BIC value corresponds to a spatial model with three dimensions and specificities. Each of the three common dimensions and the specificities are perceived with more or less weight depending on the latent classes.

The spatial configuration of the three common dimensions is shown in two-dimensional projections in Fig. 3. The sounds are considered to vary along these continuous dimensions. Fig. 4 shows an homogeneous distribution of sounds along each dimension taken independently which means that the correlations between physical characteristics and perceptual coordinates along a given dimension can thus be performed reliably.

The positions of the synthetic sounds correspond quite closely to the target position predefined by the interpolation operation. For example, the synthetic sound B, resulting from the interpolation between sounds 9 and 18, is situated quite close to the line joining these latter two positions in the Dimensions 1–2 plane. Similar relations are found for sounds A and C. We cannot evaluate sound D from this standpoint since one of its progenitors (sound 3) was removed from the sample set for the main experiment, but, in the Dimensions 1–2 plane, D would fall almost on a straight line between sounds 32 and 8. It is interesting to notice that the latter was close to sound 3 in the Dimensions 1–2 plane (Fig. 2) but far away along Dimension 3 (not presented) in the preliminary space.

The specificities indicate the extent to which each sound was perceived as being different from all the others along one or more dimensions or discrete features that are not shared by the whole sound set. The square roots of the specificity values are comparable in magnitude to coordinates of the sounds along the common dimensions. The mean of the specificity values is 22, (7) which is not very high in comparison with the range of values along dimension 1, 2 and 3 (80, 50, 55, respectively). The specificities of the synthesized sounds are not particularly higher than the others: A (15), B(0), C(26), D(24), suggesting that the listeners did not perceive a synthetic character that may have distinguished them from the original recorded sounds. Listening to and studying the sound set did not lead us to a clear idea concerning what acoustic characteristics are captured by the specificities, aside from temporal fluctuations. It should be emphasized that the listeners were not informed as to the origin of the sounds. Most of them attributed the sounds to a machine (rotating drum of a washing machine, the passing of a vehicle, a domestic appliance, etc.) or to elements of the natural environment (wind, ocean, waterfall, etc.). The sounds were often described by opposing aspects related to the sound quality, such as ''low'', ''round'', ''soft'' compared to ''windy'' or ''noisy''.

A second Monte Carlo simulation confirmed the choice of five latent classes for the chosen spatial model. The posterior probabilities indicate that the majority of the listeners clearly belong to a given class with the exception of five of them whose class belongingness is ambiguous. The number of listeners per class, the weights assigned

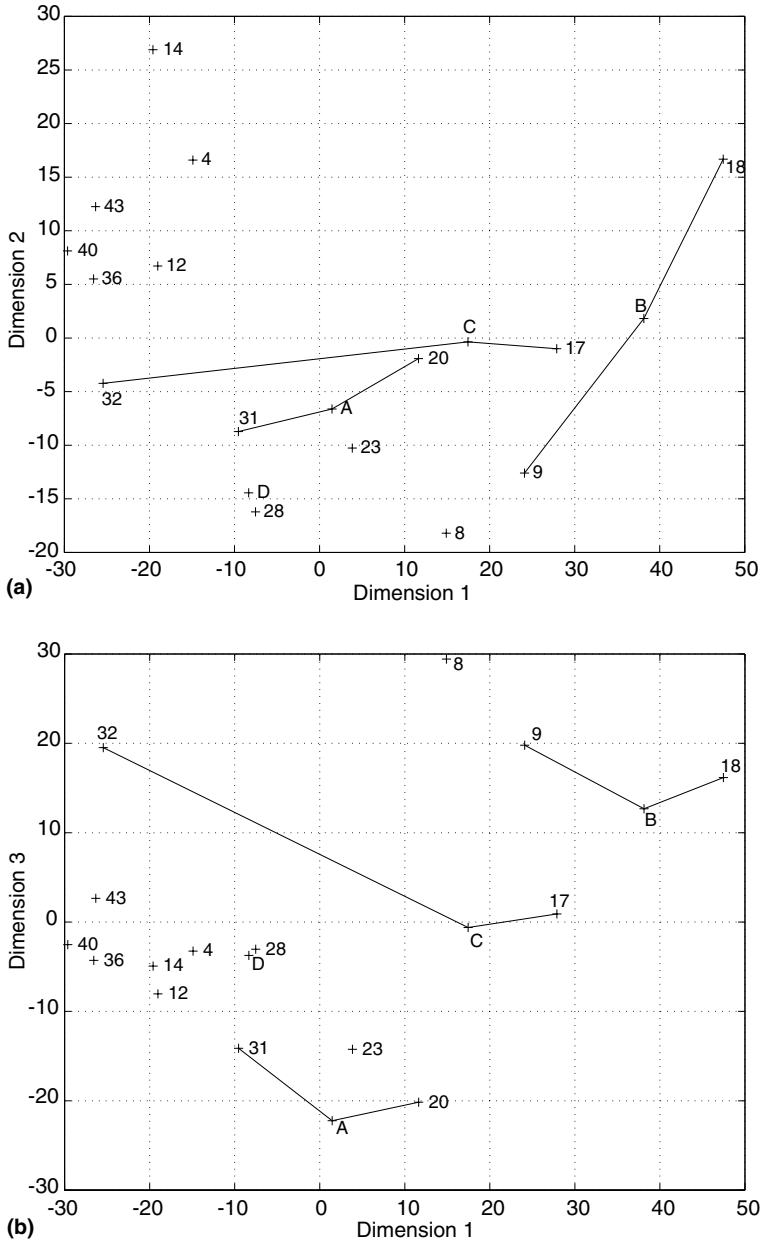

Fig. 3. Projection of the sounds onto the plane of Dimensions 1 and 2 (a) and Dimensions 1 and 3 (b) from the model with three dimensions and specificities. The sounds are labelled. The hybrid sounds are connected by lines to their progenitors, with the exception of sound D, sound 3 having been removed from the final stimulus set.

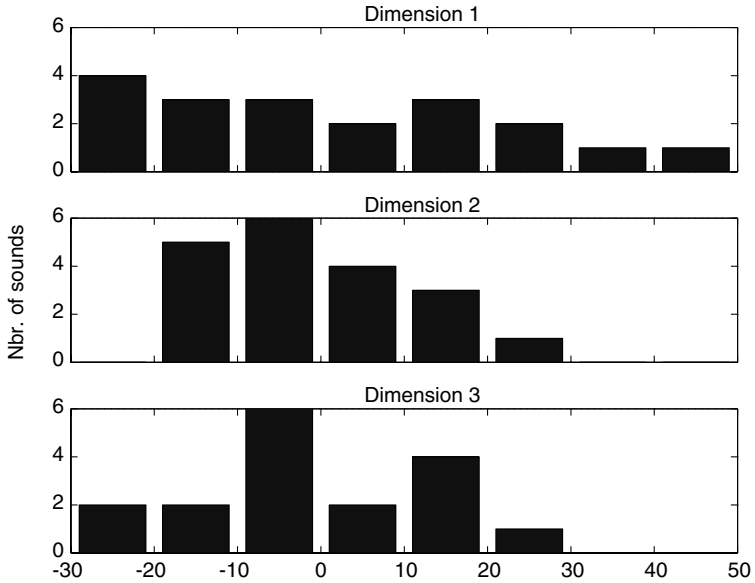

Fig. 4. Histogram representing the distribution of the sounds along each of the three perceptual dimensions.

Table 2
Number of listeners, weights on each dimension and on the set of specificities for each class

Class	#Subjects	Dimension 1	Dimension 2	Dimension 3	Specificities
1	4	1.09	1.30	1.85	2.71
2	3	0.33	0.23	0.30	0.16
3	18	1.35	1.52	0.98	0.55
4	7	1.26	0.91	1.17	1.18
5	13	0.95	1.03	0.68	0.37
Unclassed	5				
Sum	50	5.0	5.0	5.0	5.0

Unclassable listeners are those for whom the posterior probabilities did not allow clear classification. The CLASCAL programme imposes the constraint that the sum of the weights for a given dimension must equal the number of classes.

to each dimension, and the set of specificities for each class are shown in Table 2. About 60% of the listeners are found in Classes 3 and 5.

If the weights are normalized for each class, the relative weights can be compared across classes (Fig. 5). The normalized weights reveal that Classes 3 and 5 have nearly identical relative patterns of weights across dimensions and specificities. Similar weights are found on Dimensions 1, 2 and 3 for Classes 2 and 4, although they differ with respect to the weights on the specificities (very low for Class 2, high for Class 4). Class 1 clearly stands out as having a different behavior from the others with relatively lower weights on the common dimensions and a high weight on the specificities. However, this class contains only 4 of the 50 listeners.

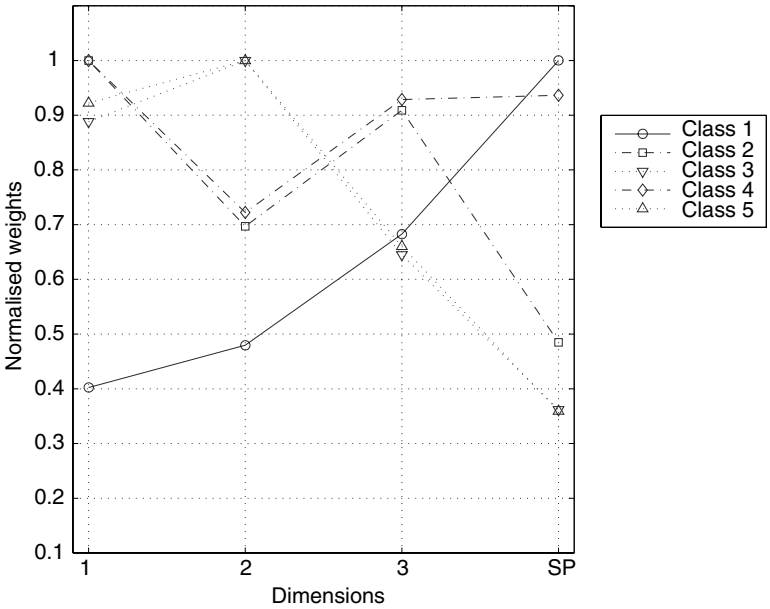

Fig. 5. Presentation of the normalized weights for Dimensions 1–3 and the specifities (SP) across the five classes obtained by the CLASCAL analysis.

The means (0.75, 0.23, 0.52, 0.58, 0.42) and standard deviations (0.11, 0.10, 0.17, 0.15, 0.15) of the dissimilarity judgments for Classes 1–5, respectively, reveal differences in the use of the rating scale. Classes 1 and 2 had extreme means (high for Class 1, low for Class 2) with small standard deviations, indicating that listeners in both of these classes used a restricted part of the scale. The other classes had intermediate means and larger standard deviations, indicating a use of the whole scale. Apart from those observations concerning subject's strategies in the use of the scale, no biographical factors could explain other differences across classes.

4. Acoustic analyses

The next step consists of determining the acoustic parameters that characterize the common perceptual dimensions. The first step in the objective characterization of the perceptual space precedes by an empirical loop consisting of listening to the sounds in terms of their relative positions in the multidimensional space in order to discover the kind of perceptual variation that corresponds to each dimension. Then hypotheses concerning the acoustic or psychoacoustic parameters are made and appropriate analytic parameters are developed that can be correlated with the coordinates in the perceptual space. A good correlation is taken as indicating a candidate objective predictor for the perceptual dimension. Most of the usual acoustic and psychoacoustic parameters have been calculated such as, sound pressure level,

loudness, loudness percentile (N_5 and N_{10}), sharpness, spectral centroid, roughness, fluctuation strength, pitch of complex sounds, using the PsySound programme [1] as well as functions developed in Matlab 4.2b. We will only present the parameters that were found to be relevant from a perceptual standpoint. The classes of objective predictors that we have retained by this procedure for the current study include: (1) the ratio of the energies of the non-harmonic part (resulting from the ventilator noise) and the harmonic part of the signal (resulting from the motor), (2) the spectral center of gravity (or spectral centroid) of the non-harmonic part, and (3) loudness (perceived level). The sound file levels were calibrated with respect to a reference, 1-kHz pure-tone signal at 44 dB SPL, in order to determine the absolute level.

4.1. Noise-to-harmonic (NHR) ratio

In order to analyse the independent contributions of the harmonic and noisy parts of the stimuli to their perception, the two parts were dissociated by an analysis/re-synthesis technique developed at IRCAM [16] that represents a sound signal as a sum of sinusoids of time-varying amplitude and frequency and a filtered noise, the spectral envelope of which can also vary over time. This technique results in synthetic sounds when the analysis data are used to reconstitute the two parts of the signal.

The program estimates the fundamental frequency, computes the fast Fourier transform (FFT) with a sliding window, estimates the spectral envelope, corrects the frequencies of the partials according to their phase, detects the spectral peaks and estimates their parameters (frequency, amplitude, phase), synthesizes the noisy part of the signal by way of a source/filter model after having subtracted the harmonic part, and synthesizes the harmonic part by inverse FFT. Once the harmonic and noise parts are separately resynthesized, the RMS level of each is evaluated for the frequency weightings A, B and C, and their ratio determined:

$$\mathrm{NHR}_{\mathrm{A,B,C}} = \frac{\mathrm{RMS}_{\mathrm{A,B,C}}(\text{noisy part})}{\mathrm{RMS}_{\mathrm{A,B,C}}(\text{harmonic part})}. \tag{1}$$

4.2. Spectral centroid (SC)

In perceptual studies on musical timbre, one of the most salient dimensions that has often been described in terms of "brightness" is strongly correlated with the spectral centroid (SC). This parameter captures certain aspects of the distribution of energy across the frequency spectrum of a sound [17,18]. It has also been found to be useful as a psychoacoustic descriptor for other types of sound sources such as cars [11]. In general, the SC is expressed in terms of a harmonic spectrum [19]. For non-harmonic sounds (vehicle noises, road noises, etc.), the SC can be determined from the mean frequency spectrum over several seconds with Welch's method [20]. In this

[1] This programme has been developed by Densil Cabrera (http://densil.tripod.com).

case, the brightness of a noise is the spectral centroid of the whole set of spectral samples. We perform this computation using the frequency weightings A, B and C. SC in Hz is thus expressed as:

$$\mathrm{SC} = \frac{\sum_{i=1}^{M} f(i) \cdot P(i)}{\sum_{i=1}^{M} P(i)}, \tag{2}$$

where $P(i)$ is the acoustic pressure level, M is the number of spectral bins in the FFT, and

$$f(i) = i\frac{f_s}{2M}, \tag{3}$$

with f_s being the sampling frequency.

4.3. Loudness (N)

Loudness is the psychoacoustic parameter corresponding to the perceived level of sound. It depends on effects of spectral masking related to the coding of sound in the cochlea as well as to as-yet-unexplained processes of integration across frequency channels. The model of loudness used here is described in ISO532B [21].

The first stage of processing in this model is a filter corresponding to the transfer characteristics of the outer and middle ears. The next stage is the computation of an excitation pattern reflecting the distribution of activity in the nerve fiber array in response to the sound. Threshold curves for pure tones masked by narrow-band noise are used to determine the excitation pattern. Then the excitation pattern is transformed from the frequency scale to a scale closer to the cochlear representation. In Zwicker's model, this scale is expressed in units of Barks, one Bark corresponding to an estimate of the width of an auditory filter or critical band. Finally, the specific loudness values N' for each of the critical bands z (in Barks) are computed as a function of the excitation E expressed in units of power. The global loudness N is then taken as the sum of the specific loudnesses. An approximation of specific loudness is given by the following relation:

$$N' \approx \left(\frac{E}{E_0}\right)^{0.23}, \tag{4}$$

where E_0 is the reference excitation level (power). The overall loudness is given by:

$$N = \int_0^{24\mathrm{Barks}} \mathrm{N}'\,\mathrm{dz}, \tag{5}$$

N being expressed in sones.

4.4. Correlations between objective predictors and perceptual dimensions

Dimension 1 corresponds to the relative balance of the harmonic (motor) and noise (ventilator) components. Fig. 6 shows the evolution of the A-weighted acoustic

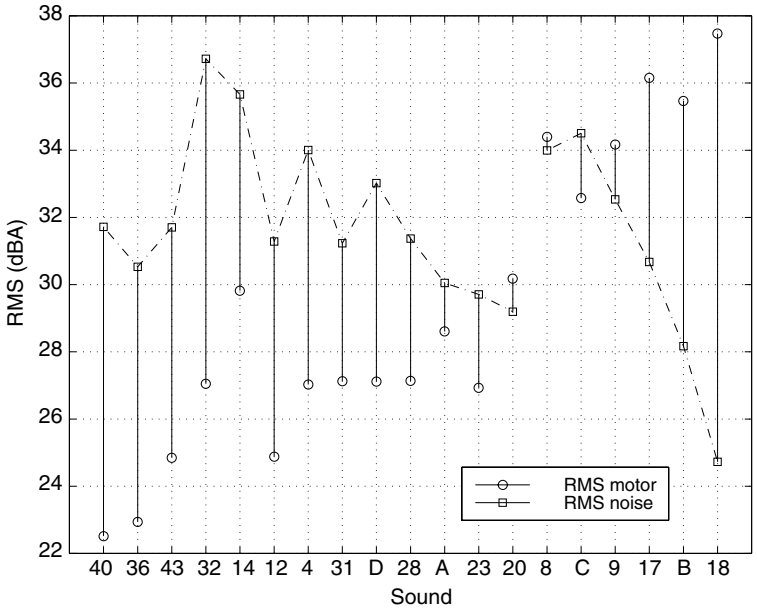

Fig. 6. Evolution of acoustic level (RMS in dBA) of the noise component (ventilator) and the harmonic component (motor) as a function of the order of the sounds along Dimension 1.

level of the harmonic and noise parts as a function of the coordinate on Dimension 1: for sounds 40–23 (on the left), the noisy part is dominant, and for sounds 20–18 (on the right, with the exception of sound C) the harmonic part is dominant. Thus the parameter NHR_A appears to be a strong candidate for this dimension. The best correlation ($r(17) = 0.97$, $p < 0.01$) is obtained with the A-weighted version of this parameter, which will be retained for the subsequent analyses (Fig. 7(a)).

For Dimension 2, the emergence of a spectral pitch led us to consider the parameter SC. The correlation between the SC computed on the entire sound and Dimension 2 coordinates is statistically significant ($r(17) = 0.58$, $p < 0.01$), although it is quite weak, explaining only 34% of the variance. Another possibility considered was the SC of each of the two parts of the sound: the noise component (SC_N) and the harmonic component (SC_H). The best correlation with Dimension 2 was found for SC_N ($r(17) = 0.73$, $p < 0.01$) (Fig. 7(b)), whereas the correlation for SC_H was not significant ($r(17) = 0.41$).

The sounds ordered along Dimension 3 vary systematically in loudness (N) ($r(17) = 0.84$, $p < 0.01$). Even though the sound sample set was selected to have a reduced range of loudness variation, the parameter N emerges as a candidate acoustic correlate (Fig. 7(c)).

Table 3 presents the correlation coefficients of each of the selected physical parameters with each of the perceptual dimensions. Each parameter explains a moderate to large portion of the variance along a single perceptual dimension: NHR_A explains 94% of the variance along Dimension 1, SC_N 53% of the variance along

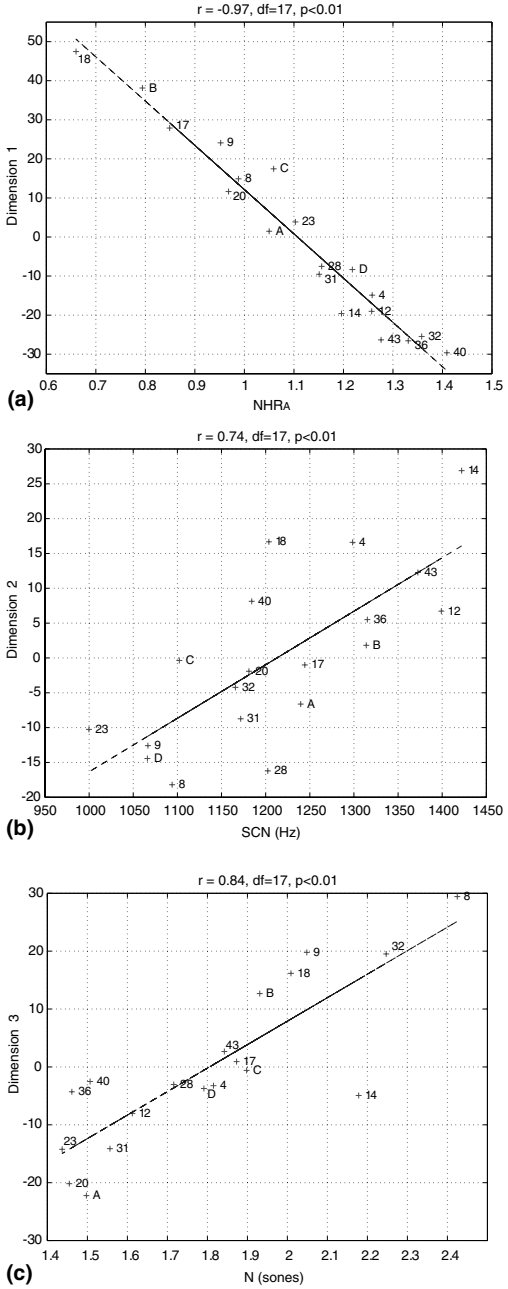

Fig. 7. Scatter diagrams and linear regression lines between the coordinates of one dimension of the perceptual space and its corresponding acoustic parameter: (a) Dimension 1 and A-weighted noise-to-harmonic energy ratio (NHR_A), (b) Dimension 2 and spectral center of gravity of the noise component (SC_N) in Hz and (c) Dimension 3 and loudness (N) in sones.

Table 3
Correlations (df = 17) among the selected physical parameters and the coordinates along the perceptual dimensions

	Dimension 1	Dimension 2	Dimension 3
NHR_A	−0.97**	0.11	−0.26
SC_N	−0.32	0.73**	−0.15
N	0.26	0.04	0.84**

** $p < 0.01$.

Dimension 2, and N 70% of the variance along Dimension 3. Note that the acoustic predictors are not significantly correlated with other dimensions, indicating their independence.

5. Preference study

The final phase of the study sought to model the preferences of listeners in terms of the perceptually relevant physical parameters selected in the previous stages of analysis (NHR_A,SC_N, N). The 19 sound samples were again presented in pairs and listeners were asked to choose the preferred sound.

5.1. Method

A preference matrix was constructed for each listener. An entry in the matrix is a binary judgment (0 or 1) indicating whether or not sound i was preferred over sound j in the pair $\{i, j\}$.

5.1.1. Listening panel

Ninety-five listeners (54 men, 41 women, average age = 30 years) with self-reported normal hearing were recruited. No specific selection criteria were used. Fifty of these listeners had participated in the dissimilarity study earlier.

5.1.2. Stimuli and apparatus

The stimuli and apparatus were identical to those in the dissimilarity study.

5.1.3. Procedure

At the beginning of the session, the listener heard all of the stimuli of the sound corpus in a randomly ordered sequence. No information concerning the nature of the sounds was given. In the training phase, 10 pairs of sound samples were chosen randomly and presented to listeners for preference judgments. Then all of the 342 pairs of different sounds in both orders (i, j and j, i) were presented in random order. On each trial, the listener heard the sound pair once and had to indicate which sound he or she preferred by clicking on the corresponding button (labeled 1 or 2) on the computer screen. The full matrix without diagonal of binary choices for each listener as well as the matrix of proportions of listeners preferences were constructed.

5.2. Results

The upper and lower triangular half matrices of the data set across listeners were compared in order to determine whether the order of presentation had an effect on preference judgments. A χ^2 test revealed no significant difference ($\chi^2(1) = 0.514$), so the upper and lower half-matrices were averaged.

A latent-class analysis was performed from two to four classes using Hope's test [22] on several different preference models. In all cases, two classes were found with 48 and 47 listeners, respectively.

BIC values were subsequently examined for various preference models (number of internal knots, order of the splines) and for all three pairs of objective predictors: NHR_A–SC_N, NHR_A–N, SC_N–N (in its current form, the program only allows the construction of two-dimensional preference maps). Only the additive model is considered, because regions without sounds in the 2-D projections make estimation of the multivariate model unreliable. The spline functions selected by BIC are third order, with one or two internal knots and two latent classes. The functions obtained are the "utility" functions that account for the contributions to preference of each objective predictor.

In Figs. 8–10, the utility of each parameter is indicated on the ordinate, the value of the predictor parameter (NHR_A, SC_N or N) on the abscissa, for both latent classes of listeners. The placement of each sound sample along the abscissa is indicated. Note that the curves are well defined by the sound corpus in all cases. Note also that

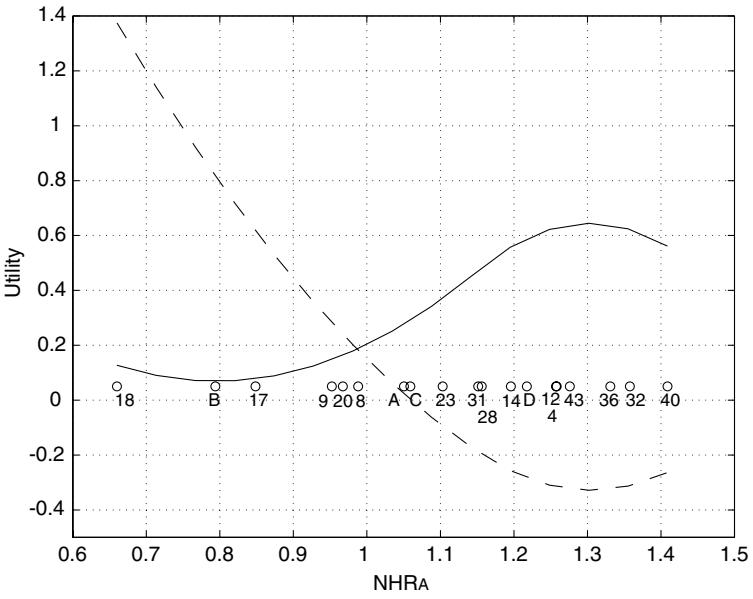

Fig. 8. Utility to preference curves for noise-to-harmonic ratio (NHR_A) for Classes 1 (solid line) and 2 (dashed line) drawn from the NHR_A–N analysis.

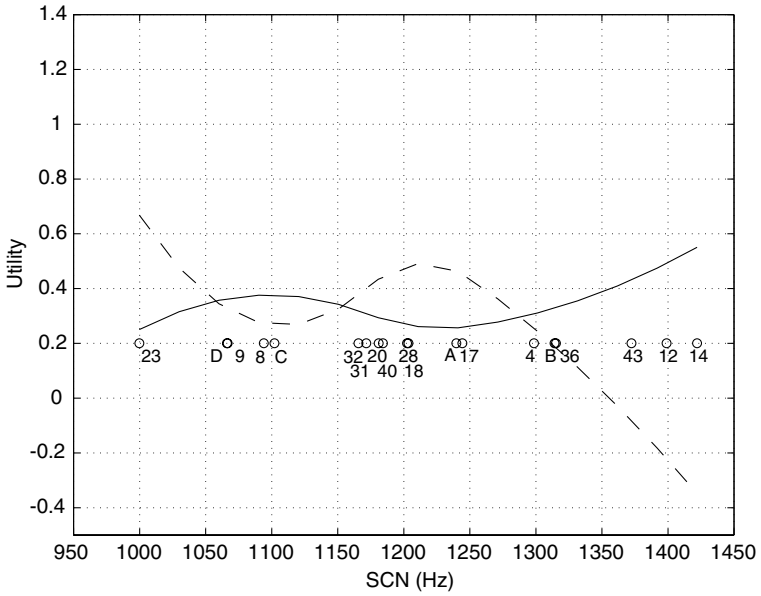

Fig. 9. Utility to preference curves for spectral center of gravity (SC_N) for Classes 1 (solid line) and 2 (dashed line) drawn from the SC_N–N analysis.

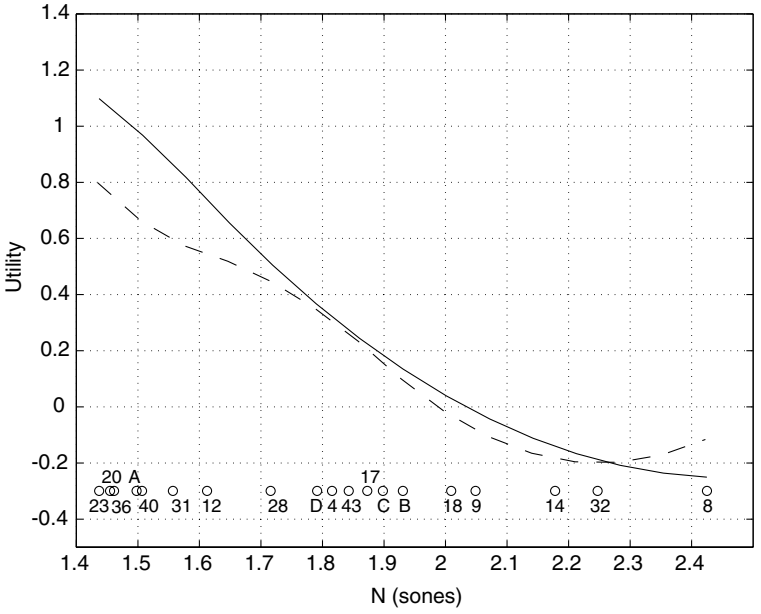

Fig. 10. Utility to preference curves for loudness (N) for Classes 1 (solid line) and 2 (dashed line) drawn from the N–NHR_A analysis.

some of the curves are not monotonic, indicating that the region of highest utility to preference is surrounded by regions of lower utility. These curves are drawn from one of the two analyses performed with each parameter (e.g. NHR taken from the 2D analysis with N). For NHR_A and N, the utility curves are very similar when paired with both other parameters (SC_N and N for NHR_A, SC_N and NHR_A for N). The utility curves for SC_N vary as a function of the second parameter for one latent class suggesting that this parameter may not be a stable predictor for some listeners.

A given sound may have quite different utility values for different predictor parameters. For example, for Class 1 the utility of sound 20 is maximal for loudness (Fig. 10), but is near the minimum for NHR_A (Fig. 8). Inversely for Class 2, sound 18 has the highest utility value for NHR_A (Fig. 8) and is near the minimum for loudness (Fig. 10). Further while sound 18 has high utility for NHR_A in Class 2, its utility is minimal for Class 1.

Globally, the utility decreases for both classes as loudness increases. However, the range of utility values as a function of loudness for Class 2 is smaller by a factor of 1.4 compared to the range for Class 1. Utility values for Class 1 decrease more rapidly in the range 1.4–1.7 sones. It seems clear that loudness is the dominant objective predictor for preference judgments in Class 1 compared to the two other parameters.

Another important factor that distinguishes the two classes of listeners is the inversion of the utility curves for NHR_A. Globally, utility increases for Class 1 and decreases for Class 2 as NHR_A increases. Listeners of Class 2 are much more sensitive to this parameter than those of Class 1. Further, for Class 2 the utility decreases as SC increases, with a local minimum near 1100 Hz and a second peak just above 1200 Hz.

No biographical factors concerning the subject panel were able to explain the differences between the two classes. In addition, subjects who had participated in the dissimilarity experiment are dispersed into the two classes, ruling out effects of previous exposure to the sound corpus.

5.2.1. Replication in another laboratory

The preference analysis was replicated on data for 94 listeners obtained in the laboratories of Electricité de France (EDF). None of these listeners had participated in the experiments at IRCAM. The experimental protocol and stimuli were identical. Again, no statistical difference was found between upper and lower triangular half-matrices. Hope's test revealed three latent classes in this data set, with 39, 42 and 13 listeners. An additional class thus appears. Proceeding in the same manner as for the IRCAM data set, the additive model was examined. The preference analysis revealed as before third-order spline functions, with one or two internal knots.

Fig. 11 shows the utility curves for the three classes with the predictor parameters N and NHR_A. Class 1 has preference judgments that are the inverse of those for Classes 2 and 3 for NHR_A, mirroring the distinction between the classes in the IRCAM data set for this parameter. Furthermore, Classes 1 and 2 have preference

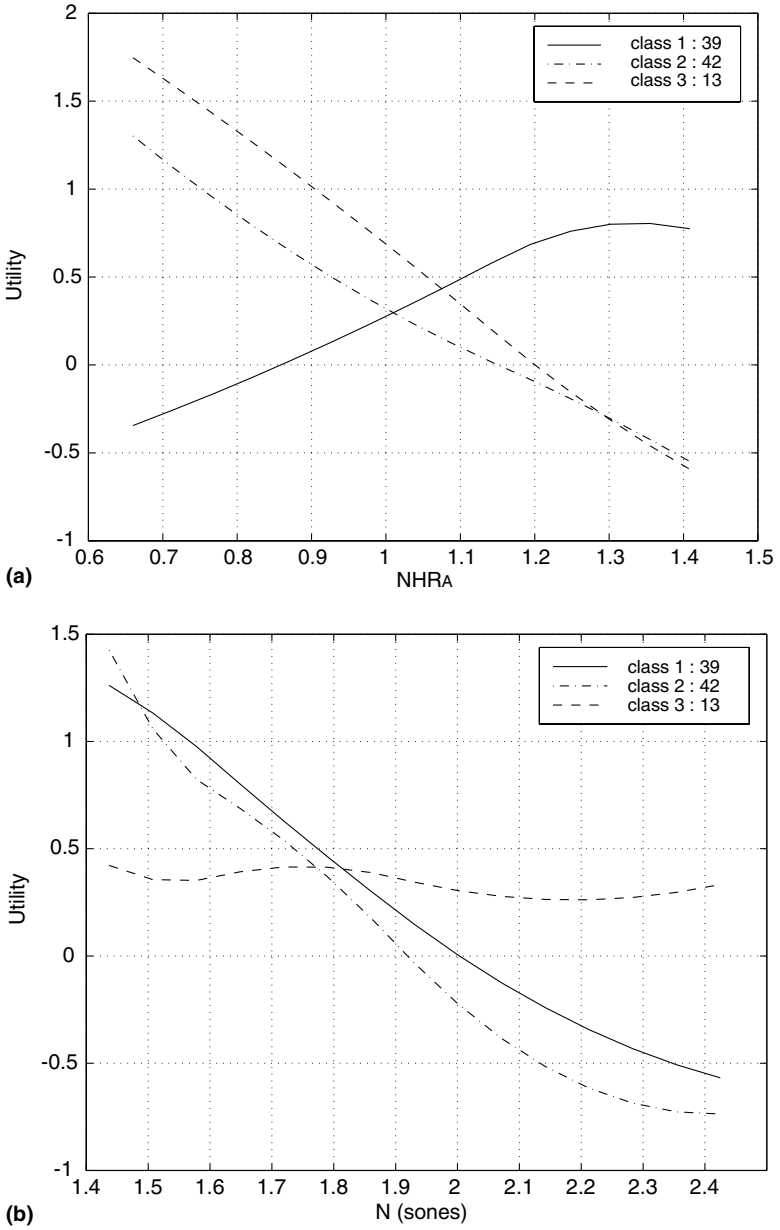

Fig. 11. Utility to preference curves for noise-to-harmonic ratio (NHR_A) (a) and loudness (N) (b) for Classes 1–3 drawn from the experiments performed at the EDF laboratories.

judgments that decrease when loudness increases, and they differ from Class 3 according to loudness for which the utility does not vary with N. Finally, the same tendencies are observed with the population tested at IRCAM.

6. Conclusion

The method applied in this psychoacoustic study characterizes the sound quality of, and preferences among, various brands and models of indoor air-conditioning units that correspond to a homogeneous class of sound sources.

On the basis of direct ratings of dissimilarity between sound samples, a set of indoor air-conditioning noises was perceptually characterized by three common dimensions which are correlated with the A-weighted noise-to-harmonic energy ratio (NHR_A), spectral center of gravity of the noise part (SC_N), and loudness (N). Globally, preference judgments depend primarily on N for one class of listeners, NHR_A and SC_N being of secondary utility for preference, whereas for another class of listeners, loudness was of weaker importance with respect to the other parameters. Further, the two listener classes were distinguished by the form of the function relating utility to preference and NHR_A. For one class, preference increases slightly with increasing energy in the noise part of the sound. However, preference was greatest for higher energy in the harmonic part and decreased as this component decreased for the other class. In the continuation of this study, it would be necessary to test a preference multivariate model against the additive model, in order to study whether the listeners' preferences can be explained by an independent contribution of the three physical parameters or a more complex interaction among them. To do this, it would be necessary in addition to create new sounds to have a more homogenous distribution in the physical parameter space.

Globally, the multidimensional analysis technique allows us to probe and precisely synthesize the continuous perceptual dimensions as a function of acoustic parameters that explain both the perceptual structure and the preferences over the set of stimuli. The advantage of this method is that it allows us to establish the relevant psychophysical relation without making a priori assumptions as to the number and nature of the dimensions, in contrast to unidimensional scaling using verbal descriptors. However, this method presupposes the existence of continuous dimensions. The analysis of the preference judgments produces a set of curves relating preference to the acoustic parameters. These curves can be considered to define the client-based specifications that situate a sound in terms of its sound quality. In our case, two different trends are revealed according to parameter NHR_A. A manufacturer would need to decide whether to satisfy one or the other population of potential clients, or to satisfy the population that is more sensitive to this parameter by keeping the lowest NHR_A value as a specification.

We have shown the method to be useful in the sound design process when it is necessary to create new sounds with respect to the dominant perceptual characteristics of a set of measured sounds. Indeed, this study showed that the synthetic sounds created within the three-dimensional space defined by the recorded sounds are in fact situated near their predicted positions, suggesting that the spatial model has a certain predictive power. It thus seems possible, when new air-conditioning units are created, to determine their positions in the perceptual space by simply measuring the three acoustic parameters that define the space (NHR_A, SC_N and N) and consequently to infer their preference with respect to the other sounds.

Acknowledgements

This research was conducted in collaboration with Electricité de France (F. Junker, E. Siekierski, C. Durquenne), the Laboratoire National d'Essai (P. Cellard, J.-M. Lambert, S. Ciukaj) and the Institut National de la Recherche Agronomique (N. Martin). The sound samples were recorded by S. Ciukaj at the LNE. The comparison data set for the preference study was collected by E. Siekierski at EDF.

Appendix A. Installation configuration of the air-conditioning units and microphone placement for the recordings performed by the LNE

VCV units (Fig. 12) were placed against a vertical reflecting plane (without touching it to prevent solid transmission of vibrations to the wall). Units not having integrated foot pedestals were placed on wooden supports at the height above the floor indicated in the manufacturer's specifications.

SPLIT units (Fig. 13) were attached with two metallic supports to the edge of the rigid ceiling panel and were abutted against the vertical reflecting plane, being separated from it by supple materials to avoid solid transmission to the wall. The height of the ceiling corresponded to the manufacturer's specifications.

For the recordings and level measurements of the noises emitted by the devices at different ventilation speeds, the two microphones were placed as indicated in Figs. 12 and 13. For the VCV units, the microphones were situated at a height of 1.2 m from the floor and at a distance of 1.2 m from the ventilation grid. For the SPLIT units, the microphones were situated at a height of 1.5 m below the ceiling, which was 1.2 m above the floor, and at a distance of 1.5 m from the wall. Both microphone placements are set to simulate the position of the ears of a person sitting in the room with the device.

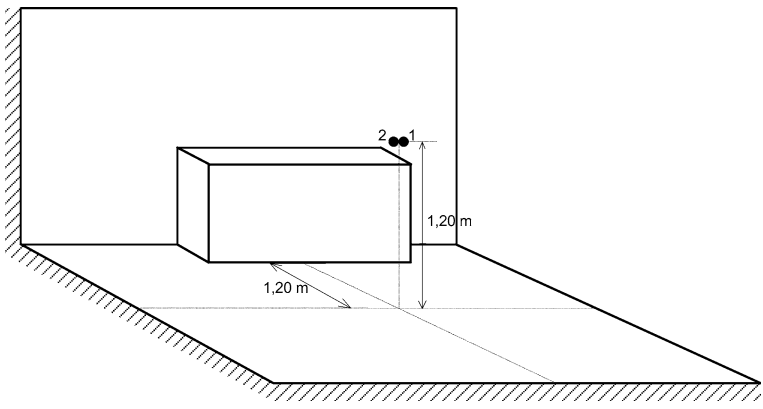

Fig. 12. Installation of VCV-type free-standing air-conditioning units. Microphones (1,2) are indicated by the black circles.

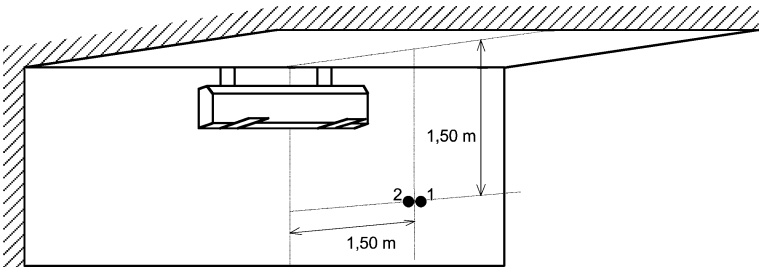

Fig. 13. Installation of SPLIT-type ceiling-mounted air-conditioning units. Microphones (1,2) were placed 1.5 m from the unit and 1.5 m below the ceiling.

Appendix B. Method of sound interpolation

The creation of sounds by interpolation requires a perfect analysis and synthesis of each pair of sounds used to create a new one by modifying and mixing their respective spectral components. First, the reference sounds were analyzed and modeled, and then interpolations between them were synthesized.

B.1. Step 1: Analysis and resynthesis of the sounds

One difficulty for the analysis and resynthesis of the recorded sounds is related to the fact that they do not only possess harmonic components but also a non-negligeable noise component. So the harmonic and noise components were separately modeled.

The harmonic part of the signal was modeled as a sum of sinusoidal components which was subtracted from the original signal leaving the noise component as a residue. This signal-processing operation was performed with the Additive program developed at IRCAM [16]. First the fundamental frequency is estimated in each 40-ms time window over the duration of the signal. Adjacent time windows overlap by 75%. For each window, an FFT is performed. The peaks in the resulting spectrum are identified and only the harmonics of the fundamental frequency estimated for that window are kept. This operation is performed by the SIEVE function. Then the possible appearance and disappearance of partials is smoothed to produce amplitude, frequency and phase functions of time for each one. This information is saved in a computer file. The resynthesis consists of an interpolation for each parameter between values in successive windows and a simple sinusoidal additive synthesis.

The fundamental frequency is about 100 Hz for the whole set of sounds. The harmonic spectrum did not extend beyond 7 kHz, and the number of partials used for resynthesis was limited to 70. With more than 70 partials, the sounds had a slightly metallic timbre which gave them a synthetic quality.

The next step involved extracting the spectral envelope of the noise component using the phase-vocoder-based program Super VP [23]. The envelope is simply the LPC amplitude spectrum of the residual noise. Super VP yields a file containing the

amplitude of each frequency bin. This envelope is then applied to a white noise to simulate the initial noise.

Finally, the synthesized harmonic and noise files are mixed with an adjustable scalar coefficient applied to each. The process is mostly automatic with the exception of the final mixing stage.

B.2. Step 2: Spectral interpolation

A requirement established at the outset was that the method be able to linearly interpolate between two arbitrary sounds and that the interpolation coefficient be chosen between 0 and 1. The process consisted of interpolating between the amplitudes and frequencies of the partials of two reference sounds. The partials in each file were identified by harmonic rank and interpolation was performed between partials having the same rank in the two files. The interpolation was performed according to the formula $I = \alpha X_i + (1 - \alpha) Y_i$, where α is the interpolation coefficient, X_i is the value of the file corresponding to sound X and Y_i, is the value of the other file. The complete process allows for the independent interpolation between the partials of the harmonic components and the spectral envelopes of the noise parts. The resulting interpolated harmonic and noise signals are mixed as in Stage 1 to obtain the final mix. Note that various trials with this method revealed that sounds with high specificity values were difficult to interpolate.

References

[1] Susini P, Perry I, Winsberg S, Vieillard S, McAdams S, Rodet X. Sensory evaluation of air-conditioning noise: Sound design and psychoacoutic evaluation. In: Proceedings of the ICA, Rome, 2001. p. 342.
[2] Siekierski E, Derquenne C, Martin N. Sensory evaluation of air-conditioning noise: Sensory profiles and hedonic tests. In: Proceedings of the ICA, Rome, 2001. p. 342.
[3] Junker F, Susini P, Cellard P. Sensory evaluation of air-conditioning noise: Comparative analysis of two methods. In: Proceedings of the ICA, Rome, 2001. p. 342.
[4] Solomon LN. Semantic reactions to systematically varied sounds. J Acoust Soc Am 1959;31:986–90.
[5] Björk EA. The perceived quality of natural sounds. Acustica 1985;57:185–8.
[6] Zeitler A, Hellbrück J. Semantic attributes of environmental sounds and their correlations with psychoacoustic magnitude. In: Proceedings of the ICA, Rome, 2001. p. 114.
[7] Osgood CE. The nature and measurement of meaning. Psychol Bull 1952;49(197–237).
[8] Chouard N, Hempel T. A semantic differential design especially developed for the evaluation of interior car sounds. Joint Meeting: ASA/EAA/DEGA, Berlin, Germany, JASA 1280 (105), 1999.
[9] Kyncl L, Jiricek O. Psychoacoustic product sound quality evaluation. In: Proceedings of the ICA, Rome, 2001. p. 90.
[10] Zwicker E, Fastl H. Psychoacoustics: Facts and models. Berlin: Springer-Verlag; 1990.
[11] Susini P, McAdams S, Winsberg S. A multidimensional technique for sound quality assessment. ACUSTICA – acta acustica 1999;85:650–6.
[12] Winsberg S, De Soete G. A latent class approach to fitting the weighted Euclidean model, CLASCAL. Psychometrika 1993;58:315–30.
[13] Winsberg S, De Soete G. A Thurstonian pairwise choice model with univariate and multivariate spline transformation. Pychometrika 1993;58:233–56.

[14] McAdams S, Winsberg S, Donnadieu S, De Soete G, Krimphoff J. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychol Res 1995;58:177–92.
[15] Smith B. PsiExp: an environment for psychoacoustic experimentation using the IRCAM musical workstation. Society for Music Perception and Cognition Conference'95, University of California, Berkeley, 1995.
[16] Depalle P, Garca G, Rodet X. Tracking of partials for additive sound synthesis using hidden Markov models. In: Proceedings of the ICASSP, 1993. p. 1225–8.
[17] Grey JM, Moorer JA. Perceptual evaluation of synthesized musical instrument tones. J Acoust Soc Am 1977;62:454–62.
[18] Carterette EC, Kendall RA. Musical communication. Recent trends in hearing research. BIS: K.S. Fastl; 1996.
[19] Krimphoff J, McAdams S, Winsberg S. Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. CFA3, Toulouse, France. J Phys 1994:625–8.
[20] Oppenheim AV, Schafer RW. Digital signal processing. New York: Prentice-Hall; 1975.
[21] Zwicker E, Deuter K, Peisl W. Loudness meters based on ISO 532 B with large dynamic range. In: Inter'Noise 85, Proceedings 1985 International Conference on Noise Control Engineering, vol. II, 1985. p. 1119–22.
[22] Hope AC. A simplified Monte Carlo signifiance test procedure. J Royal Statist Soc, Series B 1968;30:582–98.
[23] Depalle P, Poirot G. Super phase vocoder: A modular system for analysis, processing and synthesis of sound signals. In: Proceedings of the ICMC, Montreal, 1991.